A randomised controlled trial of the effectiveness of the Nuffield Early Language Intervention (NELI) – Preschool programme Evaluation Protocol

Evaluator (institution): National Foundation for Educational Research
Principal investigator(s): Dr Stephen Welbourne

Evaluation summary

Project title	A randomised controlled trial of the effectiveness of the Nuffield Early Language Intervention (NELI) – Preschool programme
Developer *(Institution)*	OxEd and Assessment (OxEd)
Evaluator *(Institution)*	National Foundation for Educational Research (NFER)
Principal investigator(s)	Dr Stephen Welbourne
Protocol author(s)	Palak Roy, Lillian Flemons, Gemma Schwendel, Stephen Welbourne, Elena Rosa Speciani and Merrilyn Groom
Trial design	Two-arm cluster randomised controlled trial with random allocation at setting level
Trial type	Effectiveness
Child age range and Key stage	3-4-year-old
Number of settings *(at design stage)*	318
Number of children *(at design stage)*	3,816
Primary outcome measure and source	Latent oral language variable formed from expressive and receptive subtests in LanguageScreen and the Renfrew Action Picture Test (RAPT)
Secondary outcome measure and source	Individual expressive and receptive subtests of LanguageScreen and the RAPT 1. LanguageScreen Expressive Vocabulary 2. LanguageScreen Listening Comprehension 3. LanguageScreen Receptive Vocabulary 4. LanguageScreen Sentence Repetition 5. RAPT information 6. RAPT grammar

Protocol version history

Version	Date	Reason for revision
1.0 [*original*]	01/08/2024	N/A

Contents

Tables

Figures

Appendix tables

Study rationale and background

Language serves as the cornerstone of education, acting as the gateway to literacy, science, and mathematics, all of which are predominantly conveyed through the medium of language. Neuroscientific research indicates that the neural mechanisms that underpin reading largely overlap with those that support language (Welbourne *et al.*, 2011); (Woollams, Halai and Lambon Ralph, 2018). The phonological representations crucial for spoken language comprehension also play a pivotal role in decoding written words, elucidating the significance of early phonological awareness in predicting later literacy success (Liberman, Shankweiler and Liberman, 1989); (Snowling *et al.*, 1994). This would suggest that early language interventions hold significant potential to shape language skills and there is a considerable body of research evidence supporting this view (Law *et al.*, 2017); (Fricke *et al.*, 2012); (Scarborough, 2009); (Marulis and Neuman, 2010). Furthermore, a recent meta-analysis focusing on early years language and communication programmes has underscored the potential for substantial positive impact on children (EEF, 2023b). However, in the UK, the evaluation of such programmes remains limited by methodological robustness, highlighting a pressing need for rigorous evaluations to bolster the evidence base. Hence, there exists a compelling argument for the development and evaluation of interventions that foster a rich language environment, through teaching speaking and listening skills, to ensure that all children have the opportunities they need to develop strong language skills, especially children from disadvantaged background.

This evaluation is also integral to the Department of Education's (DfE) Stronger Practice Hubs (SPH) policy, which aims to foster evidence-informed practice in early years education. Launched in November 2022 as part of the DfE's Early Years Education COVID-19 Recovery Package, SPHs play a crucial role in addressing the pandemic's impact on young children. They achieve this by aiding early years settings in establishing local networks and disseminating evidence-informed practices, ultimately enhancing the quality of education and care. The Educational Endowment Foundation (EEF) actively supports the SPH initiative by enhancing the evidence base for early years approaches. This involves identifying evidence-informed programmes, funding evaluation-related activities, and overseeing project delivery and evaluation, ensuring ongoing progress monitoring.

The Nuffield Early Language Intervention (NELI) Preschool (formerly known as NELI Nursery or NELI-N) is an oral language enrichment programme for all children in the year before they begin school. It builds on two previous language intervention programmes, combining the focus on children's vocabulary, narrative and active listening development in NELI and the shared reading component of PACT (Parents and Children Together). NELI is an early language intervention programme for children with weak oral language skills, delivered by teaching assistants (TAs) in reception, and has been widely evaluated. The EEF has funded two independent randomised controlled trials for NELI delivery in reception year. The efficacy trial tested two versions of the programme, a 30-week and a 20-week intervention, both of which showed positive intervention effects on pupils' language skills. 350 children in 34 schools took part in the efficacy trial which showed positive impact (Sibieta, Kotecha and Skipp, 2016); (Fricke *et al.*, 2017). In the 30-week intervention, where nursery children received the programme in the final term of nursery and continued in reception year, children made an additional 4 months' progress, compared to the 20-week intervention, delivered in

reception only where children made an additional 2 months' progress. The programme was also evaluated as an effectiveness trial with 1,071 children from 192 schools (Dimova *et al.*, 2020); (West *et al.*, 2021). This trial tested the 20-week intervention in reception, and it showed additional 3 months' progress in language skills. It was subsequently scaled up nationally with funding from the DfE. The EEF funded an independent impact evaluation of the scale-up which was evaluated using a quasi-experimental Fuzzy Regression Discontinuity (FRD) design (Smith *et al.*, 2023). 10,759 children from 356 schools took part in this study, which found that children who received the NELI programme made the equivalent of 4 additional months' progress in language skills, on average, compared to children who did not receive NELI. PACT is an early language programme for parents to deliver to children at home. The programme includes parent-child interactions that involve elements of shared book reading, explicit vocabulary teaching and retelling aspects of the story. The efficacy trial for PACT was funded by the Nuffield Foundation and involved 208 preschool children (aged 3-years) and their parents in 22 children's centres. The trial showed improvements in children's language and narrative skills at immediate post-test (Burgoyne *et al.*, 2018).

The NELI Preschool programme was developed by Charles Hulme and his research group at the University of Oxford, and is now delivered by OxEd and Assessment (OxEd), a University spinout company founded to translate the research into practical application. It is a 20-week oral language enrichment programme for 3-4-year-old children in the year before they enter formal education. The programme aims to support language development in early years settings via enrichment and targeted components. It is designed around the principles of shared book reading and guided play. The programme comprises a blend of language screening for children, online training and delivery support for practitioners, a scripted programme for in-person delivery in preschools, and supportive materials such as storybooks and digital resources. Enrichment sessions focus on whole-class activities, such as reading books, engaging in dialogical questioning and activities to support the learning of related vocabulary, while the targeted component involves small-group and one-to-one sessions tailored to children at the bottom 20-25% of the class[1] in oral language skills, incorporating activities to support and consolidate learning, enhance narrative skills and scaffolded language production.

Between 2020 and 2022, NELI Preschool was developed and then evaluated in an efficacy trial, funded by the Nuffield Foundation and conducted by Hulme's research team at the University of Oxford (West et al., 2023). 65 settings (predominantly maintained nurseries and a handful of Private, Voluntary and Independent (PVI) settings) from 7 geographical areas in England took part in the trial. All 3-4-year-old children attending preschool for 4 or more days (or half days) a week were considered eligible for the trial. In total, 1,586 preschool children across 70 preschool classrooms took part. LanguageScreen was used as a screening test. The settings were randomised to intervention or control groups, stratified by size (dichotomised by mean number of children attending), which resulted in 33 settings in the intervention group and 32 in the waitlist control group. Following screening, the 6 children in each class receiving whole-class enrichment and additional targeted support, plus 4 children randomly selected

[1] This is equivalent to selecting six children from a class of 25–30 children. In the evaluations, this has been implemented by selecting six children with the lowest oral language skills in a class or per setting – see impact evaluation design for further details.

from the remainder of each class (n = 726) received individually administered in-depth testing, in order to be able to evaluate the enrichment-only element of NELI Preschool, as well as the programme received by children identified with language weaknesses receiving additional targeted support.

In-depth testing at baseline and endline included subtests from the Child Evaluation of Language Fundamentals (CELF) Preschool II UK (Semel, Wiig and Secord, 2006) and the Renfrew Action Picture Test (RAPT) (Renfrew, 2019) and also involved assessing vocabulary knowledge taught in NELI Preschool intervention, narrative skills measured via the Renfrew Bus Story, self-regulation assessed using the HTKS-R assessment (Gonzales *et al.*, 2021) and children's behavioural adjustment to school from the Brief Early Skills and Support Index (BESSI; (Hughes *et al.*, 2015). . The preregistered primary outcome was a latent language variable defined by loadings from 8 subtests from LanguageScreen, CELF and RAPT tests.

The evaluation found that the children receiving the enrichment (whole class) component of the intervention made more progress in their language skills than control group children (Cohen's d=0.26 (0.101, 0.425), equivalent to 3 months additional progress in language development). Further analysis was undertaken for the children identified as requiring the additional targeted support (those with the lowest language scores at baseline), which showed a significant effect of the intervention on this subgroup (Cohen's d = 0.16 (0.039, 0.302), equivalent to 2 months additional progress in language development). The report highlighted this difference in the two effect sizes and recommended that the next evaluation should test two versions of the programme, one with and one without the targeted support element.

Overview of the integrated evaluation design

This effectiveness trial, funded by the EEF and DfE's SPHs, aims to build on the findings from the efficacy trial by testing whether NELI Preschool is effective at improving children's oral language skills when delivered on a larger scale across both maintained nursery settings and PVI settings. Since the efficacy trial included only a small number of PVI settings, this evaluation will ensure that at least one-third of the participating settings are PVI.

As a brief overview, we will employ a setting-randomised design. We will randomise 318 early years settings on a 1:1 basis into two groups, intervention and control, stratified by setting type (maintained nurseries or PVI settings) and setting size (one preschool class vs more than one preschool class) to ensure equal group allocation in each strata. Intervention settings will deliver the full version of NELI Preschool programme that includes both enrichment and targeted components. This is different from the efficacy trial recommendation due to several reasons. The suggestion comes from the difference in the effect sizes between the programme impact on children exposed solely to the enrichment component and those also receiving the additional targeted component. However, this comparison lacks the statistical power to draw definitive conclusion as it did not permit any statistical inferences from the differences in two effect sizes given the sample size, and considerable overlap in the confidence intervals which means these apparent differences may simply result from sampling error. Besides, proposing two programme versions raises ethical concerns regarding offering a version without the targeted component, which aims to make it more accessible to the lowest ability children. Given these considerations, the full version of the programme will be implemented in this

evaluation including both enrichment and targeted components. These components will also be evaluated separately, similar to the efficacy trial, as part of a secondary analysis.

All 3 to 4-year-old children (as of August 31, 2024) in the preschool classroom are eligible to take part in the trial. These children will be assessed by settings using the LanguageScreen prior to randomisation. LanguageScreen will be a baseline for the primary outcome, and it will also act as a screening test for the intervention settings to select children who will be targeted for additional support.

The primary outcome for this trial will be a latent language measure generated using endline LanguageScreen and RAPT. In contrast to the efficacy trial, this evaluation will not employ the CELF as CELF necessitates the involvement of speech and language therapists for one-on-one administration. Moreover, based on prior evaluation data, it indicates that a latent language factor derived from LanguageScreen and RAPT demonstrates a strong correlation with the latent factor that includes CELF. Considering this, the added workload does not justify incorporating CELF into this evaluation. This is elaborated further in the design section. The secondary outcomes for the trial will be the total raw score for each subtest of LanguageScreen and RAPT at endline. There are several additional impact analyses to ascertain the effectiveness of NELI Preschool at improving the language skills of subgroups of children. These include children in the targeted component, children who only receive the whole-class (enrichment) component, disadvantaged children (as identified by Early Years Pupil Premium or EYPP), children in the PVI settings only.

The Implementation and Process Evaluation (IPE) for this trial encompasses several research questions that explore the extent to which NELI Preschool was successfully scaled up to a greater number and variety of settings by examining fidelity of core component implementation and potential adaptations made by settings. In addition, the IPE will investigate how well the programme reaches and supports disadvantaged children, evaluates the targeted intervention's perceived effectiveness, explores contextual and moderating factors influencing implementation, and assesses the perceived impact on settings, practitioners, and children – including disadvantaged children specifically.

Key differences between the efficacy and the effectiveness trial are presented in Appendix A.

Intervention

The NELI Preschool programme was developed by the OxEd team while at the University of Oxford. The programme will be delivered by OxEd for the purpose of the effectiveness trial between January and June 2025. In addition to this, the programme will be implemented in a small number of settings between February and July 2024 to support the formative evaluation (see below).

A detailed description of the intervention in the context of the TIDieR checklist[2] is presented below.

[2] See online guide for more detail: https://tidierguide.org/.

Rationale, theory and/or goal of essential elements of the intervention

NELI Preschool is a scripted programme that aims to improve the receptive and expressive oral language skills of young children through screening and intervention delivered by trained early years practitioners. The programme was developed in response to evidence that language skills are a key determining factor for later attainment in both literacy (Hjetland *et al.*, 2020); (Hulme *et al.*, 2015) and numeracy (Chow and Ekholm, 2019); (Hornburg, Schmitt and Purpura, 2018). Moreover, children from disadvantaged backgrounds are more likely to start school with lower levels of language abilities (Guo and Harris, 2000); (Hart and Risley, 1995) (Sampson, Sharkey and Raudenbush, 2008); (Hutchinson, Reader and Akhal, 2020); (Sirin, 2005), establishing the disadvantage attainment gap that will on average widen over the course of their educational career (Hutchinson, Reader and Akhal, 2020).

There is likewise evidence to indicate that early language development is influenced by the quality and quantity of spoken language children are exposed to through interactions with their caregivers (Hoff, 2003); (Murray and Egan, 2014); (Weisleder and Fernald, 2013); (Huttenlocher *et al.*, 1991), and may mitigate the effects of disadvantage (Roulstone *et al.*, 2011); (Goodman and Gregg, 2010). Dialogic exchanges in the nursery setting may consequently provide children with opportunities for language development that they may not receive in their home environment, including in cases where English is an Additional Language (EAL) for their family. As a result, the programme seeks to provide enrichment opportunities for all children aged 3-4 years in the setting, as well as targeted intervention for children with lower abilities in an effort to close the language gap that has already emerged and would influence the child's life opportunities (Field, 2010).

Finally, the NELI Preschool programme aims to provide early years practitioners with high-quality continuous professional development (CPD) to provide more children with better support for their language development. High-quality CPD has also been found to increase satisfaction and retention in the education workforce (Fletcher-Wood and Zuccollo, 2020), which may help to address the significant recruitment and retention challenges being faced by the sector (Haux *et al.*, 2022). The developers identified the key outcomes for practitioners as increased pedagogical knowledge, confidence and motivation around supporting early language development. Self-determined types of motivation have been found to be associated with higher job satisfaction and lower levels of both job burnout and turnover intentions in the workplace (Fernet *et al.*, 2008). Confidence has also been argued to be an 'essential element' of teacher professional performance (Nolan and Molla, 2017).

Recipients of the intervention

Recipients of the intervention are 3-4-year-olds in maintained and PVI early years settings.

All children participate in the whole-class sessions. This includes younger children if they are present in the classroom and wish to join in.

In addition, three or six children per class, or six children across the setting, will be selected to attend additional small-group and one-to-one targeted sessions. Settings will select these

children as per the Targeted Group Selection Guidance provided by the delivery team.[3] These children are intended to be those with the lowest language abilities (as assessed using LanguageScreen – see below), excluding any children who are not developmentally ready to engage in small-group and/or one-to-one activities.[4] In addition, OxEd advise that only children attending the setting for 15 hours or more per week be selected for the targeted component.

Small group sessions are conducted with children selected for the targeted component in groups of three. Each child selected for the targeted component also participates in a weekly one-to-one session with a practitioner.

Physical or informational materials used in the intervention

Training

All participating settings are given access to an online training platform, as well as an online Delivery Support Hub. The online training course includes a range of texts, videos and quizzes, as well as interactive forums for comments and discussion. Both the training platform and the Delivery Support Hub provides participating practitioners with access to the delivery team, experienced practitioners and speech and language specialists who can provide support and answer any questions. The Hub also allows participants to share reflections in relation to the delivery process in their settings and tips for the wider community of NELI Preschool practitioners. Upon completing the training, participants receive a NELI Preschool practitioner certificate, accredited by the CPD Certification Service (CPDUK).

Intervention

Settings are provided with access to the online LanguageScreen App for completing the child screening assessments required for selecting the group for the targeted component of the programme. The setting will require a tablet to carry out the screening process. Tablets will be provided by OxEd for the purpose of the trial where settings do not already have one available. This will be sent to all settings prior to randomisation as LanguageScreen is administered at baseline. Settings will be able to keep this tablet for continued provision after the trial is completed.

Settings are provided with material for 20 storybooks, one for each week of the intervention. The material includes both a physical version of each book and digital slides that display the books with related questions and whole-class activities. Settings also receive a handbook with the session scripts and prompts and physical activity materials such as vocabulary flashcards. Targeted sessions use weekly session cards, story sequencing cards and press-out characters to facilitate small-world play. In addition, physical materials are provided to support good behaviour during the sessions, including a plush elephant puppet (and press-out version), best listener stickers and a listening rules poster. Additional digital materials include song files and other materials for continuous provision opportunities. Finally, practitioners are provided with guidance on selecting children for the targeted component (Targeted Group Selection Guidance), as well as the Narrative Progress Tracker template (see below for more

[3] Implementation models are described in detail in the impact evaluation design section.
[4] This may be as a result of identified or unidentified Special Educational Needs (SEN).

detail) with a list of Receptive and Expressive Language Skills Checkpoints for ages 18-24, 24-36, 36-48 and 48-60 months for use with the children receiving the targeted intervention.

Settings are provided with a parent 'home learning link' template that can be sent out to parents for each week of the intervention, to inform them of what their child is doing and to promote parental engagement in the language development process.

Settings require a TV screen or monitor to deliver the digital material for the whole-class sessions. OxEd will provide intervention settings with a screen, a wireless keyboard and a trolley for the purpose of the trial where the setting does not already have appropriate equipment available. Settings will be able to keep all equipment for continued provision after the trial is completed.

Procedures, activities and/or processes used in the intervention

Training

All practitioners delivering the intervention must have completed the NELI Preschool online training course. This course is fully asynchronous and takes 10-12 hours to complete. The training includes practical exercises and participants are encouraged to participate in the forums in both the training course and Delivery Support Hub to create online communities of practice. The first part of the training focuses on why language skills are important, developing understanding of receptive and expressive language skills, how to recognise difficulties with these, and also how to support their development. The second part of the training looks at the practical considerations and preparation necessary for effectively delivering the intervention. This includes videos of each session type being delivered according to best practice, accompanied by expert commentary. Each section of the training also has a quiz to enable participants to monitor their progress. There is a short assessment at the end of each of the two parts of the training course. Trainees must pass the assessments, in order to be awarded their certificate at the end of training. Participants are able to revisit the training material and access the Delivery Support Hub at any time.

An additional four webinars are delivered at two-monthly intervals to support practitioners over the course of the academic year. The webinars cover:

- An introduction to the programme, LanguageScreen and selecting children for Targeted Intervention
- Expanding learning from the programme beyond the sessions
- Setting targets for Week 10
- Dialogical reading, storytelling and making the most of the NELI Preschool books

While a minimum of two people per setting must complete the training and at least one must attend or watch the webinars, settings are encouraged to support as many practitioners as possible to do so, in order to create a 'whole team' approach to delivering the intervention. This approach also supports greater flexibility around covering for absent staff members as needed. Additional staff members can complete the training partway through the delivery period if there is any turnover of previously trained staff – although they must complete the training before they deliver any sessions.

There are no minimum requirements for engaging with the Delivery Support Hub.

Intervention

The intervention sessions focus on reading and dialogical questioning to support understanding of each selected book. Each book introduces four or five 'special words', which constitute the focus of a range of vocabulary learning activities over the course of the week.

The intervention involves both whole-class sessions and small-group and individual sessions for the six children selected for the additional targeted component. The setting must complete the child screening process with LanguageScreen assessment to identify the targeted group before starting delivery. LanguageScreen assessment is an adaptive assessment and must be administered one-to-one by a practitioner. The practitioner does *not* have to complete the training before starting the screening process. LanguageScreen is also intended to inform practitioners of each child's ability level so that they can better tailor their support to individual need.

In the whole-class sessions, each week begins with a dialogical reading session. The practitioner displays the digital version of the book on the screen, reading it to the children, asking them questions about it. Subsequent sessions during each week focus on special words from the books and engage children in activities to support vocabulary learning, including using flashcards and a wide range of digital resources. While a script for each session is provided, the training encourages practitioners to develop their own questions and prompts in accordance with the programme guidance around effective dialogic reading. There is, however, a prescribed structure for each session. Each week, the first session focuses on reading the book, while the following sessions focus on each of the new special words. The sessions include regular consolidation of previously learned words.

Targeted sessions focus on supporting children to develop their narrative skills by retelling aspects of the books engaged with in the whole-class sessions. In the first small-group session each week, each child takes on a role, and the practitioner guides the children to retell the story through role-play, following the workbook outline and using provided pictures, toys and character cards. The one-to-one sessions follow the same structure, but the child is asked to retell the story themselves. The remainder of the small-group sessions focus on vocabulary, with the practitioner using flashcards to revise the meaning of each of that week's 'special words' followed by a related activity.

At three points over the 20-week programme, practitioners use the Narrative Progress Tracker to assess how each of the children in the targeted group is progressing in relation to appropriate language development targets. In Week 1, Week 10 and Week 20, the practitioner transcribes how each child in the targeted group retells that week's story in the one-to-one session. They record the number of words and make any notes they deem relevant. The practitioner is then asked to highlight and date evidence of the child having achieved any of the listed Receptive and Expressive Language Skills Checkpoints, either through the transcribed story or in continuous provision. The practitioner will then discuss with their SENCo class teacher or Room Leader their reflections on the progress observed and how best to proceed with the tailored support.

Intervention providers / implementers

Training

The online training course was developed by the research team and the University of Oxford (now at OxEd). The training course forums and Delivery Support Hub are moderated by the OxEd team and speech and language specialists. Supporting webinars are hosted by the OxEd team alongside occasional experts and special guests to provide additional guidance and support.

Intervention

It is recommended that the Nursery Class Teacher or Room Leader delivers the whole-class sessions, and that an early years practitioner, teaching assistant or nursery nurse delivers the targeted sessions. This recommended approach stems from the aim of fostering teamwork and a sharing of the additional workload across different staff in the setting. However, it is up to the individual setting to decide how the session delivery will be staffed, with the only condition being that any practitioner delivering sessions must have completed the full online training course.

For the purpose of the trial, settings are asked to nominate a NELI Preschool Lead responsible for overseeing programme implementation at setting level and liaising with the evaluation team.

Mode of delivery

Training

The training is delivered online in a fully asynchronous and interactive format. The Delivery Support Hub is also online.

Intervention

The intervention is delivered in-person. Whole-class sessions are delivered to the whole class present at the early years setting at that time. The targeted sessions involve small group sessions that are delivered to a maximum of three children at a time, as well as one-to-one sessions.

Location of the intervention

The intervention is delivered in maintained and PVI early years settings. It is recommended that the targeted sessions are delivered in a quieter area of the classroom, such as a hallway, book corner or coat area.

Duration and dosage of the intervention

The intervention lasts for 20 weeks. Five whole-class sessions are delivered each week, along with three small-class sessions and one one-to-one session (per targeted child). It is recommended that whole-class sessions are delivered daily, that small-group sessions take place following the first, third and fifth whole-class sessions, and the one-to-one sessions

following the fourth whole-class session. It is at the discretion of each setting to determine their delivery schedule.

Whole-class sessions are intended to last 15-20 minutes, small group sessions 10-15 minutes and one-to-one sessions 5-10 minutes. Practitioners are encouraged to focus on the quality of interaction that takes place during the session, rather than imposing a specific duration. For example, practitioners are allowed and encouraged to strategically expand on child-led interactions relevant to the story or subject at hand.

Adaptation of the intervention

Settings can tailor delivery schedules and staffing to their own needs. In addition, sessions are scripted with flexibility for practitioners to adapt them to the cohort's ability level. In whole-class sessions, 'step-up' words and questions are provided as opportunities to further extend the vocabulary of more advanced children. 'Step-down' words and questions are likewise provided to provide children with less advanced language abilities with support and opportunities to engage. In addition, a large number of dialogical questions are supplied for each book, with the expectation that the practitioners choose their own selection. The training includes specific additional guidance around how to support children with EAL.

Practitioners are expected to follow the overall structure of the programme, to regularly complete the activities accompanying each session and to use dialogical reading strategies throughout.

Strategies to maximise effective implementation

The online and asynchronous nature of the training course means it can be completed by practitioners as and when it is convenient for them prior to starting delivery, and as many practitioners as possible are encouraged to engage with it. Practitioners can also return to the material for a refresher at any point as required. The official CPD accreditation for completing the course is intended to further motivate practitioners to engage.

The online Delivery Support Hub provides tailored expert support to the individual needs of each practitioner throughout delivery.

Participating settings also receive regular newsletters to introduce them to the next stage of the programme and remind them that they can receive support at any time via email and the Delivery Support Hub.

The programme is highly structured and scripted but with flexibility and adaptation encouraged in practice. This provides practitioners with a clear reference point when needed while allowing greater freedom where they feel confident in applying the programme principles.

The digital intervention materials are designed to be high quality and visually engaging for children. Behaviour management and listing promotion tools, such as the NELI puppet, are also embedded in the programme to encourage child engagement.

Evidence of implementation variability

The process evaluation carried out as part of the efficacy trial reported that the programme had been generally well adhered to. However, the efficacy trial sample consisted almost entirely of maintained settings, so there is no evidence to date about how effectively the programme may be implemented in PVI settings, where there are very different structures, resources and staffing available.

Theory of Change

The Theory of Change (ToC) for the NELI Preschool programme is shown in **Figure 1**. It outlines the target population of the intervention and the activities, outputs, short-term and intermediate outcomes that are intended to lead to the ultimate outcome of improving children's receptive and expressive language skills, thus providing them with a better foundation for learning in the long-term. NELI Preschool involves a training programme for early years practitioners in addition to the actual sessions that are delivered to the children, so an additional row of outputs and outcomes has been included to map how practitioners' knowledge and teaching skills may develop in parallel with the children's own language skills.

The programme consists of a whole-class enrichment component *and* a targeted intervention. These are separated out in different coloured boxes to represent the different pathways children in the targeted and enrichment sessions are anticipated to follow. The purpose of the targeted intervention is to support children with the lowest language abilities to access and benefit from the enrichment sessions alongside their peers. This mechanism relies on the following key assumptions:

- The children that settings are instructed to select for the targeted intervention, according to guidance provided by OxEd, are those who will benefit most from this component – including in terms of being able to access the level of focus and behaviour required for the intervention.
- Practitioners are able to deliver the number of sessions required and with the right children in the context of what is often quite an unpredictable nursery environment.
- Practitioners are sufficiently skilled to adapt their teaching approaches to the needs of different levels of language abilities.
- The online training course is adequate to support effective delivery of the programme, as this depends substantially on skilled teaching not just adherence to a script.
- All children are given the opportunity to actively participate in the sessions, not just those children who are already quite vocal.

While a script is provided for each session, the NELI Preschool training encourages practitioners to be proactive in tailoring discussion to their specific context and participating children. It also relies on practitioners employing a range of teaching techniques, such as scaffolding and modelling, beyond the script provided. Therefore, in order for the sessions to improve children's language skills:

- Children must engage with the sessions and remain engaged for their duration, including practising their expressive language skills through speaking.

- The skills children gain through the sessions must be transferable to their broader interactions with adults, other children and the world around them.

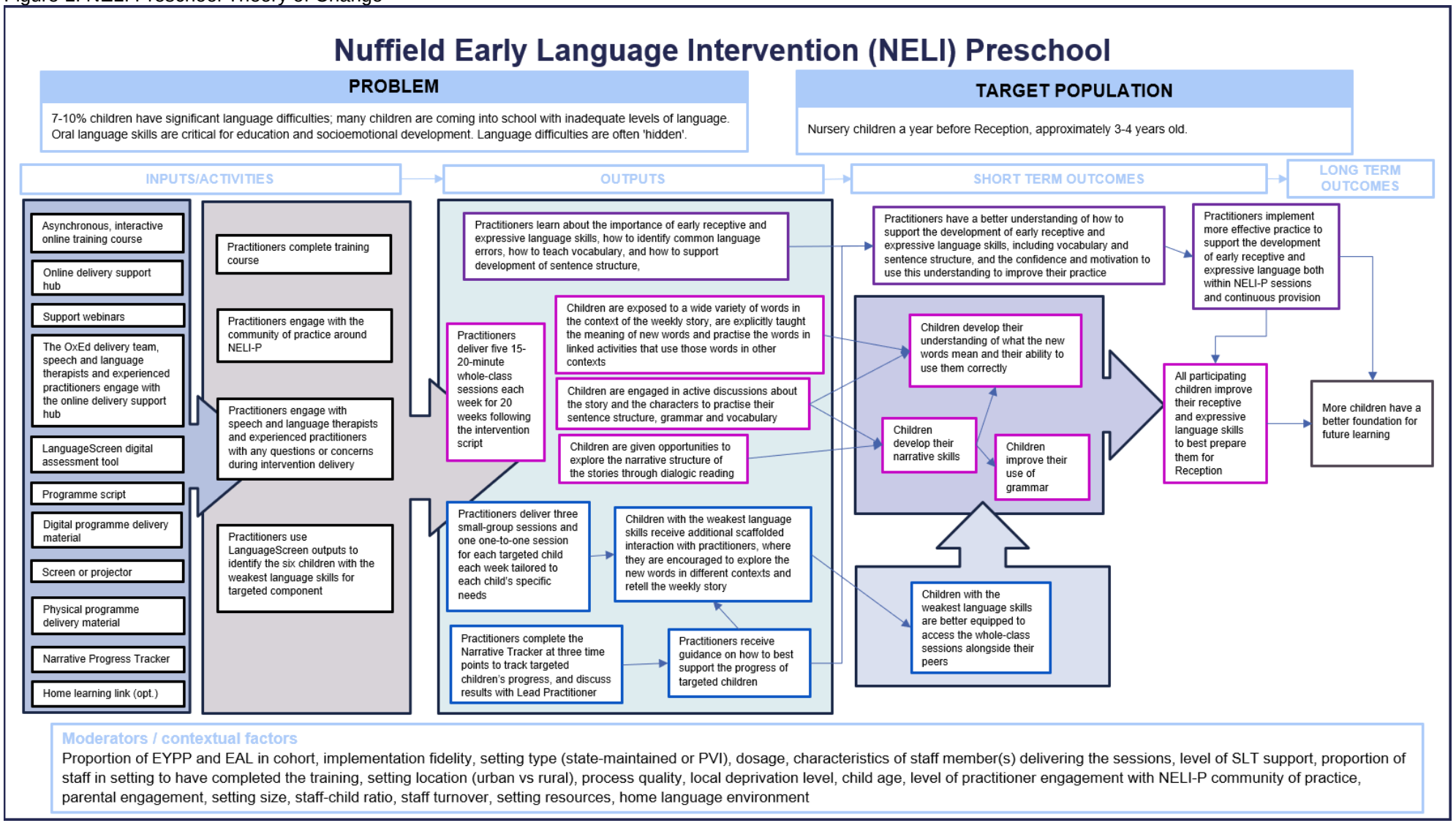
The NELI Preschool efficacy trial (West et al., 2024) provides evidence to suggest that the above holds true for the trial sample, which were overwhelmingly maintained settings. The 2024 formative evaluation of the programme (see below) will investigate whether they likewise hold in PVI settings. The effectiveness trial will further explore these elements through the surveys, interviews and observations carried out for the IPE (see below for more detail).

Practitioner outcomes in relation to increased understanding of supporting early language development relies on training and support provided as part of the programme to be sufficiently comprehensive and of high quality to engage the practitioners and enable this learning to occur. In addition, the potential for increased practitioner motivation and confidence rests primarily on the assumption that increased understanding breeds confidence and motivation – although positive feedback through interactions via the Delivery Support Hub and webinars may also have a direct influence on these factors.

Longer-term outcomes for both practitioners and children were not considered as part of the efficacy trial. Therefore, the effectiveness trial will be crucial for determining whether the following assumptions are valid – particularly through its longitudinal component:

- Increased practitioner understanding, confidence and motivation in relation to supporting early language development leads to improved practitioner practice (that is, there are no counteracting forces or constraints at play).
- Improvements in children's receptive and expressive language skills supports children to better engage with teaching and learning opportunities on a long-term basis, resulting in better academic outcomes.

Figure 1: NELI Preschool Theory of Change



Impact evaluation design

NELI Preschool implementation models

Before outlining the research questions, it is helpful to define what a preschool class is, the NELI Preschool implementation models and a number of different sample populations that will be of interest.

In the context of this evaluation, preschool class is defined as a group of children who are taught by the same teacher or room leader, or a group of children who are attending the setting the same time of the day. For example, children attending morning only sessions (class 1) and children attending afternoon only sessions (class 2).

All settings will be asked to choose their preferred NELI Preschool implementation model prior to randomisation[5]. These models, described below, are in line with the Targeted Group Selection Guidance that the intervention settings will receive.

A. **One targeted intervention group with six children** – This is the most common format where settings are single form entry and have only one preschool class. The setting will run one whole class session and will select six children for targeted intervention.

   This model will also be selected by settings with more than one preschool class, and where the setting will run a separate whole class session for each class but each class does not have at least 12 children who attend the setting for 15 hours or more (see model B for more information on this additional criterion for running more than one targeted intervention group). In this case, settings will run a separate whole-class session for each class, but select six children for targeted intervention across the setting, rather than per class.

B. **Multiple targeted intervention groups with six children per group** – This model works best in larger, multiple form entry settings where NELI Preschool can be run separately for each class. This means the setting runs the whole programme more than once, including separate targeted intervention groups, one for each class. Six children per class will be selected for targeted intervention. The additional criterion for running more than one targeted intervention group is to have at least 12 children who attend the setting for 15 hours or more in each class, as per the eligibility criteria for settings to participate in the trial (see *Participant selection* section below). In cases where there are not at least 12 such children, the setting will select six children across the setting for targeted intervention group (Model A above).

C. **One targeted intervention group with six or more children but split across multiple classes** – This model is ideal for settings that run more than one whole group sessions (one for each class), but they find the targeted intervention in Model B too intensive and unrealistic. This means settings would ideally select three children from each class to be

[5] Settings will be asked to consider all children who will be 3-4-year-old by 31st August, do not have complex special educational needs and are registered to attend the setting by the end of September 2024.

included in targeted intervention irrespective to the number of children in each class who attend 15 or more hours.

This means, the evaluation team will have prior knowledge of each setting's intention regarding the selection of children for targeted intervention. Specifically, they will know whether the setting plans to select three or six children per class or if they intend to select six children per setting for the targeted intervention. This information will be available for all settings prior to randomisation.

Analysis samples

- **S1 Ideally Targeted Children**: NFER's sample of children in a setting who should have been selected for the targeted component of NELI Preschool if the settings followed OxEd's Targeted Group Selection Guidance correctly (as per the setting's preferred implementation model A, B or C, as described above). Usually, these will be the six children within a class who have the lowest baseline language score, attend at least 15 hours per week and are not flagged as being unable to participate in small group work (i.e., due to behaviour, or complex special educational needs). The selection of children in this sample will be based on the preference of implementation model that settings provide prior to randomisation.
- **S2 Practitioner Targeted Children (intervention settings only)**: Practitioner sample of children (actually selected by the setting to receive the targeted intervention. Settings will identify these children and share this information with NFER. In the majority of settings, we expect the S2 sample to be the same as S1 if settings followed OxEd's Targeted Group Selection Guidance correctly. It will only be different from S1 if settings/practitioners applied their own discretion to select children for the targeted component.
- **S3 Enrichment-only Sample**: NFER's sample of six randomly chosen children (per class in a setting), who attend the setting at least 15 hours per week and are not part of S1 or S2.
- **S4 EYPP**: All children in the setting between 3 and 4 who are entitled to Early Years Pupil Premium (EYPP). This may include children who attend the settings for less than 15 hours a week and are part of the other samples.

Research questions

Our primary research question for this impact evaluation will be:

RQ1: How effective is NELI Preschool at improving oral language skills of 3-4 year-old children in intervention settings compared to children in control group settings?

To answer the primary research question, we will use all children in S1 and S3 in intervention and control settings.

Secondary research questions for this trial will ascertain effectiveness of NELI Preschool on the six components of the primary outcome measure. These research questions will utilise the six secondary outcomes mentioned earlier.

RQ2: How effective is NELI Preschool at improving different aspects of children's oral language skills as measured by the subtests of LanguageScreen and RAPT?

These RQs will in turn be answered for: LanguageScreen subtests of Expressive Vocabulary, Listening Comprehension, Receptive Vocabulary and Sentence Repetition, and RAPT subtests of Information and Grammar. The sample for RQ2 will be identical to the sample used for RQ1.

Additional research questions for this trial will ascertain effectiveness of NELI Preschool for subgroups of children.

As explained in the previous section on samples, S1 and S2 samples may differ in intervention settings if settings do not follow OxEd's Targeted Group Selection Guidance fully. Therefore, we propose two separate research questions to determine the programme's effectiveness on children who should have been selected to receive the targeted component of the intervention and those who were actually selected to receive the targeted component of the intervention.

RQ3a: How effective is NELI Preschool at improving the language skills of the subgroup of six children selected to receive the targeted component of the intervention (S2)?

To answer this research question, we will use all children in S2 in intervention settings and S1 in control settings. S1 in control settings is necessary as a comparison here because control settings will not have access to the intervention including LanguageScreen data which is an integral part of the intervention. Consequently, they will not be able to identify children who would have received the targeted component, hence won't have an S2.

RQ3b: How effective is NELI Preschool at improving the language skills of the subgroup of six children who should have been selected to receive the targeted component of the intervention (S1)?

RQ4: How effective is NELI Preschool at improving the language skills of children who only receive the whole-class (enrichment) component of the intervention (S3)?

RQ5: How effective is NELI Preschool at improving the language skills of disadvantaged children (as identified by EYPP; S4)?

RQ6: Is NELI Preschool effective at improving language skills of 3-4-year-old children in PVI intervention settings compared to children in PVI control settings? This research question will utilise the same group of children as RQ1 (S1 and S3) for PVI settings only.

Design

Table 1: Trial design

Trial design, including number of arms		Two-arm, cluster randomised
Unit of randomisation		Setting
Stratification variables (if applicable)		Setting type (maintained and PVI) and setting size (settings with one preschool class vs those with more than one preschool class)
Primary outcome	**Variable**	Oral language skills

	Measure (instrument, scale, source)	Latent oral language variable formed from expressive and receptive subtests in LanguageScreen and the RAPT
Secondary outcome(s)	**Variable(s)**	1. Vocabulary knowledge (LanguageScreen Expressive Vocabulary or EV) 2. Literal and inferential language comprehension and expressive language skills (LanguageScreen Listening Comprehension or LC) 3. Vocabulary understanding (LanguageScreen Receptive Vocabulary or RV) 4. Language comprehension and production (LanguageScreen Sentence Repetition or SR) 5. RAPT information (RAI) 6. RAPT grammar (RAG)
	Measure(s) (instrument, scale, source)	LanguageScreen and RAPT
Baseline for primary outcome	**Variable**	Oral language skills
	Measure (instrument, scale, source)	Latent oral language variable formed from expressive and receptive subtests in LanguageScreen
Baseline for secondary outcome	**Variable**	1. Vocabulary knowledge (LanguageScreen Expressive Vocabulary or EV) 2. Literal and inferential language comprehension and expressive language skills (LanguageScreen Listening Comprehension or LC) 3. Vocabulary understanding (LanguageScreen Receptive Vocabulary or RV) 4. Language comprehension and production (LanguageScreen Sentence Repetition or SR)
	Measure (instrument, scale, source)	LanguageScreen (individual subtest scores for outcomes 1 to 4 and overall for outcomes 5 & 6)

This trial will be a two-arm, cluster-randomised effectiveness trial with randomisation at setting level. Practitioners from both groups of settings will administer baseline LanguageScreen assessment with all eligible children (see *Participant selection* section for further details) prior to randomisation. Settings will then be randomly allocated to two groups are: intervention and control group. After randomisation, practitioners from intervention settings will receive NELI Preschool training and will deliver the intervention to children for 20 weeks. NFER will share baseline LanguageScreen assessment results with intervention settings as LanguageScreen is an integral part of the programme and the results will aid in selection of children for the targeted component. The data will include raw and standard scores for each component as well as overall standard score and percentile. Intervention settings will use this data along with OxEd's Targeted Group Selection Guidance to select children for the targeted intervention. Control settings will continue their usual practice and will not have access to LanguageScreen assessment results or NELI Preschool training or materials during the trial.

Settings will receive incentive payments as follows: Intervention settings will receive £100 following all baseline data collection, £150 on completion of NELI Preschool session delivery logs, £150 following endline assessments, paid staff time to complete NELI Preschool training

(£120 per trainee for up to 5 trainees per setting) and paid staff time to administer LanguageScreen at baseline. In addition to this, 12 case study intervention settings will receive an additional £100 for participating in NELI Preschool session observations and the two practitioners per setting who give up their time to be interviewed will each receive £50. Control settings will receive £100 following all baseline data collection, £150 following endline assessments, and paid staff time to administer LanguageScreen at baseline. Control settings will also receive £1000 at the end of the trial which can be used to purchase the NELI Preschool programme.

Both groups will receive assessment feedback reports at the end of the trial. This will include child-level assessment feedback report on LanguageScreen and RAPT.

Choice of outcome measures

Previous evaluations of NELI Preschool used a primary outcome measure that combined the CELF along with RAPT and LanguageScreen. Analysis from the NELI effectiveness trial (Dimova *et al.*, 2020) indicated that there was a strong relationship between the original latent language variable generated by CELF and RAPT, and an alternative latent language factor generated using LanguageScreen and RAPT. These two latent oral language factors were highly correlated at endline (r = 0.8). Had the original NELI effectiveness trial used LanguageScreen and RAPT, without measuring CELF, it would have reached the same conclusion, with the effect size within three-hundredths of the original effect size (Groom, Brown and Lymperis, 2023). This trial therefore uses LanguageScreen at baseline, and a latent measure formed from LanguageScreen and RAPT at endline. Please see the outcomes section for further details.

Participant selection

Eligibility Criteria for Settings

Settings have to fulfil the following requirements to participate in this trial:

- Settings must either be state-maintained or PVI nurseries. Settings registered as 'childcare on domestic premises' will also be eligible to take part in the trial, as these settings have more than four people who provide care and the settings are subject to the same requirements as nurseries for staffing ratios and staff qualification levels.
- Settings should have a minimum of fourteen 3 to 4-year-old children attending 15 or more hours per week during the 2023-2024 academic year (starting September 2023).This criterion is a proxy for the minimum number of children required for the trial in 2024-25, i.e., minimum of 12 children per setting[6].
- Settings must be willing to allocate time for a minimum of two nursery staff members to complete online NELI Preschool training, encouraging as many practitioners who work with 3 to 4-year-olds as possible to complete the online training.
- Settings must be located within the following Local Authorities from nine SPHs: Bexley, Bournemouth, Christchurch and Poole, Bracknell Forest, Brighton and Hove, Bromley,

[6] A minimum of 12 children per setting are required for trial analysis such that analysis samples include six children from the enrichment component and six from the targeted component of NELI Preschool.

Buckinghamshire, Cambridgeshire, Cornwall, Croydon, Darlington, Derby, Derbyshire, Devon, Dorset, Durham, Ealing, East Riding of Yorkshire, East Sussex, Gateshead, Greenwich, Hammersmith and Fulham, Hampshire, Hartlepool, Hillingdon, Hounslow, Isle of Wight, Kent, Kingston upon Hull, Kingston upon Thames, Lambeth, Lewisham, Medway, Merton, Middlesbrough, Milton Keynes, Newcastle upon Tyne, Norfolk, North East Lincolnshire, North Lincolnshire, North Tyneside, North Yorkshire, Northumberland, Nottingham, Nottinghamshire, Oxfordshire, Plymouth, Portsmouth, Reading, Redcar and Cleveland, Richmond upon Thames, Slough, South Tyneside, Southampton, Southwark, Stockton-on-Tees, Suffolk, Sunderland, Surrey, Sutton, Torbay, West Berkshire, West Sussex, Wandsworth, Windsor and Maidenhead, Wokingham, and York.

- Settings are not eligible if they are participating in another SPH funded programme delivered in the 2024-2025 academic year including allocation to the control group for one of the other SPH trials (e.g., Early Talk Boost, The ONE Programme, EYCP, Concept Cat, Communication Friendly Settings).
- Settings cannot take part if they are participating in another EEF-funded programme or early years evaluation in the in 2024-25 academic year. Although this is identified as an eligibility criterion, there is some flexibility for the OxEd team (in agreement with NFER and the EEF) to recruit settings to the trial even if they are part of an EEF evaluation. This is especially true where the programme in question is not very similar to NELI Preschool in terms of its recipients and/or aims, or it is viewed as a business as usual (BAU).
- Settings from the efficacy trial or those taking part in the formative evaluation cannot take part in this trial.

Settings will only be randomised if they fulfil baseline data collection requirements. These include: providing required data for the eligible children (see below) prior to baseline LanguageScreen assessment, completing the baseline LanguageScreen assessment with eligible children, uploading this data on LanguageScreen portal and completing the baseline IPE survey.

Settings that are part of a nursery chain or a nursery group are eligible to take part as an individual unit. Given each setting will be considered a separate unit for randomisation, these settings will need to agree to adhere to random group allocation. OxEd team will confirm this with each setting during the recruitment.

If settings are taking part in DfE's Early Years Professional Development Programme (EYPDP), they will still be eligible to take part in this trial as EYPDP is considered as BAU.

Eligibility Criteria for Children

All 3–4-year-old children on 31st August 2024 and those who can engage with LanguageScreen assessment (e.g., do not have complex special educational needs[7]) will be assessed with LanguageScreen at baseline and will form part of the trial. This will also include

[7] Settings are given clear guidance about how to identify children with complex special educational needs. 'Complex special educational needs' in this guidance will be described as: children who have such severe auditory, sensory, visual or behavioural difficulties that they would be unable to access a five-minute LanguageScreen assessment. Settings are strongly encouraged to refer to their SENCo and speech and language team for further guidance on suitability.

children who attend the setting for less than 15 hours a week, as some of these children may subsequently be discovered to be entitled to EYPP. Only the children with baseline LanguageScreen will be included in the trial. Settings will be asked not to assess children who were not between 3 and 4-years old on 31st August 2024 or have complex special educational needs even though they may be present in the preschool classroom. These children can join whole class sessions (if their setting is randomised to intervention group) but will not form part of the trial.

In accordance with the Targeted Group Selection Guidance and their preferred implementation model, settings will select children (usually six per class) to receive the targeted component of the programme (sample S2 above). They will use the results of the baseline LanguageScreen assessment and the Targeted Group Selection Guidance provided by the delivery team. We will mimic this selection process to select the sample of children (S1 above) who would ideally be selected to receive the targeted component of the programme. Children who attend less than 15 hours per week or are flagged as being unable to participate in small group work (i.e., due to behaviour , or complex special educational needs) will not be eligible for selection as per the Targeted Group Selection Guidance.

To select children for the enrichment-only component (S3 above), we will remove children who are in S1 or S2 samples as well as those who do not attend the setting for at least 15 hours a week. Then, we will run a simple randomisation where six of the remaining children per class in a setting will be randomly selected for follow up testing.

S4 will include all children in the setting between 3 and 4 who are entitled to EYPP and have completed baseline LanguageScreen. This may include children who attend the settings for less than 15 hours a week and are part of the other samples.

Recruitment

The delivery team at OxEd will manage the coordination of the recruitment process. Three hundred and eighteen early years settings across nine SPHs in England are required for this trial. Each SPH will have a distinct target (see Table 2 below) totalling slightly higher to 320 settings in total. Overrecruiting in certain hubs may be necessary to meet the final target. Additionally, over recruitment beyond the target is also possible to address potential high attrition rates observed in previous early years trials.

Table 2: Target number of settings for each Stronger Practice Hub (SPH)

SPH	Local Authorities	Target
Northern Lights Early Years Stronger Practice Hub North East	Darlington, Gateshead, Hartlepool, Middlesbrough, North Tyneside, Redcar and Cleveland, South Tyneside, Stockton-on-Tees, Sunderland	32
The Great North EYSPH	Durham, Newcastle upon Tyne, Northumberland	32
Yorkshire & Humber Together EY SPH	East Riding of Yorkshire, Kingston upon Hull, North East Lincolnshire, North Lincolnshire, North Yorkshire, York	38
Derbyshire and Nottinghamshire EYSPH	Derby, Derbyshire, Nottingham, Nottinghamshire	38
East of England Early Years Stronger Practice Hub	Cambridgeshire, Norfolk, Suffolk	38

Thames Valley Early Years Stronger Practice Hub	Bracknell Forest, Buckinghamshire, Hampshire, Milton Keynes, Oxfordshire, Portsmouth, Reading, Slough, Southampton, West Berkshire, Windsor and Maidenhead, Wokingham	32
London South EYSEP	Bexley, Bromley, Croydon, Ealing, Greenwich, Hammersmith and Fulham, Hillingdon, Hounslow, Kingston upon Thames, Lambeth, Lewisham, Merton, Richmond upon Thames, Southwark, Sutton, Wandsworth	38
EYE South West, Early Years Stronger Practice Hub	Bournemouth, Christchurch and Poole, Cornwall, Devon, Dorset, Plymouth, Torbay	34
Kent Early Years Stronger Practice Hub	Brighton and Hove, East Sussex, Isle of Wight, Kent, Medway, Surrey, West Sussex	38

The recruitment activities commenced from November 2023. Initial data collection, including eligibility criteria, were facilitated through an official Expression of Interest (EoI) form accessible on both the OxEd and EEF websites prior to the commencement of recruitment activities. Key tasks involve finalising webpage setup, outreach to interested settings, and collaboration with Local Authorities and other contacts in the sector. These efforts are systematically tracked by OxEd. Additionally, OxEd will also leverage their current network and connections to aid in recruitment by providing materials and organising dedicated recruitment events and webinars. Social media promotion targeting Early Years groups, nurseries, and schools will also be part of their strategy, managed in line with current practices on OxEd's social media platforms.

Once settings sign the EoI, OxEd will send a Memorandum of Understanding (MoU) to these settings. After this, OxEd will call the setting to discuss the preschool intake, attendance patterns of the children and class structure/s for each setting. In addition to this, they will also find out whether settings will require a tablet to administer LanguageScreen at baseline and whether they have a TV screen or monitor (which is required to deliver NELI Preschool if randomised to the intervention group). The headteacher or the nursery manager will sign the MOU and they will nominate a key contact person who will act as their NELI Preschool lead for the trial.

OxEd will monitor recruitment numbers to ensure that the trial has at least one-third maintained nurseries and one-third PVI settings. Although the remaining one-third could be any setting type, the aim is to have an equal proportion of each setting type in the trial. In addition to this, OxEd will aim to achieve the following targets for each SPH.

Once the settings are signed up, OxEd will share these details with NFER. As part of the baseline data collection, NFER will contact settings to distribute parent letters/parent information sheets, collect children data and liaise with the settings to administer LanguageScreen baseline assessment. Once the baseline assessment data is uploaded on LanguageScreen portal, NFER will randomise the settings to intervention or control group. For further details on randomisation, please see the *Randomisation* section below.

Outcome measures

Baseline measures

Baseline for Primary outcome

LanguageScreen[8] will form the baseline measure for this trial. Other assessments were discussed during the set-up but the pre-post correlation gain of additional assessments with the outcome/s did not justify the time and resource burden to use these as baseline.

LanguageScreen is a language assessment tool accessed via an App on a tablet. It is designed to evaluate oral language skills in young children. It is administered one-to-one with children and takes approximately five to ten minutes to complete. It was standardised using data collected from 8,273 schools and 348,944 children aged 42 to 107 months (equivalent to 3 years and 6 months to 8 years and 11 months) between 2018 and 2020 (Hulme *et al.*, 2024). It demonstrated strong reliability for the total score with a Cronbach's alpha of 0.92 and McDonald's omega hierarchical coefficient of 0.75. The individual subscales also showed reliability ranging from 0.74 to 0.80. The test–retest reliability of LanguageScreen was 0.78 (N = 9,778).

It is comprised of the following four subtests:

i. Expressive Vocabulary (24 items) assesses ability to name pictures; this provides a measure of vocabulary knowledge.
ii. Listening comprehension (16 items) assesses the ability to understand spoken stories by asking questions about three short stories being played to the child; this provides a measure of both literal and inferential language comprehension and expressive language skills.
iii. Receptive vocabulary (23 items) assesses the ability to match spoken words to pictures by asking them to match a word they hear to one of four pictures on the screen; this provides a measure of vocabulary understanding.
iv. Sentence Repetition (14 items) assesses the ability to repeat sentences verbatim by asking children to repeat the sentences they hear; this is a measure of language comprehension and production.

Children are scored correct or incorrect for each item with automated discontinuation rules. Responses are scored by the App and settings can access and download reports containing results for each child who has been assessed via a secure website. Raw and standard scores for each subtest are provided along with LanguageScreen total standard score. Standard scores are age-standardised across a sample of 348,944 children (mentioned above) and express a child's performance relative to their age.

To create the baseline measure for the trial, we will undertake Principal Component Analysis (PCA) to aggregate all the subscores and will use the first component as the measure. The main advantage of using the first component of PCA is that it will weight the individual components so as to maximise the variance captured by the measure. More details will be provided in the Statistical Analysis Plan.

Baseline for Secondary outcomes

We will use the LanguageScreen assessment to measure baseline for secondary outcomes. For the four LanguageScreen secondary outcomes (Expressive Vocabulary, Listening

[8] https://oxedandassessment.com/uk/languagescreen/

Comprehension, Receptive Vocabulary and Sentence Repetition), we will use the corresponding score from each subtest of baseline LanguageScreen. For RAPT Information and Grammar outcomes, we will use the same baseline as the primary outcome (first component of PCA, see above).

At baseline, LanguageScreen will be administered by practitioners prior to randomisation. This approach is taken to mimic usual practice outside the trial, where practitioners assess children with LanguageScreen to identify suitable subset for the targeted intervention. OxEd will share the administration support and guidance with all settings to ensure they administer the assessment for all eligible children who can access LanguageScreen. NFER (supported by OxEd) will monitor completion of baseline assessments and will receive LanguageScreen data from OxEd.

Primary outcome

The primary outcome for this trial will be oral language skills, specifically expressive and receptive language skills. It will be measured using a latent variable constructed from PCA of LanguageScreen and RAPT[9] assessments collected at follow-up. The first component from this analysis will be used as the measure.

Similar to LanguageScreen, RAPT is a one-to-one administered test. It assesses the speech and language development of children who are between 3 and 8 ½ years of age by using 10 picture cards, depicting a range of everyday scenarios, that stimulate children to give samples of spoken language that can be evaluated in terms of grammatical structures, sentence length and identifying information.

RAPT was re-standardised in 2018 where the assessment was conducted digitally with the images presented to a child on-screen. All images and scoring guidelines from digital RAPT are identical to those used in print version and therefore the re-standardised data can be used consistently across both print and digital versions. Over 40 Speech and Language Therapists assessed 882 children from mainstream schools in the UK and were aged 3–8.5 years. All children in the sample spoke English as their first language.

For the purpose of this trial, we will use the print version of RAPT.

RAPT has two subtests, both of which use the same 10 items and scored independently using scoring guidelines:

i. Information (10 items, score range 0-41), where children are asked to describe the information shown in a set of pictures; and
ii. Grammar (10 items, score range 0-39), which checks the grammar used by children, such as the use of verb tenses, while describing the information shown in a set of pictures.

Endline assessments will be independently administered by NFER Test Administrators who will visit each participating setting. Up to 160 Test Administrators will be trained to administer the RAPT and LanguageScreen in settings. We recognise that there may be a need for more than one test administration visit per setting. We have allowed for three visits per setting and,

[9] https://www.routledge.com/Action-Picture-Test/Speechmark-Renfrew/p/book/9781138586208

as far as possible, the same Test Administrator will visit each setting over the period of one week to ensure that all eligible children are tested. This will ensure high follow-up rates[10], reduce testing burden from the settings and remove any concern about bias as Test Administrators will be blind to group allocation.

NFER Test Administrators will score RAPT as per the RAPT scoring guide where a child receives individual subtest and total scores. NFER will access LanguageScreen scores from OxEd's portal similar to baseline. The RAPT scoring is quite complex. NFER will ensure all Test Administrators are adequately trained and can demonstrate accurate scoring. NFER will also develop quality control procedures to ensure high quality scoring is maintained. These procedures are currently being considered and will be fully described in the Statistical Analysis Plan.

The primary outcome latent variable will be generated using PCA that amalgamates raw scores from LanguageScreen and RAPT. PCA has an advantage over Structural Equation Modelling as it is guaranteed to capture the maximum possible variance in the variables. Further details of the PCA will be included in the Statistical Analysis Plan.

Secondary outcome

Individual subtests for LanguageScreen and RAPT assessments will be the secondary outcomes for this trial. These are: Expressive Vocabulary, Listening Comprehension, Receptive Vocabulary and Sentence Repetition from LanguageScreen, and Information and Grammar from the RAPT. Each outcome will be measured using the total corresponding raw score for the component.

Sample size

Table 3: Sample size calculations

		All children (RQ1 & RQ2)	Targeted (RQ3)	Enrichment-only (RQ4)	EYPP (RQ5)	PVI settings (RQ6)
Study parameters taken from		OxEd (West et al., 2023)	RAND (Dimova et al., 2020)	RAND (Dimova et al., 2020)	OxEd (West et al., 2023)	OxEd (West et al., 2023)
Minimum Detectable Effect Size (MDES)		0.163	0.213	0.213	0.200	0.231
Pre-test/ post-test correlations	Level 1 (child)	0.81	0.75	0.75	0.81	0.81
	Level 2 (setting)					
ICC	Level 2 (setting)	0.21	0.349	0.349	0.21	0.21
Alpha		0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8

[10] Note that the follow-up will take place only within participating settings. If a child moves to another setting, we will not be able to administer follow-up assessments.

One-sided or two-sided?		Two	Two	Two	Two	Two
Average cluster size		12	6	6	2.3[11]	12
Number of settings	Intervention	159	159	159	159	80
	Control	159	159	159	159	80
	Total	318	318	318	318	160
Number of children	Intervention	1,908	954	954	366	954
	Control	1,908	954	954	366	954
	Total	3,816	1,908	1,908	732	1,908

Sample size calculations were carried out using the PowerUpR (Bulus et al., 2021) package within Rstudio. The numbers of settings and children in the table above represent the recruited sample before any attrition, but the MDES is calculated including an allowance for 10% setting level attrition and 23% child attrition, in line with EEF EY lessons learned (EEF, 2019).

The ICC and pre-post correlations used in the power calculations and set out in Table 3 above are taken from two prior NELI trials. The NELI effectiveness trial (Dimova et al., 2020) targeted the lowest language ability children and yielded ICC and pre-post correlation estimates of 0.349 and 0.75, respectively. By contrast, the NELI Preschool efficacy trial (West et al., 2023) used all children in the class, giving a lower ICC of 0.21 (reflecting the more heterogeneous population) and a higher pre-post correlation of 0.81. The populations used in the analyses to answer RQ1 (all children), RQ5 (EYPP children) and RQ6 (children in PVI settings) are expected to be relatively heterogenous, so the parameters used in these sample size calculations are taken from the NELI Preschool efficacy trial (West et al., 2023). The populations used in the analyses to answer RQ3 (targeted) and RQ4 (enrichment-only) are expected to be rather more homogenous. Therefore, the parameters used to estimate sample size are taken from the NELI effectiveness trial (Dimova et al., 2020) to reflect this.

In the initial proposal we suggested a sample size of 300 settings would be required for this study, which we thought would give an MDES for the primary research question of 0.207. However, this was before we understood the point in the previous paragraph about the differing nature of the targeted and whole class samples. In the subsequent discussions we realised that 300 settings would be relatively well-powered for the primary analysis and that by extending it to 318 we could power the study for an MDES of 0.20 in the EYPP sample. Aiming for an MDES of 0.2 seems reasonable based on the effect size of 0.26 found in the efficacy trial (given this trial will be delivering at a larger scale and including PVIs we would expect the effect size to be a bit lower than the efficacy trial). While this meets EEF guidelines in terms of powering to the 0.20 benchmark, the trial may still be underpowered if we use the effect sizes found for targeted children in the efficacy trial, 0.16, as our estimate of expected effect size rather than the 0.26 found for enrichment only children. In reality, our primary analysis includes

[11] Based on multiplying the proportion of children in receipt of EYPP 0.0906 by the average number of children per setting - 24.4 (https://explore-education-statistics.service.gov.uk/find-statistics/education-provision-children-under-5/2022#explore-data-and-files) (West et al., 2023).

both targeted and enrichment-only children, so it is reasonable to assume the expected effect size would fall between the two values, which would be 0.21. We also undertook one further sample size calculation to revise our estimate of the sample size required to power the primary research question to an MDES of 0.20. This would require just 211 settings which represents the lowest number of settings for which it would be sensible to continue running the trial in the case of under recruitment.

Randomisation

Settings will be randomised into two arms, intervention and control, on a 1:1 basis. Randomisation will be stratified by two variables: (i) nursery setting type – maintained nursery settings and PVI settings and (ii) setting size (settings with one preschool class vs those with more than one preschool class). Randomisation will be carried out by NFER statisticians using R Code, which will be stored for reproducibility and transparency. The statistician will be provided with a minimum set of variables required to carry out the randomisation: each setting's unique ID and the two stratification variables. The results of the randomisation will be passed over to NFER's Research and Product Operations team who will liaise with settings to inform them of their group allocation.

We anticipate that baseline data sharing and LanguageScreen assessments will take place in a staggered manner. This means there will be settings that fulfil all their randomisation requirements early on. To minimise wait times for these settings and allow those allocated to the intervention group to commence NELI Preschool training, randomisation will be conducted in two batches. Specific dates for each randomisation will depend on when settings complete baseline requirements, but it is envisaged that they will be a few weeks apart in October and November 2024. Programme delivery timelines will remain consistent across both phases.

Statistical analysis

All analyses will be conducted on an intention-to-treat basis and follow the EEF's statistical analysis guidance (EEF, 2022). Analysts will not be blinded to group allocation for any of the analyses.

Primary analysis

The primary analysis will be intention-to-treat and conducted to answer RQ1. The primary outcome measure will be the latent oral language variable (extracted as the first principal component from the PCA of endline LanguageScreen and RAPT scores). It will be the dependent variable in a linear multilevel model that accounts for the clustering of children within settings as a random effect. The following covariates will be included in the model: prior attainment (extracted from the first principal component from the PCA baseline LanguageScreen scores), random group allocation (intervention vs control group) and randomisation stratifiers. For this analysis, we will use samples S1 (ideally targeted children) and S3 (enrichment sample).

Secondary analyses

The secondary analyses will be intention-to-treat and conducted to answer RQ2 research questions. The secondary outcome measures will be the total raw score for each subtest of

LanguageScreen and RAPT at endline: LanguageScreen Expressive Vocabulary (0–24), LanguageScreen Listening Comprehension (0–16), LanguageScreen Receptive Vocabulary (0–23), LanguageScreen Sentence Repetition (0–14), RAPT Information (0–40) and RAPT Grammar (0–38). We will run six linear multilevel models where each secondary outcome will be the dependent variable. As in the primary analysis model, the model will account for the clustering of children within settings as a random effect include the same covariates (the baseline measure for each secondary outcome, random group allocation and randomisation stratifiers) and use samples S1 (ideally targeted children) and S3 (enrichment sample).

Sub-group analyses

There are several analyses planned to ascertain the intervention's effect on specific subgroups of children: RQ3 to RQ5. All analyses will be intention-to-treat. RQ3 to RQ5 models will be set up the same as the primary analysis model, using the same dependent variable and covariates.

RQ3 analysis will be in the subgroup of children selected for the targeted intervention (S2: practitioner targeted children - RQ3a) and, if different, the children who should have been targeted for the intervention (S1: Ideally targeted children – RQ3b).

RQ4 analysis will be in the subgroup of the six other randomly selected children who attend the setting for at least 15 hours and are not included in the RQ3 analysis (S3: enrichment sample).

RQ5 analysis will be of the subgroup of EYPP-eligible children (S4), where EYPP eligibility was identified by settings towards the end of the trial[12]. This will determine the intervention's effect for all EYPP children in the trial, regardless of whether they attend the nursery for 15 hours per week or not and whether they are selected to receive the targeted intervention or not. In addition to the model outlined above, a second interaction model will be run that will use the primary analysis (S1 and S3) plus any EYPP-eligible children (S4) not already included in that sample. This interaction model will be set up the same as the primary analysis model but will additionally include an EYPP eligibility covariate as well as an interaction term between random group allocation and EYPP eligibility covariates.

RQ6 analysis will be in the subgroup of children that attends PVI settings and will use the same samples as the RQ1 analysis for those settings only. The model structure will be same as the primary analysis model except that the setting type covariate will not be included.

Estimation of effect sizes

We will follow the EEF's statistical analysis guidance (EEF, 2022) to calculate appropriate effect sizes for each analysis. The numerator for the effect size calculation will be the coefficient of the intervention group from the multilevel model. The effect sizes will be calculated using the total variance without covariates, as the denominator (i.e. equivalent to Hedges' g). Confidence intervals for each effect size will be derived by multiplying the standard

[12] We will collect EYPP eligibility at the start of the trial along with all children's data. Due to the rolling nature of applications for EYPP, we will also ask settings to update this information for each child towards the end of the trial to ensure we assess all EYPP eligible children.

error of the intervention group model coefficient by 1.96. These will be converted to effect size confidence intervals using the same formula as the effect size itself. We will include further details in the Statistical Analysis Plan.

Analysis in the presence of non-compliance

Compliance measure

Compliance for intervention settings will be defined as completion of NELI Preschool training and delivery of a minimum number of NELI Preschool whole class, small group and one-to-one sessions. Compliance will be measured at the setting level, using a binary variable with following thresholds. In other words, settings will be considered compliant (compliance variable equal to 1) if they fulfil each of the following criteria.

1. At least 2 staff members have completed full online training.
2. At least 75% of whole class sessions delivered.
3. At least 70% of small group sessions delivered.
4. At least 65% of individual sessions delivered.

For Criteria 1, OxEd collects NELI Preschool training data on their platform. They will share this data with NFER at the end of the trial. If a setting has at least two practitioners who completed NELI Preschool training, they will fulfil this compliance criteria.

Criteria 2–4 will be determined by the NELI Preschool session delivery logs. NFER will collect weekly data from intervention settings on the NELI Preschool sessions delivered and children's attendance at those sessions. To fulfil each criterion, settings will have to deliver (for each nursery class):

a) at least 75 whole class sessions overall (75% of overall sessions, five whole-class session per week for 20 weeks)
b) at least 42 small group sessions (70% of overall small-group sessions, three per week for 20 weeks) and
c) at least 78 individual sessions (65% of overall individual sessions, six per week for 20 weeks).

Child-level session attendance is not considered for the compliance measure although the patterns of child-level NELI session attendance will be analysed descriptively.

Control group settings will not have access to NELI Preschool, so they will all be considered compliant.

CACE analysis

We will conduct a CACE analysis for RQ1 using the compliance measure described above. We will explore the linear effect of this compliance indicator on the primary outcome measure using instrumental variable modelling. The proposed CACE analysis will be fully described in the Statistical Analysis Plan.

Additional analyses and robustness checks

Practitioner confidence is conceptualised as a short-term outcome in the ToC (see Figure 1 and Practitioner outcomes survey section in the IPE for further details). We will conduct an additional mediation analysis using the practitioner confidence measure to examine the extent improvements in children's language ability are mediated by increased practitioner confidence as implied by the ToC. This mediation analysis will be detailed in the Statistical Analysis Plan.

We will also conduct dosage analysis to determine the intervention effect on outcomes for S2 and S3 samples. These analyses will use data from the NELI Preschool session delivery logs as predictors in the models. The measures and model structure will be detailed and pre-specified in the Statistical Analysis Plan.

Missing data analysis

There is likely to be some degree of missing data for the primary outcome at endline. Where children are unavailable for testing, the reason for this will be established where possible and described in the final report. As per the EEF's statistical analysis guidance (EEF, 2022), if more than 5% of primary outcome data is missing, further missing data analysis will be undertaken. Primarily, this will seek to determine whether the primary outcome data is Missing Completely At Random (MCAR) or whether there is a degree of correlation between a missing indicator and other covariates (Missing At Random, MAR, or Missing Not At Random, MNAR). Covariates that are found predictive of missingness for the primary outcome will be added to the primary analysis model and the results of the two models compared. If the results of these two models differ, this will be indicative that data is either MAR or MNAR and sensitivity analysis may be required. Further missing data analysis will be undertaken if needed as per the EEF's statistical analysis guidance (EEF, 2022) and described in detail in the Statistical Analysis Plan.

Implementation and process evaluation (IPE) design

Research questions

Fidelity

IPE RQ1. To what extent did NELI Preschool successfully scale up to a greater (i) number and (ii) variety (i.e., both maintained and PVI) of settings?

1.1 To what extent were the core components implemented as intended?

- Practitioners engaged with training and support that equipped them to successfully deliver the intervention
- Practitioners correctly selected children for the targeted component using LanguageScreen and the Targeted Group Selection Guidance
- Settings delivered the 20-week programme in the intended format and in accordance with the script

1.2 To what extent, and in what ways, did settings adapt the intervention to better meet their own needs?

This research question will provide key information about intervention compliance for the impact evaluation and offer insights into the extent to which the training and intervention were implemented with fidelity (i.e., as intended). This will include descriptive statistics detailing the number of sessions delivered, as well as their frequency, duration and level of attendance. We are particularly interested in how the programme is implemented in PVI nurseries, given that only maintained settings were involved in the efficacy trial. We will also explore how well components that have been added since the efficacy trial work in practice – primarily the Narrative Tracker and Targeted Group Selection Guidance. We will explore variation across settings and between setting type to understand the kind of variation that is introduced by delivery in a 'real-world' context.

Child responsiveness

IPE RQ2. How well did the NELI Preschool intervention reach and support disadvantaged children (those eligible for EYPP and/or those with English as an Additional Language [EAL])?

2.1 To what extent did disadvantaged children engage with the intervention?

2.2 Were there any challenges or facilitators that applied to disadvantaged children in particular? How (if at all) did settings address these?

This research question will complement the impact analysis of any differential outcomes for children eligible for EYPP and provide insight into potential reasons behind any differences that may be observed. This question will also feed into the project's ambitions to increase understanding around how disadvantaged children can be better supported to close the attainment gap that has already emerged by this age. This is particularly relevant given the likely crossover that will occur between disadvantaged children and those assigned to the targeted intervention, for whom the programme was found in the efficacy trial to have a smaller impact compared to those in the enrichment group (West *et al.*, 2023).

How children with EAL respond to the programme is likewise of particular interest given the programme's focus on oral and literacy skills, and the additional challenges these children may face in following the sessions and having this learning reinforced at home.

IPE RQ3. To what extent can the targeted component of the intervention be seen to be fulfilling its intended purpose?

3.1 Did the Targeted Group Selection Guidance (including use of LanguageScreen) enable settings to identify the children who would benefit most from the targeted component?

3.2 Is the targeted component set up to support selected children to access the whole-class element?

3.3 To what extent does the distinction between targeted and enrichment activities and groups play out in practice?

This question will explore a key area for further research that emerged from the efficacy trial: namely that a smaller effect size (d = .16) was observed for children in the targeted group compared to those in the enrichment group alone (d = .26). The efficacy trial authors (West *et al.*, 2023) suggested that this may be due to (a) the targeted component being insufficient to

make up for the gap in learning capability between these children and their peers, or (b) the children selected for the targeted component not yet being developmentally at a stage to be able to benefit from the intervention. This question will explore both these options, as well as the extent to which the targeted component functions in accordance with its role in the ToC. It will also complement the impact analysis looking for any differential effects for children assigned to the targeted group compared to those participating only in the enrichment sessions.

Context & moderators

IPE RQ4. What are the key moderators and contextual factors that influenced how effectively the intervention was implemented in each setting?

4.1 What barriers and facilitators did settings encounter in implementing the intervention?

4.2 What (if any) is the role of the structural and process quality[13] of a setting in moderating how effectively the intervention was implemented?

4.3 What (if any) is the role of other potential moderators (see the ToC) in moderating how effectively the intervention was implemented?

This research question looks to understand any contextual or individual factors that may influence the extent to which children are able to benefit from the programme. This will help to contextualise the findings of the impact analysis, in relation to any barriers or facilitators that may have influenced the outcomes observed. This question will also help us to understand if there are any particularly significant moderators that need to be better accounted for in the ToC, and inform future programme design and implementation guidance to ensure as many potential barriers to impact are removed as is possible.

Perceived impact

IPE RQ5. What was the perceived impact of the intervention for participating (i) settings, (ii) practitioners, and (iii) children?

5.1 To what extent can practitioner outcomes (e.g., increased understanding, confidence and motivation) be seen to mediate child outcomes?

5.2 To what extent have the principles of and learning from the intervention become integrated into broader and/or ongoing practice within the settings?

5.2 What are the unintended consequences (positive or negative) of the intervention?

Providing CPD around early language development is an important objective of the programme as a mediator for better outcomes for both those children involved in the programme and more widely. This research questions helps us to understand the extent to which this plays out in practice in order to better understand this aspect of the ToC. This question will also help to explore the relative significance of practitioner-level outcomes for

[13] Process quality refers to the nature of daily interactions a child has in the setting, not just with staff but also with space and materials, other children, their families and the wider community (OECD, 2020).

child-level outcomes, or whether delivery of the programme as intended is sufficient in this respect. Similarly, this question will explore any other child outcomes that the programme may support that are not within the scope of the impact analysis – including self-regulation, which saw mixed results in the efficacy trial (West *et al.*, 2023).

In addition, we will look to identify any potential unintended negative consequences stemming from programme implementation, such as increased workload, loss of learning opportunities due to displacement, and/or a widening of the disadvantage learning gap. We will take care to differentiate these unintended consequences from any that may have emerged from taking part in the trial specifically.

Finally, this question will consider any long-term impacts of the programme for participating settings to determine the extent to which this arm of the ToC can be seen to hold true (see Longitudinal evaluation design).

Programme differentiation

IPE RQ6. What was 'business as usual'?

6.1 What was usual practice in relation to oral language development in all settings prior to the intervention? To what extent (if at all) did this differ between PVI and maintained settings?

6.2 What was usual practice in control settings in relation to language enrichment during the intervention?

This question will complement the impact analysis by looking to understand the extent to which usual practice in control and intervention settings differed, and hence the extent to which any difference in outcomes may be attributable to the intervention. It will also look at any differences in usual practice between PVI and maintained settings specifically, to contextualise any differences observed between these subgroups in the impact analysis. Finally, this question will inform our understanding of which components of the ToC represent genuine departure from what settings would already have in place.

Research methods

We are planning to use a variety of different research methods to capture the IPE data, as described below (see also Table 4 for a summary of the IPE methods). These methods were confirmed following the project IDEA workshop and set-up meetings, which offered the evaluation team the opportunity to speak with OxEd in depth about the intervention and the key areas of interest for the IPE.

Table 4: IPE methods overview

Research methods	Data collection methods	IPE dimension(s)	IPE RQ(s) addressed	Sample size and sampling criteria	Data analysis methods
Structured reviews	Structured review of online training course	Fidelity	1	Full training course	Thematic analysis
	Structured review of Delivery Support Hub activity	Fidelity	1, 4	All activity	Thematic analysis
Structured observations	Structured observation of webinars	Fidelity, context & moderators	1, 4	2 webinars (of 4)	Thematic analysis
Case studies	Structured observation of whole class & small group session delivery	Fidelity, child responsiveness, context & moderators, perceived impact	1, 2, 3, 4, 5	12 intervention settings – range of setting type, location, size, local deprivation level and proportion of EYPP and/or EAL children	Thematic analysis
	Process quality observation (SSTEW)[14]	Context & moderators	4		Thematic analysis
	Practitioner interviews	Fidelity, child responsiveness, context & moderators, perceived impact	1, 2, 3, 4, 5	Practitioner(s) responsible for delivering each of the sessions observed per setting[15]	Deductive & inductive coding; thematic analysis
Online surveys	BAU survey (baseline & endpoint)	Programme differentiation	6	NELI Preschool leads – all settings at baseline / control settings at endpoint	Descriptive statistics
	NELI Preschool Lead survey	Fidelity, child responsiveness, context &	1, 2, 3, 4, 5	NELI Preschool leads - intervention settings	Descriptive statistics

[14] The Sustained Shared Thinking and Emotional Well-being (SSTEW) Scale is a method for assessing the process quality of early years education environments and pedagogy (Siraj, Kingston and Melhuish, 2023).
[15] Note: this is a preference; in no case will practitioners be obliged to take part in an interview.

Research methods	Data collection methods	IPE dimension(s)	IPE RQ(s) addressed	Sample size and sampling criteria	Data analysis methods
		moderators, perceived impact			
	Practitioner outcomes survey (baseline & endpoint)	Perceived impact	5	Two to six practitioners per setting who completed the training, nominated by NELI Preschool lead	Pre-post control vs intervention
Attendance data	Training attendance	Fidelity, context & moderators	1, 4	Number completed in intervention settings	Descriptive statistics
	Webinar attendance	Fidelity, context & moderators	1, 4	Intervention settings	Descriptive statistics
	Session attendance	Fidelity, child responsiveness, context & moderators	1, 2, 3, 4	All eligible children in intervention settings	Descriptive statistics
	Setting attendance patterns	Fidelity, context & moderators	1, 4	All eligible children in intervention settings	Descriptive statistics

Building on the findings of the efficacy trial, the IPE will focus on the differential role and impact of the targeted component of the intervention, and the expansion of delivery to PVI settings. In addition, we will look to further develop the programme's ToC by exploring how implementation varies in response to a range of moderators and contextual factors, how effective LanguageScreen is for identifying those children who will benefit most from the targeted component, and perceived outcomes for both disadvantaged children and those with the poorest language skills. We will also be using the IPE surveys to investigate practitioner outcomes that could not be measured under Impact due to the lack of existing validated measures that would be appropriate for understanding the intended CPD outcomes of the programme.

The sequencing of the IPE activities is outlined in Figure 2 below. The baseline BAU survey will help us to understand the extent to which the programme differs from standard practice in settings, while the endpoint BAU survey allows us to determine the extent to which practice in control settings differed from practice in intervention ones. The practitioner outcomes survey at pre-trial and post-delivery allows for pre-post analysis. Conducting the structured review of the online training course ahead of the case study means findings from the former can feed into development of data collection tools for the latter. Similarly, findings from the case studies and data collection around training and support will inform development of the NELI Preschool Lead survey, which will seek to obtain an overview of the issues and experiences of settings in implementing the NELI Preschool programme.

All IPE data collection instruments will be reviewed by the trial manager, principal investigator and a consultant from RAND Europe. A minimum of two researchers will be involved in qualitative data collection, interpretation and analysis to reduce the risk of personal bias.

Figure 2: Sequencing of IPE activities

Pre-trial	During training	During delivery	Post-delivery
BAU survey (all)			BAU survey (control)
Practitioner outcomes survey			Practitioner outcomes survey
	Structured review of online training course		Structured review of Delivery Support Hub activity
		Structured webinar observations	
		Case studies	
			NELI Preschool Lead survey

1. Training and support material reviews and observations

The nature of training and support provided to practitioners will be explored through the triangulation of three different data collection processes. The data collection instruments for these activities will be informed by the programme ToC, TIDieR framework and the IPE research questions, with a particular focus on fidelity. These data collection activities will inform our understanding of the nature and accessibility of support available, as well as the frequency and nature of engagement with it.

Structured review of the online training course

An NFER researcher will complete a structured review of the online training course material. A framework will be developed based on the activities outlined in the ToC and TIDieR framework. The researcher will take notes alongside each element of this framework to record how they are communicated in practice. This review will be completed in parallel to training completion by the practitioners and will inform development of the instruments for subsequent data collection activities – particularly for the webinar and session observations.

Structured observations of webinars

An NFER researcher will attend and observe two of the four webinars that OxEd will deliver to settings over the course of the delivery period. The selected webinars will cover LanguageScreen, the Narrative Progress Tracker, Targeted Group Selection and reviewing implementation and child progress midway through the delivery period. Particular attention will be paid to how the webinars complement the online training material in terms of both additional support and consistent messaging. We will also look at the nature of practitioner engagement, including the nature of questions asked during the webinars, as well as more qualitative aspects such as level and quality of interaction, and perceived levels of satisfaction and/or concern with the programme. As these webinars will be spaced out at intervals across the delivery period, we will also note any observations relating to the evolution of these elements over the course of the programme. In addition, questions raised by attending practitioners will inform our analysis of context and moderators that informed implementation, particularly in terms of challenges and facilitators that practitioners observed.

The data collection instrument for the webinars will include the same framework as the one used for the online training course review (see above). Additional elements will be added, however, to record levels of practitioner engagement, as well as any practitioner reports relevant to other IPE dimensions such as adaptations, challenges and facilitators. Any discrepancies in information provided between webinars or with the online training course will be noted.

Structured review of Delivery Support Hub activity

Following the delivery period, an NFER researcher will carry out a structured review of any activity that occurred in the Delivery Support Hub over the course of the trial. This will enable us to assess the overall level and quality of engagement with the community of practice, which is an important element of the programme ToC. This will inform our understanding of fidelity to programme intent, as well as the potential impact of moderators to programme

implementation - as level of engagement with the community of practice was identified in the IDEA workshop as a potential moderator of the fidelity and quality of delivery.

A similar instrument to the one used for the webinars will be employed to note relevant practitioner experiences under the various IPE dimensions as well as to track the instructions and guidance provided by the NELI Preschool team and specialists, and how this compares to the other training and support material.

2. Case studies

Case studies will be carried out with 12 settings, with one visit per setting spaced over the course of the delivery period to help us develop an in-depth understanding of numerous IPE dimensions. We have opted for a sample size of 12 to allow for more in-depth explorations of each case study, including several different data collection activities, while allowing for pragmatic considerations around cost and research team capacity. This number will allow us to achieve sufficient variety in key contextual points of interest, primarily: setting type (PVI or maintained), size (big or small),[16] and local deprivation level (using the Income Deprivation Affecting Children Indices [IDACI]). This data will be collected either as part of the recruitment process or through the child data collection at baseline. We will aim for half of the case study settings to be PVIs. While this will still result in an underrepresentation of PVIs at a national level[17], it will likely still be an *over*representation of PVIs in terms of the full trial sample. This will support our focus on how well implementation of the programme extends to PVI settings specifically. Data collection will take place at staggered intervals throughout the delivery period with the aim of gathering perspectives from practitioners at different stages of the implementation process.

Case study settings will receive £100 at the end of the trial delivery period in recognition of their time given to the research.

Structured observations of whole-class & small-group session delivery

Each of the case study settings will receive a visit from an NFER researcher during the delivery period to observe delivery of one whole-class session and one-small group session. The researcher will use a structured observation tool, which will look at fidelity and adaptations to the teaching techniques outlined in the training course, adherence to the stipulated session length and structure, contextual considerations such as location and child attendance, and qualitative assessments of child engagement.[18] Particular attention will be paid to children eligible for EYPP and/or with EAL.[19] The observation tool will be piloted and refined as part of the formative evaluation. Data collected with the tool will be qualitative in nature.

[16] This will be based on the number of practitioners in the setting working with the children.
[17] There are 9,700 maintained and 21,200 PVI settings nationally (Explore Education Statistics, n.d.).
[18] This refers to the physical act of being present at the session (not playing elsewhere), and the extent to which the child appears to be listening (not distracted or talking on unrelated topics) and actively contributing (responding to questions, raising their hand, etc.).
[19] The names of these children will be identified in advance using the child data shared by each setting. The practitioner will then be asked to identify each child by name at the start of the setting, so that the observer is able to match names (and corresponding characteristics) with faces.

Observing both a whole-class and small-group session in each case study setting will also allow us to draw links between child engagement in each session type, particularly with those children for whom the targeted intervention is intended to facilitate their engagement in the enrichment component.

We will not be observing one-to-one sessions, as we feel this could be intimidating for the practitioner and/or distracting for the child to the extent that what was observed would not be a meaningful representation of standard practice.

Practitioner interviews

The practitioner that delivered each of the observed sessions will participate in a 45 to 50-minute semi-structured interview with the NFER researcher. This will result in one or two interviews per case study setting, depending on whether the same or different practitioners delivered the sessions.[20] The interviews can take place in-person on the same day as the observations or remotely via Microsoft Teams at another time of their choosing, following the visit. The interviews will cover how the practitioners found the training and implementing the programme, including the screening and targeted group selection process. They will also be used to pick up on and expand any points relating to context, fidelity or adaptation that may have emerged from the session observations in the same setting. We will ask practitioners about challenges and facilitators they have encountered, how engaged they have found the children to be in the programme and any outcomes they have perceived for either the children or practitioners (including themselves), and the extent to which outcomes for the latter have appeared to influence outcomes for the former. The majority of the interview schedule will be piloted and refined as part of the formative evaluation.

Practitioners who participate in the interviews will personally receive a £50 in recognition of the time they have given to the research. This is particularly important given that early years practitioners will often not receive any paid time outside of face-to-face time with the children.

Process quality observations

As part of the case study visits, the NFER researcher will carry out a two-hour structured observation using the validated Sustained Shared Thinking and Emotional Well-being (SSTEW) scale to assess process quality within the setting (Howard *et al.*, 2020)[21] Process quality has been consistently linked with positive early years outcomes (OECD, 2020); (Eadie *et al.*, 2022, (Sim *et al.*, 2018)., (Eadie *et al.*, 2022) and identified as key for providing an environment and conditions for the child to be ready to learn (Siraj, Kingston and Melhuish, 2023). It is likely, therefore, that process quality will be an important moderator for programme impact. Given the personnel load of conducting an observation of this kind, it is not possible

[20] Practitioner consent to participate in the interview will be sought in advance. Should the practitioner delivering the session not wish to participate in an interview, another colleague with relevant experience of the programme who is willing to participate may take their place.

[21] The SSTEW contains 14 elements, which combine as five sub-scales: (1) building trust, confidence and independence, (2) social and emotional wellbeing, (3) supporting and extending language and communication, (4) supporting learning and critical thinking, and (5) assessing learning and language. Each element is rated from 1 (inadequate practice) to 7 (excellent practice) based on the pattern of presence/absence of indicators.

to carry this out with a sample size large enough for quantitative moderator analysis. Therefore, we intend to use the SSTEW to scale to formulate a qualitative assessment of process quality in each case study setting and reflect, through triangulation with the interviews and other case study observations, on any indicators of how this may relate to successful implementation and/or perceived outcomes. Given the small sample size and reliance of setting engagement this analysis may not be considered representative of broader trends.

3. Online surveys

Business as usual (BAU)

Two online BAU surveys will be implemented: one at baseline to inform our understanding of BAU in all settings prior to programme delivery, and one at endpoint to understand what BAU looked like in control settings over the trial period. The nominated NELI Preschool Lead at all participating settings will be asked to complete the baseline survey, but only those in control settings will be asked to complete the BAU survey at endpoint.[22] The BAU survey will ask about practice around oral language development in the setting, both in terms of formalised programmes or interventions and day-to-day working. We will also ask about any related training participating practitioners may have completed in the past.

This information will help us to understand the extent to which the programme differs to standard practice in settings, as well as the extent to which the settings asked *not* to deliver the intervention can legitimately be understood to represent a 'control'. This will inform our interpretation of the findings from the impact analysis. We will also be able to establish whether their setting is maintained or PVI, so that we can explore any potential differences in programme differentiation between the two setting types.

NELI Preschool Lead Survey

The NELI Preschool Lead in all intervention settings will be asked to complete an online survey at endpoint. Like the case study interviews, this survey will cover settings' experiences of implementing the programme, the training and support made available, perceived child and practitioner outcomes, and any challenges or facilitators they encountered. Unlike the interviews, however, the survey will focus more on the structural and logistical aspects of delivery, rather than what happens within the sessions. This survey will provide a better sense of the prevalence of different experiences, including those that emerged through the case study interviews.

Practitioner outcomes survey

Practitioner outcomes are a key component of the ToC, as the programme is conceived to achieve broader and long-term impacts by acting as CPD for transferable skills around supporting early language development. As there are not currently any validated tools that would be sufficiently rigorous for measuring the intended practitioner outcomes of the programme (improved understanding of how to support early language development, as well as the confidence and motivation to do so) in the impact evaluation, we will be carrying out

[22] Relevant BAU questions (e.g. other interventions implemented over the trial period) will also be asked of intervention settings in the NELI Preschool Lead Survey.

exploratory analysis through the IPE surveys, with additional qualitative insights gathered through the case study interviews. Between two and six staff members from each setting, including the NELI Preschool Lead, will be asked to complete the same online survey at baseline and at endpoint.[23] This survey will look to assess the practitioners' confidence and motivation in relation to key areas of practice promoted by the programme: (1) explicitly teaching the meaning of new vocabulary, (2) providing opportunities for children to practice new words in a range of contexts and (3) providing opportunities for children to explore the narrative structure of stories through dialogic reading, and (4) providing scaffolded support to those children with the weakest language skills. As there are no validated measures for assessing early years practitioners' confidence and motivation in these areas, we have developed suitable scales based on existing relevant validated measures.

The confidence measure (see Appendix B) is an adapted version of the second scale of the Early Math Beliefs and Confidence Survey (EM-BCS),[24] which has been validated for assessing teachers' confidence in helping preschool children (aged 3-4 years) to learn maths (Chen and McCray, 2013).[25] The measure consists of 11 statements with a Likert scale to indicate level of agreement with each. The statements have been adapted to capture the shift in domain (from maths to language development). Each question receives a score between 1 and 5 (depending on the selected option on the Likert scale). Scores for all questions are then combined to produce a summary score between 11 and 55, with higher scores demonstrating greater confidence in helping nursery aged children develop their language skills.

The motivation measure (see Appendix B) is based on the Work Tasks Motivation Scale for Teachers (WTMST) (Fernet *et al.*, 2008), and will be refined using cognitive interviewing in the formative evaluation (see below). This measure consists of two questions for each of the five areas of practice outlined above. The first question is a multiple-choice asking about the practitioner's level of motivation, while the second will ask practitioners to rank a range of options around what motivates them. This will allow us to understand not only the degree of practitioner motivation, but also the type (intrinsic, extrinsic, etc.). This is significant as there is evidence to suggest that types of motivation with higher levels of self-determination (*intrinsic motivation*[26] and *identified regulation*[27]) are positively associated with teacher self-efficacy, and negatively associated with burnout (Fernet et al., 2008). On the other hand, types of motivation

[23] The staff members invited to complete the survey at baseline will be those whose contact details were shared via the Memorandum of Understanding.
[24] Each of the three scales produce separate scores and cannot be combined. The other two scales look at 'beliefs about nursery aged children and maths' and 'confidence in maths abilities' and hence were not deemed relevant to this trial (Chen et al., 2013).
[25] Internal consistency assessed using Cronbach's alpha with 228 preschool was 0.90 for the second scale (Chen et al., 2013). Test-retest reliability with the same group of teachers after 18 months found no significant difference (Chen et al., 2013).
[26] '*Intrinsically* motivated behaviours are engaged in for the pleasure or the satisfaction derived from performing them' (Fernet et al., 2008).
[27] *Identified regulation* is where individuals choose to do something 'because it is congruent with their own values and goals' (even if the activity is not intrinsically interesting or enjoyable) (Fernet et al., 2008).

with lower levels of self-determination (*external regulation*[28] and *introjected regulation*[29]) are negatively associated with self-efficacy, and positively associated with burnout (Fernet et al., 2008). While we will follow the same structure as the WTMST, the statements for gauging motivation type have been modified with the aim of making them more relevant to early years practitioners. This will provide some insight into the extent to which CPD around the importance of early language development may or may not motivate practitioner practice.

All surveys will be hosted on Questback and each participant will receive their unique link. Survey routing will deliver only relevant questions to minimise completion burden for practitioners. Responses will be downloaded from Questback and the data quality assured via NFER's data quality assurance (DQA) process prior to its analysis.

4. Administrative data

Training course completion

OxEd will share with NFER the number of practitioners who registered to complete the online training course in each setting, as well as the number of practitioners who completed it. This will inform and the fidelity (and compliance) analysis. Training course completion data will be at individual level, but identifiable at setting level only.

Webinar attendance

OxEd will share with NFER attendance numbers for each of the four webinars, as well as number of views of the webinar recordings. This will inform our understanding of fidelity and levels of engagement with this programme element.

Session attendance

NFER will ask all intervention settings to complete an online Session Delivery Log over the session delivery period, recording the date each session is delivered, the type of session, and which children attended it. The Session Delivery Log will be shared with settings in an Excel format using NFER's secure online portal for five weeks at a time. The settings will be asked to return the completed template every 5 weeks using the same portal and download the next template for completion over the following 5 weeks.

The Session Delivery Log will inform the fidelity (and compliance) analysis. We are also interested in understanding the extent to which those children selected for the targeted group are indeed the recipients of additional sessions, as well as the extent to which these additional sessions are also attended by children with stronger language skills – which could potentially serve to widen the gap between them.

[28] '*External regulation* occurs when behaviours are regulated to obtain a reward or to avoid a constraint' (Fernet et al., 2008).
[29] '*Introjected regulation* is when an individual puts 'pressure on themselves through internal coercion (e.g., anxiety, shame, or guilt) to make sure that a particular behaviour is performed' (Fernet et al., 2008).

Setting attendance patterns

NFER will ask all intervention settings to provide the days and total number of hours each eligible child typically attends the setting each week as part of the baseline data collection. This data will be used to understand whether attendance patterns disproportionately exclude certain children, particularly disadvantaged children, from receiving the targeted intervention.

Analysis

Qualitative data - observations, interviews & reviews

Interviews notes will be written up from video recordings as intelligent verbatim transcripts[30] and uploaded to the qualitative data analysis software MAXQDA. High-level deductive coding will be used to sort the data into relevant themes. Detailed inductive coding will then allow us to draw out the key findings under each of these themes.

Observation data will be treated qualitatively to provide a clear narrative of what the intervention looks like in practice and the key variables that influence its effectiveness. Session observation notes will be typed up and combined into a single Excel spreadsheet. Between-case inductive coding will allow us to draw out the key findings under each area of the observation tool across the sample as a whole. A separate Excel spreadsheet will support within-case analysis, where key findings from the interviews, session observations and process quality observations will be summarised for each case study. This will allow us to triangulate the data sources and better understand the context within which particular issues or perspectives emerge. The significance of process quality in relation to findings from the session observations and interviews will also be considered, drawing on both within- and between-case analysis to do so.

All deductive coding will be based on a coding frame that will be developed in advance based on the programme ToC and IPE research questions, and with input from RAND Europe's early years specialist. A minimum of two researchers will be involved in the coding process and will carry out a review of each other's work at any early stage of the process to ensure inter-coder reliability.

Quantitative data - surveys & administrative data

As outlined above, survey response data will be exported from Questback and quality assured prior to its analysis, with each data source stored in a separate file. All quantitative analysis (descriptive and inferential) will be carried out using appropriate R packages. The required analyses will be set out in detail within the SAP (see impact analysis section earlier for more detail) and will include both the descriptive statistics and inferential statistics required to answer the IPE research questions, including cross-tabulations of key potential moderators.

Data from the Practitioner Outcomes Surveys (baseline and endpoint) and the BAU surveys (baseline and endpoint) will be subject to statistical analysis to determine whether the change

[30] Intelligent verbatim transcription excludes fillers and redundancies that do not add meaning to the content to make the text more 'readable' (McMullin, 2021).

from baseline to endpoint for practitioners in intervention settings differs significantly from the same change for practitioners in control settings.

Triangulation of qualitative and quantitative data

The design of the data collection tools for each of the qualitative and quantitative components of the IPE will mutually inform each other to ensure consistency and create opportunities for complementary analysis. For example, similar questions will be asked in the case study interviews and NELI Preschool Lead survey, which will enable us to explore the detail of a particular issue as well as how perspectives on this may vary more broadly across the intervention settings. We will also develop an integrated analysis framework that will map how each of the data sources will feed into our analysis and reporting for each of the IPE research questions. Findings from each of these data sources will subsequently be examined in tandem to ensure their integration when responding to each research question in the final report. We will collate and triangulate all data sources through an analysis workshop to ensure we provide a comprehensive assessment of the implementation effectiveness and perceived outcomes of the NELI Preschool programme to inform our interpretation of findings from the IPE.

Cost evaluation design

We will design and undertake cost evaluation following the EEF's Cost Evaluation Guidance (EEF, 2023a). The primary cost categories under scrutiny will include programme costs, training costs, preparation and delivery costs, equipment and other resources costs, and the cost of other programme inputs. These costs will also include staff time required to implement the programme (including NELI Preschool training time, time to attend webinars, time to administer assessments and, time to prepare and deliver NELI Preschool sessions as well as any other additional time relevant to the programme). Costs will be grouped into prerequisite, start-up, and recurring implementation costs.

To gather cost and staff time data comprehensively, we will utilise a cost proforma specifically designed for case study settings, supplemented with additional information obtained through the NELI Preschool Lead survey and practitioner interviews. Additionally, we will tap into existing implementation monitoring and evaluation data, capturing details such as training completion, sessions delivered, and resource usage. Furthermore, we will develop a cost proforma for the delivery team to account for all aspects of programme delivery. We will ascertain the average cost per-child-per-year using estimated costs over a three-year period following the EEF's cost guidance. We will follow the ingredients method and will include the cost of all resources necessary to implement the programme. In addition to this, we will also present total time by personnel for the training, preparation and delivery of NELI Preschool.

Pre-trial delivery and formative evaluation design

The main trial will be preceded by a formative evaluation of a pre-trial delivery period running from November 2023 to July 2024, with programme delivery starting in February 2024. Given the reliance on maintained settings for the efficacy trial of NELI Preschool (West *et al.*, 2023), the aim of the formative evaluation will be to assess the feasibility of NELI Preschool implementation in PVI settings and to explore any implications for the ToC. This is especially important to explore prior to the main trial as the evidence indicates that maintained settings

are significantly different from PVI settings, such as having higher qualified staff, greater availability of additional and specialist services and a higher proportion of EYPP-eligible and SEND children (Paull and Popov, 2019); (Bonetti and Blanden, 2020).

The formative evaluation will also trial some important and high-risk components of the main trial, such as the selection of children for the targeted component of the programme, completion of NELI Preschool Session Delivery Logs, and incentive levels. Incentives equivalent to those in the main trial will be used for all data collection activities to support an accurate prediction of engagement in these elements in the main trial. All settings will receive compensation of £120 for staff time taken to complete the training, for up to two practitioners, as well as £55 for administering LanguageScreen assessment.

The formative evaluation will comprise an IPE element with 10 PVI IPE settings (there will be no impact element). These will be selected with the aim of ensuring variation across setting size, location, level of deprivation and staffing profile.[31] The NELI Preschool Lead in each IPE setting will be asked to nominate a practitioner (which could be themselves) to participate in a 45-minute remote interview. Nominations will be made on the basis that they have completed the training, were involved in the selection of children for the targeted component and are delivering at least some of the sessions.

The 10 IPE settings will also be asked to track which sessions are delivered, when, and attended by which children using the NELI Preschool Session Delivery Log provided by NFER. The purpose of this is to better understand the feasibility of asking settings to provide this information and to explore the level of detail on child attendance given both the existing pressures on participating staff members and the broadly free-play environments in which the programme is being delivered. Each setting that completes and returns their logs to NFER will receive £150 in recognition of the time taken to collect this data. In addition, there will be a short introductory call with these settings to onboard them to the evaluation and recruit for participation in the observations and interviews. We will also carry out light-touch cognitive interviewing as part of these calls to develop an appropriate practitioner motivation measure for use in the main trial.

Half (five) of the IPE settings will also be asked to facilitate in-person observations of NELI Preschool delivery, including one whole-class and one small-group session. Data collection will be staggered over the course of the 20-week delivery period to provide insight into different stages of the intervention. The observations and interviews will probe for practitioner experiences, challenges and facilitators in relation to the training, support (e.g., webinars and Delivery Support Hub), targeted group selection and delivery model, as well as use of the session delivery log. Settings will receive a 'thank you' payment of £100 for facilitating observations, and the practitioner who participates in the interview will personally receive a voucher of £45 in recognition of the time given.

Finally, we are interested in understanding the extent to which minimum attendance thresholds set for the trial may disproportionately exclude disadvantaged children who would otherwise

[31] Given the constrained timelines of the formative evaluation and the importance of timely reporting of findings, the evaluation team will aim to involve settings on a rolling basis, which will limit the extent to which the sample characteristics can be pre-determined.

have been selected for the targeted intervention, thereby contributing to a widening of the disadvantage attainment gap. This concern arises from evidence indicating that disadvantaged children tend to spend less time in early years settings compared to their more affluent peers (House of Commons Education Committee, 2019). While OxEd have sought to minimise this by setting the threshold at the number of hours of free childcare entitlement for all children, it would still ultimately be at the discretion of setting staff to decide whether a child's level of attendance was appropriate for inclusion in the targeted group.

To understand the extent of this phenomenon, we will compare the children selected with those who would have been selected based purely on LanguageScreen scores, and the extent to which those excluded as a result qualify for EYPP. In addition, we will also observe any discrepancies in selection to understand the extent to which settings follow the Targeted Group Selection Guidance in practice, as the main trial analysis relies on a consistent approach for children selection between the settings (for delivery) and NFER (for analysis). To achieve this, all 20 settings involved in the pre-trial delivery will share the list of children selected for the targeted component, and the data used to select them (LanguageScreen scores and attendance patterns). We will then use this data to pilot the selection process we intend to use for formulating the sub-group for the targeted impact analysis in the main trial. By comparing the list of children this process produces with the setting's own list, we will be able to assess the extent of variation between the two. All settings that supply the requested child data will receive a £150 'thank you' payment.

Qualitative data from the IPE element will be analysed thematically using a combination of deductive and inductive coding (see the Analysis section for the main trial above). Compliance with completion of the Session Delivery Log will be analysed using descriptive statistics.

The research team will feed back emerging findings informally to the EEF and delivery team on a fortnightly basis for the duration of the formative evaluation, to ensure any information relevant to the main trial can be acted on as quickly as possible. In addition, once all the relevant data has been collected, we will present their overall findings to the EEF and share a brief summative final report, to be included in the report of the main trial.

All participating settings will receive a short summary of key findings from NFER in July 2024 in recognition of their valuable contribution. The research team will take care to ensure that the findings shared will in no way introduce bias into the main trial.

Longitudinal evaluation design

In addition to the main trial analyses, we will carry out a longitudinal component involving administrative data analysis coupled with a light-touch IPE. This reflects the emphasis the programme ToC places on the CPD element of the programme and its aim to improve practice around language development across the setting and beyond session delivery.

To support this, we will match children's data from the main trial with National Pupil Database (NPD) to obtain each child's Early Years Foundation Stage Profile (EYFSP). We will submit an application to access this administrative data in April 2026 and expect to receive access to the data in January 2027. We will construct a composite score for each child based on the average score for all elements of the EYFSP that involve language skills of any kind. We will

repeat the primary analysis of the impact evaluation using this measure as the outcome measure instead of the latent oral language variable. All the other model parameters will remain the same as the primary analysis. Further details on this outcome measure and intended analysis will be detailed in the Statistical Analysis Plan.

The light-touch IPE component will seek to understand the longer-term impact of the intervention on practitioners and settings in terms of implementing effective language teaching techniques in continuous provision. The NELI Preschool Lead in each intervention setting will be asked to complete a short online routed survey using Questback, with questions relating to any continuation of the intervention and/or elements thereof, the nature of language provision in the setting at that time, any challenges for embedding the intervention in continuous practice (e.g., staff turnover) and other potential moderators that may have changed over that time. It will also ask about any perceived longer-term impacts of the trial for practitioner confidence, motivation and practice in relation to supporting language development, as well as any other setting-level changes that may have occurred as a result (including possible unintended consequences such as staff turnover). This data will be analysed using descriptive statistics following the same process as the quantitative data analysis for the main IPE. Analyses from the longitudinal impact and IPE will be integrated in an addendum report which is expected to be published in June 2027.

Ethics and registration

The trial will be designed, conducted and reported to CONSORT standards (http://www.consort-statement.org/consort.statement/) and registered on ISRCTN registry (http://www.controlled-trials.com/).

NFER is committed to the highest ethical standards in all of its activities and ethical considerations are embedded in its detailed quality assurance processes. This evaluation will be conducted in accordance with NFER's Code of Practice, available at NFER Code of Practice. The ethical approval for this evaluation was approved by NFER's Code of Practice during the set-up of this trial in October 2023. All of NFER's projects abide by its Code of Practice, which is in line with the Codes of Practice from BERA (the British Educational Research Association), MRA (the Market Research Association) and SRA (the Social Research Association), among others. NFER, OxEd and the EEF will work together to also ensure each organisation's policies can be applied in practice.

Ethical agreement for participation within the trial will be provided by the headteacher or the nursery manager of the setting via signing the MoU that outlines the responsibilities of all parties involved in the trial. Individual staff members selected for case study interviews will give their own agreement to participate.

Before requesting settings to share their children's data, NFER will ask the settings to distribute a parent information sheet and withdrawal form to the parents/carers of all participating children. Parents/carers who do not want their child to participate in the evaluation can inform their child's setting using the withdrawal form. Settings will not share data for the children whose parents/carers withdrew them at this stage. Parents/carers will also have the opportunity to withdraw their child from the evaluation and associated data processing at any stage of the trial, using the withdrawal form.

In addition, NFER will offer settings a chance to opt out of sharing additional data after the trial. We will do this when we approach settings to gather baseline data. This data sharing pertains to the evaluation data collected by NFER for this trial which will be shared with OxEd following the main trial report (see the data protection section for further details).

Data protection

All data gathered during the evaluation will be held in accordance with the Data Protection Act 2018 and General Data Protection Regulation (GDPR) and will be treated in the strictest confidence by NFER, OxEd and the EEF. No nursery, practitioner or child will be named in any report arising from this work, nor will we include any information that might mean that someone else could identify them.

NFER is the data controller for this evaluation and makes decisions about what personal data is used and how it is processed, in accordance with the objectives of the evaluation set by the EEF. OxEd is the data processor during the evaluation. After the evaluation report is published on the EEF website, NFER will share evaluation data with OxEd, at which point OxEd will become an independent data controller for this data. The evaluation data will also be pseudonymised and transferred to the EEF data archive, at which point the EEF will become the data controller for the archived data. More information on this data sharing and archiving is included below.

The lawful basis for processing personal data is covered by GDPR Article 6 (1) (f): Legitimate interests: the processing is necessary for your (or a third party's) legitimate interests unless there is a good reason to protect the individual's personal data which overrides those legitimate interests. A legitimate interest assessment has been undertaken. The trial fulfils one of NFER's core business purposes (undertaking research, evaluation, and information activities). It has broader societal benefits and will contribute to improving the lives of learners by providing evidence about the impact of teaching techniques used in the classroom. Research cannot be done without processing personal data, but processing does not override the data subject's interests.

NFER and OxEd have signed a data processing and data sharing agreement that will govern the collection and sharing of personal data for this trial. This agreement includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, OxEd will provide an MoU to settings, explaining the nature of the data being requested of settings, practitioners and children, how it will be collected, and how it will be passed to and shared with NFER. Two separate Privacy Notices are available: one for settings[32] and another one for parents[33].

For the purposes of the trial, OxEd will collect expressions of interest (EoI) from interested settings, gathering nursery staff contact details. Upon eligibility confirmation, Headteachers/Nursery Managers will sign an MoU and identify the NELI Preschool lead in their setting. Personal data collected includes names, contact details, and job roles for

[32] https://www.nfer.ac.uk/media/rvkauwre/neli_effectiveness_trial_nursery_staff_privacy_notice.pdf
[33] https://www.nfer.ac.uk/media/g5ehuzmc/neli_effectiveness_trial_parent_privacy_notice.pdf

Headteachers/Nursery Managers and NELI Preschool leads. Practitioner personal data for the trial will be collected directly from nurseries by Oxed via the MoU. OxEd will share the EoI and MoU data with NFER via a secure online portal for trial coordination. Contact details shared on the EoI or MoU may also be shared with the EEF and SPHs. This will enable the EEF to monitor recruitment progress and to confirm setting eligibility. SPHs may contact settings to let them know about support available through the Hub.

The NELI Preschool lead in each setting will be asked to complete a survey at the beginning and at the end of the trial period. In addition, nursery practitioners will be asked to complete an online survey about their confidence and motivation in supporting early language development (at the beginning and end of the trial). Finally. One year after the trial ends, the NELI Preschool lead from intervention settings will also be asked to complete an online survey to give their views about the longer-term impact of the programme. OxEd will share practitioner training attendance data with NFER. Practitioners' training attendance data at a nursery level will be linked to the child-level data collected as part of this trial to understand how important training attendance is for child outcomes. Additionally, specific variables from the practitioner surveys covering self-reported confidence and motivation in supporting early language development will be included in the impact analysis, which will be linked to the child-level dataset.

NFER will collect children's personal data directly from the settings. This includes full name, date of birth, class (if applicable), gender, Local Authority (LA) of home, EAL status (as perceived by nursery staff), EYPP status and attendance patterns at the nursery. NFER will share children's personal data with OxEd via OxEd's LanguageScreen software or via NFER's secure online portal. LanguageScreen assessment will be administered by practitioners (at baseline) and by NFER's Test Administrators (in-person at endline). NFER's Test Administrators will also administer RAPT paper assessments (in-person at endline). OxEd and NFER will access children's LanguageScreen scores via OxEd's LanguageScreen software.

For intervention settings, NFER will collect whether children were selected for the NELI targeted intervention and their attendance at NELI Preschool sessions via a session delivery logs, completed by nursery practitioners using NFER's secure online portal.

NFER will share children's personal data with the DfE in order for their data to be matched to the NPD. NFER will collect children's EYFSP from the NPD and use it for the longitudinal analysis.

A small number of settings will be invited to facilitate a visit from an NFER researcher involving practitioner interviews and observations of NELI Preschool sessions and usual practice. Children will be observed as part of this, but no further personal data will be collected. Some of the training webinars in which practitioners actively participate will also be observed by an NFER researcher.

Additional data sharing between NFER and OxEd

After the main EEF report is published (currently planned for March 2026), NFER will share the evaluation data with OxEd, which will enable OxEd to link it with the data they hold and undertake their own research to publish findings from the trial in scientific journals. At this point, OxEd will become an independent data controller. The evaluation data will include:

- Children's personal data (except the data collected from the NPD).
- Close-ended survey responses from the practitioner surveys covering self-reported confidence and motivation in supporting early language development. This will include nursery ID and practitioner role and qualification level.
- Anonymised data from close-ended responses from the NELI Preschool lead surveys. This will not include any personal, nursery or contextual data that could make matching or identification possible.

Data archiving and deletion

Three months after the publication of the final evaluation report (currently planned for June 2027), pseudonymised child-level data will be added to the EEF data archive, which will include practitioner's training attendance at a nursery level and specific variables from the practitioner surveys covering self-reported confidence and motivation in supporting early language development. At this point, the EEF becomes the data controller and is responsible for taking decisions about the means and purposes of processing. The EEF archive is managed by FFT Education on behalf of the EEF and hosted by the Office of National Statistics (ONS). Further information is available in the privacy notice for the EEF data archive. NFER will delete any recording and digitalised transcriptions of interviews as soon as the notes are finalised and personal data collected as part of this evaluation within one year of publication of the final report (currently planned for June 2027). This deletion process is currently planned for June 2028.

Staff/Teacher interview data and notes from the observations will not be shared with the EEF or archived. Open-response survey questions and data from the NELI Preschool lead surveys will likewise not be shared or archived.

Personnel

Table 5: Key members of the evaluation team (NFER and RAND) and the delivery team at OxEd

Name	Organisation	Role and Responsibilities
Stephen Welbourne	NFER	Project Director – responsible for overall delivery of the trial
Elena Rosa Speciani	RAND	Consultant on the evaluation side, offering advice on IPE design, trial design and implementation in early year settings and contributing towards the interpretation of trial findings
Palak Roy	NFER	Trial Manager and impact evaluation lead – responsible for the day-to-day management of the trial, delivery of the trial design
Lillian Flemons	NFER	IPE Lead – responsible for the design and delivery of the IPE
Gemma Schwendel	NFER	Trial statistician – responsible for leading quantitative analysis for the main trial
Kathryn Hurd	NFER	Research Operations Lead – responsible for overall data collection and setting communications strategy
Jo Stringer	NFER	Senior Project & Delivery Manager – responsible for day-to-day operations for the main trial, including coordinating data collection and serving as point of contact for the settings
Katherine Stoodley	NFER	Project and Delivery Manager – responsible for data collection and setting communications for the Formative Evaluation
Dr Merrilyn Groom	RAND	Consultant on the evaluation side, contributing to the IPE design, providing ongoing advice in working with early years settings and contributing towards the interpretation of trial findings
Mariela Rios Diaz	OxEd and Assessment	Project Manager of NELI Preschool Trial – responsible for the day-to-day management of the trial
George Ulmann	OxEd and Assessment	Managing Director of OxEd – responsible for overseeing trial management
Dr Gillian West	OxEd and Assessment	Director of OxEd – responsible for overseeing research aspects for the NELI Preschool trial
Prof Charles Hulme	OxEd and Assessment	CEO and Founder of OxEd – responsible for overseeing research aspects of the NELI Preschool trial

Blansh Al-Awkabi	OxEd and Assessment	Junior Research Officer – responsible for liaising with settings, recruitment and setting engagement and delivery support
Hazel Carter	OxEd and Assessment	Marketing Director – responsible for marketing and recruitment and setting engagement
Joe Lowe	OxEd and Assessment	Operations Director – responsible for logistical, strategic and operational planning and execution, and process improvement
Kevin Price	OxEd and Assessment	Support Team Operator – responsible for tech and training support
Ina Skronska	OxEd and Assessment	Support Team Operator – responsible for tech and training support
Kate Humphreys	OxEd and Assessment	Research Officer – responsible for mentoring on the training
Harley Richardson	OxEd and Assessment	Product Director of Technology and Technological Aspects – responsible for overseeing technology use and development, including, LanguageScreen and OxEd portal
Mihaela Duta	OxEd and Assessment	Data and Software Engineer – responsible for data architecture design, pipeline development and data modelling and optimization
Chris Speight	OxEd and Assessment	Data Protection Officer – responsible for monitoring compliance and advising on data protection obligations
Jenny Mason	OxEd and Assessment	Programme Delivery and Product Advisor – responsible for mentor on the training
Sarah Hearne	OxEd and Assessment	Project/Research Manager – responsible for ad hoc project and research management
Sue Lowe	OxEd and Assessment	Communications Manager – responsible for developing communication strategies and coordinate communications for recruitment

Risks

Risk	Assessment	Controls, countermeasures, and contingencies
Insufficient settings recruited to the trial	**Likelihood:** low **Impact:** high	• NFER will input into recruitment material and will work closely with OxEd. If required, our experienced operations team can assist with recruitment through a separate grant agreement. • If required, decide and monitor pre-agreed recruitment targets to identify any unfavourable trends early on to act quickly. • Setting-level randomisation will be more manageable and attractive.
Setting attrition from trial and primary analysis	**Likelihood:** low **Impact:** moderate	• Clear initial and ongoing communication with settings explaining principles and expectations. • Sign up to the trial via MoU with clear identification of requirements. • Where possible, lower testing burden on settings by NFER administering the assessments. • NFER to communicate with one key contact per setting to inform them of next steps. This will help update any changes in contact in settings. • Incentive payments to all settings upon completion of baseline and endpoint assessments.
Insufficient number of children selected	**Likelihood:** low **Impact:** high	• At the EOI stage, settings are asked to confirm whether at least 14 children are registered who meet the eligibility criteria to ensure at least 12 children can be selected for primary analysis in the following academic year. • NFER to select children for trial analyses after baseline assessment as per trial protocol. • Sample size calculation also assumes 10% setting attrition and 23% children attrition.
Intervention is not implemented as intended	**Likelihood:** low **Impact:** moderate	• Clear information provided to settings explaining the principles of the trial and expectations. • Two randomisation tranches help settings to access training sooner for those that provide baseline data early. • IPE to monitor implementation fidelity.
Difficulty in securing target response rates for IPE (either settings for case studies or individuals for interviews)	**Likelihood:** moderate **Impact:** moderate	• Communication with settings explaining research benefits. • Ongoing reminders. • Flexibility in timings of visits. • Close liaison with OxEd to support IPE engagement. • Online data collection where possible to minimise burden. • Incentive payment to settings and individual practitioners to cover for their time spent in taking part in IPE interviews.

Timeline

Table 6: Timeline of activities for the pre-trial delivery and the formative evaluation

Dates	Activity	Staff responsible/leading
Jan-Feb 2024	Recruitment of settings	OxEd
Feb-March 2024	Completion of training Administration of LanguageScreen LanguageScreen data shared with NFER IPE sample selection and onboarding Pre-populated session delivery logs shared with IPE settings Child attendance patterns and targeted group selection collected from settings Delivery begins (staggered)	Settings Settings OxEd NFER NFER NFER NFER Settings
Apr-Jun 2024	IPE activities (interviews and observations) Session delivery logs returned at 5-week intervals Informal formative feedback via email and meetings	NFER Settings NFER
Jun-Jul 2024	Analysis and reporting Presentation of findings to the EEF and OxEd Summary of findings and financial incentives shared with participating settings	NFER & RAND NFER & RAND NFER & OxEd

Table 7: Timeline of activities for the effectiveness trial

Dates	Activity (organisations responsible/leading)
August 2023 – February 2024	Project set-up (NFER, OxEd and EEF) Develop and finalise recruitment documents (NFER, OxEd and EEF)
January 2024- Mid-June 2024	Setting recruitment via EOI and MoUs (OxEd) Check setting eligibility criteria (OxEd and EEF, ongoing)
Mid-June 2024	Recruitment details shared with NFER (OxEd)
July 2024	Send keep-in-touch emails to settings with instructions for baseline data collection (NFER)
September 2024	Baseline data collection activities (NFER) • Phase 1 baseline data collection – BaU survey and Practitioner Outcomes survey • Distribute parent information sheet and withdrawal letters via settings • Phase 2 baseline data collection – Child Data • Upload child data on LanguageScreen and share QR code with settings

Dates	Activity (organisations responsible/leading)
October 2024	• LanguageScreen Assessment completed in settings (administration by practitioners) • Monitor LanguageScreen completion (NFER, supported by OxEd)
November 2024	• Randomisation in two tranches and share group allocation with settings upon LanguageScreen completion (NFER) • Intervention settings receive baseline LanguageScreen data (NFER) • Intervention settings receive Targeted Group Selection Guidance (OxEd) • Selection of targeted intervention children (intervention settings)
Mid-November – December 2024	NELI Preschool Training in settings (OxEd and settings)
January 2025	• NELI Preschool Programme Delivery begins (settings with support from OxEd) • Intervention and control settings receive payment to cover staff time for completing baseline LanguageScreen assessments (OxEd) • Intervention settings receive payment to cover staff time to complete NELI Preschool online training (OxEd)
January – June 2025	• NELI Preschool Programme Delivery (settings with support from OxEd) • Weekly NELI Preschool session delivery logs (NFER shares templates for settings to complete) • Selection of case study sample and ongoing IPE activities (NFER)
June – July 2025	• NELI Preschool Lead survey, BaU survey (control group) and Practitioner Outcomes survey (NFER shares these for settings to complete) • Endline assessments for LanguageScreen and RAPT (NFER Test Administrators)
September 2025	• Assessment feedback reports sent to settings (all LanguageScreen results and RAPT data as feedback showing them vs other settings in the trial) (NFER) • All settings receive payment for completing relevant trial activities, and liaise with control settings on final payment or order NELI Preschool (OxEd)
November 2025	First draft report (NFER)
Dec 2025 – March 2026	Report revisions and final publication of the main report (NFER and EEF)
April 2026	NPD application submitted for EYFSP (Longitudinal evaluation) (NFER)
May – July 2026	Light-touch IPE survey (Longitudinal evaluation) (NFER)

Dates	Activity (organisations responsible/leading)
February 2027	First draft addendum report (NFER)
March-June 2027	Report revisions and final publication of the addendum report (NFER and EEF)
July-September 2027	Archive data in EEF data archive (NFER)

References

Bonetti, S. and Blanden, J. (2020) *Early years workforce qualifications and children's outcomes: an analysis using administrative data*. Available at: https://www.nuffieldfoundation.org/wp-content/uploads/2020/12/Early-years-workforce-qualifications-and-childrens-outcomes.pdf (Accessed: 15 March 2024).

Burgoyne, K. *et al.* (2018) *Parents And Children Together (PACT): evaluating a parent-delivered oral language enrichment programme for pre-school children*. Available at: https://www.nuffieldfoundation.org/wp-content/uploads/2019/11/Hulme204179320-20Main20public20report_PACT.pdf (Accessed: 14 March 2024).

Chen, J.-Q. and McCray, J. (2013) *A survey study of early childhood teachers' beliefs and confidence about teaching early math*. Available at: https://earlymath.erikson.edu/wp-content/uploads/2013/12/2013-1-Chen-McCray-Survey-Study-of-Early-Childhood-Teachers-Beliefs-and-Confidence-about-Teaching-Early-Math.pdf (Accessed: 14 March 2024).

Chow, J.C. and Ekholm, E. (2019) 'Language domains differentially predict mathematics performance in young children', *Early Childhood Research Quarterly*, 46(Quarterly), pp. 179–186. Available at: https://doi.org/10.1016/j.ecresq.2018.02.011.

Dimova, S. *et al.* (2020) *Nuffield Early Language Intervention: evaluation report*. Available at: https://educationendowmentfoundation.org.uk/public/files/Nuffield_Early_Language_Intervention.pdf (Accessed: 26 June 2023).

Eadie, P. *et al.* (2022) 'Domains of quality in early childhood education and care: A scoping review of the extent and consistency of the literature', *Educational Review*, pp. 1–30. Available at: https://doi.org/10.1080/00131911.2022.2077704.

EEF (2019) *Lessons learnt from EEF early years trials: recommendations for evaluators*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-syntheses/EY_lessons_learnt.pdf?v=1630582495 (Accessed: 14 March 2024).

EEF (2022) *Statistical analysis guidance for EEF evaluations*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1709857308 (Accessed: 15 March 2024).

EEF (2023a) *Cost evaluation guidance for EEF evaluations*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1710150030 (Accessed: 13 May 2024).

EEF (2023b) *Teaching and learning: Early Years Toolkit guide*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/toolkit/Toolkit_guide_v1.2_-_2023.pdf?v=1687895771 (Accessed: 24 May 2024).

Fernet, C. *et al.* (2008) 'The Work Tasks Motivation Scale for Teachers (WTMST)', *Sage Journals*, 16(2), pp. 256–279. Available at: http://journals.sagepub.com/doi/10.1177/1069072707305764 (Accessed: 14 March 2024).

Field, F. (2010) *The Foundation Years: preventing poor children becoming poor adults*. Available at: http://www.complexneeds.org.uk/modules/Module-1.1-Understanding-the-child-development-and-difficulties/All/downloads/m01p040d/the_foundation_years_preventing_poor_children_becoming_poor_adults.pdf (Accessed: 15 March 2024).

Fletcher-Wood, H. and Zuccollo, J. (2020) *The effects of high-quality professional development on teachers and students*. Available at: https://epi.org.uk/wp-content/uploads/2020/02/EPI-Wellcome_CPD-Review__2020.pdf (Accessed: 15 March 2024).

Fricke, S. *et al.* (2012) 'Efficacy of language intervention in the early years', *Journal of Child Psychology and Psychiatry*, 54(3), pp. 280–290. Available at: https://doi.org/10.1111/jcpp.12010.

Fricke, S. *et al.* (2017) 'The efficacy of early language intervention in mainstream school settings: a randomized controlled trial', *Journal of Child Psychology and Psychiatry*, 58(10), pp. 1141–1151. Available at: https://doi.org/10.1111/jcpp.12737.

Gonzales, C.R. *et al.* (2021) 'The Head-Toes-Knees-Shoulders Revised (HTKS-R): development and psychometric properties of a revision to reduce floor effects', *Early Childhood Research Quarterly*, 56, pp. 320–332. Available at: https://doi.org/10.1016/j.ecresq.2021.03.008.

Goodman, A. and Gregg, P. (2010) *Poorer children's educational attainment: how important are attitudes and behaviour?*, *Joseph Rowntree Foundation*. Available at: https://www.jrf.org.uk/care/poorer-childrens-educational-attainment-how-important-are-attitudes-and-behaviour (Accessed: 15 March 2024).

Groom, M., Brown, E.R. and Lymperis, L. (2023) *Nuffield Early Language Intervention: addendum report*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/projects/EEF_NELI_addendum_report-final_OnWebsite.pdf?v=1684239545 (Accessed: 27 June 2023).

Guo, G. and Harris, K.M. (2000) 'The mechanisms mediating the effects of poverty on children's intellectual development', *Demography*, 37(4), pp. 431–447. Available at: https://doi.org/10.1353/dem.2000.0005.

Hart, B. and Risley, T.R. (1995) *Meaningful differences in the everyday experience of young american children*. Baltimore: Brookes Publishing. Available at: https://products.brookespublishing.com/Meaningful-Differences-in-the-Everyday-Experience-of-Young-American-Children-P14.aspx (Accessed: 15 March 2024).

Haux, T. *et al.* (2022) *Early years recruitment, retention and business planning during the Coronavirus pandemic research report: a summary of findings*. Available at: https://assets.publishing.service.gov.uk/media/62695faae90e0746cd3f0f9c/Early_years_recruitment__retention_and_business_planning_during_the_Coronavirus_pandemic.pdf (Accessed: 15 March 2024).

Hjetland, H.N. *et al.* (2020) 'Preschool pathways to reading comprehension: A systematic meta-analytic review', *Educational Research Review*, 30(3), p. 100323. Available at: https://doi.org/10.1016/j.edurev.2020.100323.

Hoff, E. (2003) 'The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech', *Child Development*, 74(5), pp. 1368–1378. Available at: https://doi.org/10.1111/1467-8624.00612.

Hornburg, C.B., Schmitt, S.A. and Purpura, D.J. (2018) 'Relations between preschoolers' mathematical language understanding and specific numeracy skills', *Journal of Experimental Child Psychology*, 176, pp. 84–100. Available at: https://doi.org/10.1016/j.jecp.2018.07.005.

House of Commons Education Committee (2019) *Tackling disadvantage in the early years: ninth report of session 2017–19*. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmeduc/1006/1006.pdf (Accessed: 15 March 2024).

Howard, S.J. *et al.* (2020) 'Measuring interactional quality in pre-school settings: introduction and validation of the Sustained Shared Thinking and Emotional Wellbeing (SSTEW) scale', *Early Child Development and Care*, 190(7), pp. 1017–1030. Available at: https://doi.org/10.1080/03004430.2018.1511549.

Hughes, C. *et al.* (2015) 'Measuring the foundations of school readiness: Introducing a new questionnaire for teachers. The Brief Early Skills and Support Index (BESSI)', *British Journal of Educational Psychology*, 85(3), pp. 332–356. Available at: https://doi.org/10.1111/bjep.12076.

Hulme, C. *et al.* (2015) 'The foundations of literacy development in children at familial risk of dyslexia', *Psychological Science*, 26(12), pp. 1877–1886. Available at: https://doi.org/10.1177/0956797615603702.

Hulme, C. *et al.* (2024) *LanguageScreen: The development, validation, and standardization of an automated language assessment app*. American Speech-Language-Hearing Association. Available at: https://pubs.asha.org/doi/10.1044/2024_LSHSS-24-00004 (Accessed: 14 June 2024).

Hutchinson, J., Reader, M. and Akhal, A. (2020) *Education in England: annual report 2020*. Available at: https://epi.org.uk/wp-content/uploads/2020/09/EPI_2020_Annual_Report_.pdf (Accessed: 15 March 2024).

Huttenlocher, J. *et al.* (1991) 'Early vocabulary growth: relation to language input and gender', *Developmental Psychology*, 27(2), pp. 236–248. Available at: https://doi.org/10.1037/0012-1649.27.2.236.

Law, J. *et al.* (2017) *Early language development: needs, provision, and intervention for preschool children from socioeconomically disadvantaged backgrounds*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/guidance/Law_et_al_Early_Language_Development_final.pdf?v=168434613 (Accessed: 15 March 2024).

Liberman, I.Y., Shankweiler, D. and Liberman, A.M. (1989) 'The alphabetic principle and learning to read', in *Phonology and reading disability: Solving the reading puzzle*. Ann Arbor, MI, US: The University of Michigan Press (International Academy for Research in Learning Disabilities monograph series, No. 6.), pp. 1–33. Available at: https://files.eric.ed.gov/fulltext/ED427291.pdf (Accessed: 15 March 2024).

Marulis, L.M. and Neuman, S.B. (2010) 'The effects of vocabulary intervention on young children's word learning: A meta-analysis', *Review of Educational Research*, 80(3), pp. 300–335. Available at: https://doi.org/10.3102/0034654310377087.

McMullin, C. (2021) 'Transcription and Qualitative methods: implications for third sector research', *VOLUNTAS*, 34(1), pp. 140–153. Available at: https://doi.org/10.1007/s11266-021-00400-3.

Murray, A. and Egan, S.M. (2014) 'Does reading to infants benefit their cognitive development at 9-months-old? An investigation using a large birth cohort survey', *Child Language and Therapy*, 30(3), pp. 303–315. Available at: https://doi.org/10.1177/0265659013513813.

Nolan, A. and Molla, T. (2017) 'Teacher confidence and professional capital', *Teaching and Teacher Education*, 62, pp. 10–18. Available at: https://doi.org/10.1016/j.tate.2016.11.004.

OECD (2020) *Quality early childhood education and care for children under age 3: results from the Starting Strong Survey 2018*. Available at: https://www.oecd.org/education/quality-early-childhood-education-and-care-for-children-under-age-3-99f8bc95-en.htm (Accessed: 15 March 2024).

Paull, G. and Popov, D. (2019) *The role and contribution of maintained nursery schools in the early years sector in England*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/912995/Frontier_Economics_MNS_report_REVISED_v2.pdf (Accessed: 15 March 2024).

Renfrew, C. (2019) *Action Picture Test*. Oxford: Routledge.

Roulstone, S. *et al.* (2011) *Investigating the role of language in children's early educational outcomes*. DFE-RR134. Available at: https://assets.publishing.service.gov.uk/media/5a7b67a5e5274a319e77f135/DFE-RR134.pdf (Accessed: 15 March 2024).

Sampson, R.J., Sharkey, P. and Raudenbush, S.W. (2008) 'Durable effects of concentrated disadvantage on verbal ability among African-American children', *Proceedings of the National Academy of Sciences*, 105(3), pp. 845–852. Available at: https://doi.org/10.1073/pnas.0710189104.

Scarborough, H.S. (2009) 'Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice', in F. Fletcher-Campbell, J. Soler, and G. Reid (eds) *Approaching difficulties in literacy development: assessment, pedagogy, and programmes*. London: SAGE Publications Ltd., pp. 23–39. Available at: https://www.researchgate.net/publication/306200560_Connecting_early_language_and_liter

acy_to_later_reading_disabilities_Evidence_theory_and_practice (Accessed: 15 March 2024).

Semel, E., Wiig, E. and Secord, W. (2006) *Child Evaluation of language fundamentalspreschool UK*. 2nd edn. Oxford: Pearson Assessment.

Sibieta, L., Kotecha, M. and Skipp, A. (2016) *Nuffield early language intervention*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/EEF_Project_Report_Nuffield_Early_Language_Intervention.pdf?v=1710077612 (Accessed: 14 March 2024).

Sim, M. *et al.* (2018) *Teaching, pedagogy and practice in early years childcare: an evidence review*. Available at: https://epi.org.uk/wp-content/uploads/2018/08/teaching-pedagogy-and-practice-in-early-years-childcare_Aug2018.pdf (Accessed: 27 June 2023).

Siraj, I., Kingston, D. and Melhuish, E. (2023) *The sustained shared thinking and emotional well-being (SSTEW) scale: supporting process quality in early childhood*. London: David Fulton Publishers. Available at: https://www.routledge.com/The-Sustained-Shared-Thinking-and-Emotional-Well-being-SSTEW-Scale-Supporting-Process-Quality-in-Early-Childhood/Siraj-Kingston-Melhuish/p/book/9781032460499 (Accessed: 15 March 2024).

Sirin, S.R. (2005) 'Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research', *Review of Educational Research*, 75(3), pp. 417–453. Available at: https://doi.org/10.3102/00346543075003417.

Smith, A. *et al.* (2023) *Impact evaluation of Nuffield Early Learning Intervention (NELI) wave two*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/NELI-wave-2-scale-up-impact-evaluation-report-FINAL_2023-09-13-094553_bzhh.pdf?v=1717592765 (Accessed: 14 June 2024).

Snowling, M. *et al.* (1994) 'The development of phonological skills', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 346(1315), pp. 21–27. Available at: https://doi.org/10.1098/rstb.1994.0124.

Weisleder, A. and Fernald, A. (2013) 'Talking to children matters: early language experience strengthens processing and builds vocabulary', *Psychological Science*, 24(11), pp. 2143–2152. Available at: https://doi.org/10.1177/0956797613488145.

Welbourne, S.R. *et al.* (2011) 'The role of plasticity-related functional reorganization in the explanation of central dyslexias', *Cognitive Neuropsychology*, 28(2), pp. 65–108. Available at: https://doi.org/10.1080/02643294.2011.621937.

West, G. *et al.* (2021) 'Early language screening and intervention can be delivered successfully at scale: evidence from a cluster randomized controlled trial', *Journal of Child Psychology and Psychiatry*, 62(12), pp. 1425–1434. Available at: https://doi.org/10.1111/jcpp.13415.

West, G. *et al.* (2023) *An efficacy trial of the Nuffield Early Language Intervention in Preschool (NELI Preschool)*. Available at: https://doi.org/10.1186/ISRCTN29838552.

Woollams, A.M., Halai, A. and Lambon Ralph, M.A. (2018) ‘Mapping the intersection of language and reading: the neural bases of the primary systems hypothesis’, *Brain Structure & Function*, 223(8), pp. 3769–3786. Available at: https://doi.org/10.1007/s00429-018-1716-z.

Appendix A: Changes since the previous evaluation

Appendix Table 1: Changes since the previous evaluation

Feature		Efficacy to effectiveness stage
Intervention	**Intervention content**	In the current evaluation, the delivery team will provide comprehensive guidance on selecting children for the targeted component. In addition to this, settings will be required to complete the Narrative Tracker template which was not part of the efficacy trial.
	Delivery model	There are no changes to the delivery mechanism, i.e., the developers will lead intervention delivery in this trial.
	Intervention duration	There are no changes in the duration of delivery between the efficacy and effectiveness trial.
Evaluation	**Eligibility criteria**	The efficacy trial involved maintained nursery settings whereas this trial will aim to have at least one-third maintained settings, one-third PVI settings and the remaining third to be of any setting type. Settings in the previous trial were from seven geographical areas and the settings in the current evaluation will be from the LAs in the nine SPH areas. The current evaluation will also exclude settings that took part in the efficacy trial, those taking part in other SPH funded programmes (in the academic year 2024-25), other EEF-funded evaluation . The target age of the children will be same as the previous evaluation although the delivery team will provide additional Targeted Group Selection Guidance which was not part of the previous evaluation.
	Level of randomisation	This evaluation will also randomise settings like the previous trial.
	Outcomes and baseline	Outcome measures are different across the two evaluations. The primary outcome measure for the efficacy trial was a latent variable from eight subtests across LanguageScreen, CELF and RAPT, whereas the current evaluation will only use six subtests from LanguageScreen and RAPT to derive the primary outcome. The baseline for the previous evaluation was a latent variable from eight subtests across LanguageScreen, CELF and RAPT whereas the current evaluation will only use LanguageScreen as a baseline measure. In the previous evaluation, LanguageScreen was administered as a screening test to identify children with the weakest language skills prior to randomisation, while CELF and RAPT were administered after randomisation. The current evaluation will also utilise LanguageScreen as a screening test to identify children with the weakest language skills and as a baseline measure, and will also be administered prior to randomisation.
	Control condition	In both evaluations, the control group will continue with their usual practice. The efficacy trial was a waiting list design whereas control group settings in the current evaluation will receive £1000 which can either be a financial payment or be used to purchase NELI Preschool programme at the end of the trial.

Appendix B: Practitioner outcomes measures for use in IPE

Practitioner motivation measure[34]

1. To what extent are you motivated to explicitly teach the meaning of new vocabulary to 3-4-year-old children in your nursery?
    a. Very motivated *(3 points)*
    b. Somewhat motivated *(2 points)*
    c. Not at all motivated *(1 point)*
    d. I'm not sure
2. Do you explicitly teach the meaning of new vocabulary to 3-4-year-old children in your nursery?
    a. Yes
    b. No
    c. I'm not sure

Ask Q3 if answer to Q2 is 'Yes'.

3. Why do you explicitly teach the meaning of new vocabulary to 3-4-year-old children in your nursery (Please rank the reasons you select in order of importance, with 1 being the most important)
    a. I enjoy doing it *(intrinsic motivation)*
    b. I think it is important for the children's intellectual development *(identified regulation)*
    c. I feel like I should, even if I don't always see the benefit *(introjected regulation)*
    d. My manager has told me to do it *(external regulation)*
    e. I don't know *(amotivation)*

Repeat the above for:

- Providing children with opportunities to practise new words in a range of contexts
- Providing children with opportunities to explore the narrative structure of stories through dialogic reading (reading 'in conversation with' the child)

[34] Note that text in italics will not be seen by the respondent.

- Providing additional scaffolded support to those children with the weakest language skills

Practitioner confidence measure[35]

I am confident in my knowledge of:

1. The level of language skills of each child that enters my class
    a. Strongly agree *(5 points)*
    b. Agree *(4 points)*
    c. Neither agree nor disagree *(3 points)*
    d. Disagree *(2 points)*
    e. Strongly disagree *(1 point)*

Likert scale repeated for questions 2-11

2. Reasonable goals for 3–4-year-olds in relation to the development of their language skills
3. The best practices and strategies for helping 3-4-year-olds to develop their language skills
4. National standards for language skills for 3-4-year-olds
5. The best ways to assess the language skills (expressive language and understanding of language) of 3-4-year-olds throughout the year

I am confident in my ability to:

1. Gauge the level of language skills of 3-4-year-olds in my class
2. Incorporate regular opportunities to learn and/or practise language skills into common preschool situations (such as art or dramatic play)
3. Plan activities to help 3-4-year-olds to develop their language skills
4. Support the language development of 3-4-year-olds when they make spontaneous comments/discoveries
5. Help 3-4-year-olds to navigate their confusion when they are developing their language skills
6. Translate assessment or screening results into curriculum plans on both a group and individual basis

[35] Note that text in italics will not be seen by the respondent.