



The Nuffield Early Language Intervention

Evaluation Report

May 2020

Sashka Dimova, Sonia Ilie, Elena Rosa Brown,
Miriam Broeks, Andreas Culora, Alex Sutherland





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.

This project was co-funded with the financial support of Intermediate Capital Group (ICG).



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:



Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP



0207 802 1653



jonathan.kay@eefoundation.org.uk



www.educationendowmentfoundation.org.uk



Contents

About the evaluator	2
Executive summary	3
Introduction	5
Methods	13
Impact evaluation.....	30
Implementation and process evaluation	43
Cost	61
Conclusion	64
References	68
Appendix A: EEF cost rating.....	71
Appendix B: Security classification of trial findings	72
Appendix C: Changes since the previous evaluation	73
Appendix D: Recruitment Documentation	74
Appendix E: Outcome Measures.....	84
Appendix F: Baseline equivalence	85
Appendix G: Measure development: analytical approach	89
Appendix H: Randomisation code, Analysis code and output	91
Appendix I: Process Evaluation: Data collection	101

About the evaluator

The project was independently evaluated by a team from RAND Europe:

Sashka Dimova, Sonia Ilie, Elena Rosa Brown, Miriam Broeks, Andreas Culora, Eleftheria Iakovidou, Natalie Picken, Susie Lee, Julie Belanger, Alex Sutherland.

The lead evaluator was initially Alex Sutherland, then Sonia Ilie.

Contact details:

Name:	Dr Sashka Dimova
Address:	RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, UK
Tel.:	+44 1223 353 329
Email:	sdimova@rand.org

Executive Summary

The project

The Nuffield Early Language Intervention (NELI) programme is designed to improve the language skills of reception pupils (aged 4–5) and involves scripted individual and small-group language teaching sessions delivered by trained teaching assistants (TAs). NELI is published by Oxford University Press and was delivered in this trial by the University of Oxford in partnership with Elklan. The 20-week intervention consists of two 15-minute individual sessions and three 30-minute small group sessions each week. The children selected to participate in this trial were the five children in each classroom who obtained the lowest scores on a school-administered app-based assessment of oral language skills (LanguageScreen). The programme focuses on developing children's narrative, vocabulary, and listening skills, while phonological awareness and letter sound knowledge are also developed during the second half of the programme. Training includes an initial two-day course for TAs followed by a half-day workshop half-way through the programme in addition to ongoing online support. Teachers also receive an initial half-day of training so that they can understand the programme and support TAs.

This project was a randomised controlled trial; 193 schools participated, 97 of which received the programme while 96 did not, the latter forming the control group. The process evaluation involved observations of the training, surveys with school staff, case studies in six schools (which used interviews with school staff), and analyses of TA training logs and online resource use. This project was co-funded with the financial support of Intermediate Capital Group (ICG). The trial took place between June 2018 and November 2019.

Key conclusions

1. Children who received the NELI programme made the equivalent of three additional months' progress in language skills, on average, compared to children who did not receive NELI. This result has a very high security rating.
2. Children who received the NELI programme made the equivalent of two additional months' progress in early word reading, on average, compared to children who did not receive NELI, in addition to four additional months' progress in language skills (as measured by the digital application LanguageScreen).
3. Children who received the NELI programme with English as an additional language (EAL) made the equivalent of three additional months' progress in language skills compared to EAL children who did not receive NELI.
4. Surveys and interviews from the process evaluation showed that schools believed the training and ongoing support provided was clear, useful, and sufficiently detailed for them to deliver the intervention effectively. Initial training attendance was high, and while top-up training attendance was lower, the vast majority of TA survey respondents made use of ongoing support such as telephone calls, webinars, or forums.
5. There was variation in the number of sessions that schools delivered to pupils and schools did not necessarily deliver the maximum number of sessions. Schools were more likely to deliver the group sessions compared to the individual sessions. TAs attending training and delivering a larger number of sessions was associated with better language outcomes for pupils.

EEF security rating

These findings have a very high security rating. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was a well-designed, two-armed, randomised controlled trial. The trial was well-powered and relatively few pupils (7%) who started the trial were not included in the final analysis. The pupils in NELI schools were similar to those in the comparison schools in terms of prior attainment.

Additional findings

Pupils who received the NELI programme made, on average, three additional months' progress in language skills compared to children in the control group who did not receive the NELI programme. This is our best estimate of impact, which has a very high security rating. As with any study, there is always some uncertainty around the result. However, even when accounting for the statistical uncertainty around the estimate, the possible impact on the primary language outcome, as measured by four separate language tests, remains positive (ranging from two to four months' progress).

The trial also found that NELI had a positive impact on two secondary outcomes: the YARC Early Word Reading test (equivalent to an additional two months' progress) and the LanguageScreen language skills assessment (equivalent to an additional five months' progress). When accounting for the statistical uncertainty around these secondary outcome results, the possible impact on the YARC Early Word Reading test ranges from a very small chance of a negligible positive impact to four additional months' progress, while the possible impact on LanguageScreen outcomes ranges from three to six additional months' progress. Children with English as an additional language (EAL) in NELI schools who received the programme also made the equivalent of three additional months' progress in language skills compared to EAL children in schools that did not receive NELI. When accounting for statistical uncertainty, this estimate of the impact on EAL learners ranges from a very small chance of a negligible positive impact to an impact as high as five additional months' progress. These findings, which are derived from the largest trial of NELI to date, are in line with the positive results found in two previous, smaller randomised controlled trials which tested the impact of NELI (Fricke et al., 2013; Sibieta et al., 2016).

Teaching assistants, teachers, and headteachers who were interviewed or surveyed as part of the process evaluation corroborate these positive findings. They perceived that NELI had a positive impact on children's language skills while TAs interviewed as part of case studies commented that they observed improvements in the vocabulary of children in their narrative and story-telling skills, in their attention and engagement levels, and in children's confidence when communicating.

Training was received positively by schools. Survey responses indicate that schools felt informed about the demands of delivering NELI and that NELI training was seen to be appropriate enough to allow TAs and teachers to deliver the intervention. Attendance at initial training was also high: 78% of TAs surveyed reported that they attended both days of initial training while 83% of teachers surveyed attended their half-day. While attendance at top-up training was lower (with only 33% of TAs reporting to have attended), the vast majority (76%) of TAs surveyed made use of some form of follow up support in the form of telephone support, webinars, and online forums. There was variation in the number of sessions that schools delivered to pupils and schools did not necessarily deliver the maximum number of sessions. The vast majority of TAs surveyed believed that both group and individual sessions benefited pupils, however, TAs surveyed preferred group sessions and noted concerns around the time taken to deliver individual sessions. High levels of implementation fidelity, in the form of attending training and delivering a larger number of sessions, was associated with better language outcomes for pupils.


A majority of headteachers, teachers, and TAs surveyed did report that it was challenging to fit NELI into the school timetable. They also reported that TAs require several specialised skills, including the ability to communicate with children and build rapport, and the ability to make the intervention engaging, in order to implement NELI. Some barriers to implementation that were noted in previous evaluations of NELI were not found to be barriers in this trial. Most schools had space to deliver NELI and most TAs felt that they had the support of headteachers and teachers when delivering the programme. TAs were expected to spend time preparing and delivering NELI, and a majority of teachers surveyed believed that NELI increased TAs working hours. However, interestingly, only a minority of TAs and headteachers surveyed perceived that TAs working hours had increased.

A majority of those surveyed (TAs) and those interviewed (teachers and TAs) had concerns about the process of selecting children to participate in NELI: 80% of TAs surveyed felt that teachers or TAs should decide which pupils should enter NELI, rather than them being selected via the LanguageScreen assessment; 67% believed that some participating pupils were unsuitable for NELI, either because they were too advanced or had learning difficulties. However, given the results from the impact evaluation, it is important to stress that pupils selected for the intervention did, on average, benefit from it.

Cost

The cost of NELI for a single-form entry school where the programme is used over three years with five children per year is £58 per child; the corresponding figure for a two-form entry school is £43 (savings are achieved in larger schools through the sharing of materials). TAs are expected to spend five hours per week or a total of 100 hours to deliver NELI for 20-weeks to five children. Additionally, TAs are expected to spend 20 hours in training, while teachers also spend half a day in training.

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P Value	EEF cost rating
Primary outcome: language skills	0.26 (0.17, 0.35)	3		1,071	P < 0.05	£ £ £ £ £

Introduction

Background

The development of children's language skills is vital to their long-term success in education and in a variety of other areas. For example, there is a well-established link between the development of language skills and literacy skills (Whitehurst & Lonigan, 1998; Scarborough, 2009). Children with more advanced language skills at the age of five are more likely to have better qualifications and subsequently be employed in adulthood compared with their peers (Feinstein & Duckworth, 2006). Likewise, children who have difficulties with language at the age of five are more likely to experience mental health issues in later life (Schoon et al., 2010).

Research has shown that oral language, the foundation of which is developed by age four, impacts on children's development of literacy, numeracy, and formal education (Duff et al., 2015; Law et al., 2013; Roulstone et al., 2011). The need for effective early intervention is also a pressing policy concern. The persistent association between groups' socioeconomic background and their educational outcomes hampers efforts to improve social mobility (Social Mobility Commission, 2017). Children from more disadvantaged backgrounds are disproportionately more likely to be affected by language difficulties when they enter school (Fernald et al., 2013; Hart and Risley, 1995). Disparities emerge at an early age. Seventy-five percent of children aged five with persistent experiences of poverty score below average in vocabulary tests compared with only 35% of those with no experience of poverty (Finnegan & Warren, 2015). The persistent gap associated with socioeconomic background and reading achievement tends to widen throughout the school experience (Fielding et al., 2007). There are also compelling financial reasons to intervene to improve language skills earlier rather than later in children's development. The cost of intervening later in young adulthood rather than early in education has been estimated at over £650 million across England and Wales. This is only the fiscal cost, which does not take into account the impact on children and families at large (Chowdry & Fitzsimons, 2016).

Given the wealth of evidence to suggest that early intervention has great potential to narrow the gap, it is surprising that few nursery and reception year programmes have been rigorously tested for impact. One exception, however, is the Nuffield Early Language Intervention (NELI).

This effectiveness trial follows on from several previous studies of the intervention (see Table 2). A pilot study conducted by the programme developers took place in 19 pre-school settings in England in 2009–2010 (Bowyer-Crane et al., 2008). This initial version of the programme lasted 30 weeks (starting in nursery then continuing into reception) and was initially developed and evaluated with funding from Nuffield Foundation and training delivered by the developers. The initial trial 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).

The Education Endowment Foundation (EEF) then funded an efficacy study examining two versions of the NELI programme: a 30-week programme delivered in three ten-week blocks starting in the final term of nursery and continuing into the first two terms of reception year in primary school, and a 20-week programme delivered in two ten-week blocks only in reception year (Sibieta et al., 2016). More detail about this efficacy evaluation can be found in Table 2. The efficacy trial used a randomised controlled design with individual pupils within nurseries randomly assigned to treatment and control groups. The overall attrition rate was 11% due to three schools withdrawing from the intervention and 27 pupils in participating schools dropping out because they changed schools between nursery and reception. The efficacy trial found positive effects for both the 30- and the 20-week NELI interventions on the primary outcome (language). The 30-week intervention had an effect size of 0.27 (significant at the 1% level) while the 20-week intervention had an effect size of 0.16 (significant at the 10% level) on the primary outcome, language skills. In the efficacy study the impact of the 30-week version, starting in nursery, was larger than the impact of the 20-week version. However, this difference was not statistically significant. Pupils who received the short 20-week programme were found to have made an additional two months' progress and those who received the 30-week programme showed four months of additional progress. The estimated effect for both versions had increased at a follow-up test conducted six months after the end of the trial suggesting that the positive impact of the intervention may increase over time. By contrast, the impact on the secondary outcome, word-level literacy skills, was not significant in either version.

These first two studies differed from the present trial in aspects of sample size and design (for example, unit of randomisation, outcome measure, Table 2). The pilot study was implemented in 15 nursery schools and included 180 children. The efficacy trial compared the 30-week and the 20-week NELI versions and was implemented in 34 schools and included 350 pupils in the trial (114 pupils in the 30-week treatment group, 121 in the 20-week treatment group, and 115 in the control group) (Fricke et al., 2013; Sibieta et al., 2016).

The effectiveness trial discussed in this report used randomisation at the school level and included 193 schools. Furthermore, the previous trials delivered the NELI sessions in smaller groups of up to four children. In order to increase power, in this trial the five lowest-performing children were selected in each classroom.

This effectiveness trial has also tested whether a scalable version of the training and resources can produce similar results to the efficacy trial. For example, the previous trials identified eligible children through the use of one-to-one testing by trained speech and language therapists (Sibieta et al., 2016). This trial also differed in the use of screening tests, some aspects of the training, and in the selection of baseline and endline outcome measures. Furthermore, in previous evaluations schools that were relatively close to the project implementation team and based in Yorkshire and the South East of England were recruited. In this trial, schools have been recruited from 13 different regions with a wide geographical distribution.

Given the promising results of earlier evaluations, it is important to seek robust evidence from a larger sample and to understand if using a more scalable version of the training and resources can produce similar results. This study represents the largest evaluation of NELI to date.

Table 2: Comparison with previous NELI trials

	Effectiveness trial (discussed in this report)	Efficacy trial (Sibieta et al., 2016)	Pilot Study (Fricke et al., 2013)
Trial design, including number of arms	Two-arm	Three-arm	Two-arm
Programme variant	20-week (Reception)	20-week (Reception) 30-week (Nursery)	30-week (Nursery)
Unit of randomisation	School-level	Child-level	Child-level
Number of schools in trial	193	34	15
Number of children in trial (at randomisation)	1,156 585 (intervention) 571 (control)	394 117 (30-week intervention) 124 (20-week intervention) 119 (control)	180 90 (intervention) 90 (control)
Regions	13 different regions across England	2 regions	1 region
Stratification variable(s) (if applicable)	Region Single/multi-form entry	N/A	N/A
Screening measure	LanguageScreen	Language screening measures (based on CELF)	Language screening measures (based on the CELF)
Baseline testing measure	Latent language measure	Language composite score	Language; Narrative; Phonological Awareness; Literacy.

Primary outcome measure(s) ¹	Latent language measure	Language composite score	Language; Spoken narrative skills;
Secondary outcome variable	YARC Early word reading; LanguageScreen	Word-level literacy skills	Phonological awareness; Literacy skills
School dropout	0.5%	9%	None
Pupil dropout	7%	9%	8%

Intervention

Figure 2 sets out the implementation logic model developed during set-up meetings and the IDEA workshop between the Education Endowment Foundation (EEF), the University of Oxford (the delivery team), and RAND Europe (the evaluation team). A small revision to the logic model is discussed in the Conclusion. The implementation logic model (Figure 2) for NELI involves: (1) a combination of individual and group sessions with a focus on narrative, vocabulary, listening plus phonological awareness and letter sound knowledge. There is an element of (2) mandatory TA training, and therefore a need to understand how well the training has worked. In addition to training, TAs are offered (3) online support during delivery on an ad hoc basis. Training and appropriate support are key prerequisites for successful delivery. The NELI structure (4) enables TAs to improve their teaching of language skills. Implementing NELI sessions also means that (5) children will improve their oral language skills and (6) their early reading skills as learning to read builds on oral language skills. The logic model considers improvements in both oral fluency and early reading skills as essential elements for (7) improved reading comprehension.

Why: The Nuffield Early Intervention (NELI) is an intensive language support programme delivered by teaching assistants (TAs) and designed to improve the spoken language ability of young children with relatively poor spoken language skills. The sessions focus on improving children's vocabulary, listening, narrative skills, and phonological awareness. NELI aims to improve children's oral language skills and early word/emergent reading skills (in particular, pre-reading skills as defined by the delivery team) with the aim of ultimately improving children's reading comprehension (as learning to read builds on oral language skills) (please see Figure 2 for more details).

Who (recipients): The programme is targeted at children with poor spoken language skills. In total 193 schools were recruited in the trial and within this the programme was delivered in 97 intervention schools to the five lowest-achieving children in each reception class based on a screening test. Schools with two reception classes had up to ten participating children.

What (physical or informational materials were used in the intervention): The intervention is delivered using resources available from Oxford University Press.² These include handbooks with detailed lesson plans, picture cards, resources that can be photocopied to use in sessions, and a puppet to support session delivery.

What (procedures, activities are being used): Prior to NELI implementation, NELI tutors delivered **training** for teachers and other school staff involved in the evaluation. In contrast to previous trials, school staff were not trained by Oxford University but by national NELI tutors who were Ekklan trained Speech and Language Therapists specialising in oral language development. Ekklan tutors also offered **support** to school staff during the programme delivery period. The NELI training comprises:

¹ More detail on the outcome measures selected is provided in the Appendix E.

² <https://global.oup.com/education/content/primary/series/nuffield-intervention/?region=international>

1. A two-day training course for TAs (mandatory). The initial training for TAs was scheduled in two consecutive days. Following feedback from previous implementation of the intervention, teachers in reception also attended the first half-day of the initial training. The half-day training for teachers was introduced so that teachers can understand the importance of the intervention and support the TAs when required (for example, by letting the TAs work with the selected children during class time). Teachers played a useful indirect role in the intervention by offering support to TAs delivering the programme.
2. Online support (optional). During implementation, Elklan tutors maintained an ongoing relationship with participating schools by providing online support in the form of telephone support, email access, webinars, and online forum. Even though the online support was not compulsory, TAs were encouraged to access online support.
3. Half-day workshop (mandatory). The half-day workshop was delivered at the end of week nine or ten. The main aim of the workshop was to discuss the implementation in the following ten weeks (weeks 11–20 of NELI), which included an additional phonological awareness component. The half-day workshop was only targeted at TAs.

All NELI sessions followed a similar structure (see Figure 1).

The sessions encouraged active participation between children and the TA. Lessons are designed around activities that support and reinforce narrative, vocabulary, listening, letter-sound knowledge, and phonological awareness skills. At the beginning of the NELI group sessions TAs reinforce previously learned vocabulary and introduce new words in a context familiar to the children. Each session includes narrative tasks, which focus on skills such as storytelling by using the newly learned vocabulary. Topics for the chosen children were selected to fit in the existing curriculum for reception. Individual sessions do not follow a particular structure, rather they are tailored to the specific need of each participating child, but the logic of NELI applies with no school variation.

In the classroom, NELI was implemented as follows:

- An initial screening test (LanguageScreen) was taken by pupils; this assessment screens pupils for their oral language skills. The use of LanguageScreen was introduced after the efficacy trial with the aim of identifying the children with poor oral language skills in each classroom in a more efficient way. It replaces the use of the CELF test, a test administered by trained speech and language therapists. For more information on LanguageScreen see the Participant selection sub-section on Children.
- The five children with the lowest language skills were then selected to take part in NELI group and individual sessions. All NELI sessions follow a similar structure and use bright and accessible materials in order to make learning more enjoyable. TAs use rewarding techniques including informal (verbal) praise and more formal praise to encourage learning in the classroom.

How: NELI sessions were delivered within the school in a quiet area (such as class, library, staff room, or dining room). TAs had the freedom to decide what a suitable place for session delivery was. Training for school staff to deliver NELI took place on-site at the school and support was also available online via telephone, email, webinars, and online forum.

When and how much: NELI was delivered during the normal classroom hours but required children selected into the programme to be taken out of classes. School staff selected individually the most appropriate classes from which participating children were to be taken out of in order to take part in NELI sessions. The programme was delivered over 20 weeks in the spring and summer terms of 2019. Each week, teaching assistants held group and individual sessions. Each week, TAs delivered three group sessions to groups of five children lasting 30 minutes each. The five children selected to participate in NELI also attended two 15-minute individual sessions each week. Group sessions and individual sessions generally did not take place on the same day.

Who (providers, implementers): NELI was initially developed by researchers from the University of Oxford (led by Charles Hulme and Maggie Snowling) together with team members from Sheffield (Silke Fricke) and York (Claudine

Bowyer-Crane). Intervention development was funded by the Nuffield Foundation.³ The Oxford University team led implementation activities (training Elklan tutors, as below) and also provided input into the design of the evaluation. In contrast to previous trials, school staff were not trained by Oxford University but by national NELI tutors. National NELI tutors were Elklan-trained⁴ speech and language therapists specialising in oral language development.⁵ The delivery team held a day of face to face training for all NELI tutors (denoted as ‘Elklan trainers’ henceforth). Following the training, support was available from the delivery team as required to answer questions from the Elklan trainers. The answers to the questions were provided to all Elklan trainers by email. Prior to NELI implementation, Elklan trainers were responsible for delivering training for teachers and other school staff involved in the evaluation. Elklan trainers also offered support to school staff during programme delivery period.

Strategies to maximise effective implementation: Following feedback from the efficacy evaluation, where there was low fidelity due to lack of support and failure to protect TAs time for programme delivery (Sibieta et al., 2016), a number of strategies were introduced to ensure effective implementation. First, the half-day training for teachers was introduced with the main aim of helping teachers understand the importance of the intervention and supporting the TAs required for delivery (for example, by letting the TAs work with the selected children during class time). Furthermore, Elklan trainers maintained an ongoing relationship with participating schools by providing ongoing support when needed.

Evidence of implementation variability: The intervention was delivered as planned with no significant variation across treatment schools. The only exception related to schools that withdrew from the study (specific details are provided in the Participant selection). Fidelity was assessed as part of this trial and will be discussed separately in the Impact evaluation section.

Figure 1: NELI session structure

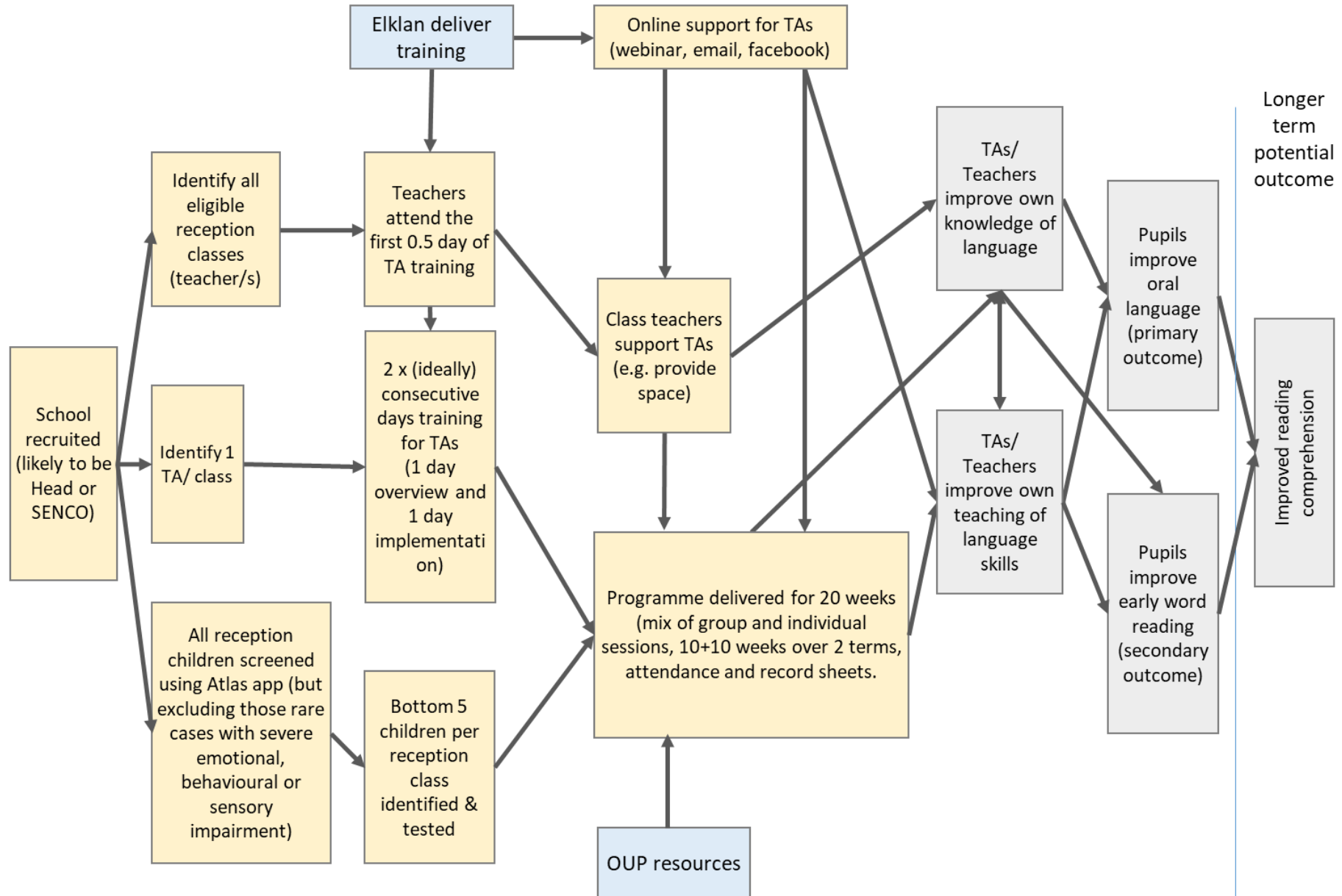
Session structure	Focus	Time (minutes)
Introduction: greeting, listening rules, and game	Introducing the child into the session	3
Reinforce vocabulary learning	Activities designed to reinforce vocabulary	5
New vocabulary	Exploring new words and talking about their meaning	5
Narrative	Support narrative and extended language skills including storytelling and retelling, and summarising stories	10
Plenary	Reviewing session content and closing session	2

³ <http://www.nuffieldfoundation.org/>

⁴ More information about Elklan can be found here: <https://www.elklan.co.uk/about>

⁵ Please note Elklan trainers are different from Elklan testers. Elklan testers were not involved in NELI delivery. All testers were blinded to allocation.

Figure 2: NELI Logic Model



Evaluation objectives

The aim of the evaluation was to assess the extent to which NELI leads to an improvement in language skills (see Outcome measures) when comparing NELI schools to business-as-usual schools. A full evaluation protocol may be found on the EEF's website (Sutherland et al., 2019).

The hypotheses tested in this trial are:

1. Hypothesis 1: The Nuffield Early Language Intervention will have a *positive* effect on the oral language outcomes of children who are in a randomly assigned intervention group compared to children in control schools.
2. Hypothesis 2: The Nuffield Early Language Intervention will have a *positive* effect on the early word reading of children who are in a randomly assigned intervention group compared to children in control schools.

Additional to the protocol and the SAP, the trial also tested the following hypothesis:

3. Hypothesis 3: The Nuffield Early Language Intervention will have a *positive* effect on the language outcomes of children who are in a randomly assigned intervention group compared to children in control schools, as assessed through the LanguageScreen measure.

In the SAP, the LanguageScreen measure analysis was included as an exploratory analysis only. This was considered appropriate as the Language Assessment is a yet-to-be-validated test developed by the delivery team. However, at the request of the developers, this was subsequently upgraded to a secondary outcome analysis and reported as such here. From a theoretical point of view it was considered appropriate to have a measure of language comprehension included as part of one of the outcomes of the trial. LanguageScreen included such a measure, focusing on listening comprehension. Further information on the LanguageScreen app and test can be found in the Secondary outcome measures section.

Ethics and trial registration

The trial has been registered on the ISRCTN (International Standard Randomised Controlled Trial Number) registry, which is used to describe Randomised Controlled Trials (RCT) and efficacy trials at inception. The trial has been assigned an ID registration number: ISRCTN12991126.

The ethics and registration processes have been in accordance with the ethics policies adopted by the University of Oxford and RAND Europe. Ethical approval for the intervention was granted by the Departmental Research Ethics Committee (DREC) within the Department of Education at the University of Oxford. The reference number for this approval is ED-CIA-18-192. Additionally, the evaluation has been reviewed and approved by RAND U.S. Human Subjects Protection Committee (HSPC).

School recruitment was carried out by the University of Oxford. All eligible schools were invited to take part in the evaluation. Before pupil data was sent to the delivery team, parents were sent an information sheet and opt-out forms by the school and parents or legal guardians had the right to opt children out of the trial at any point. The delivery team was responsible for ensuring distribution of the information sheet and collection of withdrawal forms. The forms were sent out to parents by the school after the school representatives had signed the MoU describing what was involved in the trial. The Methods section provides further detail about school recruitment. The recruitment documentation, including a Memorandum of Understanding (MoU), Information Sheet for parents, and withdrawal form, are provided in Appendix D.

RAND Europe collected consent forms for headteachers, teachers, and TAs who volunteered to participate in an interview. Furthermore, the cover page for each questionnaire survey collected as part of the study contained an informed consent sheet and data protection statements for respondents. It informed respondents that participation in the survey was entirely voluntary. Also the surveys did not collect personal identifying information such as the respondent's name, date of birth, or contact details.

None of the members of the evaluation team had any conflicts of interest in undertaking this evaluation.

Data protection

This project was set up in 2018, prior to GDPR. Therefore, in terms of fair processing of personal data, at the time the project fulfilled Condition 1 of processing personal data in Schedule 2 of the Data Protection Act (DPA) as the data subjects gave their implicit consent to participate in the study via an opt-out letter to parents at the beginning of the trial.

As currently applicable under GDPR, the delivery and evaluation teams obtained personal data from schools and pupils as data controllers. Any data sharing required was governed by the data sharing agreement signed between the funder (the EEF), the delivery team (University of Oxford), and the Evaluation team (RAND Europe). RAND Europe obtained baseline data under a data sharing agreement with the University of Oxford. RAND Europe obtained pupil outcome data from its subcontractor (Elklan) that acted as a processor pursuant to appropriate data sharing terms in its subcontract. Data obtained by Elklan was requested on the legal basis of legitimate interests. The request for data fell under the General Data Protection Regulation (GDPR), as implemented by the Data Protection Act (2018).

The evaluation team undertook measures to ensure the trial was GDPR compliant. RAND Europe adopts good industry practices regarding the protection of personal data as part of its obligations as a 'data controller' under the Data Protection Act 2018 and takes appropriate technical and organisational measures to protect personal data. Individuals targeted by the study have the right to oppose, have access to, rectify, or remove personal or sensitive personal data held by RAND Europe. In order to ensure GDPR compliance, all data has been only saved on GDPR-compliant, secure servers inside the EEA or U.K. RAND Europe is registered with the Information Commissioner's Office (ICO), registration number Z6947026 and is certified for adhering to ISO 9001:2015 quality management practices.

RAND will delete all data one year after the project ends. For the purpose of research, following the completion of the trial, the data will be shared with the EEF archive, at which point the EEF will become the data controller.⁶ The data will be shared with the EEF's archive manager and, in an anonymised form, with the Office for National Statistics and potentially other research teams. Further matching to NPD and other administrative data including KS1 may take place during subsequent research.

Project team

The intervention was developed and implemented by the University of Oxford. Elklan was a delivery partner, with Elklan trainers being trained by the delivery team and then training school staff. The overall responsibility for training school staff and providing support to all schools as per the terms of the sign-up information sheet ultimately rested with the Oxford team. The Oxford team was also responsible for organising and administering baseline testing. In addition, they provided the key indicators of implementation and compliance with the intervention. The delivery team at Oxford also provided input into the design of the trial. The delivery team at Oxford was led by Professor Charles Hulme and Dr Gillian West and also involved Professor Maggie Snowling, Ms Denise Cripps, and Dr Elizabeth Worster.

The evaluation was conducted independently by RAND Europe. The evaluator was responsible for the outcome and process evaluation, trial design (with input from the Oxford team based on prior NELI evaluation experience, and, in response to a prior NELI trial protocol by Dunne and Miller, 2017), analysis, reporting, and quality assurance of the evaluation. The Oxford team initially pre-registered the trial independently of the evaluator; this registration was then amended by the independent evaluator to reflect the final design. The evaluation team in RAND Europe comprised: Dr Sonia Ilie (current project leader), Dr Sashka Dimova (project management, field work and analysis), Miriam Broeks, Dr Andreas Culora, Elena Rosa Brown, Eleftheria Iakovidou, Dr Susie Lee, Natalie Picken (analysis), and Dr Alex Sutherland (previous project leader, formerly RAND Europe). The evaluation team also benefitted from advice and quality assurance from Dr Monica Melby-Lervåg and Dr Arne Lervåg (University of Oslo).

⁶ For more information about the EEF archive, please see:
https://educationendowmentfoundation.org.uk/public/files/Evaluation/Data_protection/Data_protection_statement_EEF_evaluation.pdf

Methods

Trial design

The trial was planned and executed as a stratified, two-arm, cluster-randomised controlled trial (cRCT) across 193 schools and a total of 240 reception classrooms recruited from thirteen geographical regions: Bristol, Cornwall, Durham, Essex, Hertfordshire, London, Manchester, North Tyneside, Northamptonshire, Surrey, Warwickshire and surrounds, Wolverhampton, and the North West of England. The areas where recruitment took place were selected based on available contacts, and to ensure an equal spread across urban and rural regions. All children in reception classes were eligible for inclusion in the trial (selection criteria are outlined below). A summary of the cRCT design can be found in Table 3.

Schools were assigned to either treatment (NELI) or control (business as usual). All schools signing up had a 50:50 chance of being assigned to the treatment group within each geographical cluster (see Randomisation section). All TAs in reception classrooms in treatment schools were eligible to receive training in the intervention.

There is only one treatment condition in this trial: that throughout the 2018/2019 academic year, schools have received training and support to implement NELI. For control schools, it was business as usual during the school year 2018/2019. As an incentive, control schools received financial compensation of £1,000. They received a payment of £500 post-randomisation and £500 upon completion of outcome testing.

Initially, every child in the participating reception classrooms was assessed by the school staff using LanguageScreen. Based on the score derived from the use of LanguageScreen, the five children in each participating reception classroom with the lowest composite language scores were selected for inclusion in the intervention and further testing. The five children were tested twice, at pre-test (before randomisation) and post-test (immediately after the programme ended). All measures selected were standardised, age appropriate, and were selected by the evaluation and delivery team in collaboration with language experts.

Table 3: Trial design

Trial design, including number of arms		Two-arm, stratified and cluster-randomised
Unit of randomisation		School
Stratification variable(s) (if applicable)		Geographic area, number of classes
Primary outcome	Variable	Improved oral language
	Measure (instrument, scale, source)	Language latent variable created from four individually administered language tests: 1. CELF recalling sentences subtest 2. CELF expressive vocabulary subtest 3. Renfrew Action Picture Test (information) 4. Renfrew Action Picture Test (grammar)
Secondary outcome(s)	Variable(s)	Improved word reading Language latent variable score
	Measure(s) (instrument, scale, source)	YARC Early Word Reading test; LanguageScreen
Baseline for primary outcome	Variable	Improved oral language

	Measure (instrument, scale, source)	Language latent variable created from four individually administered language tests: 1. CELF recalling sentences subtest 2. CELF expressive vocabulary subtest 3. Renfrew Action Picture Test (information) 4. Renfrew Action Picture Test (grammar)
Baseline for secondary outcome(s)	Variable	Improved word reading; Language latent variable score (from LanguageScreen)
	Measure (instrument, scale, source)	YARC Early Word Reading; and respectively The LanguageScreen app assessment

Participant selection

Schools

Schools meeting the following criteria were eligible for inclusion in the study:

- has not previously delivered NELI;
- has above average free school meal (FSM) eligibility;⁷
- is willing to be randomly assigned to intervention or 'business as usual' at the level of the school;
- is willing to engage with the intervention and implement it with the pupils identified by screening;
- is willing to provide child background information to the evaluation team; and
- is willing to facilitate baseline and post-intervention data collection.

Eligible schools were accepted into the trial once they had completed the required paperwork and prerequisite tasks, and after signing a Memorandum of Understanding (MoU). The MoU outlined the roles and responsibilities of all stakeholders involved. Schools' responsibilities and associated tasks were:

- distributing an information sheet, privacy notice and opt-out form to parents/carers, and collecting any opt-out forms that were returned;
- conducting the initial app-based testing of all children in participating reception classes and repeating this testing at the end of the intervention;
- accommodating and facilitating the child assessments being carried out by the Research Team at the beginning and end of the study;
- assisting the Research team in identifying any children who should be excluded from the trial according to specified criteria; and
- consenting to approximately 14 at allocation.

The MoU specified that once schools agreed to participate, the expectation was that final outcome testing of children would be allowed, even if the school were to withdraw from the intervention. All schools that fulfilled the inclusion criteria and provided the paperwork and prerequisite tasks were eligible for randomisation.

The delivery team had responsibility for school recruitment. Based on prior knowledge of likely recruitment level, they approached around 1,100 schools in the summer of 2018. A total of 207 schools expressed an interest to be involved in the NELI trial. Of these 207 schools, 14 pulled out of the trial prior to randomisation for a variety of very acceptable reasons ranging from building work and flooding to a change of headteacher. After these withdrawals, schools were issued by the delivery team with a MoU outlining the details of their involvement in the trial; 193 schools returned the signed MoUs and were therefore eligible for randomisation. All school recruitment documentation, available in ORandomisation, was carried out by the evaluation team after baseline data collection (see below). The evaluation team then communicated the allocation result to the delivery team, which passed it on to participating schools.

⁷ Around 13.7% pupils were eligible for FSM in 2018.

Of the 193 schools, 88 indicated on the MoU that they wished to enter more than one class. The implementation team provided schools with a clear description of the time commitment and costs related to enrolling more than one class to encourage schools to think about this issue before the start of the programme as opposed to during the intervention. Forty-five of these schools chose to reduce their allocation. Forty-one schools entered two classes and three schools entered three classes each. As a result, there were 240 classes enrolled on the trial with over 6,000 children considered eligible to take part. A breakdown of the number of schools and classes by geographical area is listed in Table 4.

Table 4: Geographical areas included in selection

Region	Schools	Classes
Bristol	15	19
Cornwall	20	22
Durham	17	22
Essex	22	25
Herts	10	15
London	11	14
Manchester	7	9
North Tyneside	9	11
North West*	10	11
Northamptonshire	19	24
Surrey	30	37
Warwickshire and Surrounds	10	14
Wolverhampton	13	17
Total	193	240
*Note: North West is Blackpool, Thornton-Cleveleys, Poulton-le-Fylde, and Lytham St Annes.		

Children

All children in reception classrooms within eligible schools, 6,000 children overall, were included in the screening process. Teachers had the option to exempt children with deafness, vision problems, or severe emotional and behavioural difficulties from screening. Children in the participating reception classes took part in the screening during which their oral language skills were measured using the LanguageScreen.

LanguageScreen is a language test (and accompanying digital application for the administration of the test) developed by the delivery team for use with NELI that assesses four core language skills: expressive vocabulary, receptive vocabulary, sentence repetition, and listening comprehension. LanguageScreen then generates a composite language score. LanguageScreen is administered one-to-one with children and takes approximately ten minutes to complete. Responses are scored using an app, which is run on a tablet. For each item of each test, children were scored as correct or incorrect without any discontinuation rules. The scores from the four subtests are combined in a composite score to give an estimate of each child's overall language ability.

The LanguageScreen app was in beta version at the time of the trial and full validation of the LanguageScreen measure has not yet been published.

For the purposes of this trial, the LanguageScreen test was administered by reception teachers to all eligible pupils in participating schools by a member of staff such as a teacher or TA. A composite score was generated by LanguageScreen based on the four sub-tests that make up the assessment. In those cases where there were missing subtest scores, the summed z-score for the available tests was used to form the basis of the eligibility ranking.

Based on the composite language score from LanguageScreen, the five children in each participating reception classroom with the lowest scores were selected for participation in NELI and for further baseline testing discussed below.

LanguageScreen was subsequently used post-intervention to assess children's language skills. The post-intervention assessment was administered to all reception pupils in all participating schools (in both control and intervention schools).

Outcome measures

Existing research has established that language (including expressive and receptive vocabulary), grammar, and oral skills develop as a unitary construct (McDonald, 2013). Therefore, the primary outcome is a language latent variable, measured through four separate tests, each capturing different but ultimately related aspects of language. Across screening, baseline, and outcome measures, all trial measures related to language.

The selected tests were standardised, age-appropriate, and selected by the evaluation and delivery team in collaboration with language development experts (see Outcome measures section). The selected tests have an adequate level of reliability and sensitivity^{8,9,10} and were administered in their entirety, in the same order for all pupils.

In-depth testing for trial participating children was administered by Elklan, by trained testers who were speech and language therapists. Testers attended one day of training before both baseline and outcome data collection. This tester training included information on each of the tests to be used and administration procedures, including how to upload assessment data as well as practice sessions, and was led by the delivery team.

All testers were blind to allocation. Testing was carried out individually (one-to-one) in school and the tests were administered in a fixed order. Testing time varied: on average the in-depth testing took between 15 and 30 minutes to complete at baseline and again the same at endline.

For both baseline and outcome testing, Elklan supplied RAND Europe with raw scores. These raw scores then underwent a measurement development procedure, consistent across baseline and outcome testing, which is described in the Statistical analysis section below.

Baseline measures

In-depth baseline assessments (for trial participating pupils only) consisted of the following individually administered tests:

- the Clinical Evaluation of Language Fundamentals (CELF) recalling sentences subtest (Wiig et al., 2006)—a test where children are asked to repeat a sentence back to the person carrying out the assessment;
- the CELF expressive vocabulary subtest—where children are asked to name objects or actions depicted in a set of images;
- the Renfrew Action Picture Test (RAPT) information sub-test—a sub-test of RAPT where children are asked to describe the information shown in a set of pictures; and
- the Renfrew Action Picture Test (RAPT) grammar sub-test (Renfrew, 2016)—a sub-test of RAPT that checks the grammar used by children, such as the use of verb tenses, while describing the information shown in a set of pictures (as part of the RAPT information test above).

The RAPT

Taken together, these four tests capture overall language skills and together are used to derive the baseline measure for the main outcome analysis in this trial through a confirmatory factor analysis as part of a structural equation modelling procedure outlined in the Measure development section.

⁸ CELF-Preschool 2UK: test-retest reliability ranged from 0.77 to 0.96 (for ages 3 to 3,11) and 0.74 to 0.95 (for ages 4 to 4,11). Cronbach's alphas ranged 0.77 to 0.95. Inter reliability ranged 0.95 to 0.97.

⁹ RAPT reliability: random testing revealed little difference in responses if the retest was given within a month from the original. A 3% of each set's score was affected by scoring discrepancy.

¹⁰ YARC is one of the most sensitive tests. Inter-rater reliability for early word recognition is 0.98.

Additionally, the following test was also administered:

- the York Assessment of Reading for Comprehension (YARC) early word reading test (Snowling et al., 2009)—a test that assesses pupils' single-word reading ability.

Baseline data was collected prior to randomisation.

Primary outcome measure

A post-test to assess overall language skills was administered to children immediately after the end of the programme in intervention schools. The following individually-administered tests made up the post-test assessment:

- the CELF Preschool 2UK: Recalling sentences;
- the CELF Preschool 2UK: Expressive Vocabulary;
- RAPT information; and
- RAPT grammar.

These are the same tests used at baseline (see the Baseline measures section for more information on the measures). Outcome data collection was completed by Ekklan testers who were not involved in the trial itself as delivery trainers. At no point did the evaluation team provide Ekklan testers with information about the treatment allocation of their schools. The primary outcome measure was constructed as an amalgamation of the four post-tests, again using confirmatory factor analysis as part of a structural equation modelling approach outlined in the Measure development section below.

Secondary outcome measures

Two secondary outcome measures were captured as part of this study in addition to the primary outcome measure:

- firstly, a previously validated measure, the YARC Early Word Reading test, was used to measure children's early reading skills;
- secondly, LanguageScreen was also administered to all children by the reception teachers once the intervention had completed in intervention schools.

The YARC test was scored simply as a number of correct answers to the questions (that is, words read correctly).

For the purposes of the trial analysis, LanguageScreen underwent an SEM analysis before being included in the final analysis models. This was separate from the same assessment deriving a composite score for the purposes of screening, and supported the independence of the data analysis.

Appendix E provides a summary of all the tests and subscales used in the evaluation.

Measure development

Structural equation modelling (SEM) is a broad statistical technique that is routinely used to explore the measurement structure of specific theoretical constructs as well as complex relationships between concepts (Hox et al., 1999). It encompasses statistical techniques such as factor analysis, regression analysis, and path analysis.

The main advantage of SEM over other analytical techniques is that it allows for the exploration of latent variables. A latent variable is defined as a substantive theoretical construct that is only measured indirectly, with several indicators all capturing different aspects of that construct (Muthén, 2002). SEM allows both for the inclusion of latent variables within other statistical models (for instance, as an outcome in a linear regression model) and for the estimation of a so-called factor score, which combines all underlying indicators making up the respective latent variable and their respective contributions to that variable.

This approach was used separately for the baseline measures, and for the outcome measures.

For the purposes of this trial, the latent variable is *language skills*, with each of the four tests respectively making up the baseline and outcome measure contributing to its measurement as separate continuous variables:

- The CELF Preschool 2UK: Recalling Sentences (RS)—this test is scored out of 37, as a sum of all points attached to correct (or partially correct) answers to the 13 test questions potentially shown to children.
- The CELF Preschool 2UK: Expressive Vocabulary (EV)—this test is scored out of 40, as a sum of all points attached to correct answers to the 20 test questions potentially shown to children.
- RAPT information (RAI)—this test is scored out of 40, as a sum of all points attached to correct answers to test questions, which vary between the ten picture items shown to children.
- RAPT grammar (RAG)—this test is scored out of 38, as a sum of all points attached to correct answers to test questions, which vary between the ten picture items shown to children.

A confirmatory factor analysis model was then separately estimated for the baseline and the outcome measure respectively under the SEM framework. All SEM analysis was carried out in Stata 15 using the `sem` command.

SEM is an iterative estimation procedure that allows for the confirmatory factor analysis model to be modified as long as the overarching theoretical assumptions about any relationships between indicators still hold. In SEM and confirmatory factor analysis, model fit is assessed by means of a series of fit indices.

Rules of thumb as to what constitutes good or acceptable fit exist, usually indicated by cut-off values for each respective fit index. In particular, the most often reported indices of fit are the Comparative Fit Index, CFI, which compares the fit of a target model to the fit of a null (or independent) model, with good fit indicated by values above 0.9, and the Root Mean Square Error of Approximation, RMSEA, which adjusts for model parsimony, with good fit indicated by values below 0.08.

Additionally, a chi-squared statistic, which assesses the overall fit and discrepancy between the sample and the fitted covariance matrices, is also usually reported. Non-significant values are indicative of good model fit although established methodological literature suggests that at high samples (like the samples in this trial), the chi-squared statistic will almost always be significant. Indeed, there is some disagreement in the field as to cut-off points and the indices to be reported; literature (for example, Hooper et al., 2008) has explored in simulation studies both how fit indices are liable to lead to misinterpretation of model fit (Kenny, 2015) or can sometimes disagree with each other (Lai and Green, 2016).

This means that SEM (and confirmatory factor analysis) measurement models are flexible, but also that the driving force behind any changes to the most parsimonious model in search of good model fit must be robust theoretical assumptions.

The main theoretical assumption relevant for the development of the language skills measures here is that the four underlying indicators (coming from the four tests above) are likely to be correlated to each other, seeing as they capture aspects of the same latent construct, language skills. As a result, it is reasonable to include such correlations between the four variables in the SEM analysis at both baseline and endline, respectively. Full details of the SEM models used are presented in Appendix G.

Sample size

The initial power calculations were based on the information provided in the Invitation to Tender and the subsequent set-up meeting with the delivery team and the EEF. It was assumed that there were on average five pupils per school who would be eligible for the programme and that with 200 schools, we would be collecting data on an estimated 1,000 pupils for this evaluation.

Estimation of ICC

The pre-intervention ICC was estimated from LanguageScreen data from both control and intervention schools by using the `loneaway` command in Stata. The data was available post-randomisation thus our original power calculations were not based on the ICC derived from the screening data (for more information please look at Table 5). We have not made any adjustments when calculating the ICC as adjusting for covariates leads, in general, to smaller ICCs. Based on LanguageScreen data, the ICC within schools is 0.117 with 95% confidence intervals of 0.089 to 0.145. The ICC for the main analysis is based on the factor score derived from the language latent outcome measure and is 0.349, based on the 192 schools in the trial and analysis.

MDES calculations

At protocol stage, our initial power calculations, based on the parameters from the previous efficacy trial of NELI, assumed that 50% of level 1 variance was explained by pre-test scores. The assumed level 2 (schools) explained variance was zero. We also assumed an alpha of 5% and an intended 80% power to detect effects. We used two-level clustered designs and based our calculations on two values for the intra-cluster correlation (ICC) (15% as per the initial trial design and 11% to show the impact of reducing ICC). Power and minimum detectable effect size (MDES) calculations were performed using PowerUp! (Dong and Maynard, 2013). Using the parameters set in the study protocol and with equal allocation to treatment and control, the MDES was evaluated at d equal to 0.193 for ICC of 15% and 0.178 for ICC of 11%.

At randomisation there were 193 schools and an average of just over six pupils per school in the trial. As above, the school-level ICC from LanguageScreen was estimated at 11.7%. With the achieved sample and ICC of 11.7%, the MDES estimated for the overall trial at randomisation was $d = 0.176$.

At analysis stage, 192 schools provided data, with an average of 5.7 children per school. The school-level ICC (explanation below) derived from the primary outcome measure was 34.9%. This resulted in an at-analysis MDES of 0.257, larger in size than both the protocol (MDES = 0.193 for an assumed ICC = 15%; Sutherland et al., 2019), and also larger than at randomisation.

As per the SAP, the study is not powered for meaningful subgroup analysis. The MDES reported in Table 5 for the EAL subgroup is estimated for the EAL-only sample (that is, as a separate subsample). Estimating the MDES for the EAL subgroup using the same parameters but under a moderator framework (estimated in PowerUp! as the primary MDES) would yield a smaller MDES of 0.192 because in the moderator power calculations for a cluster-randomised trial with a binary moderator at the individual level, the between-school variance does not impact the power calculations.¹¹ All MDES calculations were undertaken in PowerUp!¹²

Table 5: Sample size calculations

		Protocol	Randomisation	Analysis	
		Overall	Overall	Overall	Subgroup: EAL
MDES		0.193	0.176	0.257	0.338
Pre-test/ post-test correlations	level 1 (pupil)	0.71	0.71	0.75	0.75
	level 2 (class)	NA	NA	NA	NA
	level 3 (school)	0	0	0	0
Intracluster correlations (ICCs)	level 2 (class)	NA	NA	NA	NA
	level 3 (school)	0.15	0.117	0.349	0.349
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		5	6.13	5.6	5.6 (2.9 EAL)
Number of schools	intervention	100	97	96	96 (61 EAL)

¹¹ Spybrook, J., Kelcey, B., and Dong, N. (2016) 'Power for Detecting Treatment by Moderator Effects in Two- and Three-Level Cluster Randomized Trials', *Journal of Educational and Behavioral Statistics*, 41 (6), 605–627.
<https://doi.org/10.3102/1076998616655442>

¹² Dong, N. and Maynard, R. (2013) 'PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasiexperimental design studies', *Journal of Research on Educational Effectiveness*, 6, 24–67. doi:10.1080/19345747.2012.673143

Number of pupils	control	100	96	96	96 (64 EAL)
	total	200	193	192	192 (125 EAL)
	intervention	500	585	530	530 (176 EAL)
	total	1000	1156	1071	1071 (359 EAL)

Randomisation

Allocation to treatment and control schools was conducted in Stata by RAND Europe on 2 November 2018 and included 193 schools. Schools were recruited by the delivery team. Randomisation took place after the baseline testing was completed or scheduled to take place, but the delivery team was not informed of allocation until after baseline testing had been formally completed. Schools were notified of their allocation on 19 November.

In preparation for randomisation, we examined the distribution of schools by region and number of classes in school, concluding that stratification by region and number of classes would be appropriate. Randomisation was therefore stratified by (1) geographical location and (2) the number of classes being put forward for intervention by the school. The latter was to ensure that schools with higher numbers of pupils were not allocated to treatment or control unevenly.

Strata were constructed from regions (geography-based strata, with the thirteen regions described in Table 4) and from number of classes within a school (one entry school vs. multiple-entry schools). Having uneven numbers of schools with one and/or multiple number of classes would mean there is a higher probability that treatment or control groups would be unequal in terms of size.

To deal with unequal treatment fractions we used the command `randtreat` and the option `misfits(global)` in Stata (Carril, 2017). No other randomisation procedure was used. The randomisation script is included in Appendix H.

Table 6 below shows actual allocations by region and prior attainment. In total 96 schools and 121 classes were allocated to the control condition, while 97 schools with total of 119 classes were allocated to the intervention condition.

Table 6: Allocation of schools to treatment or control by region and prior attainment

Region		Allocation			
		Control		NELI school	
		N schools	N classes	N schools	N classes
1	Bristol	7	8	8	11
2	Cornwall	10	11	10	11
3	Durham	9	12	8	10
4	Essex	11	13	11	12
5	Herts	5	8	5	7
6	London	5	6	6	8
7	Manchester	4	5	3	4
8	North Tyneside	4	5	5	6
9	North West*	5	5	5	6
10	Northamptonshire	10	13	9	11
11	Surrey	15	19	15	18
12	Warwickshire and Surrounds	5	8	5	6
13	Wolverhampton	6	8	7	9

	Total	96	121	97	119
* North West is Blackpool, Thornton-Cleveleys, Poulton-le-Fylde, and Lytham St Annes.					

Statistical analysis

The outcome analysis was undertaken on an intention-to-treat (ITT) basis. The analysis included all randomized schools in the groups to which they were randomly assigned initially, regardless of the treatment they actually received, withdrawal post-randomisation, or any deviations in programme implementation. This principle is essential to ensuring non-biased intervention effect estimation and compares outcome means for the treatment and comparison groups as they resulted from the randomisation procedure. Therefore, the ITT approach is inherently conservative as it captures the averaged effect of offering the intervention, regardless of whether or not the participants comply with assignment.

Analysis was not blind to allocation, however analysis bias was minimised by the provision of explicit Stata analysis code (and output) and the later depositing of data for potential future replication work. Full details of the analysis code and relevant output are included in Appendix H.

Primary analysis

The primary goal was to examine whether participation in NELI produced improvements in language, as measured with an underlying latent variable. Therefore, the primary outcome analysis was based on structural equation modelling (SEM) whereby a latent language construct will represent the main outcome measure as well as baseline measures. The latent construct will be evaluated across the two points in time.

The starting point of the SEM analysis was the development of a conceptually-driven measurement model that describes the underlying measurement structure of the language construct. The measurement model will be assessed for fit (how well the observed variables combine to capture the underlying language construct) by using the recommendations of Hu and Bentler (1999) for cut-off criteria for fit indexes to evaluate model fit (Hu and Bentler, 1999).

This procedure was applied to both the pre-test and post-test language construct. Once model fit was established, a factor score was generated for both the pre-test (baseline) and post-test (outcome) measure separately. Each of these two factor scores therefore captured the language latent variable in such a way that it could be included in the EEF-preferred multilevel model specification, as follows.

The primary outcome analysis consisted of model (1) below, for outcomes of pupils nested in schools:

$$Y_{ij} = \beta_0 + \text{NELI}_j\tau + Z_j\beta_1 + X_{ij}\beta_2 + u_j + e_{ij} \quad (1)$$

where Y_{ij} is the language factor score for child i in school j ; NELI_j is a binary indicator of the school assignment to intervention [1] or control [0]; Z_j is a vector of school-level characteristics, here the two stratifying variables of geographical location and number of classes per schools (as used for randomisation); X_{ij} represents characteristics at pupil level (pupil i in school j), specifically the baseline language factor score; u_j are school-level residuals ($u_j \sim i.i.d N(0, \sigma_u^2)$) and e_{ij} are individual-level residuals ($e_{ij} \sim i.i.d N(0, \sigma_e^2)$). Equation (1) is known as a 'random intercepts' model because $\beta_{0j} = \beta_0 + u_j$ is interpreted as the school-specific intercept for school j and $\beta_{0j} \sim i.i.d N(\beta_0, \sigma_u^2)$ is random (that is, it is assumed to be random).

Our target parameter, and the coefficient upon which the main result of the trial draws, is τ , which is the average effect of the intervention on pupil outcomes in intervention schools compared to those in control schools after accounting for the effect of the baseline assessment and the two stratification variables.

All analyses were performed in Stata, versions 15.1 onwards (Heß, 2017).

Secondary analysis

The protocol and statistical analysis plan set out two secondary outcome measures: first, an early word reading score derived from the YARC test, and second a language score derived from LanguageScreen. The LanguageScreen measure was not included in the protocol and statistical analysis plan as a secondary outcome measure but added in at the request of the delivery team as a secondary outcome in this trial. The results of these analyses are presented in turn for the two outcomes in the Impact evaluation section.

To estimate the impact of NELI on language captured through the YARC test we used a two-level multilevel model to account for clustering of pupils in schools. This is in line with the EEF's stated preference for using multilevel modelling for clustered trial designs. The multilevel modelling framework can flexibly handle complex variation within/between schools (Hox, 1998; Snijders, 2005; Snijders and Bosker, 1994).

This secondary outcome analysis consisted of model (2) below, for outcomes of pupils nested in schools, which matches the approach for the primary outcome analysis.

$$Y_{ij} = \beta_0 + \text{NELI}_j \tau_{YARC} + Z_j \beta_1 + X_{ij} \beta_2 + u_j + e_{ij} \quad (2)$$

where Y_{ij} is the YARC score for child i in school j ; NELI_j is a binary indicator of the school assignment to intervention [1] or control [0]; Z_j are school-level characteristics, here the two stratifying variables of geographical location and number of classes per schools (as used for randomisation); X_{ij} represents characteristics at pupil level (pupil i in school j), specifically the baseline YARC assessment; u_j are school-level residuals ($u_j \sim i.i.d N(0, \sigma_u^2)$) and e_{ij} are individual-level residuals ($e_{ij} \sim i.i.d N(0, \sigma_e^2)$).

For this secondary outcome analysis τ_{YARC} is the average effect of the intervention on pupil outcomes in treatment schools compared to those in control schools when using the YARC measures for pre-test and post-test. This constitutes a secondary outcome of the trial.

To estimate the impact of NELI on language captured through LanguageScreen, we also used a two-level multilevel model, as per the above specification.

$$Y_{ij} = \beta_0 + \text{NELI}_j \tau_{ATLAS} + Z_j \beta_1 + X_{ij} \beta_2 + u_j + e_{ij} \quad (3)$$

where Y_{ij} is the LanguageScreen score for child i in school j ; NELI_j is a binary indicator of the school assignment to intervention [1] or control [0]; Z_j are school-level characteristics, here the two stratifying variables of geographical location and number of classes per schools (as used for randomisation); X_{ij} represents characteristics at pupil level (pupil i in school j), specifically the baseline LanguageScreen score; u_j are school-level residuals ($u_j \sim i.i.d N(0, \sigma_u^2)$) and e_{ij} are individual-level residuals ($e_{ij} \sim i.i.d N(0, \sigma_e^2)$).

For the LanguageScreen secondary outcome analysis, τ_{ATLAS} represents the average effect of the intervention on pupil outcomes in treatment schools compared to those in control schools when using the measure for pre-test and post-test. This constitutes a secondary outcome of the trial. We note that due to different timing of data collection between the primary outcome measure and the LanguageScreen measure at outcome testing, data for the LanguageScreen measure is available for a smaller sample of pupils than the primary outcome analysis. LanguageScreen testing was undertaken with all eligible pupils in participating schools; however, for the purposes of the secondary outcome analysis focuses, only NELI-participating pupils are included in the analysis. Given the nature of the outcome (secondary) and of the measure (intervention-intrinsic), no missing data imputation was undertaken and results in the Secondary analysis section are reported on the smaller sample.

Analysis in the presence of non-compliance

Given the complex nature of the programme, there are various indicators that could be used to capture the dosage and fidelity. Our suggested strategy to determine compliance was to create a dummy indicator primarily focusing on training and the amount of NELI sessions delivered in the school. The compliance measure agreed with the delivery team involved scoring the following key programme elements: share of eligible school staff attending NELI training, proportion of group NELI sessions delivered, and number of individual NELI sessions delivered (Table 7).

Some of these elements were weighted as more important. The provision of individual NELI sessions to children was coded as the most important element, with x4 the weight of initial training and x1.5 the weight of group sessions being delivered. The register of attendance revealed perfect compliance on the initial training element. Data on teacher attendance and Part 2 TA training (highlighted in red in Table 7) were not available and ultimately not included in the final fidelity score calculation. In relation to delivery, there was substantial variation in the proportion of group and individual sessions delivered. This indicates differences in the quality of programme implementation (or, potentially, issues documenting the number of individual and group sessions for each child). The present study did not have in place practices to externally record the number of NELI sessions completed. The fidelity record was completed by the TA directly and this was considered a reasonable proxy and the best practical option for the collection of such data.

Given the missing data (highlighted in red in Table 7), the fidelity score was calculated out of a maximum of 66. A threshold score of 80% was applied so that any school scoring 52 or higher on the fidelity measure was considered a 'high complier', indicating a high degree of implementation quality. This threshold was chosen as it allowed for some variation of provision and training attendance while applying a relatively high standard of fidelity.

Missing data on any measure was scored as zero. While this is likely to have skewed the fidelity measure downwards, it is consistent with the ITT approach that seeks to provide a conservative estimate of the effect of the intervention overall.

Table 7: Fidelity measure development

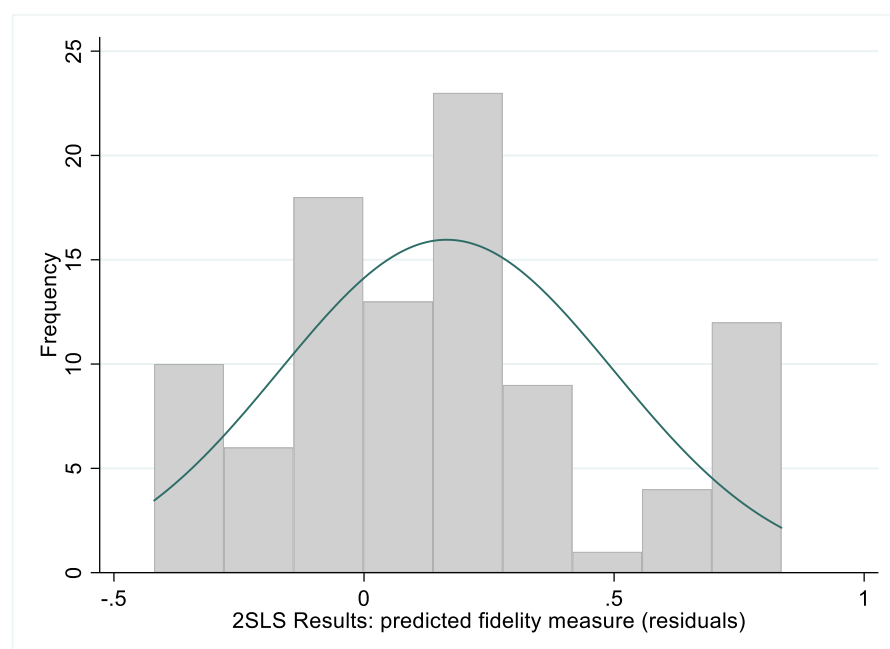
Adherence to the programme				
Measure	Data source	Tasks	Score	Weighted score
Attendance at all NELI training sessions for TAs	Register of attendance	a. Attendance at Part 1 TA training, day 1		
		• 50%–80% of TAs attend	1	1
		• 81%+ of TAs attend	1	2
		b. Attendance at Part 1 TA training, day 2		
		• 50%–80% of TAs attend	1	1
		• 81%+ of TAs attend	1	2
		c. Attendance at Part 2 TA training		
		• 50%–80% of TAs attend	1	1
		• 81%+ of TAs attend	1	2
			Max. score	9
Attendance at NELI session for teachers	Register of attendance	a. Attendance at Teacher training, day 1		
		one or more Reception teachers attend	1	1
			Max. score	1
Number of group NELI sessions completed	Fidelity record	a. Number of group sessions at 1 st half term		
		• 70–90% of group sessions completed	1	2
		• 91%+ of group sessions completed	1	4
		b. Number of group sessions at 2 nd half term		
		• 70–90% of group sessions completed	1	2
		• 91%+ of group sessions completed	1	4
		c. Number of group sessions at 3 rd half term		
		• 70–90% of group sessions completed	1	2
		• 91%+ of group sessions completed	1	4
		d. Number of group sessions at 4 th half term		
		• 70–90% of group sessions completed	1	2
		• 91%+ of group sessions completed	1	4
			Max. score	24
Number of individual NELI sessions completed	Fidelity record	a. Individual sessions at 1 st half term		
		• 70–90% of individual sessions completed	1	3
		• 91%+ of individual sessions completed	1	6

		b. Individual sessions at 2 nd half term		
		• 70–90% of indivial sessions completed	1	3
		• 91%+ of individual sessions completed	1	6
		c. Individual sessions at 3 rd half term		
		• 70–90% of individual sessions completed	1	3
		• 91%+ of individual sessions completed	1	6
		d. Individual sessions at 4 th half term		
		• 70–90% of group sessions completed	1	3
		• 91%+ of group sessions completed	1	6
			Max. score	36
Maximum score achievable for fidelity			70	
New maximum score achievable			66	

Only 11 schools emerged from the analysis as ‘high compliers’ and therefore an analysis using a compliance dummy (0/1) was not undertaken as this would have been substantially underpowered.

Given imperfect compliance, we drew on an instrumental variable (IV) approach using two-stage least squares (2SLS) estimation to recover the treatment effect for those who complied with assignment. The first stage estimated if the assignment to NELI pushed schools to take up treatment (the first stage regresses treatment assignment on compliance, as defined above). This provided an estimate of the compliance rate. Results for the first stage will report the correlation between the instrument and the endogenous variable as well as an F test. The second stage of the IV estimation predicts the outcome using the compliance rate estimated in the first regression (for more information, see Angrist and Krueger, 1991; Angrist 2006). The predicted scores (from the above procedure) for treatment schools are illustrated in Figure 3.

Figure 3: Distribution of predicted fidelity score



Missing data analysis

Missing data can arise from item non-response or attrition of participants at school and pupil levels. Even though it is important to include all data, it can be problematic to apply the intention to treat principle if we are not able to complete follow up testing for all randomised schools. To deal with missing data we completed the following steps.

Firstly, the ideal solution to the problem is to avoid missing data altogether. We attempted to follow up with all randomised schools even if they had withdrawn from allocated treatment. As indicated in the Attrition section, one

school withdrew from the trial entirely and two withdrew from the intervention only but were engaged in the process of outcome testing.

We explored attrition across trial arms as a basic step to assess bias (Higgins et al., 2011). We provided cross-tabulations of the proportions of missing values on all baseline characteristics (as detailed in the previous section on Analysis in the presence of non-compliance, at both pupil and school level) as well as on the primary outcome measures.

To assess whether there are systematic differences between those who drop out and those who do not—and thus whether these factors should be included in analysis—we model missingness at follow-up as a function of baseline covariates, including treatment. The analysis model for this approach mirrors the multilevel level model given above (pupils clustered in classes) but the outcome is a binary variable identifying missingness (yes/no).

For less than 5% missingness overall, a complete-case analysis might suffice (that is, assuming data is Missing Completely at Random—MCAR). Given that for pupils present at endline we have complete data on the outcome measure, we did not carry out any imputation. Had this not been the case, we would have checked results using approaches that account for missingness but that rely on the broader, weaker Missing at Random (MAR) assumption and subsequently used Full-Information Maximum Likelihood (FIML) because FIML can be estimated in a single model and simulation studies show that it can reduce bias as well as MI (for a discussion of FIML vs MI see Allison, 2003).

Subgroup analyses

The study was not powered for meaningful subgroup analysis. However, with the previous efficacy study (Sibieta et al., 2016) suggesting that children who speak English as an additional language were more receptive to NELI, a focus on the EAL subgroup is warranted. Therefore, as an exploratory modelling approach, EAL is incorporated into the analysis as a binary moderator variable, taking the value [1] if a child is EAL and [0] otherwise. The EAL indicator will then be interacted with treatment allocation to assess the conditional impact of NELI on EAL pupils. As this analysis is exploratory and underpowered, we do report point estimates and confidence intervals transformed into effect sizes but do not report significance tests/p-values.

While it is usual for an EEF trial to report on pupils eligible for Free School Meals (FSM), and indeed the SAP listed this as a subgroup analysis, this was not actually possible in this trial given the lack of FSM data availability for the age-group in the trial. With FSM the primary population of interest in EEF trials, a separate addendum report will be produced exploring the effect of NELI on FSM pupils only. This will make use of linked NPD data (to be obtained by the evaluation team) and published separately as soon as data is forthcoming.

Sensitivity analysis

As a sensitivity analysis we undertook multilevel modelling for clustered designs for each of the four components of the latent variable construct. Language was therefore iteratively assessed by substituting these measures as outcomes into equation 1 (see Primary analysis section). For each of the four components (expressive vocabulary, recalling sentences, information, grammar), the raw post-test score was the outcome measure and the raw pre-test score was entered in the model as a baseline measure of language and therefore a covariate.

Details of the analysis code, and associated output tables for this analysis as well as for the primary and secondary outcome analysis, are available in Appendix H.

Estimation of effect sizes

We used the effect size for cluster-randomised trials given in the EEF evaluators' guidance, adapted from Hedges (2007):

$$s^* = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Where s^* represents the clustered effect size, $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean difference between intervention groups adjusted for baseline characteristics (that is, the coefficient from the multilevel model), and the denominator is an estimate of the pooled variance. A 95% confidence interval for the ES that takes into account the clustering of pupils in schools is reported.

Implementation and process evaluation

The logic model illustrated in the Intervention section (Figure 2) identified the core components of the intervention, key principles against which to measure fidelity, expected moderators and mediators, and the linkages between these elements. The aim of the Implementation and Process Evaluation (IPE) was to address each of these elements through a range of methodologies and drawing on varied data sources. The process evaluation addressed the following questions:

RQ1: Was the intervention implemented with fidelity in the intervention classrooms?

RQ2: What factors and initial conditions appear to explain variation in fidelity of implementation?

RQ3: What appear to be the necessary conditions for success of the intervention?

RQ4: What were the barriers to delivery?

In addition to the above RQs, which were defined at the protocol stage, the process evaluation addressed the following questions:¹³

RQ5: To what extent did NELI, as implemented in the trial, lead to specific outcomes such as improved knowledge and teaching of language?

RQ6: To what extent does NELI differ from usual practice?

Research methods

To answer the IPE research questions a mixed-methods approach was applied, drawing on a combination of:

- **Online surveys with school staff** (headteachers, teachers, teaching assistants). The surveys collected data on usual practices, attitudes, perceptions, and language-skills related activities in the classroom from school staff. Survey questions were tailored to the respondents in each group (treatment and control). All surveys were completed online onto the SmartSurvey platform. Survey data was collected at two different time points: headteachers were surveyed at baseline in November and post-intervention in July; in addition to headteachers, surveys were distributed to teachers and TAs post-intervention in early July. More information on the surveys including topics covered can be seen in 0.
- **Interviews with school staff.** School staff from case study schools were invited for semi-structured in-depth interviews to get a detailed understanding of their experiences in the trial. In-depth interviews with those who have delivered the intervention have helped determine the significance of the drivers of change as identified in the logic model. The interview protocols are available in 0.
- **Observation of TA/teacher training.** The observation of the staff training captures the quality of the training events and offered an insight of school staff engagement with the programme. Researchers captured information from the observations using semi-structured protocols.
- **Observations of tester training.** These observations served two main purposes: first, they offered insight and information about the tester training approach; second, the observations explored if there was consistency in the training for testers responsible for baseline and those for outcome test administration. Information from the tester training observations was captured in the form of note-taking.
- **Attendance records.** The attendance records addressed fidelity directly by keeping record of attendance at the initial two-day training event for school staff. The same information for the second half-day training event was not available.

¹³ The additional research questions were added as it is important to present comprehensive evidence of the usual practices in the schools and on the perceived outcome of NELI.

- **TA logs and online TA support records.** These logs capture number of group and individual sessions delivered by each TA.

Pre-intervention methods

The IPE approaches before the intervention included the observation of training of Elklan testers and a survey of headteachers.

The observation of the training for testers provided insights into reasons for any differences in pupil data obtained, as they may derive from differences in test administration. Twenty-five assessors were trained for NELI on 5 and 6 September 2018. The delivery team led this training and provided a positive atmosphere throughout the day by engaging in pair discussions with trainees. We also observed the training of Elklan testers responsible for outcome testing in the summer of 2019. This provided an insight into any possible variations in testers' training. It also provided insights into reasons for differences in pupil data obtained that may derive from differences between baseline and outcome administration of tests.

The online headteacher survey served to collect information on usual practices in both intervention and control schools as well as attitudes, perceptions, and language-skills-related activities in classrooms. The survey also allowed for the examination of motivations for joining the trial and headteachers' understanding of the interventions.

During-intervention methods

IPE methods applied during the implementation phase included observations of TA and teacher training by Elklan and monitoring of training attendance records. The observation of the training events addressed implementation fidelity directly (with training being a key part of the intervention). The observations were conducted in the form of note-taking. No video recordings of training were undertaken, as per the initial protocol. The monitoring of attendance also related to fidelity and relied on data being shared with the evaluators by the delivery team and trainers, which was successfully achieved. The attendance records provide information on attendance at the initial two-day training. If school staff were not able to attend the initial training it recorded if the school received the initial training later. Attendance records for the second half-day event of training were not available.

Post-intervention methods

Once the implementation phase was complete, the IPE used the following approaches to explore experiences of participation, perceived outcomes, and implications of participation as well as control school practices. The methods included online surveys of teachers and TAs, case studies with selected schools, analysis of TA logs, and online support for TAs.

The online surveys were administered to teachers, TAs, and also headteachers in both control and intervention schools, with questions to intervention group staff relating to NELI and questions to control group staff related to potential contamination and usual practices. The headteacher survey mirrored the questions asked of teachers and TAs, but for intervention headteachers also included additional questions relating to the costs associated with participation in NELI; for intervention teachers the questions also asked about the increases in working hours and associated costs.

The case studies were undertaken in six schools to allow for the in-depth exploration of the engagement with, and commitment to, NELI. A random sample of schools was drawn and schools were given the option to opt out of the case studies. Case study schools then took part in interviews. Interviews lasted between 20 and 35 minutes and took the form of a guided conversation using a semi-structured interview guide. With the participants' consent, interviews were audio-recorded for the purpose of writing up accurate notes on the interview. Interviews were analysed thematically according to the questions explored by the researcher conducting the interview. Findings were discussed between team members in terms of their relationship to insights from the surveys and observations. The topics covered included questions about teachers' and TAs' knowledge and pedagogy change as a result of their involvement in NELI, attitudes towards the intervention and detailed feedback of teachers' and TAs' perceptions of the usefulness of the online support and resources provided as part of NELI. To provide clarity on the source of statements and expressed perspectives, throughout the report, interview evidence is referenced in the form of 'TAS#', 'TS#', or 'HTS#' where the first initials indicate the role of the individual (teaching assistant, teacher, or headteacher)

while S# identifies case study school (S1 is School 1). Individuals interviewed have been anonymised in relation to specific points discussed and views expressed for the purposes of respecting informed consent.

Finally, TA logs, which were documents filled out by TAs to monitor the amount of NELI activities being undertaken with participating pupils, were analysed. A vast majority of TAs used Excel to log instances of group and individual sessions. The delivery team's earlier experience of TAs being able to complete these logs well was also observed in this trial. Additionally, aggregate data on online support for TAs (available through a combination of webinar, email access, and a Facebook page) was explored to understand average levels of TA engagement. Table 8 illustrates the range of methods used and how they relate to each of the IPE research questions and relevant aspects of the NELI logic model.

Table 8: IPE methods overview

Research methods	Data collection methods	Participants/data sources	Data analysis methods	Research questions addressed (updated IPE RQ)	Implementation/logic model relevance
Surveys (pre/post)	Online survey	Headteachers	Quantitative descriptive analysis, including frequency counts, proportions, measures of central tendency	RQ6	Documenting counterfactual and any interventions that may impact on child language.
Surveys (post)	Online survey	Headteachers		RQ1	Exploring experience of schools in delivery. Reporting on costs associated with delivering NELI.
		Teachers		RQ1; RQ2; RQ3; RQ4; RQ5	Exploring experience of schools in delivery and impact. Reporting on costs associated with delivering NELI.
		TAs		RQ1; RQ2; RQ3; RQ4; RQ5	Exploring experience of schools in delivery and impact. Reporting on costs associated with delivering NELI.
Observations of training	Observation	Elklan trainers	Thematic analysis	RQ1; RQ2	Documenting quality of training.
		Teachers/TAs		RQ1; RQ2	Documenting quality of training and TA/teacher responsiveness.
Case studies	Interviews	Headteachers	Thematic analysis; Within-case analysis; Cross-case analysis	RQ1; RQ2; RQ3; RQ4; RQ5	Exploring experience of schools in delivery.
		Teachers		RQ1; RQ2; RQ3; RQ4; RQ5	Document the quality of intervention and exploring perceptions of programme impact and drivers of change.
		TAs		RQ1; RQ2; RQ3; RQ4; RQ5	Document the quality of intervention and exploring perceptions of programme impact and drivers of change.
Review of TA logs	Document analysis	Logs	Quantitative descriptive analysis	RQ1; RQ2	Documenting dosage of intervention/compliance.
Review of records of online support for TAs	Document analysis	Records	Quantitative descriptive analysis	RQ1; RQ2	Documenting dosage of intervention/compliance.

Costs

Average marginal costs per pupil per year were estimated in two stages: first, from a discussion and data provided by the delivery team and then, second, using one of the IPE surveys to query school staff if they had spent extra hours working while implementing NELI. Average marginal costs per pupil per year over a three-year period were estimated in line with the convention followed by other EEF projects (EEF, 2016). The EEF cost rating is included in Appendix A. Headteachers were also asked if other school staff, such as teachers, had been required to work longer hours. These estimates together provided a general indication of the costs involved with implementing NELI. We recognise that in the absence of complete data on costs from all schools (not all schools responded to the cost-querying survey), costs may differ between schools—particularly between one-form-entry and multi-form-entry schools.

The following cost categories were explored using questions to school staff:

- direct financial costs, including direct costs paid for training and materials if the intervention were to be offered without subsidy; and
- direct time costs, including staff time.

Timeline

The trial timeline is illustrated in Table 9.

Table 9: Trial timeline

Dates	Activity	Staff responsible / leading
June 2018	IDEA workshop	RAND Europe
Jun–Aug 2018	Recruiting schools and teachers	Oxford
Sep 2018	Elklan testers training day	Oxford
Sep 2018	Opt-out forms to be sent to parents	Schools
Sep–Oct 2018	LanguageScreen testing by TAs	Schools
Oct 2018	Analysis of 5 lowest scorers per class	Oxford
Oct–Nov 2018	Individual testing of 5 lowest scorers per class	Elklan
Nov 2018	Randomisation	RAND Europe
Nov 2018	Training of TAs to administer intervention	Elklan. RAND Europe observing
Jan–Jul 2019	Programme Implementation	Schools
Jan 2019	Survey 1 of headteacher/SENCO in all schools	RAND Europe
Feb 2019	Completion of Statistical Analysis Plan	RAND Europe
April and June 2019	Distribution of newsletter to schools	Oxford
July 2019	Compilation and distribution of record of access of online support for TAs	Elklan
July 2019	Outcome testing	Elklan/RAND Europe
July 2019	Survey 2 of headteacher/SENCO	RAND Europe
July 2019	Survey of TAs and teachers in all schools	RAND Europe
July 2019	Interviews with TAs/Teachers in case study schools	RAND Europe
Nov 2019	Final EEF report	RAND Europe

Impact evaluation

Participant flow including losses and exclusions

The participant flow diagram is presented in Table 10. The diagram shows that 207 schools were recruited in total. Before randomisation 14 schools withdrew for various reasons including significant staffing changes or difficulties (five schools), GDPR concerns (three schools), due to OFSTED threat and being downgraded (two schools), building work and flooding (one school), or with no clear explanation (two schools). One of the recruited schools did not meet the eligibility criteria for conflict of interest as it had been already running NELI in early years as part of the Language First project.

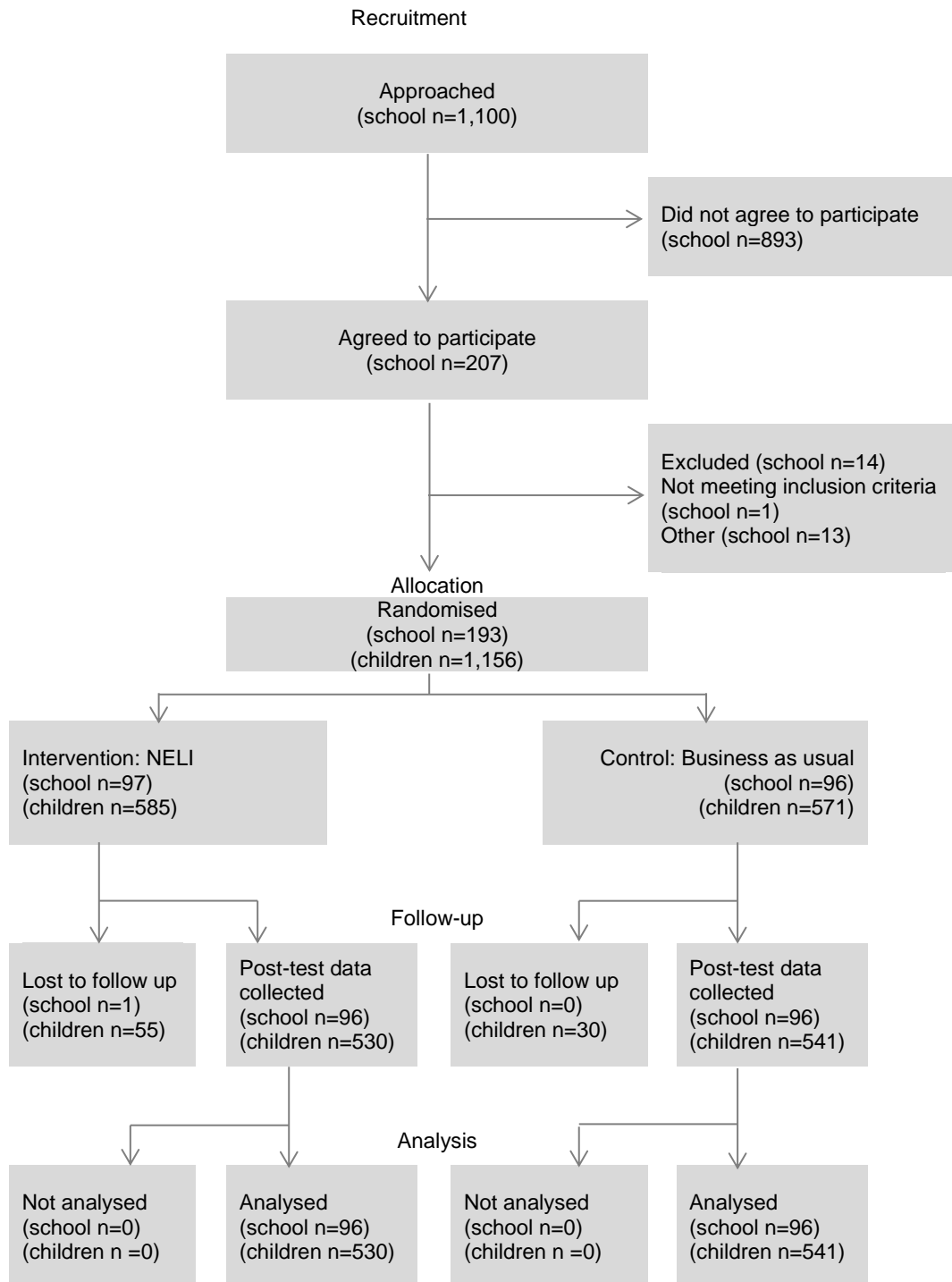
In total 193 schools remained in the trial for randomisation. Parent information sheets with opt-out forms were sent to all schools in the trial. Schools were asked to circulate these sheets to all parents with children eligible to take part. All recruited schools had confirmed by 19 September 2018 that parent information sheets had been sent out. In depth assessment of children eligible for NELI began on 15 October. Baseline data was collected for 1,156 children.

Of the 193 schools randomised, 97 schools (585 children) were assigned to the intervention group, while 96 schools (571 children) were assigned to the control group.

After the school-level randomisation, one of the 97 intervention schools dropped out. The reason for drop-out reported to the delivery team revolved around the project lead leaving the school and as a consequence the school had a lack of staff capacity to support NELI implementation.

At the time of analysis, data had been collected from 192 schools and 1,071 children.

Table 10: Participant flow diagram



Attrition

Only one intervention school withdrew entirely from the trial (stopped implementing NELI and did not participate at outcome testing). It is not included in the final analysis and has been listed under the 'lost to follow-up' category in the participant flow diagram. No control schools dropped out from the trial.

Baseline data was available for 585 children in 97 intervention schools and 571 children in 96 schools. At analysis stage, data was available for 530 children in 96 intervention schools and 541 children in 96 control schools. In total, three schools withdrew at least from the intervention post-randomisation, including one which withdrew from the trial entirely, as indicated in the participant flow diagram. Two schools withdrew due to staffing capacity issues and one due to a lack of funds. Testing was organised in the schools withdrawn from the intervention, but not in the school withdrawn entirely from the trial (Table 11).

The overall ratio between pupils analysed and pupils randomised is 1,071 to 1,156—a retention rate of 92.7% and an attrition rate of 7.3%. In the control group, the retention rate is 94.7% (an attrition rate of 5.3%) with 541 pupils in the analysis from the 571 randomised. In the intervention group, the retention rate is 90.6% (an attrition rate of 9.4%) with 530 pupils analysed of the 585 randomised to the intervention condition.

The main reasons children were not tested were the following:

- pupils had left the school (77% of children not tested);
- pupils were absent at both the initial and top up training (14% of children not tested);
- the school dropped out and did not want to take part in follow up testing (8% of children not tested); and
- parents did not allow their children to be tested (1% of children not tested).

Table 11: Pupil level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	585	571	1156
	Analysed	530	541	1071
Pupil attrition (from randomisation to analysis)	Number	53	30	83
	Percentage	9.1%	5.3%	7.3%

Pupil and school characteristics

Table 12 shows the baseline distribution of school and pupil characteristics across the control and intervention schools using all pre-randomisation data available on each respective variable. To assess the balance of pupil characteristics at baseline in accordance with EEF and CONSORT¹⁴ guidelines we provide descriptive tables of pupil characteristics in the control and treatment schools. Given appropriate randomisation procedures were followed, any differences between control and treatment groups at baseline will be by definition due to chance, and classical statistical testing is therefore unnecessary and imbalance may be gauged from the differences in means. There is a convention in some disciplines that a ten percentage point or larger difference in intervention and control means at baseline constitutes 'imbalance' and would justify including those measures in sensitivity analyses.

There were no meaningful differences in terms of FSM eligibility distribution with the proportion of FSM eligible children in the intervention schools at 34.05% and at 33.81% in the control schools. The two groups are also balanced in terms of share of schools rated 'inadequate' (4.76% in the intervention and 4% of schools in the control group). There are some imbalances in the share of schools rated as 'good' or 'requires Improvement' but this could be due to missing

¹⁴ <http://www.consort-statement.org/checklists/view/32-consort/510-baseline-data>

data on OFSTED rating. The two groups are balanced at reception class level in terms of gender, with a difference of only 4.2 percentage points, with more girls in the control group than in the intervention group. English as an additional language (EAL) was balanced between the intervention and control groups. The control and intervention groups also show balance in terms of the baseline testing undertaken at the start of the study.

The distributions of the means for the variables are provided in Table 12. Although we routinely argue that a statistical testing of mean differences between intervention and control groups is redundant immediately post-randomisation—as any differences will, by definition, have occurred by chance—we have followed What Works Clearinghouse (WWC) guidelines in reporting effect sizes for the mean differences.¹⁵ Based on the WWC guidelines, differences between groups are assessed based on a rule that states that intervention and comparison groups are equivalent at baseline when the corresponding effect sizes are smaller than 0.25 standard deviations. In this trial, the baseline comparisons of means yield very small effect sizes, well below 0.25, as illustrated in Table 12. We therefore conclude that the randomisation resulted in intervention and control groups that are comparable in terms of school characteristics as well on children's characteristics and language attainment.

Table 12: Baseline characteristics of groups as randomised

School-level (categorical)	National-level mean	Intervention group		Control group	
		n/N* (missing)	Count (%)	n/N (missing)	Count (%)
OFSTED Rating					
1: Outstanding		84/97 (11)	13.10%	75/96 (21)	18.67%
2: Good		84/97 (11)	77.38%	75/96 (21)	64%
3: Requires Improvement		84/97 (11)	4.76%	75/96 (21)	13.33%
4: Inadequate		84/97 (11)	4.76%	75/96 (21)	4%
School proportion eligible for FSM		87/97 (10)	34.05%	80/96 (16)	33.81%
School-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
KS2 Maths	105	85/97 (12)	104.118 (2.805)	80/96 (16)	103.713 (2.887)
KS2 Reading	104	85/97 (12)	104.882 (2.929)	80/96 (16)	104.362 (2.803)
Class-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
Gender					
Male		1,543/2,993	51.5%	1,414/2,897	48.8%

¹⁵ What Works Clearinghouse Standards Handbook v4.
https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf, (page 14).

Female		1,450/2,993	48.5%	1,483/2,897	51.2%	
English as an additional language		505/2,993	17%	497/2,897	17%	
Pupil-level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)	
Gender						
Male		336/585	57.4%	304/571	53.2%	
Female		249/585	42.6%	267/571	46.8%	
Pupil-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size
Age (in months)		585/585	52.188 (3.49)	571/571	52.37 (3.51)	0.052
LanguageScreen		585/585	-1.28 (0.699)	571/571	-1.27 (0.678)	0.045
CELF expressive vocabulary		585/585	10.33 (6.279)	571/571	10.48 (5.905)	0.023
CELF recalling sentences		585/585	8.00 (6.288)	571/571	7.87 (6.233)	0.030
RAPT Information		585/585	19.11 (7.863)	571/571	20.09 (7.359)	0.128
RAPT grammar		585/585	11.69 (6.868)	571/571	12.11 (6.565)	0.059
YARC early word		585/585	0.61 (2.987)	571/571	0.48 (2.37)	0.053
*Note: n refers to the sample size of a specific group (for example, intervention schools) while N refers to the total sample size (for example, all schools in the trial).						

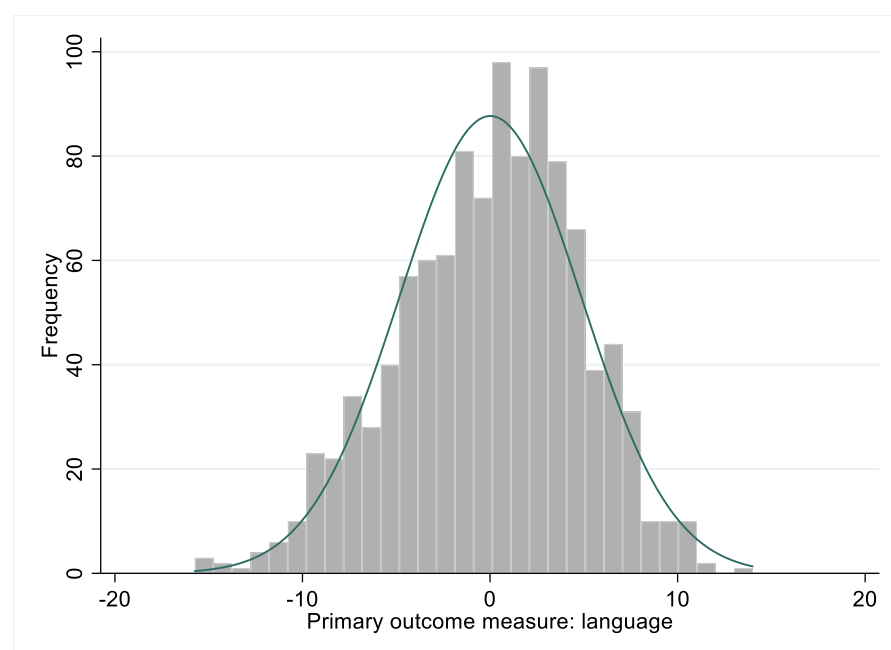
Outcomes and analysis

Primary analysis

This evaluation uses as a primary outcome measure the standardized language skills variable derived as outlined in the Measure development section (that is, a factor score derived from a confirmatory factor analysis on the four individual language sub-scores at baseline and corresponding factor score at endline). For pupils in the final analysis (1,071 children with observable data on all variables) the language skills variable had an overall mean of 0.02 and standard deviation of 4.83. The range for the analytical sample was from -15.75 to 14.00. The distribution of the primary outcome measure is illustrated in Figure 4.

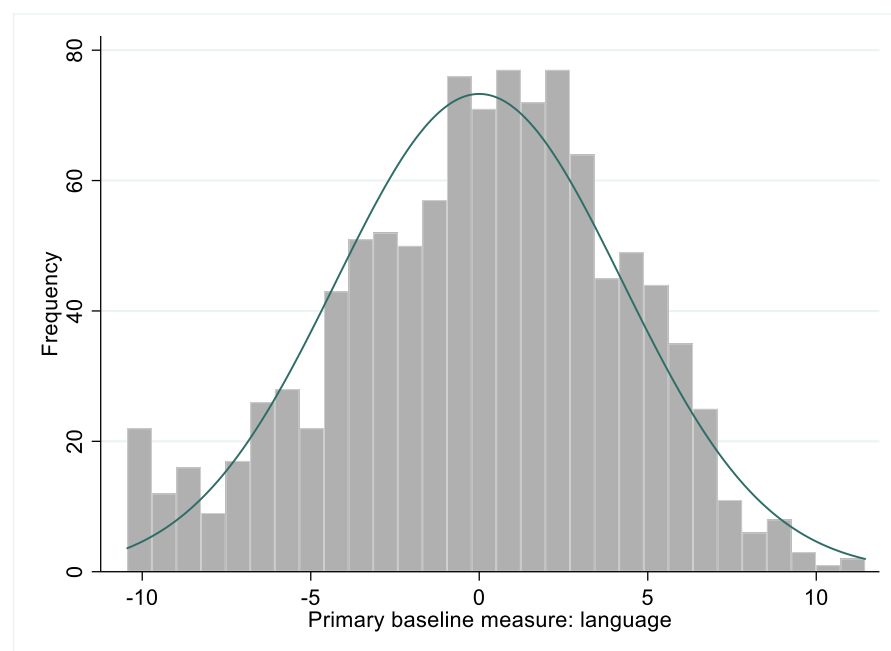
Although not perfectly normal, the parametric tests used in the analysis of data are robust to the slight skew illustrated by the data and therefore the variable was used without any further modification or transformation as the outcome variable for the primary analysis. A higher score on this variable indicates a better language skills outcome.

Figure 4: Primary outcome measure—distribution of language outcomes at endline



The primary language outcome measure was correlated to the language baseline measure, with $r = 0.75$. The language baseline measure distribution is illustrated in Figure 5.

Figure 5: Distribution, primary baseline measure



As outlined in the Methods section, the primary analysis took an intention-to-treat approach, with both stratification variables included in the randomisation also present in the analysis (that is, region and number of classes within a school). Multilevel models were used to assess the effectiveness of the intervention so as to account for the clustering of children in schools. The details of the model for the primary outcome analysis are reported in Table 13.

The raw mean for the intervention group was larger than the raw mean for the business-as-usual control group. As Table 13 illustrates, the raw, or unadjusted, difference in means was 1.

In the multilevel models that accounted for the clustering of pupils in schools and also accounted for the language pre-test, this difference was statistically significant with $p = 0.000000633$. The adjusted difference in means obtained from the multilevel model was 1.243. The effect size associated with this adjusted difference in means is 0.26 (Table 14).

Table 13: Primary outcome analysis results

			Intervention group		Control group		
Outcome	Unadjusted differences in means (I-C)	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Primary outcome: Language	1	1.243	530 (0)	22.83	541 (0)	23.39	23.11

Table 14: Effect size estimation, primary outcome analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Primary outcome: Language	530 (0)	0.527 (0.11, 0.93)	541 (0)	-0.473 (-0.88, -0.064)	1,071 (530, 541)	0.26 (0.17, 0.35)	p < 0.05

The conclusion of the primary outcome analysis is therefore that NELI had a positive and statistically significant impact on pupils' language skills.

Secondary analysis

The protocol and statistical analysis plan set out two secondary outcome measures: first, an early word reading score deriving from the YARC test, and second a language score deriving from LanguageScreen. The results of these analyses are presented here in turn for the two outcomes.

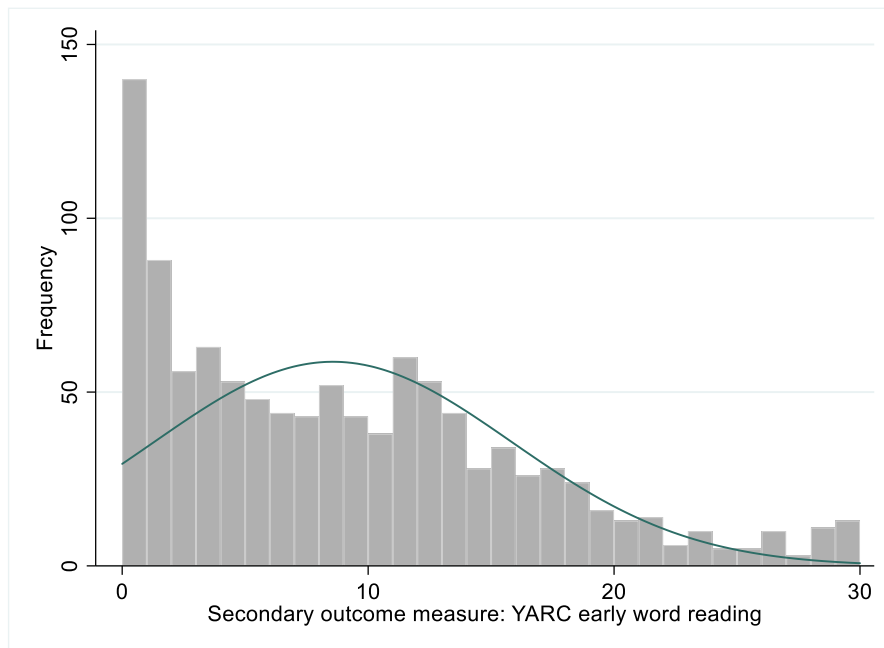
Early word reading (YARC)

Single-word reading as measured through the YARC Early Word Reading test was the first secondary outcome to be considered in the secondary outcome analysis.

Pre-test YARC data was available for 1,156 pupils; post-test YARC data was available for 1,071 pupils. Therefore, 1,071 pupils with both baseline and outcome data were included in the YARC secondary outcome analysis.

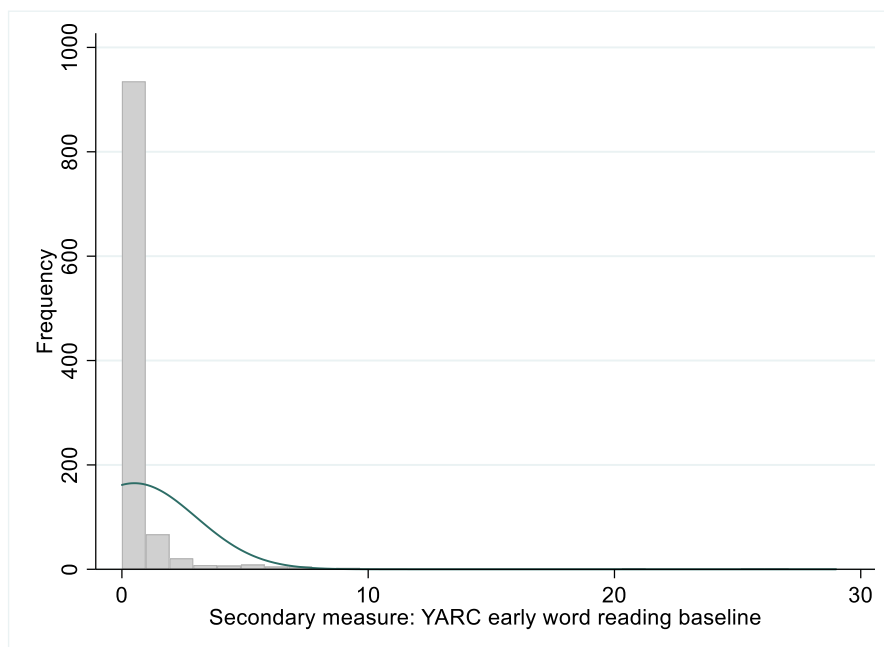
For the analytical sample, early word reading scores had a mean of 0.51 and standard deviation of 2.50 at baseline and a mean of 8.57 and standard deviation of 7.28 at endline. The distribution of the early word reading score at endline is illustrated in Figure 6. The outcome variable was not perfectly normally distributed, with 13.1% of pupils obtaining a total score of 0 on this test. There was therefore a large floor effect for this variable.

Figure 6: Secondary outcome measure—distribution of YARC early word reading at endline



The distribution of the early word reading score at baseline is illustrated in Figure 7. At baseline, 87% of pupils in the analysis sample scored zero in the YARC early word reading test indicating significant non-normality. The floor effect for this measure was expected given the age of pupils and the nature of the test. The baseline scores were used in the analysis as previous research shows that even test scores at the floor contain real information about pupil performance (Catts et al., 2009).

Figure 7: Secondary measure—distribution of YARC early word reading at baseline



While the distribution of the outcome YARC measure sufficiently approximated to normality so that the parametric tests employed in the analysis are robust, the distribution of the baseline YARC measure did not allow for this conclusion and therefore cannot be considered a normally-distributed continuous variable. To avoid non-normality affecting the final estimates, the baseline YARC measure was entered into the analysis as a categorical variable; Table 15 illustrates the recoded variable and the balanced achieved at baseline between the intervention and control groups. Although this reduces the variance in the pre-test data, given that the baseline serves to provide an indication of the overall early reading level of pupils, this is a sufficiently robust approach.

Table 15: Baseline YARC measure—recoded

Original YARC Score	Recoded value	Proportion of all pupils	Intervention group	Control group
0	1	87.30%	88.49%	86.14%
1	2	6.26%	4.91%	7.58%
2 and higher	3	6.44%	6.60%	6.28%

At endline, the unadjusted mean difference between control and intervention groups was 0.86, as reported in Table 16.

Analysis was carried out as per the Methods section, again using a multilevel model to account for the clustering of pupils in schools and for the pre-test under an intention-to-treat approach. The adjusted mean difference between the intervention and control groups was 1.09 (Table 17).

The effect size associated with this adjusted mean difference was 0.15.

Table 16: Secondary outcome analysis—early word reading

			Intervention group		Control group		
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Secondary outcome: YARC	0.86	1.09	530 (0)	56.85	541 (0)	48.86	52.81

Table 17: Effect size estimation, secondary outcome analysis—early word reading

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Secondary outcome: YARC	530 (0)	9.00 (8.36, 9.65)	541 (0)	8.15 (7.56, 8.74)	1,071 (530, 541)	0.15 (0.03, 0.34)	p = 0.020

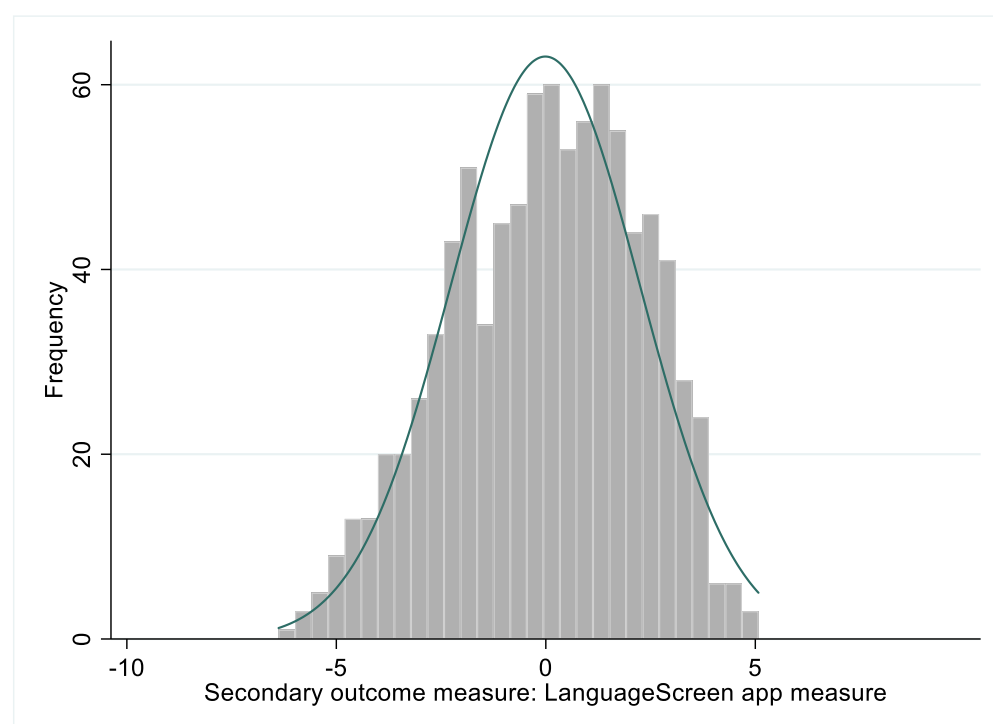
Therefore, the results of this secondary analysis indicate that NELI had a positive and statistically significant impact on pupils' early word reading language skills.

LanguageScreen measure

The second secondary outcome measure in this trial was also language skills, but assessed through the LanguageScreen app and measure. As detailed previously, the LanguageScreen measure captures—through the app—four core language skills: expressive vocabulary, receptive vocabulary, sentence repetition, and listening comprehension. The LanguageScreen score as derived from the procedure described in the Measure development

section was used in this analysis. A higher score is indicative of better core language skills. The distribution of this measure is provided in Figure 8, sufficiently normal for the application of the parametric tests to follow.

Figure 8: Secondary outcome measure—LanguageScreen histogram



Data for 904 pupils was available at both baseline and endline (the analytical sample for the LanguageScreen secondary outcome analysis). The pre-test language skills variable had a mean of -0.06 and standard deviation of 1.88. The post-test language skills variable had a mean of -0.01 and standard deviation of 2.26. The unadjusted mean difference between the intervention and control group was 0.05 (Table 18).

The analysis also took the form of multilevel models that accounted for the clustering of pupils into schools and all stratification factors used at randomisation, as per the intention-to-treat approach. The adjusted mean difference between the intervention and control group was 0.77 (Table 19) and was statistically significant at $p < 0.05$. The effect size associated with this difference was 0.358 (Table 19).

Therefore, the results of this analysis indicate that NELI had a positive and significant effect on language skills as captured by the LanguageScreen test in the sub-sample with data.

Table 18: Secondary outcome analysis—LanguageScreen

Outcome	Intervention group				Control group		Pooled variance Population variance (if applicable)
	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Secondary outcome: LanguageScreen	0.614	0.802	444 (0)	4.702	460 (0)	5.319	5.016

Table 19: Effect size estimation, secondary outcome analysis—LanguageScreen

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Secondary outcome: LanguageScreen	444 (0)	0.302 (0.099, 0.504)	460 (0)	-0.312 (-0.523, -0.100)	904 (444, 460)	0.358 (0.22, 0.47)	p < 0.05

Analysis in the presence of non-compliance

The results of the fidelity analysis using the continuous fidelity measure outlined in the Methods section suggest a statistically significant ($p=0.038$) effect of NELI compliance on the main language outcome (Table 20). In other words, increased compliance with NELI (i.e. attendance of TAs and teachers at training and delivery of minimum of 70% of sessions) leads to better language outcomes for pupils.

Table 20: Analysis in the presence of non-compliance—main outcome

Outcome	Compliance measure				
	Coefficient (CI)	Standard error	Z	p value	N
Language: EAL group	1.07 (0.062, 2.08)	0.515	2.08	0.038	1,071

Missing data analysis

The variables included in the analysis do not display patterns of missing data; more than 95% of data is present and analysable:

- outcome measure: 100% of data available for tested pupils; and
- EAL status (for subgroup analysis only): 98.2% of data available.

Therefore, no imputation or other adjustment for missingness was carried out.

Subgroup analyses

The trial protocol did not set out a subgroup analysis. However, the statistical analysis plan included a subgroup analysis by pupils' status as speaking English as an additional language. The rationale for the inclusion of EAL children is linked to the fact that earlier efficacy studies of the NELI programme have suggested that these children were more receptive to the intervention (Sibieta et al., 2016). The statistical analysis plan acknowledges that this analysis would most likely be underpowered and as a result the analysis is exploratory only and would not report statistical significance levels for the results.

The analysis tested the hypothesis that NELI had a positive effect for EAL pupils in the intervention group compared to EAL pupils in the control group. Results are reported in Table 21 and in Table 22. At endline, there were 359 pupils that were identified as EAL: 33.21% of analytical sample pupils in intervention schools and 33.83% of analytical sample pupils in control schools.

The exploratory EAL subgroup analysis was therefore undertaken for the main outcome measure only. An interaction variable was created, interacting the EAL indicator with intervention allocation to assess the conditional impact of NELI on EAL pupils.

Table 21: Subgroup analysis—pupils with English as an additional language (EAL)

Outcome			Intervention group		Control group		Pooled variance
	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Language: EAL group	0.352	1.10	176 EAL (0)	22.02	183 EAL (0)	21.55	21.78

To obtain the effect size, the above difference in means was divided by the pooled variance in the estimation sample (Table 21 and Table 22).

Table 22: Effect size calculation—subgroup analysis, pupils with English as an additional language (EAL)

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Language: EAL group	176 EAL (0)	-2.22 (-2.92, -1.53)	183 EAL (0)	-2.57 (-3.25, -1.89)	1,071 (530, 541) (176, 183 EAL)	0.236 (0.04, 0.43)	0.000

This analysis therefore indicates that NELI had a positive effect on EAL pupils' language skill development in the intervention group compared to the control group.

Using the same analysis, the additional hypothesis deriving from the efficacy trial about the potential differential effect of NELI on EAL pupils can be investigated—that is, whether NELI had a significantly higher effect on EAL pupils in the treatment group compared to non-EAL pupils in the treatment group (Table 23 and Table 24).

Table 23: Subgroup analysis—EAL and non-EAL pupils in the intervention group

			Intervention group: EAL		Intervention group: non-EAL		Pooled variance
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Language: EAL group	-4.113	-0.008	176 EAL (0)	22.02	351 non-EAL (0)	17.71	19.83

To obtain the effect size, the above difference in means was divided by the pooled variance in the estimation sample.

Table 24: Effect size calculation—subgroup analysis, pupils with English as an additional language (EAL)

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention)	Hedges g (95% CI)	p-value
Language: EAL group	176 EAL (0)	-2.22 (-2.92, -1.53)	351 non- EAL (0)	1.89 (1.45, 2.33)	1,071 (530, 541) (176 EAL, 351 non-EAL)	-0.002 (-0.15, 0.14)	0.981 (for the intervention allocation variable)

This exploratory and underpowered analysis suggests that there was no differential effect of NELI on EAL pupils compared to non-EAL pupils.

Implementation and process evaluation

In the following sections we discuss the findings in relation to implementation compliance, implementation fidelity, perceived impact, and usual practice. Compliance deals with the question, ‘Did schools do it?’ Fidelity considers the question, ‘Did schools do what they were supposed to do (to the appropriate pupils and in the way they were supposed to)?’ The subsequent section on outcomes discusses the extent to which NELI activities, as delivered in the trial, led to specific implementation outcomes (that is, TA improved knowledge and teaching of language). Finally, the usual practice section looks at the extent to which NELI differs to what was happening before NELI was implemented (at baseline) and what was being delivered in control schools (at endline).

It should be highlighted here that the findings reported in this section draw largely on insights drawn from analysis of questionnaire surveys, interviews, observations of teacher and TA training, and TA logs, which represent respondents’ self-reports and perceptions. Therefore, these may not be representative of the study sample as a whole. While the same question was often asked over time in multiple surveys, it was not possible to track individual respondents across time given that cross-survey tracking information was not collected from respondents.

Finally, as part of the IPE, individuals were asked about ways to improve NELI. Where relevant to the larger IPE research questions we have included these in the main body of the report.

Compliance

Box 1: Compliance key findings

Responses by TAs and teachers to their respective endline surveys and data from interviews with TAs from selected case study schools show that:

1. Attendance at initial training sessions was high, both in terms of self-report (from surveys) and in terms of logs from Oxford (as shown in the Compliance section). Survey responses show that four fifths (approx. 80%) of TA and teacher survey respondents attended the mandatory initial training.
2. Attendance at the top-up training was notably lower. About only one third of TAs responding to the survey reported having attended this training. However, use of ongoing support by TAs was relatively high. A notable majority (approx. 80%) of TA survey respondents reported having made use of online resources and ongoing support by Elklan trainers available to them throughout the intervention.
3. Delivery of the required number of individual and group sessions was moderate, with evidence of wide variability across settings.

Crucial to successful delivery are the non-negotiables of NELI: attendance at the initial training, attendance at the top-up training, and delivery of NELI sessions with high fidelity. As discussed in the Impact evaluation section, compliance with the NELI programme was measured as a combination of attendance at training sessions and number of NELI sessions delivered to pupils in schools. This section describes these two elements in further detail. For more information please see the Analysis in the presence of non-compliance section.

Attendance at training sessions and use of ongoing support while delivering NELI

Survey data suggests there was relatively high compliance with initial training attendance

Attendance at the initial training was systematically recorded by Elklan trainers who were working closely with the delivery team. Compliance with the initial two-day training was 100% because the delivery team did not let any school deliver the programme without the needed training. In six cases, where schools were unable to make it to the training, a delivery trainer was sent to the school to deliver. Attendance at the second training event was not recorded.

Survey data, even though incomplete, provided information as to whether TAs attended the initial or alternate training as a replacement and recorded attendance at the top up training. Responses to the teacher and TA endline surveys reveal that attendance at NELI training sessions was, overall, high. TAs were expected to attend two days of mandatory training—day one provided an overview of the programme, day two focused on implementation. In the endline survey, the majority of TAs responding to the survey reported attending day one (83%, 38 out of 46) of NELI’s

initial training; a slightly higher majority reported having attended day two (89%, 41 out of 46). Of these TAs, 36 (out of the 46 responding to the survey, 78%) reported having attended both days, meaning that some TAs attended only one day of training rather than both. Furthermore, there were only a few cases of no training attendance. Three TAs reported having attended none of these training days but they reported having received alternative training or ongoing support. Only one TA responding to the endline survey said that they had not attended any of the training sessions nor received any form of top-up training or accessed any form of ongoing support. Nonetheless, given the low incidence of this scenario and the overall reported high attendance it can be inferred that compliance levels to training attendance were relatively high for TAs.

A similar trend is observed in the case of teachers, who were invited to attend half a day of *day one* of the TAs training (attendance was not mandatory for teachers but seen as a key success factor in the logic model). In the endline teacher survey the majority of teachers (83%, 46 out of 55) reported having attended the training. Hence, it can also be concluded that compliance to training attendance was, overall, high for teachers—based on endline survey responses. This is particularly important when considering the role of teachers in providing support to TAs to deliver NELI. During case study interviews, two TAs noted that it was important that teachers attended part of their training to help the teachers develop their understanding of what is expected from TAs for NELI (TAS5, TAS6).

Survey responses show low compliance with top-up training attendance, but high use of continuous support

In addition to the initial training, there was a *mandatory one-day face-to-face top-up training* for TAs delivered midway through the intervention (after ten weeks). This is additional to the initial NELI set-up and therefore is included in the revised logic model in Figure 17. This training focused on phonics since the second half of NELI delivery also places more focus on this aspect. Responses to the endline TA survey show notably lower levels of compliance with this element of the intervention compared to initial training. Considerably fewer TAs responding to the survey—about a third (33%, 15 out of 46)—reported having attended the top-up training. This may raise some concerns given the importance of the training, as voiced by one TA from a selected case study school:

'A lot of it was reassurance [at the top-up training]. You were asking her to reassure you, really, if you are doing it correctly and obviously it is very different for everybody because we are working with different children and it is important to know that we are all following the same format. So when we came together after the first initial ten weeks, we got together before we started the second ten weeks. That was really good and everyone came together then in the third training day and it was good that people had the opportunity to chat and see what was working really well and obviously what was more difficult for other people' (TA1S3).

The top-up training therefore provided a chance to clarify questions and share and obtain learning from others. While exemplified by one case, the training most likely also provided reassurance—to those attending the training more generally—on whether TA practice was appropriate and in compliance with how NELI should be implemented.

In addition to the mandatory training sessions, TAs had access to online resources and support throughout the programme delivery. While non-mandatory, this was formal ongoing support in the form of telephone support (from Elklan trainers), webinars, and online forums. TA survey responses show relatively high use of these resources by TAs, with a notable majority (76%, 35 out of 46) reporting having made use of these resources. Furthermore, as part of the case study interviews, TAs were asked about their use of ad hoc ongoing support. TAs from different case study schools reported making use of these resources, showing that the experience was widespread across the selected case study schools. More specifically, four TAs (from four different case study schools) noted that they received support via email (TAS5, TAS2, TA1S3, TA2S4), four TAs (from three different schools) made use of support provided through webinars (TAS5, TAS2, TA1S4, TA2S4), two made use of the Facebook page (TAS5, TAS2), and another two also reported having met with their trainers (TAS6, TA2S3). All of these TAs noted that it was important to have access to these forms of support to ensure appropriate implementation and delivery of NELI. Given the high use of continuous support and the high levels of compliance observed, attendance at the top-up training may have been less integral to the logic model than had previously been considered.

TA logs suggest moderate delivery of group sessions with wide variability

A core component of NELI is the delivery of three 30-minute group sessions and two individual 15-minute sessions per week. In England, school terms have 12 to 13 weeks. In this trial, the 20-week NELI programme was implemented across two terms, with ten weeks delivered in each term. Compliance with delivery of group and individual sessions

are discussed below in light of TA log data analyses for the selected six case study schools. This data was available per half term.

For group sessions, high compliance would entail delivering approximately 30 group sessions per term or 15 sessions per half term. TA log data analyses showed that compliance to the delivery of group sessions was moderately high but there was wide variability across settings. The median number of group sessions per half a term varied between 11 and 14 sessions, and ranged between 3 and 17 group sessions. There were no substantial differences in the number of attended group sessions by half term. The number of attended sessions was slightly higher in the second half term (three more sessions, that is, 14) compared to the fourth half term where the median number of attended sessions was 11. Based on interview data, teachers and TAs generally expressed mixed views about the appropriateness of the structure and length of group sessions although, on the whole, feedback on group sessions was positive with many considering them useful for the development of children's language skills. Additional details on views on group sessions are provided under the Programme delivery section.

In terms of individual sessions, high compliance would require delivering about 12 to 14 individual sessions per half term.¹⁶ Given the high frequency of the sessions, it is typical for children to miss some of the individual sessions. Analyses of the TA logs recording the number of sessions received by each pupil show that compliance with the delivery of individual sessions was moderate, although some variation across half terms and case study schools was observed. According to the log records across the different schools and half terms, the number of sessions ranged from none to 14. Overall, the number of individual sessions was higher in the first half term (median was eight) in comparison to the other half terms (where the median was five in the second term and six in the third and fourth terms). The logs data suggests that fidelity—in terms of the number of individual sessions attended—was relatively moderate in the first half-term (57%) but dropped over the period of the trial (42% in the second to 35% in the third half-term). This may be explained by the findings from interviews as discussed in the Programme delivery section. The preference for group sessions expressed by some interviewees may have resulted in practitioners choosing to prioritise the delivery of more group sessions at the expense of individual ones as the trial progressed. Additionally, in the training TAs had been advised to prioritise group over individual sessions.

We explore reasons for why group and individual sessions may not have been delivered as intended and implications for development of NELI in more detail in the Implementation fidelity section below.

Implementation fidelity

Box 2: Implementation fidelity key findings

Responses by TAs and teachers to their respective endline surveys and data from interviews with TAs from selected case study schools show that:

1. Schools felt that training was appropriate enough to allow TAs and teachers to deliver the intervention and were largely positive about the level of information they received about the demands of delivering NELI.
2. The majority of surveyed TAs and interviewed teachers and TAs had concerns about the process of selecting children to participate in the NELI intervention. A notable majority (approx. 80%) of TA survey respondents felt that teachers or TAs should decide which pupils enter the NELI programme.
3. Some aspects of delivery were noted as being negatively or positively associated with implementation, with time restrictions and perceptions of individual sessions being barriers to implementation, while space, the Oxford University Press resources, and support from senior staff were noted as factors that supported effective delivery.

To understand better how NELI worked in practice, we identified several elements where fidelity would be of key importance. When discussing these elements we focus on quality, content, appropriateness (for example, of training or pupil selection), ease of delivery (enablers and barriers), and presence of any adaptations or deviations from the

¹⁶ The number of individual sessions varies by the number of weeks in a half-term.

original design of the intervention (as opposed to, for example, the number of sessions delivered, which is captured under the Compliance section). The elements discussed using these lenses are: TA and teacher training, pupil selection, and elements of programme delivery associated with positive or negative aspects of implementation. These elements were selected based on the intervention's logic model and on previous evaluations of NELI. These are discussed in turn below.

Perceived training quality and levels of preparedness to implement NELI

If an intervention is to be successful, training to deliver that intervention must be successful. Overall, IPE data suggests that TAs and teachers were largely positive about the quality of the training they received. Similarly, teacher and TA endline survey and interview responses suggest that the majority of those responding to these IPE data collection activities felt prepared to start delivering NELI following the training they received.

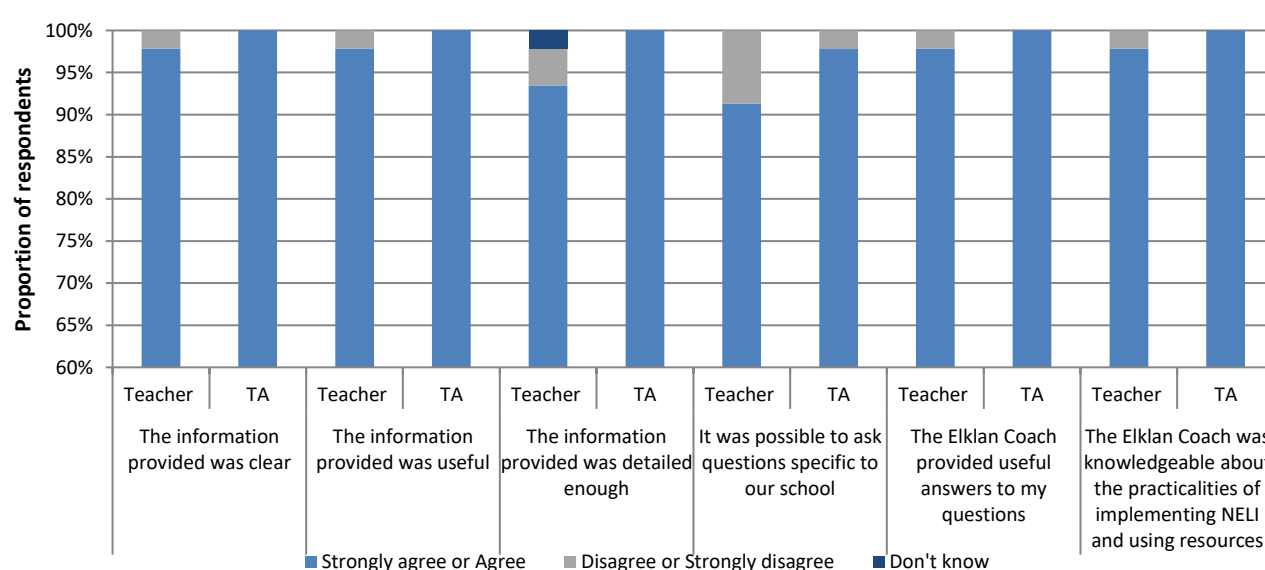
Perceptions of training

Both TAs and teachers were asked about the clarity, usefulness, and detail of the training (initial and/or top-up training)¹⁷ that they received. Teachers were asked specifically about day one of the initial training, while TAs were asked about their views in general for both days of initial training that was available to them. As shown in Figure 9, the vast majority of survey respondents reported positive views on the training with TAs being slightly more positive in their responses than teachers across the different aspects of the training they were asked about.

As shown in Figure 9, all 45 TAs responding to the endline survey agreed or strongly agreed that training was clear (out of 45: 14 strongly agreed, 31 agreed), useful (18 strongly agreed, 27 agreed), and detailed enough (29 strongly agreed, 16 agreed). This was echoed by teachers responding to the endline survey: the vast majority of teacher survey respondents also reported that the training was clear (45 out of 46: 21 strongly agreed, 24 agreed), useful (45 out of 46: 20 strongly agreed, 25 agreed), and detailed enough (43 out of 46: 21 strongly agreed, 22 agreed). A large majority of TAs and teachers also strongly agreed or agreed that it was possible to ask school-specific questions during the training (44 out of 45 TAs and 42 out of 46 teachers). However, four teachers disagreed or strongly disagreed with this statement.

TAs and teachers were also asked about whether the Elklan trainers provided useful answers to their questions, and whether they perceived Elklan trainers were knowledgeable about the practicalities of implementing NELI and using resources. Once again, TA and teacher reports were positive in this regard (as shown in Figure 9) with all TAs responding to the survey agreeing with these statements and all teachers, except one, similarly agreeing.

Figure 9: Endline survey responses—teacher and TA level of agreement to statements regarding the training sessions and ongoing support received



¹⁷ For TAs, this survey question did not distinguish between initial training (day one or day two) or the top-up training (day three mid-way NELI delivery).

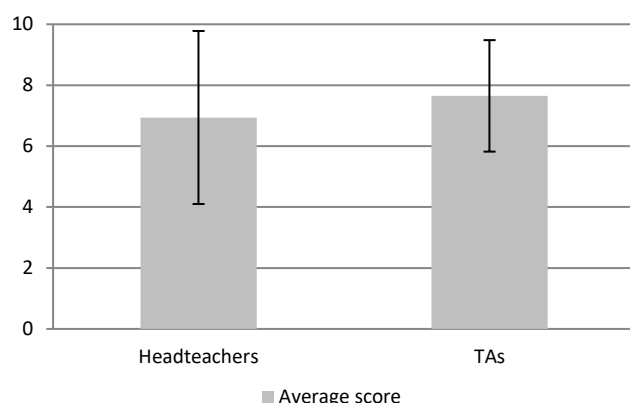
Note: Percentages on y-axis start at 60% to better show variation between responses. Teacher endline survey question 8, n = 46; TA endline survey question 10, n = 45.

The interviews largely corroborate this information. Interviewed TAs, teachers, and headteachers thought that the training was effective overall. They noted that the training was useful and informative in preparing TAs to deliver NELI in reception class (TAS5, TS2, TS6, TAS2, TA1S3, TA2S3, TA1S4, TA2S4, HTS5). Two interviewed TAs noted that the training was the right duration—in their opinion two days of training allowed practitioners to gain in-depth information on the aims and the format of the intervention (TAS6, TAS2). Another two TAs from different case study schools reported that the initial training sessions had a good structure and relevant topics were covered, which helped them transform their teaching practice (TAS1, TAS5).

Reported levels of preparedness to implement NELI

At endline, headteachers and TAs were asked about how prepared they felt to implement NELI following the information they received at the start of the trial. Headteachers were asked to rate their *school's* level of preparedness and their level of understanding of school requirements on a scale of 1 to 10 (with 1 indicating 'completely unprepared' and 10 indicating 'fully prepared'). TAs were asked a similar question to rate their level of 'preparedness for implementing NELI in their *classroom*'. In both instances, headteachers and TAs, on average, reported feeling closer to fully prepared than unprepared with TAs reporting feeling slightly higher levels of preparedness than headteachers (average scores of 6.94 for headteachers and 7.65 for TAs). As shown in Figure 10, there was less variability in views on preparedness among TAs compared with headteachers.

Figure 10: Average preparedness score for delivering NELI (with standard deviation)

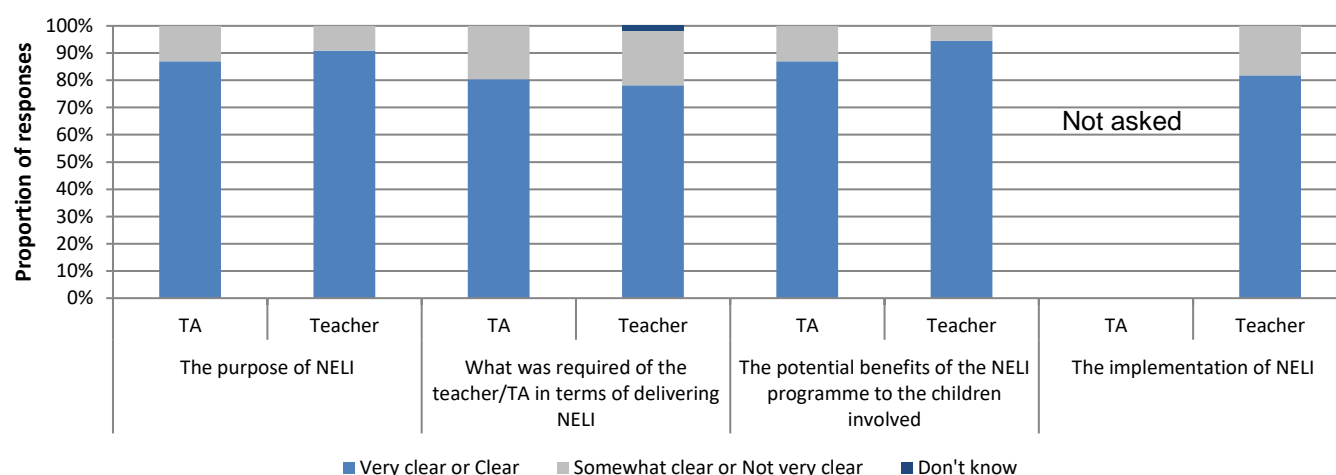


Question: 'How would you rate your overall preparedness for implementing the NELI programme? Please rate your preparedness from 1 to 10 with 1=completely unprepared; 10=fully prepared.' Headteacher endline survey question 30, n = 34, TA endline survey, n = 46.

In addition, in the endline surveys teachers and TAs were asked about *how clear* they were about the tasks entailed to implement NELI, the purpose of the programme, and potential benefits of NELI following the information they received about the programme. As shown in Figure 11, overall, the majority of teacher and TA survey respondents reported that they had clarity on both tasks and potential benefits; in all cases, more than 70% of survey respondents reported that these aspects were very clear or clear to them. There were no notable differences in the proportion of positive responses across the different aspects between teachers and TAs with most clarity reported for knowing about 'the purpose of the programme' and 'the potential benefits of NELI to children'. Only teachers were asked about whether they were clear about 'the implementation of NELI': the majority of teacher survey respondents (81.8%, 45 out of 55) said that they were very clear or clear about this (Figure 11).

Similarly, when asked about their understanding of the aims of NELI, all interviewees demonstrated having an accurate understanding of what the intervention sought to achieve. Frequently mentioned descriptions of NELI included: 'intensive, regular and daily [language intervention] aiming to improve pupil's language' (HTS1); 'help expand vocabulary' (TS6); 'guiding pupils to build narrative' (TAS2, TA2S3); 'help improve literacy' (HTS5), 'improve spoken language' (HTS1, TS2, TS5), 'phonics' (TS6, TS5), 'communication skills' (TS6, TS2, TA2S3) and 'confidence' (TAS5).

Figure 11: Endline survey response—teacher and TA perceptions about the communicated clarity of the following NELI elements



Note: Teacher endline survey, n = 55; TA endline survey question 7, n = 46. Wording variation between teacher and TA survey for second item: Teachers were asked, 'What was required of the teacher in terms of delivering NELI?' while TAs were asked, 'What was required of the TA in terms of delivering the programme?'

Compared to survey responses on the question about training perceptions (discussed in the previous section, Figure 9) responses to the question in Figure 11 show slightly higher proportions of *less positive* views voiced by respondents. For example, 20% of teacher survey respondents (11 out of 55) were only 'somewhat clear' or 'not very clear' about 'what was required of the teacher/TA in terms of delivering NELI'. This was also the statement for which a higher proportion of TAs reported having less clarity (19.5%, 9 out of 46 TAs) compared to the other statements. Despite this seemingly being the experience of a minority of survey respondents, participants in IPE activities identified a number of barriers to implementation; these are discussed in detail within the Implementation fidelity section.

Taken together, the fact that schools felt informed about the demands of delivering NELI and that NELI training was seen to be appropriate enough to allow TAs and teachers to deliver the intervention suggest that the training elements of the intervention were successful. The fact that training was so highly attended (see Compliance section) further suggests that NELI training was a core element of successful implementation.

Pupil selection and screening

A key aspect of the NELI logic model is that appropriate children are identified for inclusion in the intervention. To assess the appropriateness of the child selection, TAs were asked to report their agreement (or lack of) with two aspects of the process: (1) whether the children selected using LanguageScreen were suitable for the NELI intervention and (2) whether teachers and/or TAs should decide which pupils are entered into the programme.

According to IPE data, the majority of surveyed TAs and interviewed teachers and TAs reported concerns about the process of selecting children to participate in the NELI intervention. While this does not constitute objective evidence that the selection mechanism was problematic, school staff's perceptions of it, as expressed directly in the IPE surveys and interviews, are valid and contribute to an understanding of these key stakeholders' experiences with the intervention.

Figure 12 illustrates that 37 out of 46 TAs (80.4%) strongly agreed or agreed that teachers or TAs should decide which pupils enter the NELI programme. This was also reported by some interviewees (as shown by the quote below from TA2S4) who suggested that they were best placed to select children to take part in the programme. Several teachers and TAs articulated that if they were able to choose the children themselves, they would have chosen the children who have stronger learning skills (for example, concentration, listening) and language skills compared to the children that were selected (TS6, TAS6, TA1S3, TA1S4, TA2S4).

Overall, just over two thirds (67.4%, 31 out of 46) of TAs strongly agreed or agreed that some participating pupils were unsuitable for NELI, either because they were too advanced or had learning difficulties. Additionally, one out of ten open responses in the TA endline survey and two out of sixteen in the teacher survey reported concerns about

the selection of children. These concerns included unsuitability for EAL children (one TA and one teacher) and the inability to withdraw selected pupils deemed unsuitable by teachers/TAs prior to the beginning of the programme (one teacher).

This concern about the process for selecting children was also reflected in interviews with teachers and TAs. When asked about the selection of children, participants articulated contradictory views on whether the right children were selected to participate in NELI. Some respondents agreed that the children selected were suitable to receive the intervention. These views were informed by the perceived positive impact that NELI had on children's language development that participants observed (TAS5, TS2, TAS2, TA2S3, TS5, HT5).

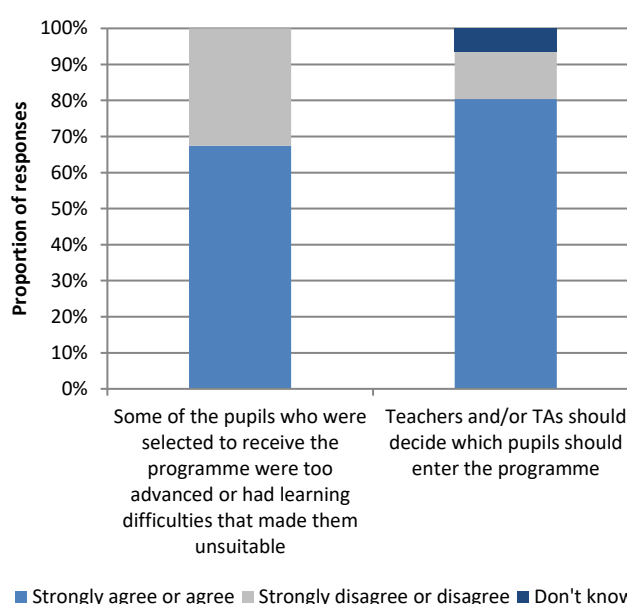
However, some interviewees voiced concerns on the appropriateness of the children selected. They explained that this was because some children selected through the LanguageScreen screening were of the lowest ability in the reception class and NELI did not have a large impact on them (TS6, TAS6, TA1S3, TA1S4, TA2S4). They added that if school staff were able to select the children to participate in NELI themselves, they would have chosen the children who have a higher ability than the ones selected (TS6, TAS6, TA1S3, TA1S4, TA2S4), as voiced by one interviewee:

'I think it would be so much better if we were able to pick [...] children. It did have some impact on the children that participated but I feel that there are other children in the class that it would have more of an impact than those children. The selection of children was made through an app called ATLAS¹⁸ and I think it chose all the low [attaining] children [...] and they are all SEN and EAL children. But I think that if it was someone who does actually speak English, children but it is low attaining, I think that it would have made more progress than someone who [was selected and] doesn't speak any English' (TA2S4).

Interviewee TA2S4 also highlights that some of the selected children were EAL pupils, which was reported by one other TA (TA1S4), noting that English-speaking pupils might have benefited more from NELI. However, this view was not widely reported by TAs or teachers in the case study schools. Three interviewees also identified that there were other children in the reception class who could have benefited from NELI (TS2, TAS2, TAS6).

However, given the results from the impact evaluation, it is interesting to note that pupils selected for the intervention did in fact benefit from the intervention, including those that were identified as having EAL. This is of key importance for future roll out of the programme as the selection of pupils may be adapted if NELI is scaled up. As noted, teachers and TAs may want to select different pupils despite the impact evaluation suggesting that the programme is effective at improving outcomes for eligible pupils.

Figure 12: TA perceptions of the selection of children into the NELI programme in the endline survey



TA endline survey question: 'Regarding the selection of children into the NELI programme, please indicate how far you agree or disagree with the following statements.' n = 46.

¹⁸ Please note LanguageScreen was known as ATLAS to school staff

Programme delivery

As specified in the Compliance section, the NELI intervention was intended to be delivered via three 30-minute group sessions and two individual 15-minute sessions per week. In this trial, the 20-week NELI programme was implemented across two terms between January and July, with ten weeks delivered in each term. Overall, data from the IPE suggests that NELI was delivered as intended, with the majority of headteachers (73.5%, 25 out of 34), teachers (69.0%, 38 out of 55), and teaching assistants (78.3%, 36 out of 46) reporting no concerns. Some aspects of delivery were noted as being negatively or positively associated with implementation, with time restrictions and perceptions of individual sessions being barriers to implementation, while space, the Oxford University Press resources, and support from senior staff were noted as supporting effective delivery. These aspects are discussed in turn below.

Time restrictions: preparation, session delivery, and workload

In the endline surveys and interviews, teachers and TAs were asked to provide their views on whether the NELI programme created any issues around the provision of time for delivery and whether the programme created any workload issues for staff. From this, it emerged that the NELI programme created a number of problems around time restrictions, particularly the time required to prepare for NELI sessions, the time taken to deliver the sessions, and the increased workload resulting from the NELI programme.

First, time emerged as an overarching issue in relation to the time required to prepare for NELI sessions. Participating settings were provided with the necessary materials to implement NELI, but some additional work was required prior to their use. When asked about the initial preparation and set up, all interviewees (n = 12) noted that they received the resources to deliver NELI in a box which arrived at the school. Three interviewees noted that the TAs had to spend time to prepare and cut the resources before delivering NELI (TA1S3, TS6, TAS6). They also noted that this task sometimes took longer than expected. Two TAs suggested that more time to prepare, plan, and sort out the resources should be provided (TAS2, TA1S3), as shown below:

'I think what would have been helpful if the resources had already been prepared and sorted out; that would have saved us a day out of school. To appreciate time is of the essence at school. We had time out of school to prepare the resources and that was lovely but we had other interventions before where the resources and the whole box just arrived and everything was already there. That would have been easier. The resources did not arrive late but then need to be sorted out, photocopied and cut up. You couldn't just pick up the programme and go to the designated area and deliver it. We had a lot of preparing to do before we were able to do that. The materials were already there but they had to cut them up. My colleague and I put them to envelopes according to the sessions (for example, session 1) in order to be able to deliver it' (TA1S3).

One TA noted that she felt overwhelmed at the beginning of NELI implementation because of the amount of preparation and planning required from her side to deliver NELI (TA2S3). However, she mentioned that the initial challenges she faced were overcome as the intervention progressed:

'Prepared me well but until you actually do it ... I felt overwhelmed in the beginning with the paperwork; I can understand that was probably it was a trial and there was a specific amount of paperwork for the assessment to be done but after that it was quite easy to follow up the intervention. We had the handbook that tells you what to do and it is pretty straightforward in that sense. Until you do an intervention and you get used to it how it gets, it is hard' (TA2S3).

The second theme around time that emerged within the IPE was the time taken to deliver the sessions. Endline surveys reveal that the majority of TAs (73.9%, 34 out of 46) strongly agreed or agreed that the duration of NELI sessions was adequate, while 12 out of 46 (26.1%) strongly disagreed or disagreed with this. The mixed views from the surveys were corroborated by interviews conducted with TAs and teachers. Three interviewees expressed concerns on the length of some NELI sessions and activities (TS6, TAS6, TA1S4). They commented on the need to extend some of the sessions and activities due to questions raised by the children (TS6, TAS6) and the large number of activities included in each session:

'Yes, the intervention is good but I think that in some sessions there are too many activities to fit in that session. Sometimes you find yourself rushing to fit the activities into one session where you could shorten the activities and spend more time with the child. Sometimes you find yourself saying "I need to do this, I need to do that, I need to do this" and it sometimes can be stressful' (TA1S4).

One interviewee noted that there was limited time to conduct the evaluation of pupil progress at the end of each session (TA2S3) explaining that this was due to the busy school timetable. Indeed, a prominent issue emerging from the TA, teacher, and headteacher endline surveys was the difficulty faced by practitioners in fitting NELI sessions into the timetable. Endline surveys show that the majority of headteachers (55.9%, 19 out of 34), teachers (70.9%, 39 out of 55), and TAs (56.5%, 26 out of 46) agreed to a great extent or moderate extent that this was difficult. Just two headteachers, three teachers, and no TAs reported that this was not an issue at all. Several interviewees also noted that busy school timetables were a key barrier to the smooth implementation of NELI (HTS1, TAS5, TS6, TA1S3, TA2S3, HTS5, TS5).

Another perceived barrier to delivery was excessive session length—especially, as two interviewees noted, for children of the lowest ability (TS6, TA2S4). This highlights an overlap between the perceived issue of session length and the perceived unsuitability of NELI for selected children:

'Maybe if they could make the individual sessions once a week rather than twice a week and the group sessions could be 15 minutes—because the children don't listen for 30 minutes. Because, as we said before, some of them are SEN kids' (TA2S4).

A number of teachers and TAs also explained that a key barrier to implementation was the large amount of time required to deliver the individual sessions (TAS5, TAS6, TAS2, TA1S3, TA2S3, TA2S4, TS6). While this was often framed in relation to the demands placed on staff workload (TAS5, TA2S4), one interviewee (TS6) also mentioned that time spent delivering individual sessions may be to the detriment of other children in the class, as shown below:

'NELI was so prescriptive and it took such a big amount of time. It negatively affected not every child but I can think of a handful of children that they would have benefited from additional interventions but they just didn't receive ... [other interventions] because NELI was so time consuming' (TS6)

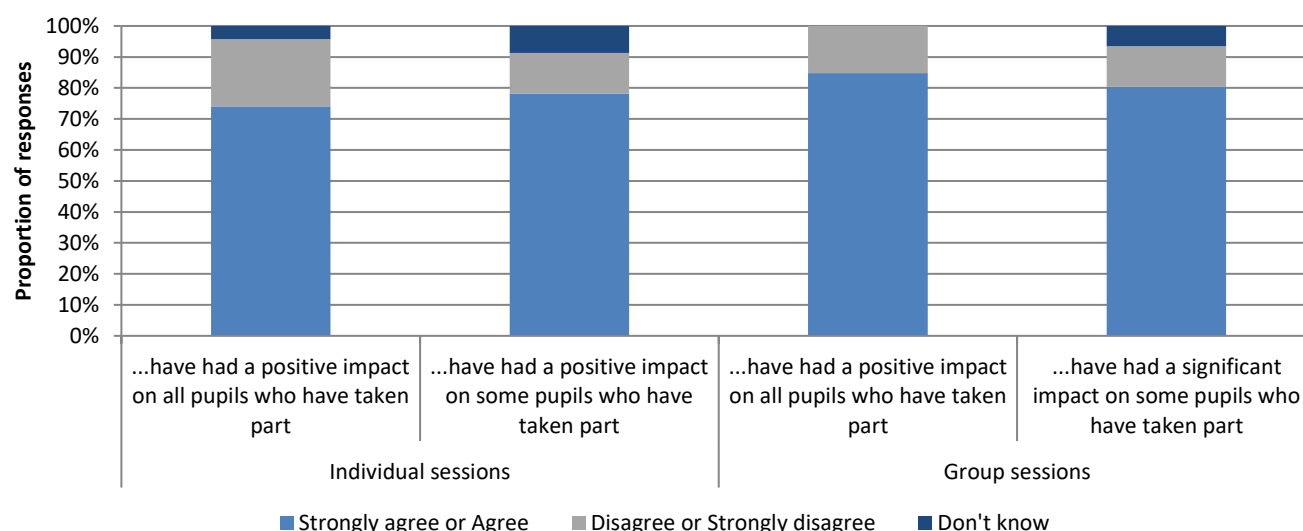
This interviewee expressed that they would like the potential benefits of NELI and time with TAs to be extended to more of the class, rather than to a single pupil.

Group and individual NELI sessions

Interviewees were asked to comment on the usefulness of group and individual sessions for delivery of the NELI programme. Given the fact that there was a notable variation in how these sessions were delivered (see Compliance section for more details) we looked at schools' perceptions of these sessions in more detail. The Programme delivery section on time restrictions above has already illustrated that NELI sessions (especially individual sessions) were too long for many practitioners. This section will focus more on why this may have been the case, looking in more detail at the perceived usefulness of group and individual NELI sessions for delivering the programme.

As shown in Figure 13, the majority of TAs agreed that individual and group NELI sessions had a positive impact on all or some of the pupils that had taken part in the intervention. High proportions of TAs strongly agreed or agreed with the statements that *group sessions* benefited all (85%, 39 out of 46) or some pupils (80%, 37 out of 46). Interviewees generally expressed positive views on the usefulness of group sessions. Respondents observed that group sessions were having a positive effect on participating children's learning and language skills (TAS5, TAS6, TS2, TAS2, TA1S3, TA2S3, TA1S4).

Figure 13: TA respondents' perceptions of the impact of individual and group NELI sessions



Note: TA endline survey, n = 46.

One TA explained that sometimes the children needed more time to develop their understanding of specific elements in the group sessions (TAS6) or that at the beginning of the intervention children found it difficult to focus during the group sessions, but this issue was overcome as they got used to the schedule (TAS5). TAS6 explains that they would sometimes dedicate more time to group sessions to ensure that pupils understood the content of the session:

'I think on the one hand it is enough time because if it starts running any longer, it would might have put people off. But personally, sometimes I would spend a bit more time—longer sessions in order to make sure that the children understand and they are getting it. I don't know if it is just my group. When we first started, you would not believe it but words that they couldn't understand I would go over again to look over. They seem to take a while to understand things. When there were words that I felt that they still needed to practice, I would look them over again and it sometimes took longer as they didn't always understand' (TAS6).

The above quote (TAS6) indicates a slight deviation from the implementation plan where the TA perceived that the structure or format of the group sessions was not appropriate.

In contrast, respondents were less enthusiastic about individual sessions: compared to group sessions, fewer TAs strongly agreed or agreed with the statements that *individual sessions* benefited all (74%, 34 out of 46) or some pupils (78%, 36 out of 46). Interviewees expressed mixed views around the usefulness and appropriateness of individual NELI sessions. Practitioners liked that TAs could target specific children and adapt the sessions according to pupils' needs through the individual sessions (TAS6, TS2, TA1S3, TA1S4). One TA noted that the individual sessions could be used to supplement skills developed in the group sessions, and could be helpful if a child misses one of the group sessions:

'The individual sessions are very useful because you target that child more if they haven't understood that in the group session. If the child has missed the group session you could catch up in the individual sessions. I think that the individual sessions were very useful because you are helping the child to understand more' (TA1S4).

Some interviewees identified elements of individual sessions that impeded the effective delivery of the NELI programme. As shown in the Staff cost section, the time taken to deliver individual sessions was deemed by many to be onerous. One interviewee (TS6) mentioned that time spent delivering individual sessions may be to the detriment of other children in the class. This interviewee expressed that they would like the potential benefits of NELI and time with TAs to be extended to more of the class, rather than to a single pupil. Another interviewee (TAS2) expressed concerns that individual sessions might have negatively impacted the development of participating pupils as they missed other reception class activities. Furthermore, there seemed to be a perceived lack of guidance in the handbook around how to deliver individual sessions, which made it harder to run these sessions. A teacher from one school (TS2) also noted that they had encountered difficulties in engaging children in the individual sessions compared to the group sessions. In the excerpt below, TS2 corroborates this by reporting that the participating children enjoyed

the group sessions more than individual sessions as this enabled collaborative working, which seemingly stimulated higher engagement:

'The individual sessions were difficult sometimes in terms of engaging children because they like working with one another in the bigger groups. So maybe the individual time could have been shorter or a little bit different. I think that was the feedback I got. They really enjoyed the group sessions but the individual sessions could be different in order to engage children a bit more' (TS2).

The lack of fidelity—and problems surrounding individual session delivery—may, therefore, reflect the perception that they were more difficult to implement or were less beneficial compared to group sessions. Although feedback was limited, this was raised as an issue and may indicated an area where training could be improved, particularly relating to how teachers and TAs might adapt the programme in order to deliver the individual sessions within busy school timetables.

Resources: suitable spaces for delivering NELI sessions and usefulness of Oxford University Press resources

A major focus of the IPE data collection activities was to assess whether TAs were provided with suitable resources and conditions to deliver group and individual NELI sessions. These are key elements and contextual factors on the NELI logic model as poor conditions or resources could result in NELI being delivered with less fidelity. According to data collected as part of the IPE, the majority of settings had space to deliver NELI and felt that resources were appropriate for delivery. We discuss these in more detail below as they may be key factors that support effective implementation of NELI.

In the endline surveys, headteachers, teachers, and TAs were asked about whether there were difficulties finding a separate space within the school to deliver individual and group NELI sessions. Finding space was not problematic for the majority of practitioners. Indeed, only a minority of headteachers (35.3%, 12 out of 34), teachers (47.3%, 26 out of 55), and TAs (37.0%, 17 out of 46) agreed to a great or moderate extent that it was difficult to find a separate space to deliver NELI. Furthermore, 11 out of 34 (32.4%) headteachers, 18 out of 55 (32.7%) teachers, and 19 out of 46 (41.3%) TAs did not agree with this statement at all, indicating mixed feelings among practitioners.

Many interviewees noted that they had a space to deliver NELI, either a quiet area (for example, class, intervention room, library, staff room, dining room; TS6, TAS6, TS2, TAS2, TA1S3, TA2S3, TA1S4) or a corridor space near the reception class (TS5). Two of these respondents (TA2S3, TS6) also noted that they had more than one space to deliver NELI, in case the preferred option was not available.

Where issues were noted, they mainly related to the fact that the allocated space was not appropriate for the delivery of NELI due to the distractions in that area (TAS5, HTS5):

'We didn't have a dedicated space. That was in a corridor and I had to move my time according to when the corridor was free in order to deliver NELI. The corridor was usually available at the same time every day and I managed to keep it at the same time' (TAS5).

In the endline survey, TAs were asked to comment on the provision of resources and materials for delivering NELI sessions. TA responses around the availability and appropriateness of programme resources were generally positive, as shown in Figure 14. Indeed, the vast majority of TAs (87.0%, 40 out of 46) agreed that they had all the information and resources that they needed to implement NELI. There was less unanimity among TAs around views on the appropriateness of the resources for use with the selected children. While the majority of TAs (71.7%, 33 of the 46) agreed with this statement, 12 out of 46 (26.1%) TAs disagreed with this.

Several interviewees expressed positive views on the quality of the materials provided to the school (HTS1, TAS5, TA1S3, TA1S4, TA2S3, TS2, TA1S4, TA2S4).

One interviewee (HTS5) noted that staff initially felt overwhelmed by the materials and resources received. However, the respondent noted that the usefulness of the training, combined with the collaborative approach to learning about delivering NELI among the TAs, allowed these resources to be effectively understood and utilised. It appeared in this case that once the user familiarised themselves with the materials provided, implementation of NELI became easier.

Some interviewees voiced concerns about the quality and appropriateness of the materials provided to settings.

Three interviewees (two from the same school) mentioned that the illustrated cards sometimes were not at an appropriate level of difficulty or the pictures themselves were not relatable to child's previous experience (TS6, TAS6, TAS2). Some of the resources were beyond the experiences of children and therefore it was difficult for children to understand the meaning of some pictures. As a result, teaching assistants were spending additional time explaining the pictures rather than discussing a child's experiences in relation to the cards.

These practitioners reported that these resources did not feel tailored to the pupils participating in the programme, making it harder for these children to understand some topics. In particular, the resources were not tailored to children from more socially deprived areas as cards featured, for example, holiday pictures of a palm beach. Interestingly, these sentiments, though minor, reflect concerns around selection of pupils (see pupil selection and screening section) and may stem from the same feelings that NELI is more suitable for some pupils than others.

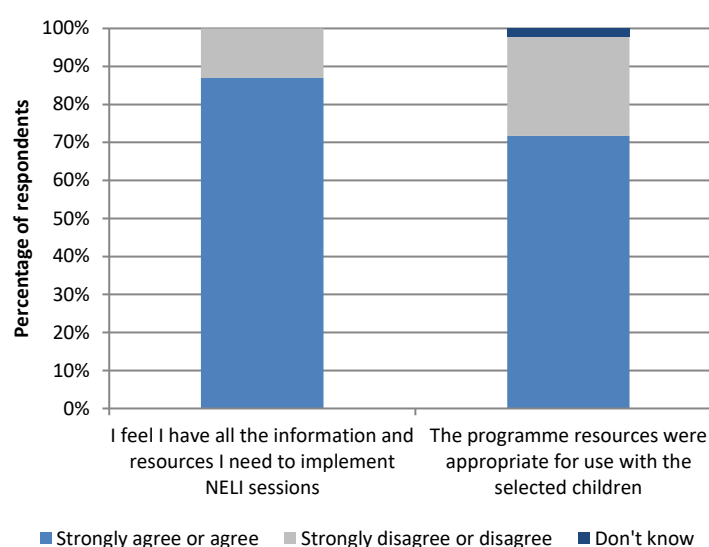
Based on the above, it is clear that the majority of respondents felt happy with the usefulness of the NELI resources and within-school provisions for delivering NELI. This runs slightly contrary to compliance findings that found that not all sessions were delivered as intended. However, taken in conjunction with the findings on time, it suggests that poor conditions and resources did not exacerbate problems with delivery. In fact, the majority of respondents felt these elements supported successful delivery.

Support for TAs in delivering NELI

The previous evaluation of the NELI programme (Sibieta et al., 2016) suggested that one of the key barriers to implementation was a lack of support for TAs from senior staff. However, results from the endline surveys reveal that the majority of TAs felt that they had the support of headteachers and teachers when delivering NELI.

Importantly, the vast majority of TAs (89.1%, 41 out of 46) responding to the endline survey reported that they strongly agreed or agreed with the statement that they have the support of colleagues in delivering NELI sessions. Interviewees noted that the main type of support TAs received was in terms of securing time for NELI preparation by dedicating time in the TA's daily schedule and by organising cover for TAs if needed (HTS5, TA1S4, TS5, HTS1). Interviewed

Figure 14: TA responses to survey questions about the supply and appropriateness of resources provided for NELI delivery



Note: TA endline survey, n=46.

headteachers and teachers considered the preparation and delivery of NELI as an activity that they needed to work around to ensure they could fit this within standard job duties. This required consideration due to busy school timetables. While introducing NELI might have meant additional work or planning, participants did not necessarily perceive this negatively. In fact, school staff noted that they made an active effort to fit NELI in their school timetable (HTS1, TS2, HTS5) because they perceived that NELI would have a positive impact on pupils.

TA respondents were also asked if they felt supported by other staff in delivering NELI, namely Ekklan trainers. In fact, ongoing support from Ekklan trainers was perceived positively by TAs and teachers with all 46 TAs strongly agreeing or agreeing with the usefulness and knowledge of trainers in answering questions and providing information on the practicalities of implementing NELI and using resources. Again, teachers supported this view, with 45 out of 46 (97.8%) strongly agreeing or agreeing with both statements. Interviewees also reported that they valued the ongoing support provided by the trainers; this helped them deliver the intervention smoothly (TAS5, TAS2, TA1S4, TA2S4, TAS6, TA1S3, TA2S3).

In contrast to previous findings, the majority of TAs felt supported by senior staff. This may reflect changes that have been made to NELI training and early communication of the requirements of the intervention. As the previous section on Attendance at training sessions and use of ongoing support while delivering NELI has noted, schools felt informed about the demands of delivering NELI from the start. Taken together, this suggests that senior staff support is a critical element of successful implementation and that one way to encourage this is to ensure high quality training and effective communication of intervention requirements.

Perceived outcomes associated with NELI

Box 3: Perceived outcomes key findings

Responses by TAs and teachers to their respective endline surveys and data from interviews with TAs from selected case study schools lead to the following conclusions:

1. IPE data suggests that the intervention had a positive impact on TAs implementing NELI. TAs, teachers, and headteachers identified a positive impact in terms of improved knowledge of language and of teaching.
2. Survey responses show that participants were slightly more positive about the impact that NELI would have on EAL learners than on pupils with SEN or other children in the classroom. Some interviewees considered that NELI was least suitable for pupils with SEN.

The NELI logic model specifies that the programme's route to impact is linked, first, to improving the knowledge and, consequently, the teaching of language skills of TAs and teachers (see Figure 2). By achieving this improvement it is expected that pupil oral language will improve (thanks to improved language instruction).

While the Impact evaluation section has discussed pupil-level outcomes on specific language measures, the following section looks at other key outcomes using data collected as part of the IPE. Specifically, it explores the impact of NELI training on TA outcomes and perceptions of impact on pupils in the classroom. These are discussed in turn below.

The impact of NELI on teaching assistants

Interviewed TAs, teachers, and headteachers identified that the programme had a positive impact on TAs in terms of improved knowledge on language instruction and confidence.

Headteachers and teachers commented on the impact that NELI had on TA skills, and TAs themselves corroborated these views. There were overall positive views expressed on the impact that NELI had on TAs (HTS1, TAS5, TAS6, TAS2, TA2S3, TA2S4, HTS5, and TS5). TAs were considered to have developed new ways of teaching (HTS1, TAS5, TAS6, TS2, TAS2, TA2S3, HTS5); developed their relationship with children (TAS5, TS5); improved their listening and communication skills (TAS5, TS2), and their management of children's behaviour (HTS5). Data from the interviews indicated that NELI was perceived to have helped increase the confidence of TAs (TA2S4, HTS5, TS5, TAS5, TA2S3).

One TA commented on how her teaching practice had changed as a result of participating in NELI. She commented that prior to using NELI she was hesitant to use certain words with pupils as she thought pupils would not understand

them. However, through NELI she realised the importance of exposing children to a wide vocabulary. NELI improved the TAs knowledge of language instruction and provided the tools (teaching skills) to better communicate with pupils:

'Some of the vocab use I felt that I cannot use it because they will not understand that ... I thought that some things might be a bit apprehensive and they will not understand them but there was no need to think that because if the children are not exposed to these words, they will not know them. I think that I shouldn't felt that way. I did get a lot out of the definitions of the words. I struggled to explain what things mean; so that was really helpful—explaining things in simpler terms' (TA2S3).

Three other interviewees noted that TA's have put into practice the knowledge and skills acquired through NELI with their reception class more widely (TS2, TAS2, HTS5).

Nonetheless, three interviewees reported that they did not observe any impact of NELI on the knowledge and skills of TAs (TA1S3, TS6, TA1S4). In two instances this was because the TA already had experience in supporting language development.

Some interviewees noted that one factor that might facilitate positive impact of NELI on TAs is the creation of strong relationships between TAs and children (TS6, HTS5) that was facilitated through the regularity of the sessions. For example, interviewees from different schools noted that the regularity of NELI and having small group sessions helped build pupils' enthusiasm (TS2), confidence (TS2, TAS2, TS6), and strengthen the relationship between children and between children and TAs (HTS5, TS6).

The skills requirements of staff to deliver NELI/early language interventions

It has been noted in previous research that practitioner skill is considered critical to the quality of early years education (Martinez-Beck and Zaslow, 2006). Findings from the endline surveys and interviews reflect this, with respondents reporting that TAs and teachers require a number of specialised skills in order to implement NELI/early language programmes.

Responses to endline surveys among headteachers, teachers, and TAs reveal a strong consensus among the practitioners that highly capable TAs and teachers are required for implementation of most early language programmes, with the vast majority of headteachers (97.0%, 33 out of 34), teachers (96.4%, 53 out of 55), and TAs (87.0%, 40 out of 46) reporting that they strongly agreed or agreed with this statement.

These endline survey responses were reinforced by findings from the interviews. Respondents reported a number of specific skills required by TAs to deliver the intervention, including the ability to communicate with children and build relationships/rapport (TS6, TS2, TA1S4, TS5), time keeping (TAS5, TA1S4, TA2S4), and the ability to make the intervention engaging and enjoyable for the children (TS2, TA2S3, TA1S4). Other skills noted included organisation skills (TAS5, TS5), having patience (TA2S3), being creative/inspirational (TS6), having good knowledge of phonics (TS6), and being role models in terms of language, speaking, and listening (TS5).

In contrast, only one teacher and two TAs noted that TAs do not need to have prior skills or training on working with children in small groups in order to effectively implement NELI (TS2, TAS2, TA1S3). TAS2 noted that as long as TAs were able to follow a scheduled lesson plan, they should be able to deliver NELI successfully.

In sum, IPE data strongly suggests that staff felt that NELI had the potential to be particularly beneficial for TAs and educational staff with emerging knowledge about children's language skill development, while at the same time, the majority of respondents felt strongly that TAs and teachers require a number of specialised skills in order to implement NELI/early language programmes. Overall, the above evidence suggests that while a minimum skill set is needed to deliver NELI successfully, NELI training itself is a good way of improving TA language development knowledge, teaching skills, and confidence.

Perceived outcomes on pupils

While the Impact evaluation section has discussed pupil-level outcomes on specific language measures, the following section looks at perceptions of NELI's impact on pupils in the classroom, including what areas of language showed the most improvement and varying impact for different groups of pupils. These are discussed in turn below.

Overall, interviewed participants corroborated impact findings, observing that NELI had positive impact on children's language and communication skills. All interviewees from the six case study schools noted that they observed that NELI had positive impact on at least *some* of their pupils (HTS1, HTS5, TAS5, TS6, TAS6, TA1S4, TA2S4, TS2, TAS2, TA1S3, TA2S3, TS5). This view was shared among headteachers, teachers, and TAs. A teacher expressed being 'impressed' and a TA from a different case study school being 'surprised' by the progress that pupils made after receiving NELI (TS5, TA2S3).

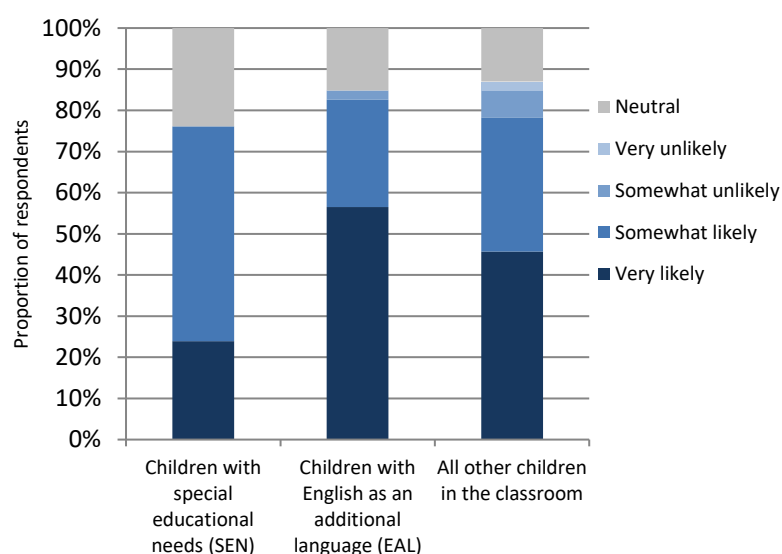
Similar to the above, teachers and TAs from other case study schools also commented that they observed improvements in the vocabulary of the children (TS5, TS6, TAS2, TA2S3). Similarly, interviewees also identified improvements in the narrative and storytelling skills of children (TS5, TS6, TAS2, TA2S3) and in their overall communication and listening abilities (TS2). In addition, interviewed teachers and TAs noted that the attention skills (TS2), engagement levels (TAS5, TA1S3), and enthusiasm to participate in class (TS6, TAS6) of children increased as a result of NELI. Many teachers and TAs also observed children's confidence to communicate improved (TAS5, TAS6, TS2, TAS6, TA1S3, TA2S3). For example, the following quote summarises the views from staff from case study School 2 on the impact they observed NELI had on pupils receiving it:

'I have noticed that a lot of my children have been more confident now that they have taken part in NELI to speak out in group situations more. Improvement on communication, understanding, listening and attention skills' (TS2).

Nonetheless, these *views were not uniformly* held and one teacher interviewed raised the fact that while they observed improvements in children's language skills, the teacher was not sure whether this was a result of NELI or because of the phonics programme they were teaching in that class (TS5). Interestingly, children receiving NELI were also benefiting from an additional phonics programme at their school. This links to what 'usual practice' at schools is, which is discussed in the section on Usual practice in more detail.

Interestingly, participants noted that NELI had varying impact on different groups of pupils. Overall, IPE data shows that participants reported that NELI had less impact on pupils with Special Educational Needs (SEN) compared to the impact of the programme on EAL learners. As shown in Figure 15, in the endline survey 76.1% (35 out of 46) of TAs agreed that it was 'very likely' (11 out of 46) or 'somewhat likely' (24 out of 46) that NELI would lead to improvement in the oral language and vocabulary skills of *SEN pupils*. This is a slightly *lower proportion* compared to the number of TAs reporting it was likely that NELI would have a positive impact on *EAL learners*: 82.7% (38 out of 46) of TAs responding to the survey reporting it was 'very likely' (26 out of 46) or 'somewhat likely' (12 out of 46) that NELI would lead to improvements in the oral language and vocabulary of EAL learners (Figure 15).

Figure 15: Respondents views on the likelihood that NELI will lead to improved oral language and vocabulary for particular groups of children



TA endline survey, $n = 46$. All survey respondents were asked to report their views on the following statement: 'Regarding the impact of NELI so far, please indicate how likely it is that NELI will lead to improved oral language and vocabulary for the following groups of children.'

With regards to the impact of NELI on phonics (a precursor to reading), similar to the above, variation in TA's reports on the potential impact on phonics by pupil group is observed: a *slightly lower proportion* of TAs (69.6%, 32 out of 46: 'very likely', 8 out of 46; 'somewhat likely', 24 out of 46) reported positive views on the impact NELI could have on phonics for SEN pupils compared to its impact on EAL learners (73.9%, 34 out of 46: 'very likely', 18 out of 46; 'somewhat likely', 16 out of 46).

A similar trend was observed in the responses of *teachers* in the endline survey on whether NELI would lead to improvements in the *oral language and vocabulary skills* of pupils (see Figure 15). Notably *fewer* teachers reported that it was 'very likely' or 'likely' that NELI would lead to improvements for SEN pupils (in total 56.3%, 31 out of 55: 'very likely', 13 out of 55; 'likely', 18 out of 55) than for EAL learners (in total 81.8%, 45 out of 55: 'very likely', 22 out of 55; 'likely', 23 out of 55).

Teachers and TAs expressed similar views to the above during interviews. While some reported observing some improvements on SEN pupils' skills (TAS2, TS6), there were overall *less positive* views on the impact that NELI had on *SEN pupils* as compared to other pupils (TS6, TA2S3, TA1S4). As noted in the

Pupil selection and screening section, some interviewees considered that NELI was not appropriate for SEN pupils due to these pupils' learning difficulties inhibiting their ability to match the learning progress of other non-SEN pupils (TS6, TA1S3, TA1S4).

In conclusion, data from the IPE evaluation sustains the possibility that the programme's route to impact is achieved through improved knowledge and consequently the teaching of language skills of TAs and teachers. Respondents also noted that pupil oral language improved, though there was a belief that the programme would have less benefit for SEN pupils. Interestingly, this mirrors earlier discussions on pupil selection, where respondents felt that the programme was more or less suitable for some pupils. As before, this will be of key importance for future roll-out of the programme as the selection of pupils may be adapted if NELI is scaled up.

Usual practice

Box 4: Usual practice key findings

Responses by TAs and teachers to their respective endline surveys and data from interviews with TAs from selected case study schools show that:

1. The use of language programmes is a very active space in schools participating in the trial. Both intervention and control schools used other language programmes prior to, and during, the implementation of the NELI trial.
2. One third of control schools implemented a language programme while the trial was ongoing, and there were two instances suggesting minor contamination (two control headteachers reported that they used NELI during the trial year).
3. Overall, the other programmes used by control schools (or in intervention schools alongside NELI) were similar to NELI; these were mostly delivered by TAs in small groups or in individual sessions (a mix of these were reported), lasted between 'five to ten weeks', and were delivered frequently (for example, in every session or once a day).

For the purposes of this evaluation, usual practice was explored in two ways: looking at the context of what schools were doing prior to delivering NELI (at baseline), and also at what schools were doing instead of delivering NELI (in control schools) or alongside NELI (in intervention schools). All of these support an understanding of context and help understand what elements are in place that might facilitate delivery of the programme as well as an understanding of how active schools are in terms of establishing the counterfactual. These will be discussed in turn below.

Usual practice in all schools at baseline

Control and intervention headteachers were surveyed at baseline before programme implementation (November 2018) and asked about the use of other language programmes (excluding NELI) in their schools. Slightly less than half of intervention headteachers responding to the survey (47%, 18 out of 38) and slightly over a third of control headteachers (37%, 21 out of 57) reported that their school was using at least one programme that aimed to improve the language skills of children in reception classrooms. Over a third of control schools (39%, 7 out of 49) were taking in part in two, while two of the schools that responded were implementing three or more language programmes for reception classrooms. We do not have systematic data across all schools, but it seems reasonable to infer that schools

allocated to the control and intervention group were implementing similar language programmes prior to starting NELI. Box 5 provides details on the specific programmes they mentioned.

Box 5: Baseline survey responses—other language interventions mentioned by headteachers

Intervention schools

In the baseline survey, 18 out of 38 intervention headteachers reported their school used one of the following language programmes:

- Wellcomm, Elklan, Talk Boost, Over and Over, Blast, Word Aware, Chatter boxes and Talk partners;
- One headteacher reported using NELI prior to the start of the trial in addition to other programmes.¹⁹

Control schools

In the baseline survey, 21 out of 57 control headteachers reported their school used the following language programmes:

- WellComm, Elklan, Talk Boost (Early Years Talk Boost), BLAST, Time to Talk, Blank level groups, Lego Therapy, Read Write Inc., Specialist SALT, Talking Success, Pre-Teaching Vocabulary, Communication Cook Book, Pearsons Early Language Booster, Speech Link, and Word Aware
- Two headteachers reported programmes they did not know the name of, that is, responded 'N/A' when asked the name of the programme.

This suggests that there was a relatively strong culture of early language intervention in schools. To investigate this further, we looked at the perceptions of staff in intervention settings to understand schools' perceptions of early language programmes in schools: headteachers, teachers, and TAs were asked about their views on early language programmes in their respective endline surveys. There was almost unanimous agreement among headteachers (100%, 34 out of 34 headteachers), teachers (95%, 52 out of 55 teachers), and TAs (100%, 46 out of 46) that early language programmes involving small group interactions are more beneficial than whole-group sessions.

Usual practice in control schools during the trial

Below we present what 'business as usual' looked like in terms of language practices for children in reception classrooms in the control schools. We also report on 'contamination', in other words, whether the control schools accessed NELI resources. Given the strong culture of early language intervention in schools, the evaluation also looked at what schools in the intervention arm were doing in addition to NELI. Interestingly, this was a very active space, with schools using a number of additional programmes to support early language development.

In the endline surveys, headteachers and teachers were asked about what language programmes other than NELI their schools used while the trial was ongoing.

As detailed in Box 6, 17 out of 49 control headteachers (35%) reported that they were taking part in at least one programme that aimed to improve the language skills of children in reception classrooms at the time of the NELI trial. Based on the surveys, it appeared that it was also common for intervention schools to continue implementing other language programmes. At endline, 6 out of 34 headteachers from intervention schools reported using other language programmes in addition to NELI. We do not have systematic data across all schools, but it seems reasonable to infer that intervention schools have reduced the number of other language programmes in the school, and potentially means that delivering NELI was time consuming and staff had less time to continue implementing other language programmes.

In addition, when asked about language training, 28% of TAs surveyed at endline (13 out of 46) said they received training other than for NELI over the course of the 2018/2019 school year (while NELI was implemented), another 28% (13 out of 46) that this was received over the course of the prior school year (2017/2018), and another third (30%, 14 out of 46) that they received training before the 2017/2018 school year. Nonetheless, slightly more than a

¹⁹ Implementation of NELI prior to the start of the trial was exclusion criterion for participation. We consider this as a reporting error.

third of TA survey respondents stated they had received *no prior training* (35%, 16 out of 46). This is corroborated by interview responses: interviewees noted that other interventions used at their school (prior or during the trial) shared similarities with NELI and that these other interventions mainly reinforce and complement NELI but do not replicate it (TAS5, TS6, TS2, TAS2, TA2S3, TA1S4, HTS5, TS5).

It should be noted that when asked about what language interventions were being implemented in control schools two headteachers from the control schools reported using NELI. Even though we cannot rule out a risk of contamination we believe that this is more likely to be a response error as implementation of NELI prior to the start of the trial was one of the criteria precipitating exclusion.

Box 6: Endline survey responses—other language interventions mentioned by headteachers and teachers

Intervention schools

- In the endline survey, 6 out of 34 intervention headteachers reported their school used the following language programmes: Wellcomm, Elklan, Talk Boost, Over and Over, Time to Talk, and Chatterway.
- In addition, in the endline survey, 13 out of 55 intervention teachers reported using programmes apart from the usual curriculum. Apart from those mentioned by headteachers they listed WordAware, Easy Peasy (elements of language), Philosophy for Children, Helicopter Stories, First Call, Hanen strategies, Talking Tales, and 'advice taken from Early Excellence visits'.

Control schools

- In the endline survey, 17 out of 49 control headteachers reported their school used the following language programmes: WellComm, Elklan, Talk Boost, BLAST, Time to Talk, Blank levels, Lego Therapy, RWI (Read Write Inc) phonics, Speech and language therapist/specialist (SALT), Beanstalk Story Starters, Colourful Semantics, Phonological awareness programme, Speech Link, Sound linkage, Listen with Lola, and PORRIC. (Three headteachers reported programmes they did not know the name of, that is, responded 'N/A' when asked the name of the programme.)
- Two headteachers reported using NELI.

The other language programmes used in control and intervention schools shared some similarities with NELI

Some of the other programmes control and intervention headteachers reported using exhibited some similarities to NELI. In what follows we focus on the characteristics of programmes reported by control headteachers since these were similar to the programmes reported by intervention headteachers (and teachers). Overall, the other programmes used by control schools (or in intervention schools alongside NELI) were similar to NELI; these were *mostly* delivered by TAs, in small groups or in individual sessions (there was a mix of these), lasted between 'five to ten weeks', and were delivered frequently (for example, in every session or once a day). In the responses of control headteachers at baseline, 15 different interventions were mentioned plus an additional two that were not explicitly named. In terms of *who delivers the programmes*, from these 17 different interventions, nearly half were said to be delivered by TAs (7 out of 15), six were said to be delivered by both TAs and teachers, with the remaining interventions being delivered by Speech and Language therapists (SALT) either alone or alongside 'SALT TA'. In terms of *mode of delivery*, seven (of the 17) reported interventions were delivered in 'small groups', similar to NELI. Apart from these, six were said to be delivered through 'one-to-one' instruction, one with the 'whole group', while Wellcomm was listed as using all three modes of delivery. Interestingly, Wellcomm was the most commonly reported programme reported by control headteachers at baseline.

In terms of the *duration* of the programmes, headteachers were asked whether the programmes lasted 'five to ten weeks', '11 to 20 weeks', or 'more than 20 weeks'. The majority of the programmes were said to last between 'five to ten weeks'. However, some of them were reported to have different durations by different respondents, suggesting that the programmes can be delivered in different formats. (Once again, Wellcomm stands out by being reported as a programme that had all three durations.)

In regard to the *delivery frequency/dosage*, no clear patterns emerged. Respondents reported that five out of the 17 programmes were delivered 'once a day', five 'once a week', and three 'in every session'. In addition, four of the reported programmes (BLAST, Lego Therapy, Talk Boost/Early (Years) Talk Boost, and Welcomm) were said to be delivered with a different frequency by different respondents.

In the endline survey, a similar picture emerged from control headteacher's responses. In total, 16 different language programmes were listed plus an additional three that were not explicitly named. *There were six programmes that had not been mentioned in the baseline survey, namely, Colourful Semantics, Listen with Lola, Beanstalk Story Starters, Sound Linkage, Phonological awareness programme, and PORRIC.*

Overall, control headteachers' responses suggest that some of the other programmes that were delivered in control schools had some similarities to NELI. Most similarity is seen in terms of the person in charge of delivering them; like in NELI, most programmes were also delivered by TAs or combined by TAs and teachers. Most of the programmes (across baseline and endline surveys) were described as delivered in 'five to ten weeks', but there were also a few lasting '11 to 20 weeks'—suggesting some degree of similarity with NELI in terms of length and/or duration of delivery. In terms frequency, NELI in this evaluation was organised into three 30-minute group sessions and two individual 15-minute sessions delivered throughout the week. Control headteacher responses therefore suggest that the other programmes delivered in their schools were somewhat similar to NELI in terms of duration; a considerable number of programmes were described as delivered 'once a day'.

Of those language interventions reportedly being used (based on survey responses), to the knowledge of the evaluation team only two have been found to have positive effects on pupil language outcomes. Early Talk Boost is a nine-week intervention delivered by early year practitioners. An evaluation of Early Talk Boost found that it can be effective in advancing the language skills of socially disadvantaged children and improve their school readiness (Reeves et al., 2018). An evaluation of LEGO therapy showed that it can lead to improvements in social communication skills (Lindsay et al., 2017). This suggests that even though there were more reading programmes running in control schools, the lack of evidence suggests that there was not enough activity that was 'unique' to control schools. Since 'usual practice' in both control and intervention schools was similar in terms of the use of other language programmes we may conclude that the situation in both arms of the trial was comparable. Therefore, it can be said that the Impact Evaluation of this evaluation was assessing the impact of 'other language programmes + NELI' compared to 'other language programmes' (rather than NELI compared to no language intervention at all). The statistically significant positive results of this evaluation show that there were larger improvements in language outcomes for pupils in the intervention group than for those in the control group, even in the presence of other language interventions (which represents usual practice).

Cost

This section discusses average costs of participation in NELI, split between single- and multi-form-entry schools.

For one-form-entry schools, the average cost directly attributable to participation in NELI is estimated at £870 (£725 plus VAT) per school (Table 25). This cost reflects the purchasing of NELI materials, including the manuals needed to deliver the intervention, and training to TAs to deliver NELI. For two-form-entry schools that can share the NELI materials, the cost is £1,290 (£1,075 plus VAT).

Of note, following the EEF guidelines, the cost estimate does not include the cost for resources that schools are expected to be already equipped with and are necessary for implementing NELI (for example, whiteboards or teachers, unless an additional new teacher is needed to deliver NELI).

This cost assumes that five pupils per classroom receive the intervention from one TA, that is, five pupils per year in single-form-entry schools, and ten pupils per year in two-form-entry schools.

Table 25: Cumulative per-school total cost of delivering NELI over three years

	Year 1	Year 2 (cumulative)	Year 3 (cumulative)
NELI—one-form-entry school	£870	870	870
NELI—two-form-entry school	£1,290	£1,290	£1,290

The cost breakdown for a single-entry school is provided in Table 26. The total cost per pupil per year averaged over three years of intervention delivery is therefore £58.

Table 26: Cost of delivering NELI for one-form-entry classes

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
NELI—one-form-entry school	Materials and training	£870 (£725+VAT)	£870	£58 (£48.33+VAT)

The cost per pupil per year for materials may be lower in multi-form-entry classes where TAs can share the NELI materials, as shown in Table 27. Once the TA is trained and provided with materials, the NELI intervention can be delivered repeatedly as much as needed, at no extra cost. The total estimated cost of delivering NELI for two-form-entry classes over a three year period is £1,290 (£1,075 plus VAT) per school (Table 27). This equates to an estimated cost of £43 per pupil per year averaged over three years.

Table 27: Cost of delivering NELI for two-form entry classes

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
NELI—two-form-entry school	Materials and training	£1,290 (£1,075+VAT)	£1,290	£43 (£35.83+VAT)

Staff cost

Staff time and costs were not included in the marginal cost calculation. Staff spent additional time in training, delivering NELI group or individual sessions, or preparing for delivery. TAs were expected to spend on average five hours per week for 20 weeks to prepare and deliver NELI, in addition to existing workload. Specifically, it was expected that TAs would spend 30 minutes to prepare for each group session. In total, 1.5 hours were needed to deliver three group sessions in a week. Each individual session took 15 minutes of TAs' time. Altogether, 2.5 hours of TA time were needed to deliver two individual sessions for five separate children. Also, it was expected that TAs would spend one additional hour to prepare for delivery. Overall, TAs were expected to spend 100 hours to prepare and deliver NELI. These calculations do not take into consideration time spent in training.

To ensure staff are capable to deliver NELI, each TA needs to attend 2.5 days of training or a total of 20 hours of training. Taking this into consideration, one TA is required to spend around 120 hours to deliver NELI to five children over the period of 20 weeks.

A further aspect related to time and cost which emerged strongly from the IPE was the increased workload resulting from the NELI programme. In the endline surveys, headteachers, teachers, and TAs were first asked to report their agreement (or lack of) with the statement that NELI had increased the working hours of TAs. Interestingly, just 10 out of 46 (21.7%) TAs agreed with this to a great or moderate extent, with 19 out of 46 (41.3%) not agreeing with this at all. While the majority of teachers (65.5%, 36 out of 55) did agree with this to a great or moderate extent, just 11 out of 34 headteachers (32.4%) agreed with this to a great or moderate extent, with 15 out of 34 (44.1%) not agreeing with this at all. Thus, it seems that the increased workload for TAs as a result of the NELI programme was perceived as a major problem for a majority of teachers, but only a minority of TAs and headteachers in the survey sample.

Interestingly, many interviewees (including a number of TAs) noted that the delivery of NELI had increased the workload of TAs (TAS5, TS6, TAS2, TA1S3, TA1S4, TA2S4, TS5). They explained that this resulted in TAs not being able to support teachers in other activities in reception (TAS5, TS6, TAS2, TA1S3, TA1S4, TA2S4, TS5). Four interviewees added that sometimes TAs had to work overtime to complete NELI tasks (for example, narrative assessment) or catch up on their other work (TS6, TAS2, TA1S3, TAS4):

'I haven't worked overtime unless I often had to take the narrative assessment home and do that in my own time. I have to say that a lot of my other duties had to go to the one side, and I had to focus on NELI more. My time after school has all been taken by either photocopying for NELI or getting all the activities and the resources ready. For a lot of my other jobs I wasn't able to do' (TAS2).

Importantly, one teacher and TA from the same school noted that they did not have to work overtime due to the help the teacher provided to plan ahead and effectively timetable NELI and other reception activities (TAS5, TS5). The teacher supported the TA when needed, which avoided the need for the TA to work overtime to effectively deliver NELI, indicating that teacher support for session delivery is important for smooth and efficient programme delivery:

'I had quite a bit to do. I was quite lucky that the teacher was helping in the class, so it wasn't a lot of things to do but sometimes I was struggling a bit to get the time to do it. This is obviously an issue because you are also a TA in the class as well and you have other things to do to help the teacher with children with other needs' (TAS5).

'She didn't have to do additional hours however that came out from the school timetable. In addition to the five sessions, I also had to allocate her in additional time within the school timetable so she would have time to plan sessions and prepare resources. That happened within the school time so effectively I lost my TA for part of that time. I was down to one TA assistant' (TS5).

Survey respondents were also asked to comment on the effect of NELI on the workload of other school staff. While a majority of teachers (67.3%, 37 out of 55) agreed to a great or moderate extent that delivering NELI had increased the workload of other staff, this was not the case for headteachers (44.1%, 15 out of 34) or TAs (23.9%, 11 out of 46). The higher proportion of teachers who perceived NELI to increase the workload of other staff may reflect the perceived increased workload of teachers in supporting TAs to deliver NELI, as shown in the quotes above (TAS5, TS5).

Issues with time restrictions were also noted as part of the efficacy trial, with TAs finding it difficult to devote enough time to NELI (Sibieta et al., 2016). Further investigation has revealed that while the majority of TAs and teachers feel that the duration of sessions is appropriate, fitting these sessions into busy school timetables is difficult. It may be that schools would benefit from having additional information on the amount of preparation time needed for sessions as well as additional training on how to adapt some of the sessions to fit them into busy school timetables (see further discussion of this below). Overall, the evidence indicates that through the investment of staff time to prepare and deliver NELI sessions, and the increased workload of TAs and other staff that may result from this, the NELI programme could result in increased staff costs for schools, although these costs may be minimised if sufficient support is provided to TAs.

Conclusion

Key conclusions	
1.	Children who received the NELI programme made the equivalent of three additional months' progress in language skills, on average, compared to children who did not receive NELI. This result has a very high security rating.
2.	Children who received the NELI programme made the equivalent of two additional months' progress in early word reading, on average, compared to children who did not receive NELI, in addition to four additional months' progress in language skills (as measured by the digital application LanguageScreen).
3.	Children who received the NELI programme with English as an additional language (EAL) made the equivalent of three additional months' progress in language skills compared to EAL children who did not receive NELI.
4.	Surveys and interviews from the process evaluation showed that schools believed the training and ongoing support provided was clear, useful, and sufficiently detailed for them to deliver the intervention effectively. Initial training attendance was high, and while top-up training attendance was lower, the vast majority of TA survey respondents made use of ongoing support such as telephone calls, webinars, or forums.
5.	There was variation in the number of sessions that schools delivered to pupils and schools did not necessarily deliver the maximum number of sessions. Schools were more likely to deliver the group sessions compared to the individual sessions. TAs attending training and delivering a larger number of sessions was associated with better language outcomes for pupils.

Impact evaluation and IPE integration

Interpretation

This evaluation was the largest evaluation of NELI ever undertaken. The evaluation aimed to test the effectiveness of NELI at scale—in a large number of schools from a range of regions in England. This cluster randomised controlled trial differed in design from previous evaluations where randomisation was conducted within-school, at the pupil level. The school-level cluster randomisation in this trial minimised the risk of contamination between the intervention and control conditions. Associated with an intention-to-treat approach that sees the analysis undertaken as per the initial randomisation approach, regardless of the uptake and implementation fidelity of the intervention, this trial provides a robust and conservative estimate of the impact of NELI.

The evaluation finds that NELI has a positive and statistically significant effect on pupils' language skills. This includes both the primary outcome measure (derived from a latent language variable drawing together four separate language tests, two CELF tests, and two RAPT sub-tests) and the two secondary outcome measures (an early word reading test, YARC, and a digital app and test, LanguageScreen, also developed by the NELI developers). In exploratory analyses, the evaluation also finds a positive and statistically significant effect of NELI on the language skills of pupils who have English as an additional language (EAL) compared to EAL pupils in the control group. It does not find that the impact of NELI on EAL pupils is different to the impact on all pupils (via the main analysis).

Overall, headteachers, teachers, and TAs in participating schools who were interviewed as part of the implementation and process evaluation corroborate these impact findings. These participants report a perception that NELI had a positive impact on children's language as well as communication skills. In the six case studies exploring the in-depth experiences and perceptions of headteachers, teachers, and TAs, all interviewees noted that they had observed a positive impact of NELI on at least some of their pupils, a view shared between the different school roles. Some teachers and TAs also commented that they had observed improvements in the vocabulary of the children and in narrative and storytelling skills. Many teachers and TAs also observed an improvement in children's confidence to communicate.

Nonetheless, these views were not uniformly held, with some teachers reporting that their perceived impact of NELI on pupils with special educational needs (SEN) may not have been high, or as high, as for non-SEN pupils. The impact evaluation does not include a subgroup analysis for pupils with SEN, an issue addressed in the Future research section below. Further, the impact on pupils not selected for NELI is currently unknown; in the context of findings

concerning some reconfiguring of TA time to allow for the implementation of NELI this remains an issue to be explored further, as also addressed in the Future research section below.

IPE data shows that NELI was implemented as intended in the vast majority of schools with TAs and other school staff being clear about NELI's structure and requirements for delivery but that schools did not necessarily reach the maximum recommended number of sessions (either group or individual) to be delivered to pupils. The IPE further disaggregates these results in relation to the different aspects of implementation fidelity. For instance, the training was perceived by participating school staff to be useful and good preparation for programme implementation; nevertheless, delivery of the required number of individual and group sessions was moderate, with evidence of wide variability across settings. Overall, NELI was well received by participating headteachers, teachers, and teaching assistants. The impact evaluation mirrors this evidence and finds that higher levels of implementation fidelity lead to better language outcomes for pupils (on the main outcome measure).

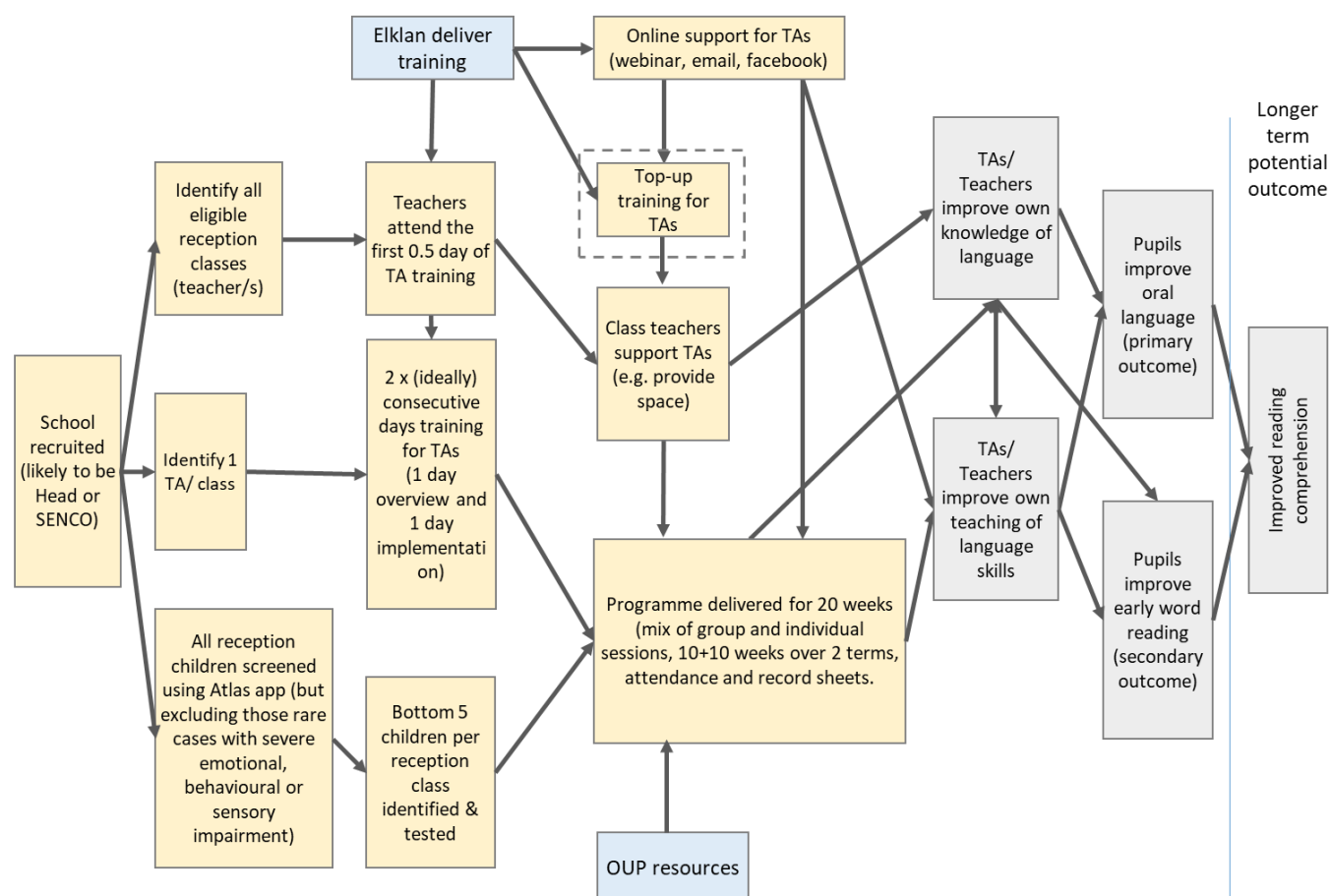
The per-pupil per-year cost as calculated in this trial (over three years of NELI) is approximately £43 to £58, depending on the size of the school and the potential to share resources between the TAs implementing NELI. This includes the costs for training and NELI materials including manuals for delivery. Compared to two other EEF-evaluated early-years interventions, EasyPeasy (£25 cost per pupil, Robinson-Smith et al., 2019) and Family Skills (£143.37, Husain et al., 2018), NELI's per pupil costs are in the middle range of the above two interventions.

The time requirements associated with participation in NELI were in the region of 120 hours for the 20-week intervention period for TAs. This included approximately 20 hours attending training and a total of 100 hours for the delivery of the individual and group sessions as well as planning and assessment. Interestingly, TAs reported less of a time-related burden on their own practice as a result of NELI than did teachers. A majority of teachers reported a high workload for TAs, although TAs themselves and headteachers were more positive about the additional work required as a result of implementing the intervention. In the case study schools, teachers commented that the provision of time and space to TAs to undertake all NELI-related activities was beneficial, as was collaboration between the teachers and TAs.

Evidence to support the logic model

The results of the evaluation overall provide support for the original logic model for the intervention and provide evidence on all key aspects of the implementation (with limitations noted below). One modification to the logic model (in Figure 16) relates to the addition of one training session during the implementation phase. This does not constitute a substantial departure from the usual NELI procedures and implementation approaches. Although evidence from the implementation and process evaluation suggests that it was attended less well than the original training sessions (which saw *all* schools participate), it was still attended by a sufficiently high number of schools to warrant its inclusion in the logic model.

Figure 16: Revised logic model (additional top-up training highlighted)



Limitations and lessons learned

This effectiveness trial was, overall, highly successful, with a very high retention rate at school level, a very high retention rate at pupil level, and good quality data on both the impact and the implementation and process evaluation components resulting from good engagement from participants. In relation to attrition at the pupil level, the 7% attrition poses minimal threats to the internal validity of the trial. In terms of external validity, the current trial recruited schools from all over England in contrast to the previous evaluation where schools were recruited from two regions.

We acknowledge several limitations of the trial. First, the fidelity measure, drawing on implementation logs completed by the TAs, revealed a lot of variation between schools in terms of number of group and individual sessions completed. We found that the majority of the schools are not considered 'high' compliers and the IPE data also provided corroborating evidence of wide variability across schools. Although we did not explore different fidelity thresholds, the distribution of the fidelity measure suggests that, overall, fidelity was moderate with few schools attaining the maximum. Despite this, the main outcome analysis (and the fidelity analysis) in this trial finds positive and significant effects of NELI on pupils' language and early word reading suggesting that even in the absence of perfect implementation fidelity, NELI can lead to improved language outcomes.

A final limitation was the non-complete response to the IPE surveys meaning that the IPE findings build upon evidence provided only by those schools and staff that engaged in the surveys, in addition to the case studies and observations also part of the IPE. Not all control school headteachers provided evidence as to their practices in relation to the use of other language programmes in reception classrooms. Higher response rates would have allowed for more definitive conclusions as to the counterfactual and any programmes that may impact on children's language being delivered in control schools. We suggest that this focus on control schools' activity is expanded in future trials.

Future research and publications

Further to the findings above, future research could focus on the impact of NELI for specific groups of pupils, including pupils with special educational needs and pupils with specific socio-economic disadvantage (for example, eligible for free school meals).

Additionally, it would be relevant to explore distal outcomes for language, literacy, and reading comprehension (as per the logic model) to understand any retention of the impact of NELI on pupils' language skills. A follow-up study exploring the delayed outcome on language has already been commissioned by the EEF. The follow-up testing will be conducted by Elklan (a delivery partner in this trial), but procedures have been put in place to ensure the independence of the data being collected at follow-up and the analysis will be undertaken by the independent evaluators.

Also, given the fact that one of the key barriers to implementation was the high demands placed on staff workload, as suggested by the IPE, it would be interesting to assess in the future if variants of NELI with fewer group or individual sessions could be equally effective for pupils while minimising the impact on teachers' and TAs' time.

Additionally, it would be useful to explore the impact of NELI on pupils not selected for the intervention from within the same classes. This should relate to both the impact on the class as a whole, and the impact on the next five pupils with low language skills in the NELI-implementing classes (that is, the next five pupils after the initial trial target pupils). The latter analysis would explore any negative impact on the pupils just missing out.

From a trial design perspective, the estimated ICC at analysis stage based on the language latent variable is significantly higher, $ICC = 0.349$, in comparison to the ICC estimated at randomisation stage, which was based on LanguageScreen data. Although this specific trial was still able to detect a significant effect (with an at-analysis MDES of 0.257), it is worth noting for any future trials in this field that standard ICC assumptions generally applying to primary school trials may not fully apply.

Separately, and building on its use in this trial as a secondary outcome measure, the LanguageScreen measure of language would benefit from further validation prior to wider use that may potentially reach outside NELI implementation.

Finally, in any future research we propose to increase the focus on business-as-usual in control schools, where appropriate, given that evidence from this evaluation suggests a busy intervention space with many language-focused programmes simultaneously implemented. Given this, and the intention-to-treat design of the evaluation, the trial is likely to have provided conservative estimates of NELI's positive and significant impact on pupils' language skills.

References

- Allison, P.D. (2003) 'Missing Data Techniques for Structural Equation Modeling', *Journal of Abnormal Psychology*, 112 (4), p. 545.
- 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), pp. 422–432.
- Bowyer-Crane, C., Fricke, S., Schaefer, B., Lervåg, A. and Hulme, C. (2017) 'Early Literacy and Comprehension Skills in Children Learning English as an Additional Language and Monolingual Children with Language Weaknesses', *Reading and Writing*, 30 (4), pp. 771–790.
- Carril, A., 2017. Dealing with misfits in random treatment assignment. *The Stata Journal*, 17(3), pp. 652–667.
- Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M. and Mendoza, K. (2009). 'Floor Effects Associated with Universal Screening and Their Impact on the Early Identification of Reading Disabilities', *Journal of Learning Disabilities*, 42 (2), pp. 163–176.
- Chowdry, H. and Fitzsimons, P. (2016) 'The Cost of Late Intervention: EIF Analysis 2016', London: Early Intervention Foundation.
- Dong, N. and Maynard, R. (2013) 'PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies', *Journal of Research on Educational Effectiveness*, 6 (1), pp. 24–67.
- Duff, F. J., Reen, G., Plunkett, K. and Nation, K. (2015) 'Do Infant Vocabulary Skills Predict School-Age Language and Literacy Outcomes?', *Journal of Child Psychology and Psychiatry*, 56 (8), pp. 848–856.
- Education Endowment Foundation (2016) 'A Marked Improvement? A Review of the Evidence of Written Marking', EEF and University of Oxford.
https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Marking_Review_April_2016.pdf
- Feinstein, L. and Duckworth, K. (2006) 'Development in the Early Years: Its Importance for School Performance and Adult Outcomes' (Wider Benefits of Learning Research report no. 20), Centre for Research on the Wider Benefits of Learning, Institute of Education, University of London.
- Fernald, A., Marchman, V. A. and Weisleder, A. (2013) 'SES Differences in Language Processing Skill and Vocabulary Are Evident at 18 Months', *Developmental Science*, 16 (2), pp. 234–248.
- Fielding, L., Kerr, N. and Rosier, P. (2007) *Annual Growth for All Students: Catch-Up Growth for Those Who Are Behind*, New Foundation Press.
- Finnegan, J. and Warren, H. (2015) 'Ready to Read: Closing the Gap in Early Language Skills So That Every Child in England Can Read Well', London: Save the Children.
- Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C. and Snowling, M. J. (2013) 'Efficacy of Language Intervention in the Early Years', *Journal of Child Psychology and Psychiatry*, 54 (3), pp. 280–290.
- Hart, B. and Risley, T. R. (1995) *Meaningful Differences in the Everyday Experience of Young American Children*, Paul H Brookes Publishing.
- Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs', *Journal of Educational and Behavioral Statistics*: <https://doi.org/10.3102/1076998606298043>
- Heß, S. (2017) 'Randomization Inference with Stata: A Guide and Software', *Stata Journal*, 17 (3), pp. 630–651.
- Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L. and Sterne, J. A. (2011) 'The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials', *BMJ*, 343, p. d5928.

- Hox, J. (1998) 'Multilevel Modeling: When and Why', in *Classification, Data Analysis, and Data Highways*, Berlin: Springer (pp. 147–154).
- Hox, J. J., Bechger, T. M., van den Wittenboer, G. L. H. and Glopper, C. D. (1999) 'The Validity of Comparative Educational Studies', *Educational Measurement: Issues and Practice*, 18 (3), pp. 18–26.
- Hu, L. T. and Bentler, P. M. (1999) 'Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives', *Structural Equation Modeling*, 6 (1), pp. 1–55.
- Husain, F., Wishart, R., Marshall, L., Frankenberg, S., Bussard, L., Chidley, S., ... and Morris, S. (2018) 'Family Skills Evaluation Report and Executive Summary', report to the Education Endowment Foundation.
- Hutchison, D. and Styles, B. (2010) *A Guide to Running Randomised Controlled Trials for Educational Researchers*, Slough: NFER.
- Kenny, D. A. (2015) 'Measuring Model Fit': retrieved from <http://www.davidakenny.net/cm/fit.htm>
- Lai, K. and Green, S. B. (2016) 'The Problem with Having Two Watches: Assessment of Fit when RMSEA and CFI Disagree', *Multivariate Behavioral Research*, 51 (2-3), pp. 220–239.
- Law, J., Todd, L., Clark, J., Mroz, M. and Carr, J. (2013) 'Early Language Delays in the UK', London: Save the Children.
- MacDonald, M. C. (2013) 'How Language Production Shapes Language Form and Comprehension', *Frontiers in Psychology*, 4, p. 226.
- Martinez-Beck, I. and Zaslow, M. (2006) 'Introduction: The Context for Critical Issues in Early Childhood Professional Development', in *Critical Issues in Early Childhood Professional Development*, edited by Zaslow, M. and Martinez-Beck, Baltimore: Brookes (1:1- 16).
- Muthén, B. O. (2002) 'Beyond SEM: General Latent Variable Modeling', *Behaviormetrika*, 29 (1), pp. 81–117.
- Renfrew, C. E. (2016) 'The Renfrew Language Scales: Action Picture Test', Speechmark.
- Reeves, L., Hartshorne, M., Black, R., Atkinson, J., Baxter, A. and Pring, T. (2018) 'Early Talk Boost: A Targeted Intervention for Three Year Old Children with Delayed Language Development', *Child Language Teaching and Therapy*, 34 (1), pp. 53–62.
- Robinson-Smith, L., Menzies, V., Cramman, H., Wang, Y. L., Fairhurst, C., Hallett, S., ... and Siddiqui, N. (2019) 'Easypeasy: Learning Through Play, Evaluation Report', report to the Education Endowment Foundation.
- Roulstone, S., Law, J., Rush, R., Clegg, J. and Peters, T. (2011) 'Investigating the Role of Language in Children's Early Educational Outcomes'.
- Scarborough, H.S., Neuman, S. and Dickinson, D. (2009) 'Connecting Early Language and Literacy to Later Reading (Dis)Abilities: Evidence, Theory, and Practice', *Approaching Difficulties in Literacy Development: Assessment, Pedagogy and Programmes*, pp. 23–38.
- Schoon, I., Cheng, H., Gale, C. R., Batty, G. D. and Deary, I. J. (2010) 'Social Status, Cognitive Ability, and Educational Attainment as Predictors of Liberal Social Attitudes and Political Trust', *Intelligence*, 38 (1), pp. 144–150.
- Sibieta, L., Kotecha, M. and Skipp, A. (2016) 'Nuffield Early Language Intervention: Evaluation Report and Executive Summary', Education Endowment Foundation.
- Snijders, T. A. and Bosker, R. J. (1994) 'Modeled Variance in Two-Level Models', *Sociological Methods and Research*, 22 (3), pp. 342–363.
- Snijders, T. A. B. (2005) 'Power and Sample Size in Multilevel Modeling', in B. S. Everitt and D. C. Howell (eds), *Encyclopedia of Statistics in Behavioral Science*, Chichester: Wiley (vol. 3, pp. 1570–1573).

- 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'.
- Social Mobility Commission (2017) 'Time for Change: An Assessment of Government Policies on Social Mobility 1997–2017', London: Social Mobility Commission.
- Sutherland, A., Dimova, S. and Belanger, J. (2019) 'Trial Evaluation Protocol: The Nuffield Early Language Intervention (NELI)', London: Education Endowment Foundation.
- Wiig, E., Secord, W. and Semel, E. (2006) 'Clinical Evaluation of Language Fundamentals–Preschool–2, UK Standardisation' (CELF-P2 UK).
- Whitehurst, G. J. and Lonigan, C. J. (1998) 'Child Development and Emergent Literacy', *Child Development*, 69 (3), pp. 848–872.
- Vandenberg, R. J. and Lance, C. E. (2000) 'A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research', *Organizational Research Methods*, 3 (1), pp. 4–70.







Appendix A: EEF cost rating

Figure 17: Cost Rating

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 B: Security classification of trial findings

OUTCOME: *Language latent variable*

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition					
5 	Randomised design	≤ 0.2	0-10%	5				5
4 	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3 	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%					
2 	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1 	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0 	No comparator	≥ 0.6	$>50\%$					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	Allocation sequence described clearly; trial arms balanced on key observables (all ES are less than or approximately equal to 0.05 with the exception of RAPT)
Threat 2: Concurrent Interventions	Low	Concurrent interventions are documented clearly in the IPE but there is no evidence suggesting differential uptake
Threat 3: Experimental effects	Low/Moderate	Usual practice data pre and post randomisation is presented. Although some control schools report implementing NELI, this is likely a reporting error as opposed to genuine contamination
Threat 4: Implementation fidelity	Low/moderate	Implementation data clearly reported along with compliance analysis. Moderate delivery of actual sessions compared to developer expectations but this is accounted for by compliance analysis.
Threat 5: Missing Data	Low	Total missing data is low and does not vary by trial arm. No analyses accounting for missing data are reported.
Threat 6: Measurement of Outcomes	Low	Outcome tests are justified and those that are not yet fully validated are noted. Testing of primary outcome was blind to allocation
Threat 7: Selective reporting	Low	Study is registered and the protocol and SAP are followed; some minor deviations but these are clearly noted and justified

- **Initial padlock score:** five Padlocks – A large, secure cluster-randomised trial with MDES at randomisation of 0.176 and low attrition of 7.3%.
- **Reason for adjustment for threats to validity:** 0 Padlocks – No adjustment
- **Final padlock score:** initial score adjusted for threats to validity = five Padlocks

Appendix C: Changes since the previous evaluation

		Pilot Study	Efficacy trial	Effectiveness trial (this trial)
Intervention	Programme variant	Main NELI intervention	Main NELI intervention	Main NELI intervention
	Intervention content	Pupil individual and group sessions	Pupil individual and group sessions	Pupil individual and group sessions
	Delivery model	Teacher/TA	Teacher/TA	Teacher/TA
	Intervention duration	30 week (Nursery)	20 week (Reception) 30 week (Nursery)	20 week (Reception)
Evaluation	Authors	Fricke et al., 2013	Sibieta et al., 2016	Dimova et al.
	Trial design, including number of arms	Two-arm	Three-arm	Two-arm
	Unit of randomisation	Child-level	Child-level	School-level
	Number of schools in trial	15	34	193
	Number of children in trial (at randomisation)	180 90 (intervention) 90 (control)	394 117 (30-week intervention) 124 (20-week intervention) 119 (control)	1,156 585 (intervention) 571 (control)
	Regions	1 region	2 regions	13 different regions across England
	Stratification variable (s) (if applicable)	N/A	N/A	Region Single/multi-form entry
	Screening measure	Language screening measures (based on the CELF)	Language screening measures (based on CELF)	LanguageScreen
	Baseline testing measure	Language; Narrative; Phonological Awareness; Literacy.	Language composite score	Latent language
	Primary outcome measure(s) ²⁰	Language; Spoken narrative skills;	Language composite score	Latent language
	Secondary outcome variable	Phonological awareness; Literacy skills	Word level literacy skills	YARC Early word reading
	School dropout	None	9%	0.5%
	Pupil dropout	8%	9%	7%
	Control condition	N/A	Business-as-usual	Business-as-usual

²⁰ More detail on the outcomes measures selected is provided in the 0

Appendix D: Recruitment Documentation

Memorandum of Understanding



Nuffield Early Language Intervention Trial

Memorandum of Understanding between Schools, The University of Oxford, Elklan, and The Rand Corporation

Elklan
Parkside
Station road
St Mabyn Cornwall
PL30 3BN
01208 841450

Henrietta McLachlan
henrietta@elklan.co.uk

Alex Hall
alexelklan@gmail.com

The University Oxford
Department of Education
15 Norham Gardens,
Oxford,
OX2 6PY
01865 284096

Professor Charles Hulme
charles.hulme@education.ox.ac.uk

Dr Gillian West
gillian.west@education.ox.ac.uk

Rand Europe
Westbrook Centre
Milton Road,
Cambridge,
CB4 1YG
01223359329

Dr Julie Belanger
belanger@rand.org

Dr Alex Sutherland
asutherl@rand.org



Memorandum of Understanding (MOU)

Agreement to participate in the Nuffield Early Language Intervention (NELI) Study

Please sign and return this MOU to:

Alex Hall
Elklan
The Old Mill
Shillingstone Lane
Okeford Fitzpaine
Dorset DT11 0RB

OR

Sign, scan and return electronically to Alex Hall: alexelklan@gmail.com

School Name:

School Address:

.....

.....

School Telephone:

Headteacher/Principal:

Email address:

In most cases only one Reception class per school will be entered into the trial. However, please indicate how many Reception classes you would like to enter into the trial and we will try to accommodate you if possible.

Number of eligible Reception classes

Number of children in each Reception class

NB: Payment to schools randomized to the control condition is per school irrespective of number of classes taking part. Priority will be given to classes with over 20 children.

We commit to the implementation and evaluation of the Nuffield Early Language Intervention (NELI) as set out above.

Named contact for LANGUAGESCREEN testing
(this is the name of the person to whom the LANGUAGESCREEN tablet computer should be sent)

Headteacher/Principal name:

Headteacher/Principal signature:

Date:

Parental Opt Out Form



Professor Charles Hulme
Department of Education,
15 Norham Gardens,
Oxford, OX2 6PY
Direct Line: 01865 284096
E-Mail: charles.hulme@education.ox.ac.uk

OPT OUT FORM

If you **DO NOT** want your child to participate in the research, The Nuffield Early Language Intervention Study, please fill out the form below and return it to the school by [dd/mm/yyyy]. If we do not receive an opt-out form from you by this date, your child may be invited to take part in this study, as described in the accompanying information sheet.

I, the undersigned, hereby DO NOT give permission for my child to take part in the study titled The Nuffield Early Language Intervention Study.

Name of child: _____

Name of parent/guardian: _____

Name of School: _____

Signature: _____ Date: _____

The Nuffield Early Language Intervention (NELI) trial

The Nuffield Early Language Intervention (NELI) is an evidence-based oral language intervention for children in Reception class who show weaknesses in their oral language skills. These children are at risk of experiencing difficulty with reading. The programme aims to improve spoken language skills such as vocabulary knowledge, story-telling and listening skills. It is delivered over 20 weeks in Reception by teaching assistants (TAs) in small groups, comprising five children.

The aim of this trial is to evaluate the effects of the NELI on the development of spoken language skills (vocabulary knowledge, story-telling and listening skills). The results of this research will make an important contribution to understanding what works in improving support for children who show weakness in their oral language skills and are at risk of experiencing difficulty with reading.

The trial is being carried out by The University of Oxford (the Research team) and will be independently evaluated by RAND Europe. The trial is a randomised controlled trial. This means that half the schools, chosen at random, will receive the intervention (the NELI Programme). These are the Intervention Settings. The other half will not receive the intervention. These are the Control Settings. Once the intervention (the NELI Programme) is completed, outcomes from parents and children in the Intervention Settings will be compared to those in the Control Settings to find out whether the intervention has made a measurable difference. The random allocation of settings is essential to the evaluation, as it is the best and most rigorous way to find out the effects of the NELI Programme on children. It is therefore important that settings understand, and consent to, the random allocation process.

Practitioners from the Intervention Schools will be trained to deliver the Programme in November 2018 and will deliver the Programme during the Spring and Summer terms 2019. Delivery of the Programme is mandatory for Intervention Settings. Control Settings will receive a payment of £1000 per school on completion of the study. They may spend this on the Nuffield Early Language Intervention (NELI) once the study is completed.

Training provided to schools to support delivery of the NELI programme

In the trial ELKAN will train teachers and teaching assistants (TAs) from Reception classes in the NELI Programme. Five children identified from each class will then receive the programme over a 20 week duration. The NELI programme will be delivered by the trained TAs. The trial is funded by the Education Endowment Foundation (EEF).

The training comprises a two-day course for TAs and a half day course for Teachers. TAs will be provided with a half day face to face workshop and four webinars during the NELI programme. Members of the Research team may be present to observe one of the training sessions as part of the independent evaluation. The TAs will be invited to join a closed Facebook group and will also have access to email support from the ELKAN tutor who trained them.

At the beginning of the study:

- The Research team will contact ALL settings to gather data about participating children (names, dates of birth and Unique Pupil Numbers) once parents have been sent opt-out consent. This will help with the evaluation and enable linkage to the National Pupil Database.
- The Research Team will send a short app-based language assessment (ATLAS) on a tablet computer to each school setting.
- Schools will carry out this initial 10-minute language assessment (ATLAS) on all children in participating reception classes.
- The Research Team will use this information to identify the five children in each class who would most benefit from the NELI and will undertake more detailed assessment with these five children.
- This screening and assessment process will be repeated at the end of the study and again 1 year later to assess progress. The individual assessments of the children who have taken part in the NELI programme will be undertaken by the independent evaluator.

Responsibilities

Elklan will:

- Provide two days' training to one TA from each intervention class.
- Provide training for the reception teacher of each intervention class during the first morning of the two day TA training course.
- Provide four webinars and one face-to-face workshop for TAs from intervention classrooms during the intervention period.
- Provide support to TAs through emails (up to 1 per week per TA) and Facebook group activity over the intervention period.

The Research Team will:

- Be the first point of contact for any questions about the research.
- Coordinate the initial pupil screening and assessments.
- Conduct the random allocation of settings to either the Intervention Group or the Control Group and inform the school of the outcome.
- Collect and analyse the data from the study (including pupil outcome data, online surveys of TAs and headteachers on implementation, and interview data from TAs and headteachers in a small number of schools).
- Ensure all members of the Research Team are appropriately trained and have full Disclosure and Barring Service (DBS) clearance.
- Publish a report on the findings of the study and disseminate the research findings.
- Ensure all schools (experimental and control) will receive project results, including specific results for their own school at the end of the study (Summer 2019).

The Schools will:

- Distribute the information sheet, privacy notice and opt-out form to parents / carers and collect any opt-out forms that are returned.
- Conduct the initial app-based testing of all children in participating Reception classes and repeat this testing at the end of the intervention.
- Accommodate and facilitate the child assessments being carried out by the Research Team at the beginning and end of the study, this will take approximately three days in each of Autumn 2017 and July 2018.
- Assist the Research team in identifying any children who should be excluded from the trial according to specified criteria.
- Consent to randomised allocation.

Additionally, schools allocated to the Intervention group will:

- Release one TA from each intervention reception classroom to attend 2 days of training between 26th and 30th Nov, 2018.
- Release one teacher from each intervention reception classroom to attend a ½ day of training between 26th and 30th Nov, 2018 (the first morning of the 2 day training course for TA's).
- Ensure the trained TAs deliver the NELI intervention over 20 weeks in Spring and Summer terms, commencing w/c 7th Jan 2019 to complete programme prior to post-trial assessments:
 - Three 30-minute sessions per week delivered to groups of five children.
 - Two 15-minute individual sessions per week for each of the five children.
- Release the trained TAs to receive support through:
 - Participation in webinars during weeks 3, 6, 12/13 and 21 of the implementation.
 - Participation in a half day face-to-face workshop in week 9 of implementation.
- Complete a fidelity log (supplied by the Research Team) detailing topics covered and child attendance at sessions.
- Provide a contribution of £280.00 (excluding VAT) + £3.85 delivery charge towards the cost of classroom resources for NELI.
- Complete a short (10-15 minutes) online survey (headteachers and TAs delivering the intervention) in each of Autumn 2018 and July 2019. Headteachers and TAs (only in selected schools) will also be invited to respond to 30-45 minute telephone interview in July 2019 on a voluntary basis.
- Agree to a further follow up assessment of all children in participating Reception classes at a 1 year follow up.

Use of data

All data, including children's test responses and any other pupil data, will be treated with total confidentiality. Data from ATLAS and other child assessments will be administered and analysed by the Research Team. Named data will be matched with the National Pupil Database and shared with the EEF. No individual setting, parent or child will be identified in any report arising from the research. The study has received full ethical approval from the Department of Education Ethics Committee at The University of Oxford. Please see the Privacy Notice at the end of this document for further detail on how personal data will be collected and used.

Timescales

Milestone	Date
Initial screening by schools of all children in all classrooms, using ATLAS (Schools)	September 2018
Five children in each classroom individually assessed (Research Team)	October 2018 (scheduled for w/c 15th October)
Classrooms assigned to experimental or control groups (Research Team)	November 2018 (scheduled for 19th November)
Elklan Tutors provide 2 days training to Teaching Assistants (TAs) in experimental classrooms. Teachers also attend for first half day.	November 2018 (scheduled for w/c 26th November)
TAs deliver NELI over 20 weeks	Spring and Summer terms 2019.
Elklan Tutors provide support for TAs	Spring and Summer terms 2019
Post assessment of children (Research Team)	July 2019
1 year follow up of children (Research Team)	July 2020

Privacy Notice

The Research Team conducting the NELI trial take privacy very seriously and in light of the new General Data Protection Regulations (GDPR) sets out in this privacy notice how we will use and look after the personal information that we collect from the children and schools taking part in the trial.

The Research Team encompasses The University of Oxford who will assess the children before the start of the trial and RAND Europe who are appointed as the independent evaluator and who will be assessing the children after the trial has ended. The Research Team are the data controllers of the data they collect and are responsible for the processing of any personal data that is collected from the children taking part in the trial within both intervention and control group settings. The Research Team have put appropriate security measures in place to keep personal data secure and to prevent any unauthorised access to or use of it.

Results for each child will be kept strictly confidential. Only a code number will be used to identify children, and any written records will be kept in a locked filing cabinet in the University of Oxford. Results from each child will be combined with the results from other children for statistical treatment and not used in any way that would identify children. Only anonymised data will be shared with research partners. In addition, the University of Oxford will provide schools with a summary of standard scores from the language assessments for the children in their participating classes and a description of what these scores mean.

What personal data will be collected?

Personal data means any information about an individual from which that individual can be identified. The Research Team will collect, use, and store this personal data for each child in the trial in order to be able to match children's pre- and post-trial assessments and determine the efficacy of the trial.

During the intervention, participants may contact the University of Oxford to request deletion of their personal data. After this period, participants should contact the FFT, who will be responsible for any personally identifiable data held under their contract.

How will this personal data be used?

The Research team will only use personal data for the purposes outlined in the table below, which describes all the ways that children's personal data will be used in the trial.

Personal Information Data			
Data (what will be collected?)	Source (who will collect it?)	Use (who will use it?)	Purpose (what will it be used for?)
Pupil data: Name, Unique Pupil Number (UPN), Free School Meal status, DoB, Gender, Special Educational Needs, English as an Additional Language.	University of Oxford	University of Oxford	Trial evaluation, academic publications
		RAND Europe	Independent evaluation of trial, academic publications
Pupil data: ATLAS app screening data from all pupils pre- and post-trial	University of Oxford	University of Oxford	Trial evaluation, academic publications
		RAND Europe	Independent evaluation of trial, academic publications
Pupil data - Pre-trial language testing of 5 children in each class allocated to programme	University of Oxford, via Eiklan testers	University of Oxford	Trial evaluation, academic publications
		RAND Europe	Independent evaluation of trial, academic publications
Pupil data - Post-trial language testing of 5 children in each class allocated to programme	RAND Europe, via ACER testers	RAND Europe	Independent evaluation of trial, academic publications
		University of Oxford	Trial evaluation, academic publications
Headteacher, teacher and teaching assistant post-trial survey data	University of Oxford	University of Oxford	Trial evaluation, academic publications
		RAND Europe	Academic publications
Pupil-level data for language assessments	University of Oxford	Schools	Results of children's language assessments shared with their school.
School data - List of participating schools by region	University of Oxford	RAND Europe	Randomisation of schools
		University of Oxford	Intervention implementation

Storage of data after the end of the trial:

Research data (not sensitive/personal data) will be kept securely by the University of Oxford for a minimum of three years according to University Policy. Additionally, all data collected by the University of Oxford and RAND Europe will be entered into the EEF's data archive at the end of the trial (NB: FFT is the organisation appointed to manage EEF's data archive). This includes the pupil outcome from subsequent external tests, all of which will be linked with the NPD

data. The data will also be stored on the UK Data Archive. FFT will be responsible for all data archiving and for the management and security of the anonymised trial data.

School staff survey preamble

For study purposes

Your school is one of the 200 schools in England taking part in the evaluation of the Nuffield Early Language Intervention (NELI). The aim of this evaluation is to understand the effects of NELI, an intervention aimed at improving spoken language abilities of children with relatively poor spoken language ability during Reception Year. The intervention is implemented by the University of Oxford. RAND Europe has been commissioned to conduct an independent evaluation of this intervention. The evaluation is funded by the Education Endowment Foundation (EEF), an independent charity who fund research into 'what works' in improving educational practice.

Staff in the participating schools is being asked to participate in the evaluation, including this online survey. This survey will take approximately 10 minutes to complete.

This survey is anonymous and will remain confidential. However, we are keeping track of which school each survey entry comes from, so we can monitor completion rates.

Participation in the evaluation of the programme is entirely voluntary. Therefore, you may refuse to complete this survey without penalty. Please click "Next" if you would like to complete the survey.

For GDPR purposes (separate screen)

RAND Europe is collecting data on the implementation of the Nuffield Early Language Intervention on the basis of legitimate interest as we have been contracted by the Education Endowment Foundation to evaluate this intervention. Your names will not be collected but to the extent to which your data becomes identifiable, we will maintain your data confidentially and securely and analyse only for the purpose of evaluating the Nuffield Early Language Intervention. We consider the likelihood of your data becoming identifiable as low.

Please do not provide any personal sensitive data in this survey, such as your political persuasion. If personal sensitive data is provided in the survey, RAND Europe will delete it before analysis.

RAND Evaluation of the Nuffield Early Language Intervention (NELI)

Evaluation Purpose.

200 schools in England are taking part in the evaluation of the Nuffield Early Language Intervention (NELI). The aim of this evaluation is to understand the effects of NELI, an intervention aimed at improving spoken language abilities of children with relatively poor spoken language ability during Reception Year. The intervention is implemented by the University of Oxford. RAND Europe has been commissioned to conduct an independent evaluation of this intervention. The evaluation is funded by the Education Endowment Foundation (EEF), an independent charity who fund research into 'what works' in improving educational practice.

Evaluation results will be used to provide information to Oxford University, stakeholders such as the Education Endowment Foundation and the Department for Education, other schools and the wider research community about the extent to which similar interventions can improve spoken language ability. This consent form regards your consent to be included in the phone interview as part of the evaluation.

Staff in all participating schools is being asked to participate in the evaluation.

Procedures: what do I need to do? What is being asked of me?

We will conduct 20-30 minute phone interviews with selected school staff members to explore experiences delivering NELI. Should you consent to take part in an interview, you will be contacted to find a suitable date and time for this interview. A researcher from the RAND Europe team will contact you at the agreed time by phone to ask you about your experiences with the intervention. We would like to record the conversation to make sure we accurately represent your responses. The recordings will be deleted at the end of the study.

Confidentiality.

Data used for and arising from the evaluation will be kept **confidential**, and will only be used for the purpose of the evaluation. RAND reports will contain summaries, and responses will never be presented in any way that would permit readers to identify you. No one at Elklan or the participating schools will have access to interview responses that include respondents' names, school names, or other information that could potentially be used to identify individuals.

Participants.

Interviews will only be conducted with select Elklan training staff.

Benefits and Costs.

There are no costs to you for participating in the evaluation. By participating in the evaluation, you will be providing valuable information that may help schools decide whether these policies are effective and worth pursuing.

Compensation.

You will not receive financial compensation for your participation in the evaluation.

Risks.

There is minimal risk in this evaluation. Only RAND evaluators on the project will have access to any of the information collected and all data will be processed in accordance with the UK data protection laws. Having noted this, there is a small risk that confidential information about you is released to the public.

Voluntary Participation.

Participation in the evaluation of the intervention is entirely voluntary. You may refuse to participate or refuse to answer specific questions at any time without penalty. You will be told if any important new information is found during the course of this evaluation that may affect your wanting to continue.

Questions?

If you have any questions about this evaluation, please feel free to contact the evaluation team direct at NELI@rand.org or by calling 01223 353 329.

You may also contact the RAND Human Research Protection Program at 001-866-697-5620 or hspcinfo@rand.org for more information, to inquire about your rights as a research subject, or to report any research-related problems. If possible, when you contact the Committee, please reference Study #2018-0501-AM03.

Your Signature and Consent

If you are happy to proceed with the interview, please sign below and return a copy to the research team.

Please tick:

1. I confirm that I have read and understood the information sheet for the above study.

☐

2. I agree to being interviewed.

☐

3. I agree to the interview being audio recorded.

☐

4. I agree to the use of anonymised quotes in publications.

☐

Name of participant

Date

Signature

Appendix E: Outcome Measures

Outcomes included in previous EEF evaluation and suggested for the current evaluation

Scale	Relevant sub-scale	In the Sibieta et al., study?	This trial
CELF Preschool 2^{UK}	CELF expressive vocabulary	Yes – positive effect , as part of a composite measure (at 10% level).	Primary outcome
RAPT	Information	Yes – positive effect , as part of a composite measure (at 10% level). However, no effect of the subscale on its own.	Primary outcome
RAPT	Grammar	Yes – positive effect , as part of a composite measure (at 10% level). However, no effect of the subscale on its own.	Primary outcome
YARC early word reading test	Early word reading	Yes – no effect , as part of a composite measure or individually.	Secondary outcome t

Outcomes not used in the previous EEF evaluation but suggested for the current evaluation

CELF Preschool 2^{UK}	CELF recalling sentences	Not included –it was used in previous NELI study for screening (Fricke et al., 2013)	Primary outcome
Other	ATLAS	No	Secondary outcome – Agreed during the IDEA workshop

Outcomes used in previous EEF evaluations but not suggested for this one

YARC early word reading test	Letter knowledge	Yes – no effect , as part of a composite measure or individually
Other	Spelling test	
YARC adaptation²¹	Listening comprehension	Yes – positive effect , as part of a composite measure (at 10% level). However, no effect of the subscale on its own.

²¹ Children listen to recordings of four short stories and answer questions about them. The listening comprehension test is a measure created by the NELI project developers, based on the York Assessment of Reading Comprehension test, YARC. The stories and questions are published in the YARC test with a reference as follows: 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.

Appendix F: Baseline equivalence

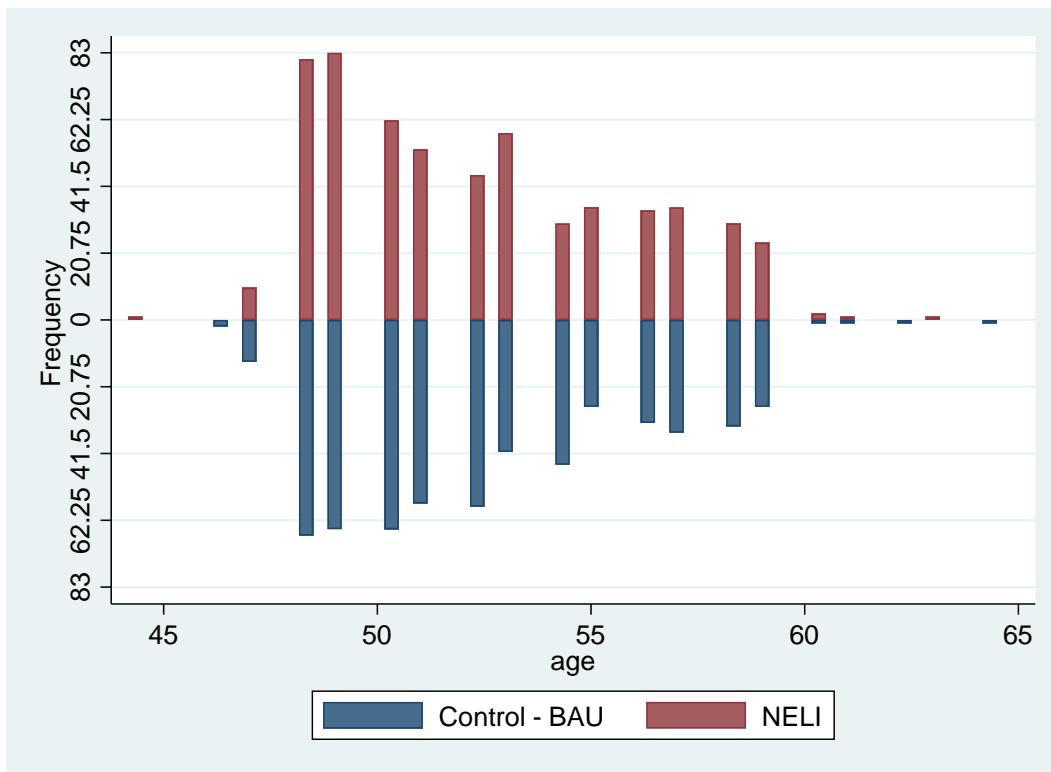


Figure 18: Distribution of age in months

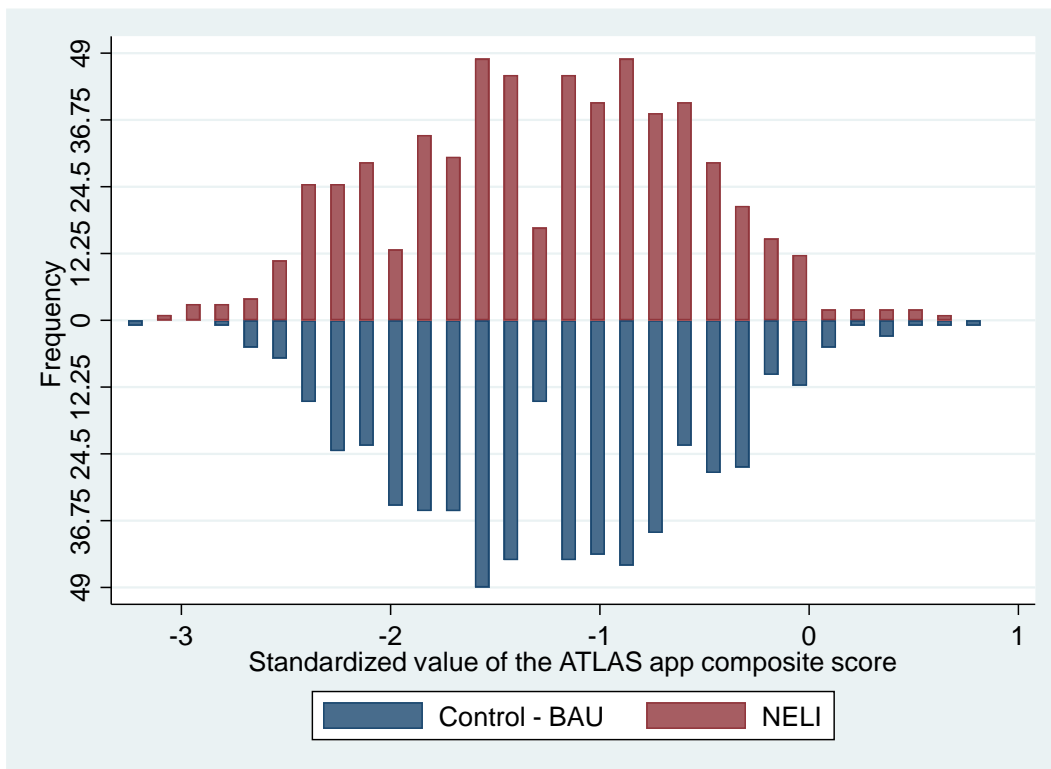


Figure 19: Distribution of ATLAS composite score at baseline

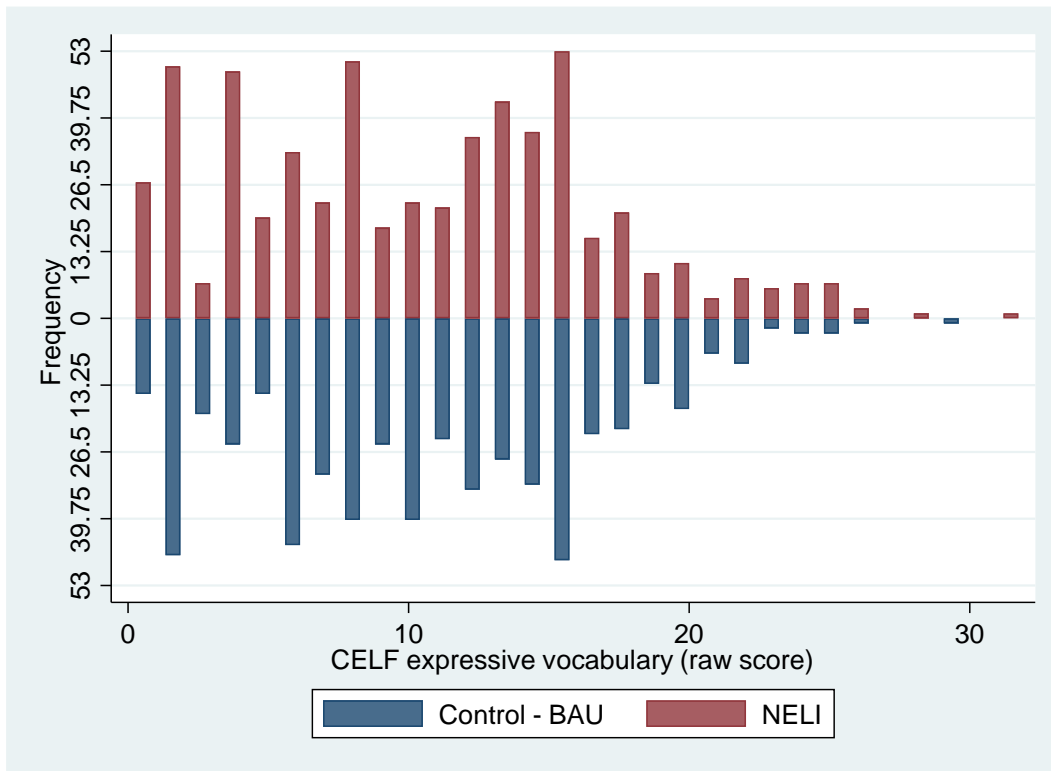


Figure 20: Distribution of CELF expressive vocabulary raw score at baseline

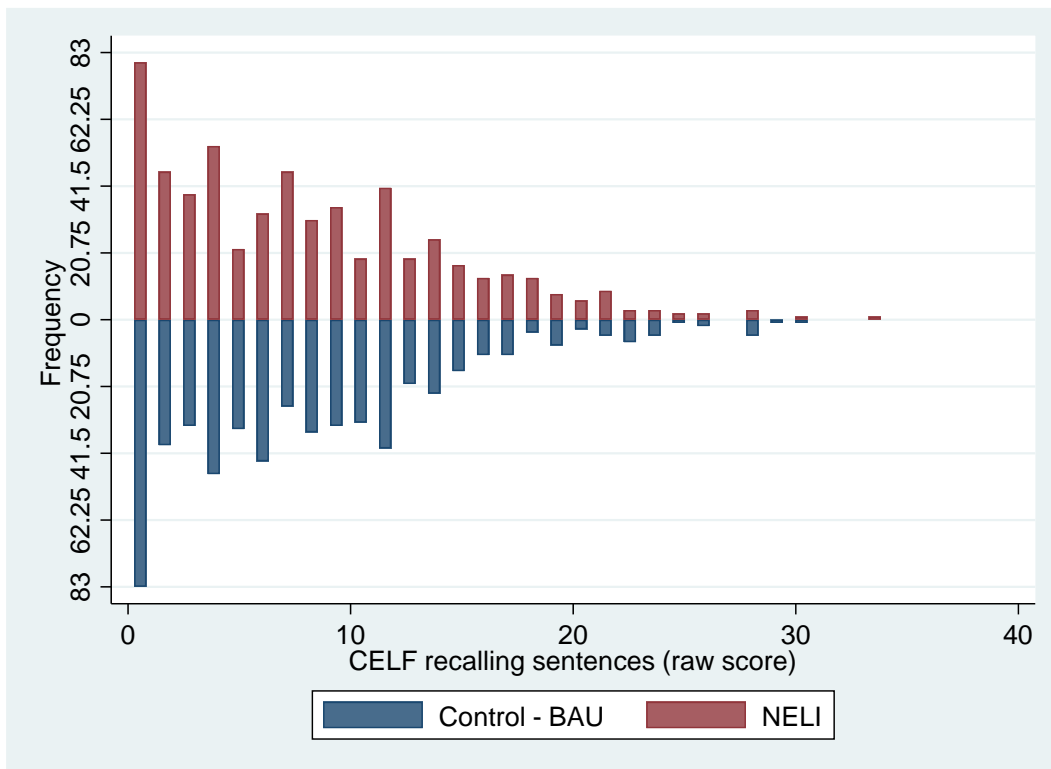


Figure 21: Distribution of CELF recalling sentences raw score at baseline

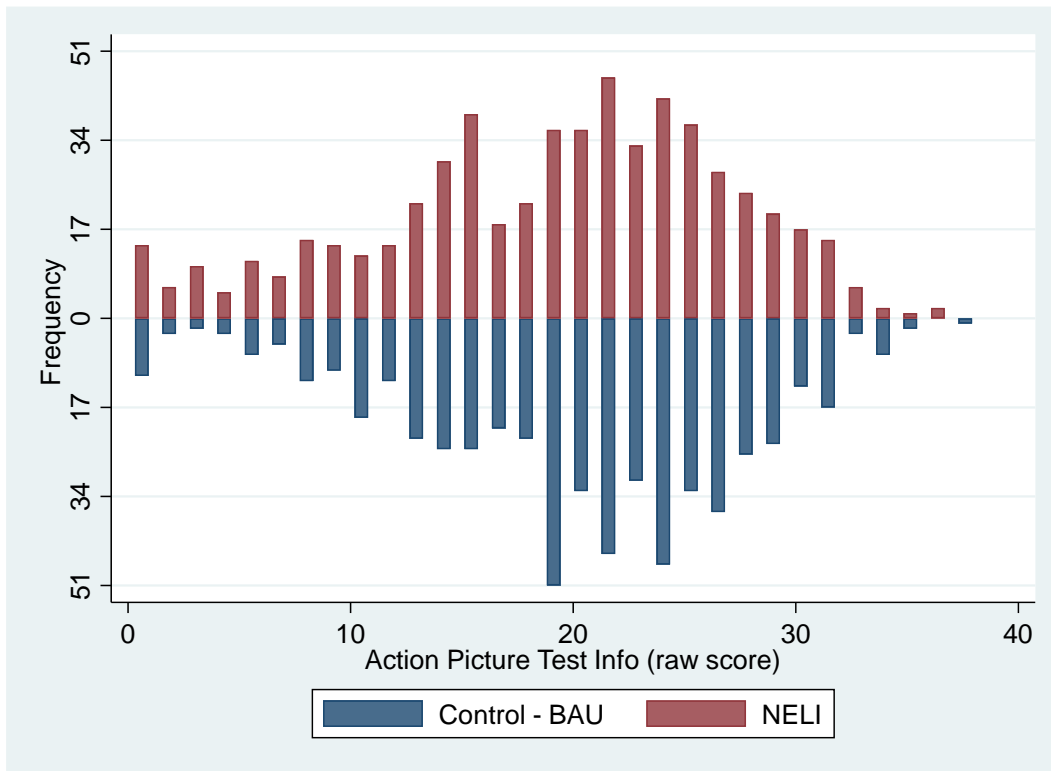


Figure 22: Distribution of RAPT Information at baseline

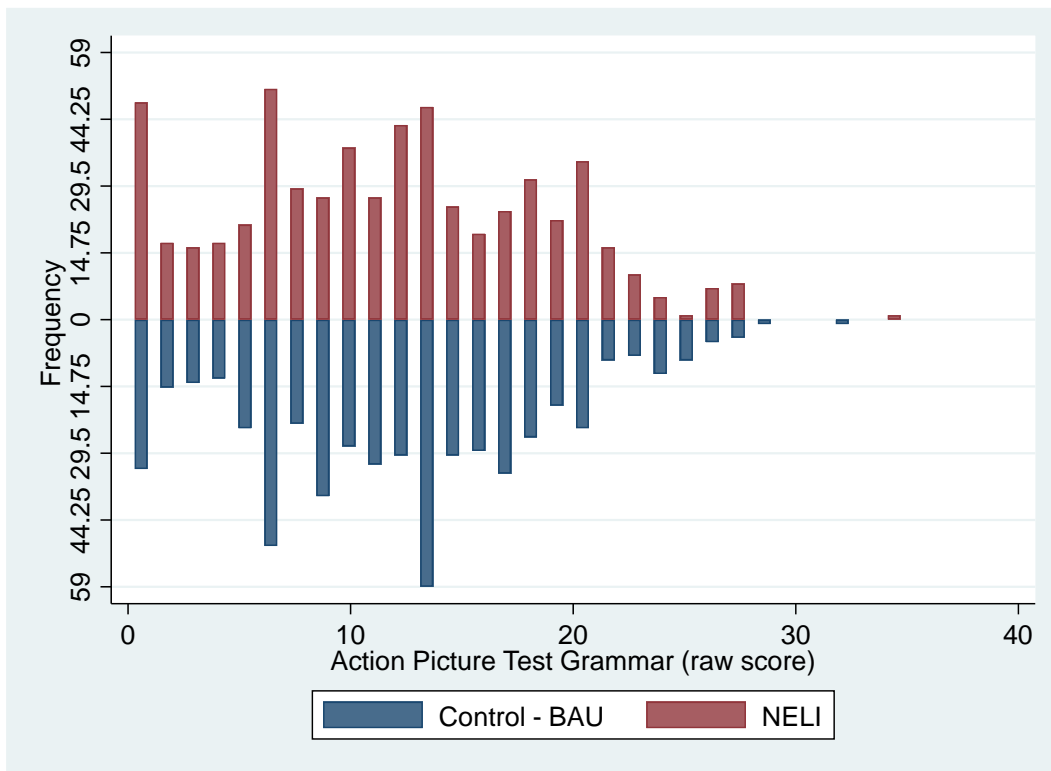


Figure 23: Distribution of RAPT Grammar at baseline

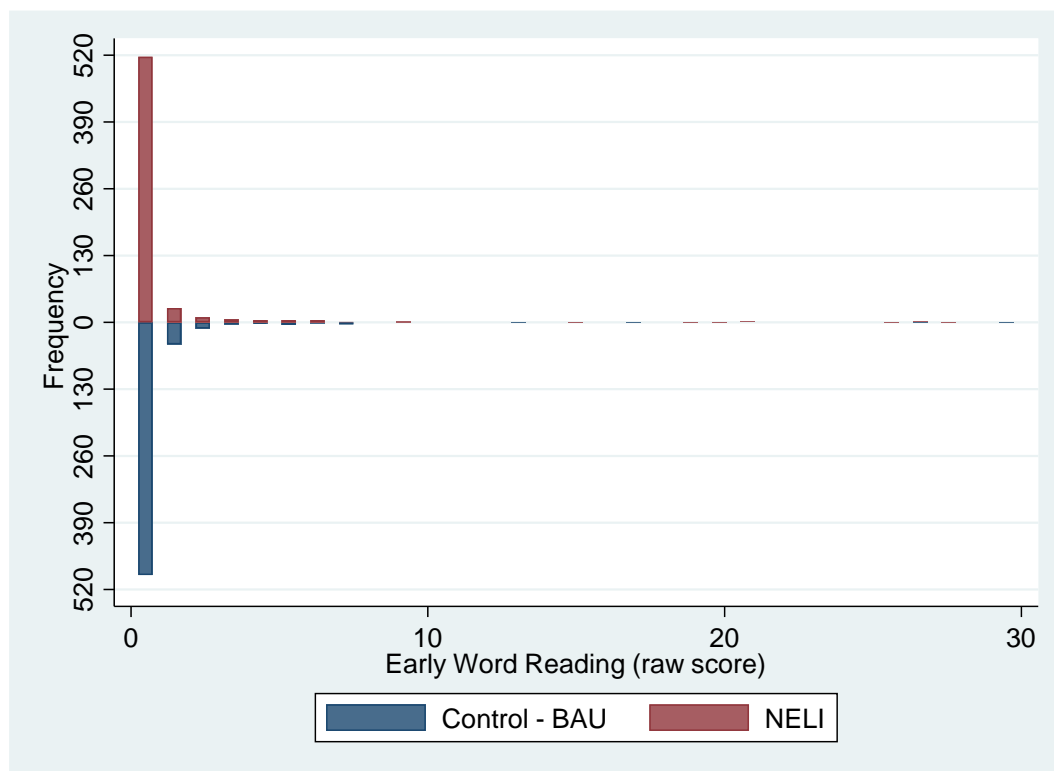


Figure 24: Distribution of Early Word Reading at baseline

Appendix G: Measure development: analytical approach

Baseline measure of primary outcome

At baseline, the original sem model to derive the main baseline measure was:

```
sem (baseline_language_latent -> RS_baseline EV_baseline RAI_baseline RAG_baseline),
latent(baseline_language_latent)
```

The initial model fit was below accepted thresholds, but only in terms of the RMSEA: CFI = 0.980; RMSEA=0.147.

The model was therefore modified, adding one covariance, fitting with the theoretical assumption outlined above.

```
sem (baseline_language_latent -> RS_baseline EV_baseline RAI_baseline RAG_baseline),
latent(baseline_language_latent) cov(e.RS_baseline*e.RAI_baseline)
```

Model fit achieved after this one modification overall met acceptable thresholds, CFI = 1.00; RMSEA = 0.017.

A factor score for the baseline language variable was then estimated.

This had an overall mean of 0 and standard deviation of 4.26 for the whole sample of children with baseline data, with a range from -10.45 to 11.45. This variable will be used in the final analysis.

Primary outcome measure

At endline, the original sem model to derive the main language skills outcome measure was:

```
sem (outcome_language_latent -> RS_endline EV_endline RAI_endline RAG_endline),
latent(outcome_language_latent)
```

The initial model fit was below accepted thresholds, again only in relation to the RMSEA: CFI = 0.972; RMSEA=0.164.

The model was therefore modified, adding one covariance only, in line with the theoretical assumption outlined above.

```
sem (outcome_language_latent -> RS_endline EV_endline RAI_endline RAG_endline),
latent(outcome_language_latent) cov(e.RAI_endline*e.RAG_endline)
```

Model fit achieved after this one modification was just slightly outside acceptable limits for the RMSEA only: CFI = 0.994; RMSEA=0.10. However, further modifications to the model were deemed unnecessary, as they would address the model fit, but add complexity to the model. Inspection of other model fit indices such as the TLI=0.963 (with rules of thumb suggesting a cut-off of 0.9 for acceptable fit) suggested sufficiently good model fit was achieved. Fuller explanations as to approaches when fit indices do not concur exist (Lai & Green, 2016) and suggest that models should not be discarded in the case of disagreement, but the underlying data investigated. This was undertaken by exploring all underlying item distributions, pairwise correlations and alternative models (not tabled) resulting in the decision to retain the current model.

A factor score for the primary outcome language variable was then estimated.

For all pupils with outcome data (1,097), this measure had an overall mean of 0 and standard deviation of 4.84 for the whole sample of children with outcome data, and a range from -15.75 to 14.00. This variable constitutes the primary outcome measure in this trial.

YARC measure

The YARC test was scored simply as a number of correct answers to the questions. This follows established procedures in previous studies that have used the test to explore pupils' language skills (Bowyer-Crane et al., 2017).

At baseline, the YARC measure had a mean 0.51 and standard deviation of 2.50 for the 1,017 pupils that took part in the in-depth testing. At endline, the YARC measure had a mean of 8.52 and standard deviation of 7.27 for the 1,097 pupils with data.

Given that each item in the YARC test was a binary variable (correct/incorrect) answer, and that the confirmatory factor analysis approach employed above relies on continuous measures as inputs, it was not appropriate to apply it to the YARC test, as this would underestimate the covariances between respective YARC test items. Alternative approaches were pursued, including item response theory (IRT), a psychometric approach falling broadly within the sem framework, though relying on different assumptions. An IRT-derived factor score capturing YARC-specific reading attainment had a correlation of 0.99 with the raw YARC score at both baseline and endline, and therefore the substantially more parsimonious model of score addition was retained as the YARC measure. The Secondary outcome discussion in the impact evaluation results further detailed issues around the YARC measure.

LanguageScreen

A similar confirmatory factor analysis approach was employed for the LanguageScreen variable, again at baseline and at outcome. The raw inputs were the raw scores on the four separate language skills assessed as part of the measure.

Both baseline, and outcome sem models for the LanguageScreen measure met acceptable model fit criteria with one modifications: (baseline, CFI, RMSEA); endline CFI, RMSEA).

The measure was collected separately to the other tests, as it is part of the intervention. Therefore, data is available

At baseline, the LanguageScreen composite measure had a mean of 0 and standard deviation of 1.96 for the 1,136 pupils who had full data on all LanguageScreen components. At endline, the measure had a mean of -0.01 and standard deviation of 2.00 for the 909 pupils that were tested and had full data on all components.

Appendix H: Randomisation code, Analysis code and output

Randomisation code

```
randtreat, generate(treatment) replace strata(num_area classes) misfits(global) setseed(872402)
```

num_area is a variable identifying the geography-based clustering;
 classes is a variable identifying whether a school is single or multi-form entry
 The seed is a random number

Primary outcome analysis

```
xtmixed outcome_language treatment baseline_language i.strata || SchoolCode:, mle var
```

```
. xtmixed outcome_language treatment baseline_language i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log likelihood = -2701.1921
Iteration 1: log likelihood = -2701.165
Iteration 2: log likelihood = -2701.165
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs      =      1,071
Group variable: SchoolCode       Number of groups    =       192
```

```
Obs per group:
      min =      2
      avg =     5.6
      max =     15
```

```
Wald chi2(27)      =    1392.49
Prob > chi2        =     0.0000

Log likelihood = -2701.165
```

outcome_language	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	1.243285	.2304533	5.39	0.000	.7916046	1.694965
baseline_language	.8379291	.024014	34.89	0.000	.7908624	.8849957
strata						
BristolBristol	-.737991	.8921076	-0.83	0.408	-2.48649	1.010508
Cornwall	-.0966919	.6234474	-0.16	0.877	-1.318626	1.125243
CornwallCornwall	.8915845	1.043458	0.85	0.393	-1.153556	2.936725
Durham	.0524421	.6905817	0.08	0.939	-1.301073	1.405957
DurhamDurham	-.8536207	.7958724	-1.07	0.283	-2.413502	.7062606
Essex	.4505366	.6114376	0.74	0.461	-.747859	1.648932
EssexEssex	-.3468428	.9289677	-0.37	0.709	-2.167586	1.4739
Herts	1.204848	.9021405	1.34	0.182	-.5633153	2.973011
HertsHerts	-.2488563	.7715645	-0.32	0.747	-1.761095	1.263382
London	.655251	.7648261	0.86	0.392	-.8437805	2.154283
LondonLondon	-.4211774	.9381545	-0.45	0.653	-2.259926	1.417572
Manchester	.4173743	.8741217	0.48	0.633	-1.295873	2.130621
ManchesterManchester	-.9675274	1.043554	-0.93	0.354	-3.012856	1.077801
North Tyneside	-.0812587	.7906164	-0.10	0.918	-1.630838	1.468321
North TynesideNorth Tyneside	-.4937858	1.094216	-0.45	0.652	-2.63841	1.650839
North West	1.158572	.7245931	1.60	0.110	-.2616044	2.578748
North WestNorth West	.5116209	1.432352	0.36	0.721	-2.295737	3.318979
Northants	-.2338378	.6389601	-0.37	0.714	-1.486176	1.018501
NorthantsNorthants	-.5183021	.8189711	-0.63	0.527	-2.123456	1.086852
Surrey	.5180151	.5966082	0.87	0.385	-.6513154	1.687346
SurreySurrey	-.4384391	.6993507	-0.63	0.531	-1.809141	.9322631
Warks	.7319002	.7939008	0.92	0.357	-.8241167	2.287917
WarksWarks	.0235903	.9309872	0.03	0.980	-1.801111	1.848292
Wolverhampton	-.1008266	.7358697	-0.14	0.891	-1.543105	1.341452
WolverhamptonWolverhampton	.0929442	.8138219	0.11	0.909	-1.502117	1.688006
_cons	-.6672592	.4958777	-1.35	0.178	-1.639162	.3046432

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	.875136	.2610836	.4876773	1.57043
var(Residual)	8.379171	.3995123	7.631615	9.199955

LR test vs. linear model: chibar2(01) = 18.95 Prob >= chibar2 = 0.0000

Secondary outcome analysis: YARC

```
xtmixed YARC_outcome treatment i.YARC_baseline_new i.strata || SchoolCode:, mle var
```

```
. xtmixed YARC_endline treatment i.YARC_baseline_new i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0:  log likelihood = -3477.0786
Iteration 1:  log likelihood = -3477.0388
Iteration 2:  log likelihood = -3477.0388
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs   =      1,071
Group variable: SchoolCode       Number of groups =       192

Obs per group:
      min =          2
      avg =         5.6
      max =         15

Wald chi2(28)    =      347.73
Prob > chi2     =      0.0000

Log likelihood = -3477.0388
```

YARC_endline	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	1.088924	.4677568	2.33	0.020	.1721376	2.00571
YARC_baseline_new						
2	4.582033	.8009792	5.72	0.000	3.012143	6.151923
3	12.80992	.7900403	16.21	0.000	11.26147	14.35837
strata						
BristolBristol	-3.017692	1.812821	-1.66	0.096	-6.570756	.5353711
Cornwall	-2.035084	1.266837	-1.61	0.108	-4.518039	.4478704
CornwallCornwall	-.158521	2.105767	-0.08	0.940	-4.285748	3.968706
Durham	-3.171693	1.404712	-2.26	0.024	-5.924877	-.4185085
DurhamDurham	-2.171313	1.612319	-1.35	0.178	-5.3314	.9887744
Essex	-1.936655	1.245798	-1.55	0.120	-4.378374	.505065
EssexEssex	-1.432121	1.88547	-0.76	0.448	-5.127574	2.263332
Herts	.151931	1.837508	0.08	0.934	-3.449519	3.753381
HertsHerts	-.3286799	1.560757	-0.21	0.833	-3.387707	2.730347
London	.432012	1.556422	0.28	0.781	-2.618518	3.482542
LondonLondon	-5.25588	1.88094	-2.79	0.005	-8.942455	-1.569306
Manchester	-2.558793	1.771104	-1.44	0.149	-6.030092	.9125062
ManchesterManchester	-5.329419	2.105948	-2.53	0.011	-9.457002	-1.201837
North Tyneside	-.8151111	1.608519	-0.51	0.612	-3.96775	2.337527
North TynesideNorth Tyneside	-1.789655	2.215311	-0.81	0.419	-6.131584	2.552273
North West	1.396864	1.476113	0.95	0.344	-1.496263	4.289992
North WestNorth West	-4.666719	2.891386	-1.61	0.107	-10.33373	1.000294
Northants	1.249011	1.299173	0.96	0.336	-1.297322	3.795344
NorthantsNorthants	-2.06331	1.652094	-1.25	0.212	-5.301355	1.174735
Surrey	-.7349499	1.214161	-0.61	0.545	-3.114662	1.644762
SurreySurrey	-1.565168	1.417436	-1.10	0.269	-4.343292	1.212956
Warks	2.186608	1.614049	1.35	0.176	-.9768692	5.350085
WarksWarks	-1.840416	1.881279	-0.98	0.328	-5.527655	1.846823
Wolverhampton	-4.508084	1.498343	-3.01	0.003	-7.444783	-1.571386
WolverhamptonWolverhampton	-2.032576	1.64301	-1.24	0.216	-5.252815	1.187664
_cons	8.244461	1.010322	8.16	0.000	6.264267	10.22466

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	3.351913	1.064447	1.798795	6.246025
var(Residual)	35.93061	1.709219	32.73203	39.44176

LR test vs. linear model: chibar2(01) = 15.81 Prob >= chibar2 = 0.0000

Secondary outcome analysis: LanguageScreen

```
xtmixed outcome_atlas treatment baseline_atlas i.strata || SchoolCode:, mle var
```

```
. xtmixed outcome_atlas treatment baseline_atlas i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -1734.939

Iteration 1: log likelihood = -1734.9339

Iteration 2: log likelihood = -1734.9339

Computing standard errors:

Mixed-effects ML regression

Group variable: SchoolCode

Number of obs = 904

Number of groups = 172

Obs per group:

min = 1

avg = 5.3

max = 12

Wald chi2(27) = 566.34

Prob > chi2 = 0.0000

Log likelihood = -1734.9339

outcome_atlas	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	.8020846	.1449607	5.53	0.000	.5179669	1.086202
baseline_atlas	.6877811	.0325442	21.13	0.000	.6239956	.7515665
strata						
BristolBristol	-.9426455	.5399609	-1.75	0.081	-2.000949	.1156584
Cornwall	.152844	.3899479	0.39	0.695	-.6114399	.917128
CornwallCornwall	-.4063789	.6329717	-0.64	0.521	-1.646981	.8342229
Durham	.0160692	.4166017	0.04	0.969	-.8004552	.8325936
DurhamDurham	-.3061664	.5014654	-0.61	0.542	-1.28902	.6766877
Essex	.3195605	.389019	0.82	0.411	-.4429027	1.082024
EssexEssex	-.5049874	.5635236	-0.90	0.370	-1.609473	.5994986
Herts	.784687	.6061414	1.29	0.195	-.4033283	1.972702
HertsHerts	-.1225869	.4707386	-0.26	0.795	-1.045218	.8000437
London	-.0351776	.5138445	-0.07	0.945	-1.042294	.9719392
LondonLondon	-1.433976	.5631764	-2.55	0.011	-2.537782	-.3301709
Manchester	.649041	.5250457	1.24	0.216	-.3800297	1.678112
ManchesterManchester	.3467205	.638097	0.54	0.587	-.9039267	1.597368
North Tyneside	.4140439	.4865293	0.85	0.395	-.5395361	1.367624
North TynesideNorth Tyneside	-.9922365	.6782072	-1.46	0.143	-2.321498	.3370251
North West	.8188347	.4457106	1.84	0.066	-.0547421	1.692411
North WestNorth West	2.0892	.8738762	2.39	0.017	.3764338	3.801965
Northants	.0882856	.3972012	0.22	0.824	-.6902145	.8667857
NorthantsNorthants	-.3364645	.4947421	-0.68	0.496	-1.306141	.6332122
Surrey	.1624853	.3713386	0.44	0.662	-.5653249	.8902955
SurreySurrey	.0000544	.4298027	0.00	1.000	-.8423435	.8424523
Warks	.5451873	.4818251	1.13	0.258	-.3991726	1.489547
WarksWarks	-1.43107	.8548456	-1.67	0.094	-3.106536	.244397
Wolverhampton	-.0441927	.5242565	-0.08	0.933	-1.071717	.9833312
WolverhamptonWolverhampton	-.7202335	.4952622	-1.45	0.146	-1.69093	.2504626
_cons	-.3508809	.3076992	-1.14	0.254	-.9539602	.2521985

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	.360932	.09565	.2147088	.6067376
var(Residual)	2.445725	.1275888	2.208015	2.709026

LR test vs. linear model: chibar2(01) = 26.12

Prob >= chibar2 = 0.0000

Sub-group analysis

```
xtmixed outcome_language ib10.EALT baseline_language i.strata || SchoolCode:, mle var
```

EALT	Freq.	Percent	Cum.
EAL-Control	183	17.09	17.09
EAL-Intervention	176	16.43	33.52
Non-EAL-Control	348	32.49	66.01
Non-EAL-Intervention	351	32.77	98.79
Unknown-EAL-Control	10	0.93	99.72
Unknown-EAL-Intervention	3	0.28	100.00
Total	1,071	100.00	

```
. xtmixed outcome_language ib10.EALT baseline_language i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log likelihood = -2698.7518
Iteration 1: log likelihood = -2698.7211
Iteration 2: log likelihood = -2698.7211
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs   =      1,071
Group variable: SchoolCode       Number of groups =        192
```

```
Obs per group:
      min =         2
      avg =        5.6
      max =        15
```

```
Wald chi2(31)    =    1405.80
Log likelihood = -2698.7211      Prob > chi2      =     0.0000
```

outcome_language	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
EALT						
EAL-Intervention	1.109708	.3629283	3.06	0.002	.3983816	1.821034
Non-EAL-Control	-.1409424	.3084777	-0.46	0.648	-.7455475	.4636626
Non-EAL-Intervention	1.102016	.3306972	3.33	0.001	.4538614	1.750171
Unknown-EAL-Control	-2.31091	1.045163	-2.21	0.027	-4.359393	-.2624279
Unknown-EAL-Intervention	1.276489	1.759548	0.73	0.468	-2.172161	4.725139
baseline_language						
	.841772	.0261597	32.18	0.000	.7904999	.8930441
strata						
BristolBristol	-.7100884	.89044	-0.80	0.425	-2.455319	1.035142
Cornwall	-.0771677	.6244749	-0.12	0.902	-1.301116	1.146781
CornwallCornwall	.9066256	1.040787	0.87	0.384	-1.13328	2.946531
Durham	.0678154	.6923032	0.10	0.922	-1.289074	1.424705
DurhamDurham	-.8273844	.7958716	-1.04	0.299	-2.387264	.7324953
Essex	.5930823	.6141486	0.97	0.334	-.6106267	1.796791
EssexEssex	-.3300279	.9273345	-0.36	0.722	-2.14757	1.487514
Herts	1.220809	.8999375	1.36	0.175	-.5430357	2.984654
HertsHerts	-.2644378	.7694141	-0.34	0.731	-1.772462	1.243586
London	.7045385	.7654946	0.92	0.357	-.7958034	2.20488
LondonLondon	-.410106	.9359986	-0.44	0.661	-2.244629	1.424417
Manchester	.4133606	.8713916	0.47	0.635	-1.294536	2.121257
ManchesterManchester	-.9742586	1.041368	-0.94	0.350	-3.015302	1.066785
North Tyneside	-.0519854	.7917308	-0.07	0.948	-1.603749	1.499778
North TynesideNorth Tyneside	-.4688412	1.093565	-0.43	0.668	-2.612188	1.674506
North West	1.151814	.7237726	1.59	0.112	-.2667544	2.570382
North WestNorth West	.5452218	1.426252	0.38	0.702	-2.25018	3.340624
Northants	-.2421803	.6367637	-0.38	0.704	-1.490214	1.005853
NorthantsNorthants	-.5246887	.8169202	-0.64	0.521	-2.125823	1.076445
Surrey	.5662288	.5959404	0.95	0.342	-.6017929	1.734251
SurreySurrey	-.4392467	.6971609	-0.63	0.529	-1.805657	.9271636
Warks	.7977714	.7944531	1.00	0.315	-.759328	2.354871
WarksWarks	.1043433	.9289286	0.11	0.911	-1.716323	1.92501
Wolverhampton	-.0840916	.7352063	-0.11	0.909	-1.525069	1.356886
WolverhamptonWolverhampton	.1023582	.8122171	0.13	0.900	-1.489558	1.694275
_cons	-.5586867	.5310794	-1.05	0.293	-1.599583	.4822098

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	.8604444	.2604935	.4753654	1.557464
var(Residual)	8.348029	.3984664	7.602468	9.166707

```
LR test vs. linear model: chibar2(01) = 18.40      Prob >= chibar2 = 0.0000
```


Fidelity analysis

```
xtmixed outcome_language fidelity_new baseline_language i.strata || SchoolCode:, mle var
```

```
. xtmixed outcome_language fidelity_new baseline_language i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -2712.228

Iteration 1: log likelihood = -2712.225

Iteration 2: log likelihood = -2712.225

Computing standard errors:

Mixed-effects ML regression	Number of obs	=	1,071
Group variable: SchoolCode	Number of groups	=	192
	Obs per group:		
	min	=	2
	avg	=	5.6
	max	=	15
	Wald chi2(27)	=	1289.82
Log likelihood = -2712.225	Prob > chi2	=	0.0000

outcome_language	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fidelity_new	1.071318	.5151045	2.08	0.038	.061732	2.080905
baseline_language	.8283574	.0244367	33.90	0.000	.7804623	.8762525
strata						
BristolBristol	-.4639313	.9722699	-0.48	0.633	-2.369545	1.441683
Cornwall	.2161074	.6755639	0.32	0.749	-1.107973	1.540188
CornwallCornwall	1.272173	1.15392	1.10	0.270	-.9894678	3.533814
Durham	.3254098	.7398501	0.44	0.660	-1.12467	1.775489
DurhamDurham	-.7850811	.8641347	-0.91	0.364	-2.478754	.9085917
Essex	.577862	.6521479	0.89	0.376	-.7003245	1.856048
EssexEssex	-.2158878	1.018831	-0.21	0.832	-2.21276	1.780985
Herts	1.214865	.9584924	1.27	0.205	-.6637454	3.093476
HertsHerts	-.2340218	.8384391	-0.28	0.780	-1.877332	1.409289
London	.7739378	.8155996	0.95	0.343	-.8246079	2.372484
LondonLondon	.0237467	1.022956	0.02	0.981	-1.98121	2.028704
Manchester	.2270024	.9295928	0.24	0.807	-1.594966	2.048971
ManchesterManchester	-.673348	1.147937	-0.59	0.557	-2.923263	1.576567
North Tyneside	.2323704	.8446839	0.28	0.783	-1.42318	1.88792
North TynesideNorth Tyneside	-.4660486	1.186568	-0.39	0.694	-2.791679	1.859582
North West	1.283139	.7769814	1.65	0.099	-.2397169	2.805994
North WestNorth West	1.317802	1.560829	0.84	0.399	-1.741366	4.37697
Northants	-.1384039	.6841442	-0.20	0.840	-1.479302	1.202494
NorthantsNorthants	-.4392945	.8909709	-0.49	0.622	-2.185565	1.306976
Surrey	.676832	.639344	1.06	0.290	-.5762591	1.929923
SurreySurrey	-.3595813	.7599871	-0.47	0.636	-1.849129	1.129966
Warks	.861561	.8430299	1.02	0.307	-.7907474	2.513869
WarksWarks	-.0935415	1.018515	-0.09	0.927	-2.089795	1.902712
Wolverhampton	-.0577402	.7816482	-0.07	0.941	-1.589743	1.474262
WolverhamptonWolverhampton	.3954205	.8982577	0.44	0.660	-1.365132	2.155973
_cons	-.1950806	.5175601	-0.38	0.706	-1.20948	.8193186

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	1.23667	.2938931	.7761875	1.970338
var(Residual)	8.343851	.396917	7.601073	9.159214

LR test vs. linear model: chibar2(01) = 35.04 Prob >= chibar2 = 0.0000

Sensitivity analyses

Expressive vocabulary

```
. xtmixed EV treatment EV_baseline i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0:   log likelihood = -3159.0255
Iteration 1:   log likelihood = -3158.9914
Iteration 2:   log likelihood = -3158.9914
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs      =      1,071
Group variable: SchoolCode      Number of groups   =      192

Obs per group:
    min =          2
    avg =         5.6
    max =         15

Wald chi2(27)      =      1101.00
Prob > chi2       =      0.0000

Log likelihood = -3158.9914
```

EV	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	1.483898	.351455	4.22	0.000	.7950589	2.172737
EV_baseline	.775119	.0252318	30.72	0.000	.7256657	.8245724
strata						
BristolBristol	-1.022433	1.35988	-0.75	0.452	-3.687748	1.642883
Cornwall	.6467653	.9516661	0.68	0.497	-1.218466	2.511997
CornwallCornwall	-.9397072	1.593894	-0.59	0.555	-4.063682	2.184267
Durham	-.6432372	1.056098	-0.61	0.542	-2.713151	1.426677
DurhamDurham	-1.852089	1.214352	-1.53	0.127	-4.232176	.527997
Essex	-.5155227	.9355184	-0.55	0.582	-2.349105	1.31806
EssexEssex	-.7984703	1.416881	-0.56	0.573	-3.575506	1.978565
Herts	.1465732	1.379189	0.11	0.915	-2.556587	2.849733
HertsHerts	-.8380996	1.176653	-0.71	0.476	-3.144297	1.468097
London	-1.422074	1.168237	-1.22	0.223	-3.711776	.8676288
LondonLondon	-2.509382	1.426531	-1.76	0.079	-5.305332	.2865681
Manchester	-1.564078	1.331299	-1.17	0.240	-4.173376	1.045219
ManchesterManchester	-.653106	1.590848	-0.41	0.681	-3.77111	2.464898
North Tyneside	.0810946	1.208354	0.07	0.946	-2.287236	2.449425
North TynesideNort..	-.4415149	1.66905	-0.26	0.791	-3.712793	2.829763
North West	1.724187	1.107022	1.56	0.119	-.4455374	3.893911
North WestNorth West	.4603127	2.184004	0.21	0.833	-3.820257	4.740882
Northants	-2.215158	.9749407	-2.27	0.023	-4.126007	-.3043093
NorthantsNorthants	-2.905816	1.24635	-2.33	0.020	-5.348617	-.4630143
Surrey	.1641506	.9117586	0.18	0.857	-1.622863	1.951165
SurreySurrey	-.0137361	1.067108	-0.01	0.990	-2.105229	2.077757
Warks	.3832691	1.213006	0.32	0.752	-1.994179	2.760717
WarksWarks	-.9967984	1.420212	-0.70	0.483	-3.780363	1.786766
Wolverhampton	-1.364726	1.123313	-1.21	0.224	-3.566379	.836926
WolverhamptonWolve~n	-1.4866	1.2396	-1.20	0.230	-3.916171	.9429707
_cons	7.879824	.7993657	9.86	0.000	6.313096	9.446552

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	2.002621	.6220695	1.089411	3.681336
var(Residual)	19.73775	.9448159	17.97016	21.67921

LR test vs. linear model: chibar2(01) = 16.39 Prob >= chibar2 = 0.0000

Recalling sentences

```
. xtmixed RS treatment RS_baseline i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0:  log likelihood = -3272.1882
Iteration 1:  log likelihood = -3271.6915
Iteration 2:  log likelihood = -3271.6915
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs    =      1,071
Group variable: SchoolCode      Number of groups  =       192

Obs per group:
    min =          2
    avg =         5.6
    max =         15

Wald chi2(27)    =     1114.37
Prob > chi2      =       0.0000

Log likelihood = -3271.6915
```

RS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	.5224842	.3478417	1.50	0.133	-.159273	1.204241
RS_baseline	.8333908	.0262961	31.69	0.000	.7818514	.8849301
strata						
BristolBristol	.540707	1.29105	0.42	0.675	-1.989704	3.071118
Cornwall	-.182391	.9627246	-0.19	0.850	-2.069296	1.704514
CornwallCornwall	-.5466498	1.507331	-0.36	0.717	-3.500964	2.407664
Durham	2.40751	1.071229	2.25	0.025	.3079385	4.507081
DurhamDurham	-.3796701	1.188133	-0.32	0.749	-2.708367	1.949027
Essex	1.200508	.9453174	1.27	0.204	-.6522796	3.053296
EssexEssex	2.296658	1.372909	1.67	0.094	-.3941945	4.987511
Herts	1.330745	1.400035	0.95	0.342	-1.413273	4.074763
HertsHerts	.7626314	1.141942	0.67	0.504	-1.475534	3.000797
London	.5971967	1.183146	0.50	0.614	-1.721727	2.91612
LondonLondon	.6302245	1.381819	0.46	0.648	-2.07809	3.338539
Manchester	2.024428	1.344673	1.51	0.132	-.6110836	4.659939
ManchesterManchester	-.2467187	1.506923	-0.16	0.870	-3.200234	2.706797
North Tyneside	2.016523	1.219879	1.65	0.098	-.3743966	4.407442
North TynesideNorth..	-.48452	1.61127	-0.30	0.764	-3.642551	2.673511
North West	.6822532	1.117508	0.61	0.542	-1.508022	2.872528
North WestNorth West	.9402531	2.073527	0.45	0.650	-3.123785	5.004291
Northants	-1.07351	.9851005	-1.09	0.276	-3.004272	.8572512
NorthantsNorthants	-.8949323	1.197936	-0.75	0.455	-3.242844	1.452979
Surrey	-.0471684	.9219952	-0.05	0.959	-1.854246	1.759909
SurreySurrey	-.0639372	1.041038	-0.06	0.951	-2.104335	1.976461
Warks	.9700982	1.227935	0.79	0.430	-1.436609	3.376806
WarksWarks	.396965	1.353241	0.29	0.769	-2.255338	3.049268
Wolverhampton	-.7965028	1.13812	-0.70	0.484	-3.027177	1.434171
WolverhamptonWolve~n	-.0507117	1.191839	-0.04	0.966	-2.386674	2.285251
_cons	6.684977	.801271	8.34	0.000	5.114515	8.255439

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	.8736395	.5851423	.2350815	3.24673
var(Residual)	25.55359	1.209993	23.28877	28.03867

LR test vs. linear model: chibar2(01) = 2.76 Prob >= chibar2 = 0.0482

RAPT Information

```
. xtmixed RAI treatment RAI_baseline i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0:  log likelihood = -3138.3949
Iteration 1:  log likelihood = -3138.3695
Iteration 2:  log likelihood = -3138.3695
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs      =      1,071
Group variable: SchoolCode       Number of groups   =       192

Obs per group:
      min =          2
      avg =         5.6
      max =         15

Wald chi2(27)      =      705.01
Prob > chi2       =      0.0000

Log likelihood = -3138.3695
```

RAI	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	1.032766	.3452765	2.99	0.003	.3560367	1.709496
RAI_baseline	.4814155	.0203565	23.65	0.000	.4415175	.5213134
strata						
BristolBristol	-1.127823	1.333723	-0.85	0.398	-3.741873	1.486227
Cornwall	1.464578	.9353652	1.57	0.117	-.3687036	3.297861
CornwallCornwall	2.162405	1.560117	1.39	0.166	-.8953679	5.220177
Durham	-1.039658	1.035428	-1.00	0.315	-3.069059	.9897437
DurhamDurham	-1.780915	1.191037	-1.50	0.135	-4.115306	.5534751
Essex	1.343722	.9164403	1.47	0.143	-.452468	3.139912
EssexEssex	-.0208309	1.389484	-0.01	0.988	-2.744169	2.702507
Herts	3.079373	1.350854	2.28	0.023	.4317477	5.726999
HertsHerts	-.1309596	1.154441	-0.11	0.910	-2.393622	2.131703
London	1.556991	1.14648	1.36	0.174	-.6900695	3.804051
LondonLondon	-1.626273	1.401978	-1.16	0.246	-4.374101	1.121554
Manchester	-.5968632	1.309821	-0.46	0.649	-3.164064	1.970338
ManchesterManchester	-2.682905	1.560394	-1.72	0.086	-5.741221	.3754114
North Tyneside	-1.754555	1.184557	-1.48	0.139	-4.076244	.567133
North TynesideNorth..	-.0245501	1.637045	-0.01	0.988	-3.233099	3.183999
North West	1.811582	1.085759	1.67	0.095	-.3164674	3.939632
North WestNorth West	1.52957	2.141852	0.71	0.475	-2.668384	5.727523
Northants	.4253386	.9577725	0.44	0.657	-1.451861	2.302538
NorthantsNorthants	-.8291374	1.22535	-0.68	0.499	-3.230779	1.572504
Surrey	1.283184	.8940533	1.44	0.151	-.4691284	3.035496
SurreySurrey	.0036663	1.046867	0.00	0.997	-2.048156	2.055489
Warks	1.743889	1.189337	1.47	0.143	-.5871694	4.074948
WarksWarks	1.38072	1.392818	0.99	0.322	-1.349152	4.110592
Wolverhampton	.2864457	1.103014	0.26	0.795	-1.875422	2.448314
WolverhamptonWolve~n	.8033764	1.216913	0.66	0.509	-1.581729	3.188482
_cons	15.40591	.8522419	18.08	0.000	13.73555	17.07628

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	1.925747	.5714637	1.07648	3.445026
var(Residual)	18.99292	.9015672	17.30559	20.84476

LR test vs. linear model: chibar2(01) = 19.05 Prob >= chibar2 = 0.0000

RAPT Grammar

```
. xtmixed RAG treatment RAG_baseline i.strata || SchoolCode:, mle var
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0:   log likelihood = -3194.9308
Iteration 1:   log likelihood = -3194.9185
Iteration 2:   log likelihood = -3194.9185
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs   =      1,071
Group variable: SchoolCode      Number of groups =       192

Obs per group:
      min =          2
      avg =         5.6
      max =         15

Wald chi2(27)    =      675.14
Prob > chi2     =      0.0000

Log likelihood = -3194.9185
```

RAG	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treatment	1.803068	.3747403	4.81	0.000	1.068591	2.537546
RAG_baseline	.5769344	.0240276	24.01	0.000	.5298412	.6240276
strata						
BristolBristol	-1.53763	1.462209	-1.05	0.293	-4.403506	1.328247
Cornwall	-.0349105	1.009365	-0.03	0.972	-2.013229	1.943408
CornwallCornwall	1.190614	1.712594	0.70	0.487	-2.16601	4.547237
Durham	.8490139	1.11802	0.76	0.448	-1.342266	3.040294
DurhamDurham	-.2086202	1.302489	-0.16	0.873	-2.761452	2.344212
Essex	.3512614	.9906108	0.35	0.723	-1.5903	2.292823
EssexEssex	-1.474828	1.518177	-0.97	0.331	-4.450401	1.500745
Herts	.1499181	1.464243	0.10	0.918	-2.719945	3.019781
HertsHerts	-1.545028	1.259466	-1.23	0.220	-4.013536	.9234793
London	1.370397	1.237963	1.11	0.268	-1.055966	3.79676
LondonLondon	-.1347961	1.52804	-0.09	0.930	-3.129699	2.860107
Manchester	-.528579	1.414306	-0.37	0.709	-3.300568	2.24341
ManchesterManchester	-.5305248	1.711557	-0.31	0.757	-3.885115	2.824065
North Tyneside	-2.127065	1.281627	-1.66	0.097	-4.639007	.3848775
North TynesideNorth..	-2.007205	1.788625	-1.12	0.262	-5.512845	1.498434
North West	.7306597	1.174551	0.62	0.534	-1.571418	3.032737
North WestNorth West	-.8322574	2.347901	-0.35	0.723	-5.434059	3.769544
Northants	.1679159	1.035289	0.16	0.871	-1.861212	2.197044
NorthantsNorthants	-.2249323	1.339963	-0.17	0.867	-2.851211	2.401346
Surrey	1.229406	.966623	1.27	0.203	-.6651406	3.123952
SurreySurrey	-.8228058	1.141138	-0.72	0.471	-3.059395	1.413783
Warks	.235607	1.289498	0.18	0.855	-2.291763	2.762977
WarksWarks	.7844432	1.524215	0.51	0.607	-2.202964	3.771851
Wolverhampton	-.6174806	1.191405	-0.52	0.604	-2.952592	1.717631
WolverhamptonWolve~n	.3132677	1.330846	0.24	0.814	-2.295142	2.921677
_cons	10.27691	.859441	11.96	0.000	8.592435	11.96138

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
SchoolCode: Identity				
var(_cons)	2.55323	.6946212	1.498013	4.351754
var(Residual)	20.84456	.9955301	18.9819	22.89001

LR test vs. linear model: chibar2(01) = 23.78 Prob >= chibar2 = 0.0000

Appendix I: Process Evaluation: Data collection

Survey Data

Survey platform All responses have been treated confidentially and stored by SmartSurvey. RAND Europe have transfer data securely from SmartSurvey. Survey respondents have had an opportunity to complete survey online on multiple occasions if required. Furthermore, we gave respondents a sufficiently long data collection window (between three to four weeks) to respond in order to provide as much flexibility for the respondents as possible.

Table 28: Online Survey Collection Activity

Data from	Time of collection	Topics
Head teaches/ – All schools	November 2018	<ul style="list-style-type: none"> - Motivations for joining the study - Preparedness and understanding of school requirements for joining the study - Other interventions/programmes for language taking place in Reception classes
TAs in intervention schools	June 2019	<ul style="list-style-type: none"> -Views on effectiveness of the training and other support and resource materials received for NELI - Support from teachers/school received - Perceived barriers and facilitators to success of intervention - Perceived impact of NELI
Teachers in intervention schools	June 2019	<ul style="list-style-type: none"> - Perception of 0.5 day of training received for this intervention - Support provided to the TAs (letting TAs take children out of class, providing them with space) - Perceived barriers and facilitators to programme delivery - Perception of the usefulness of the programme
Head teacher/ – intervention schools	Early July 2019	<ul style="list-style-type: none"> - Any change in literacy practices over the course of the academic year - Perceived barriers and facilitators to success of intervention - Data on costs associated with NELI implementation
Head teacher/ – control schools	Early July 2019	<ul style="list-style-type: none"> - Any change in literacy practices over the course of the academic year

Interview Protocols

NELI Head Teacher	Answers
Context and overview of language needs	
1. To start, could you tell me a bit about your role in the school?	
2. What is your role in relation to NELI?	
3. What is your overall understanding of the aims of NELI	
4. General views on the principle of early years interventions	
5. Are there any other language interventions targeted at children in Reception? 6. What are these? 7. And how do they sit with NELI? (e.g. reinforce it; replicate it; are separate from it)	
Reason for joining NELI	
8. Could you please describe what motivated you and your school to take part in the NELI intervention? [Previous research has shown that some factors influencing the decisions by school to participate in the programme are the following: personal interest; schools' aspirations; the needs of children;	
9. In your view, did the school have the capacity to deliver the programme; reputation of the institutions delivering the programme?	
10. Do you have previous experience with similar programmes in the last years	
Experience with NELI	
11. How would you describe your overall experience with the NELI intervention? <ul style="list-style-type: none"> What was the extent of involvement (light/heave touch; one-off tasks)? What was your extent of preparation and planning needed to deliver NELI (e.g. things to consider: timetabling; school/classroom space; pressure on school timetable) 	
12. In your view did the school receive sufficient support during implementation?	
Resources	
13. How would you describe the resources/facilities available to deliver the NELI intervention in your Reception class? <ul style="list-style-type: none"> How did you describe the quantity of the resources/ facilities available to deliver NELI? How did you describe the quality of the resources/ facilities available to deliver NELI? <ul style="list-style-type: none"> Was anything that was missing regarding resources/facilities? What was that and how did this affect the delivery of the NELI intervention at your Reception class? What resources/facilities enabled the delivery of the intervention in your Reception class? 	
14. Based on your experience to date, what do you feel are the key enablers and challenges to delivering the NELI intervention at your school? (<i>what worked really well and at the same time what worked less well during implementation of the NELI intervention in your school? could you please provide us with specific examples focusing on up to 3 key ones</i>)	
Effectiveness and impact of NELI intervention at your school	
15. Is your school likely to continue with the programme after the trial?	

16. If this trial was to be again at your school, what would you change and why in relation to the following aspects? (<i>what are the two or three most important things that need to happen?</i>)	
<ul style="list-style-type: none"> • Preparation and training of teachers and TAs • Resources/facilities provided to your school • Staff engagement 	
17. Are there any further comments that you would like to add?	
18. Do you have any suggestions for the implementation of the intervention in the future?	

NELI Teacher	Answers
Context and overview of language needs	
19. To start, could you tell me a bit about your role in the school?	
20. What is your role in relation to NELI? <ul style="list-style-type: none"> • Any history of delivering language intervention 	
21. Describe any language interventions that took place prior to NELI (How these compares to NELI?)	
22. What is your overall understanding of the overall aim of NELI? (prompt: How familiar you are with NELI?)	
23. General views on NELI <ul style="list-style-type: none"> • Too soon/too late for children • Does intervention provide the right structure/an approach • How does what is taught in intervention complements, conflicts or replicates language classes in Reception • What do they think about the focus of the intervention- Active listening, vocabulary, phonics (Do you think is appropriate for this group of children?) 	
Training and Preparedness for NELI	
24. Did you attend day one of the initial NELI training <ul style="list-style-type: none"> • What did you think of the <u>overall structure</u> of the training? (prompts: duration, topics covered and if these were suitable for the pupils?) • Challenges and what worked well 	
Experience of delivering NELI	
25. Deciding who delivers NELI <ul style="list-style-type: none"> • How selected/recruited were the TAs? How many TAs in your Reception class were involved in the delivery of NELI? • Views on the appropriateness of TAs delivering programme (if so, level of TA important? E.g. Advanced TAs) 	
26. Resource implications <ul style="list-style-type: none"> • Staff Time: To what extend the TA(s) in your Reception class needed additional support from you and/or the head teacher to deliver NELI? How would you describe the support provided by you the head teachers of the school to the TAs in order to deliver NELI in your Reception class? • In how far has the use of NELI intervention affected the TAs in your Reception class workload? (Prompt: Did they have to work overtime? Or did it ease their workload? How did they perceive the additional work, if any?) • School time (e. g. fitting sessions in school day, week, term) • Dedicated spaces to deliver group and individual sessions 	

Experience with NELI	
27. Based on your experience to date, what do you feel are the key enablers and challenges to delivering the NELI intervention in your Reception class? (<i>what worked really well and at the same time what worked less well during delivery of the NELI intervention in your Reception class? could you please provide us with specific examples focusing on up to 3 key ones</i>)	
28. Selection of children In regards to the selection of children to take part in NELI, what's your opinion regarding that? Was the sample of children suitable for the programme and its activities? (What's your opinion about the selection of children who took part in NELI?)	
Effectiveness and impact of NELI intervention	
29. Impact on children <ul style="list-style-type: none"> pupils' oral language outcomes (e.g. oral language skills, vocabulary, narrative, listening and general language; early word reading outcomes) how, if at all, the implementation of the NELI intervention has impacted different groups of pupils (SEN/ EAL, all children) in any way? Wider impact of NELI (e.g. on pupil's soft skills, other classes); Overall school performance (e. g. absence in classes) In your view, how, if at all, the intervention affected the kids that did not receive the intervention? 	
30. TA Impact <ul style="list-style-type: none"> How, if at all, the implementation of the NELI intervention has impacted the way TAs in your Reception class work in any way? What type of skills the TA(s) in your Reception class developed throughout the programme? What types of skills you think are needed to deliver NELI? In your view, how confident the TAs in your Reception class are in delivering the intervention? 	
31. Are there any further comments that you would like to add?	
32. Do you have any suggestions for the implementation of the intervention in the future?	

NELI Teaching Assistant	Answers
Context and overview of language needs	
33. To start, could you tell me a bit about your role in the school?	
34. What is your role in relation to NELI?	
35. Describe any language interventions that took place prior to NELI (How these compares to NELI?)	
36. What is your overall understanding of the overall aim of NELI? (prompt: How familiar you are with NELI?)	
37. General views on the principle of early years interventions	
Training and Preparedness for NELI	

<p>38. What type of training did you receive for the implementation of the NELI intervention in your Reception class? (this should be a two day training; followed by top up training and ongoing support)</p> <ul style="list-style-type: none"> • Which aspects of the training have been <u>most useful/least useful</u> for delivering NELI in your Reception class? Were there any aspects of the training that you would <u>change</u>? (in terms of effectiveness of delivery of the training / modules, content) • What did you think of the overall <u>structure of the training</u>? (prompts: duration, topics covered and if these were suitable for the pupils?) • Can you identify any gaps of the training? • To what extent do you feel the training prepared you to <u>deliver/apply</u> the NELI intervention in your Reception class? 	
<p>39. Ongoing support</p> <p>40. Description of support. Check if they have received any of the following</p> <ul style="list-style-type: none"> • Email support • Online forums • I CAN manual • Impact of support/lack of support on delivery • Gaps in support and how they would improve it <p><u>Other forms of support</u></p> <ul style="list-style-type: none"> • Source (e. g. school, teacher, other TAs), Impact of support on delivery How would you describe the support you received from Head Teachers and teachers in your school to deliver NELI in your Reception class? • Gaps and improvements 	
Delivery Structure	
<p>41. <u>Individual sessions</u></p> <ul style="list-style-type: none"> • 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? • Preparation time [2 hours in total a week for] • 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. corrective feedback; modelling; highlighting; prompting; scaffolding; multisensory learning; progressing; consolidating] 	
<p>42. <u>Group Sessions</u></p> <ul style="list-style-type: none"> • Preparation time [should be 2 hours in Reception] • Duration of sessions [30 minutes] • How pupil progress is monitored in individual and group sessions 	
<p>43. <u>Factors that affect delivery</u></p> <ul style="list-style-type: none"> • NELI related (NELI 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 30 minutes) • Pupil related factors (appropriateness of pupils to intervention, attendance, engagement) • Parent related factors (level of engagement with child's education) 	
Experience with NELI	

44. In how far has the use of NELI intervention affected your workload? (Prompt: Did you have to work overtime? Or did it ease your workload? How did you perceive the additional work, if any?)	
45. Based on your experience to date, what do you feel are the key enablers and challenges to delivering the NELI intervention in your Reception class? (what worked really well and at the same time what worked less well during delivery of the NELI intervention in your Reception class? could you please provide us with specific examples focusing on up to 3 key ones)	
46. In regards to the selection of children to take part in NELI, what's your opinion regarding that? Was the sample of children suitable for the programme and its activities? (What's your opinion about the selection of children who took part in NELI?)	
Resources	
47. How would you describe the resources/facilities available to deliver the NELI intervention in your Reception class? <ul style="list-style-type: none"> • How did you describe the quantity of the resources/ facilities available to deliver NELI? • How did you describe the quality of the resources/ facilities available to deliver NELI? <ul style="list-style-type: none"> ○ Was anything that was missing regarding resources/facilities? What was that and how did this affect the delivery of the NELI intervention at your Reception class? ○ What resources/facilities enabled the delivery of the intervention in your Reception class? 	
Effectiveness and impact of NELI intervention	
48. Impact on pupils: How, if at all, the implementation of the NELI intervention has impacted the following <ul style="list-style-type: none"> • Language & literacy (e.g. oral language skills, vocabulary, narrative, listening and general language; early word reading) • Wider impacts (e.g. soft skills, other classes) • Overall school performances (e.g. missing out on classes) • The sustainability of impact – long-term or short-term • If none, why not? • Who has the programme impacted most and least? Reasons for this. How, if at all, the implementation of the NELI intervention has impacted different groups of pupils (SEN/ EAL, all children) in any way? 	
49. How, if at all, the implementation of the NELI intervention has impacted the way you work in any way? <ul style="list-style-type: none"> • What type of skills you developed throughout the programme? • What type of skills you think are needed to deliver NELI? • Do you feel confident in delivering the intervention in your Reception class? 	
50. If this trial was to be again at your Reception class at your school, what would you change and why in relation to the following aspects? (<i>what are the two or three most important things that need to happen?</i>)	
Other Comments	
51. Are there any further comments that you would like to add?	
52. Do you have any suggestions for the implementation of the intervention in the future?	

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> 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 <https://educationendowmentfoundation.org.uk>



The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

<https://educationendowmentfoundation.org.uk>

 @EducEndowFoundn

 Facebook.com/EducEndowFoundn