# Impact Evaluation of Nuffield Early Language Intervention (NELI) Wave Two

Evaluation Report

September 2023

Andrew Smith, Ruth Staunton, Aarti Sahasranaman, Jack Worth

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

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

# Contents

# About the evaluator

This independent impact evaluation of the Nuffield Early Language Intervention (NELI) wave 2 scale-up was undertaken by the National Foundation for Educational Research (NFER). The evaluation team was led by Jack Worth, Lead Economist.

The NFER is the leading independent provider of education research and holds the status of Independent Research Organisation (IRO) from UK Research and Innovation (UKRI). Our unique position and approach deliver evidence-based insights designed to enable education policymakers and practitioners to take action to improve outcomes for children and young people. Our key topic areas are accountability, assessment, classroom practice, education to employment, social mobility, school funding, school workforce and systems and structures. As a not-for-profit organisation, we re-invest any surplus funds into self-funded research and development to further contribute to the science and knowledge of education research.

**Contact details:**

# Acknowledgements

## ONS SRS publication status and statistical results

# Executive summary

## The project

The Nuffield Early Language Intervention (NELI) is designed to improve the language skills of reception pupils (aged four to five) and involves scripted individual and small group language teaching sessions delivered by school staff, usually teaching assistants (TAs). The 20-week intervention consists of two 15-minute individual sessions and three 30-minute small group sessions each week, delivered to the three to six pupils with the weakest language skills. The sessions focus on improving pupil's vocabulary, active listening, and narrative skills and in the second ten weeks include a small additional element (three minutes only) focusing on developing phonological awareness and letter–sound knowledge.

As part of the Department for Education's (DfE) efforts to support education recovery as a consequence of the Covid-19 pandemic, £9 million—and then an additional £8 million—was provided to make NELI available to state-funded schools with reception pupils. In the 2020/2021 academic year, approximately 6,500 schools registered to receive NELI (wave one) and in 2021/2022 about 4,000 additional schools registered (wave two). An independent implementation and process evaluation (IPE) of both waves of the NELI programme has been completed by RAND Europe and the reports are publicly available. Although an impact evaluation of the wave one roll-out was planned, school closures mandated due to the Covid-19 pandemic prevented this evaluation from being commissioned.

This study, conducted by the NFER, evaluated the impact of NELI delivered at national scale (as part of the wave two national implementation) on pupil's oral language skills using a quasi-experimental Fuzzy Regression Discontinuity (FRD) design. Five hundred forty-six schools (19,212 pupils) agreed to take part in the evaluation and had completed baseline testing with their reception pupils. Not all schools that agreed to take part returned all data necessary for the evaluation; the analysis therefore used data from 356 schools (10,759 pupils). Schools registered to receive NELI in wave two between May and July 2021; there was a further round of recruitment in October 2021. Delivery of NELI to pupils took place from January 2022. Schools were invited to take part in the evaluation in June 2022 and data collection was completed by September that year.

*Table 1: Key conclusions*

| Key conclusions |
|---|
| 1.   Pupils who received the NELI programme made the equivalent of four additional months' progress in language skills, on average, compared to pupils who did not receive NELI. This result has a moderate to high security rating. |
| 2.   Subgroup analysis found pupils eligible for free school meals (FSM) who received the NELI programme made an additional seven months' progress in language skills, on average, compared to pupils eligible for FSM who did not receive the programme. |
| 3.   Subgroup analysis found pupils with English as an additional language (EAL) who received the NELI programme made an additional four months' progress in language skills, on average, compared to pupils with EAL who did not. However, the sample of pupils for this subgroup was small and potentially not sufficient to confidently interpret the level of impact. |
| 4.   Exploratory analysis highlighted that the effect of receiving NELI was greater for pupils whose TA delivered more of the programme's group sessions compared with pupils whose TAs delivered fewer group sessions. |
| 5.   Exploratory analysis found that the effect of receiving NELI was greater for pupils in schools where more than 50% of TAs had attended at least one training session compared to pupils in schools where fewer than 50% of TAs had attended between zero and three training sessions. |

## EEF security rating

The headline finding has a moderate to high security rating. This was an impact evaluation of NELI, which tested whether the programme worked under everyday conditions, outside of a trial, in a very large number of schools during a national roll-out of the programme to support education recovery. The evaluation used a quasi-experimental Fuzzy Regression Discontinuity (FRD) design, and the primary outcome was well-powered. Thirty-one percent of the pupils who started the evaluation were not included in the final analysis because their school did not implement outcome testing. There were also some imbalances in characteristics (for example, pupils with SEND or EAL) between the pupils who received the programme and those who did not, which were considered when assigning the security rating.

## Additional findings

The logic model developed by RAND Europe as part of the linked process evaluation identified 11 theorised outcomes of the DfE-funded scale-up of NELI. This impact evaluation of wave two of the national scale-up focused on one of these outcomes: pupil's immediate improvements in language skills. Pupils who received the NELI programme made, on average, four additional months' progress than those who did not receive the programme. This is our best estimate of impact, which has a medium to high security rating. As with any study, there is always some uncertainty around the result: the range of impacts for this programme include smaller positive effects of up to two months' additional progress and larger positive effects of up to six months' additional progress.

This national scale-up evaluation is the final stage in the EEF's 'evaluation pipeline' (EEF, 2023) following a pilot study (Fricke et al., 2013, funded by Nuffield Foundation), an efficacy trial (Sibieta, Kotecha, and Skipp, 2016), and an effectiveness trial (Dimova et al., 2020). The effectiveness trial found pupils made approximately three months of progress in oral language skills when they received NELI and language was measured using a latent variable of four standardised assessments. The language screening tool linked with NELI, LanguageScreen, was a secondary outcome and pupils were found to make an additional four months' progress in oral language skills when measured by this assessment. This study used LanguageScreen to analyse impact on pupil's language outcomes and reports a similar finding to that of the effectiveness trial suggesting the average impact of the programme was maintained when delivered at national scale and when an online training and support model was adopted.

Subgroup analyses were undertaken to look at the impact of NELI on pupils with EAL and pupils eligible for FSM. The former made an additional four months' progress in language skills, on average, compared to pupils with EAL who did not receive NELI. The sample for this subgroup was small and potentially not sufficient to confidently interpret the impact finding. However, the result is consistent with findings from the effectiveness trial which found the programme was, on average, equally beneficial for pupils with EAL as for all pupils who received NELI. Pupils eligible for FSM who received the NELI programme made an additional seven months' progress in language skills, on average, compared to FSM-eligible pupils who did not receive the programme. While the effectiveness trial did not estimate the impact of NELI on pupils eligible for FSM due to limited permissions to link data, the longitudinal follow up of the pupils who received NELI as part of the effectiveness trial (Groom et al., 2023) did find evidence of a positive effect of NELI on FSM-eligible pupil's language outcomes based on an exploratory analysis.

During delivery of NELI, schools completed brief surveys sharing information on their use of the programme. Analysis implemented as part of the independent process evaluation of the scale-up found that on average schools suggested they had delivered 30 NELI group sessions (or 10 weeks) by July 2022 with wide variation in the number of group sessions delivered. The sample of schools in this impact evaluation on average reported they delivered 33 group sessions (or 11 weeks). Analysis highlighted that the effect of receiving NELI was greater for pupils whose TAs delivered more of the group sessions compared with pupils whose TAs delivered fewer sessions. Therefore, even when a lower than intended dosage of NELI was delivered, there was, on average, a positive impact on pupils' language outcomes.

*Table 2: Summary of impact on primary outcome*

| Group/outcome | Effect size (95% confidence interval) | Estimated months' progress | EEF security rating | No. of pupils | P value | EEF cost rating |
|---|---|---|---|---|---|---|
| All pupils LanguageScreen standardised score | 0.29 (0.12, 0.47) | 4 | 🔒🔒🔒🔓🔓 | 4476 | <0.001 | Not evaluated |
| FSM LanguageScreen standardised score | 0.56 (0.14, 0.99) | 7 | n/a | 688 | 0.009 | Not evaluated |
| EAL LanguageScreen standardised score | 0.29 (-0.00, 0.62) | 4 | n/a | 853 | 0.079 | Not evaluated |

# Introduction

## Background

The Covid-19 pandemic disrupted language and literacy development in early childhood (Francis, 2022) with potential downstream effects on educational attainment and employability into adulthood. Research undertaken by the University of York, the Education Policy Institute, and the National Institute of Economic and Social Research to examine the impact of Covid-19 on primary school starters suggested that 96% of schools surveyed reported being concerned about their pupils' language and communication skills due to the pandemic (Bowyer-Crane et al., 2021). Although this disruption impacted all pupils, pupils with English as an additional language (EAL) appeared to experience disproportionate disruptions to their literacy learning (Tracey et al., 2022).

There is good evidence that the Nuffield Early Language Intervention (NELI) improves the oral language skills of pupils. An EEF-funded individually randomised controlled efficacy trial conducted in 34 early years settings in 2016 demonstrated that two versions of the NELI programme—a 30-week programme starting at the end of nursery and continuing in the reception year and a 20-week programme offered in reception only—had positive impacts on the language skills of pupils. Children receiving the more expensive 30-week version experienced the equivalent of about four months of additional progress whereas those receiving the 20-week version experienced about two months of additional progress (Sibieta et al., 2016; Fricke et al., 2013). This difference, however, was not found to be statistically significant, which may indicate that there was in fact no difference between the two interventions or that the trial was not powered to detect one. These positive findings were replicated in a larger EEF-funded cluster randomised controlled effectiveness trial in 193 schools where only the 20-week version of the programme was tested (Dimova et al., 2020; West et al., 2021). Children who received the NELI programme made, on average, an additional three months' progress in language skills compared to children who did not receive NELI. Children with EAL who received NELI also made the equivalent of three additional months' progress in language skills compared to EAL children who did not receive NELI.

Recognising the evidence in support of NELI to narrow the gap in language skills for disadvantaged pupils, the Department for Education (DfE) committed £9 million to make the 20-week NELI programme available to state-funded primary schools with reception pupils (DfE and Ford, 2021) as part of the government's Covid-19 recovery efforts. In wave one, rolled out in 2020/2021, NELI was offered at no cost to around 6,500 primary schools. Although an impact evaluation of the wave one roll-out was planned, school closures mandated because of the Covid-19 pandemic prevented this evaluation from being commissioned. Given the demand for the NELI programme in wave one and the continued impact of the second and third Covid-19 lockdowns in England on access to suitable programmes for nursery children who were expected to progress to reception in the following academic year, the DfE committed a further £8 million for wave two of the national roll-out of NELI. In wave two, rolled out in 2021/2022, access to NELI was expanded to approximately a further 4,000 primary schools at no cost to schools. An independently commissioned implementation and process evaluation (IPE) of both waves of the NELI programme has been completed and the reports are now publicly available (Disley et al., 2023a; Disley et al., 2023b).

While NELI demonstrated a positive impact in randomised controlled trials (RCTs), its impact when delivered at national scale was not known. The aim of this scale-up evaluation, therefore, was to assess the impact of NELI in everyday conditions, outside of a trial. We therefore adopted a quasi-experimental (QED) approach using a Fuzzy Regression Discontinuity (FRD) design to assess the impact of the NELI wave two scale-up on children's oral language outcomes. The implementation of the national scale-up, and selection of pupils to receive NELI implied a data structure in support of this approach (this is described in detail in the Participant Selection and Fuzzy Regression Discontinuity sections).

## Intervention

This was a scale-up impact evaluation of wave two of the national roll-out of the Nuffield Early Language Intervention. The NELI logic model is described in detail in the report of the IPE of the NELI wave two scale-up (Disley et al., 2023b).

**Name**

Nuffield Early Language Intervention (NELI).

**Why (Theory/Rationale)**

Strong oral language skills in children lay the foundation for the development of literacy and numeracy skills (Duff et al., 2015; Law et al., 2013; Roulstone et al., 2011), essential for subsequent academic attainment and eventually employability in adulthood (Feinstein and Duckworth, 2006). The link between the development of language and literacy skills is well established (Whitehurst and Lonigan, 1998; Scarborough et al., 2009). It is also widely accepted that the process of language learning begins early in the preschool period and that support for those with poor language skills should be provided as early as possible (Whitehurst and Lonigan, 1998; Scarborough et al., 2009; Hulme and Snowling, 2015). There is also a policy imperative to intervene early as children from disadvantaged backgrounds are disproportionately more likely to suffer from language difficulties as they enter school (Fernald et al., 2013). These disparities tend to widen throughout the school experience, ultimately hampering efforts to improve social mobility.

NELI aims to improve the oral language ability of reception pupils with relatively poor spoken language skills with a longer-term goal of improving pupils' reading comprehension (as literacy builds on oral language skills). A significant body of evidence suggests that NELI has a positive impact on the language skills of children receiving the programme (Sibieta et al., 2016; Dimova et al., 2020; Bowyer-Crane et al., 2008; Fricke, et al., 2013; 2017; West et al., 2021). The programme has also been shown to be effective in improving the language skills of children with EAL, a group thought to be impacted by Covid-19 due to the reduced opportunity to hear and speak English. This evaluation assessed the impact of NELI when it was delivered at scale.

Based on the NELI implementation logic model (Dimova et al., 2020), the key activities of the intervention include:

- delivery of training to TAs and teachers along with continued remote support offered through the course of intervention delivery;

- the 20-week intervention focusing on vocabulary, listening, and narrative skills and, in the final ten weeks, also on phonological awareness delivered in the form of group and individual sessions; and

- support for teachers with screening and identifying eligible pupils for the intervention.

TAs' and teachers' knowledge of language and teaching of language skills is expected to improve through participation in training and because of the support received. By receiving NELI sessions, pupils' oral language and early reading skills are expected to improve, ultimately leading to the potential longer-term outcome of improved reading comprehension.

**Who (Recipients)**

Reception pupils (four to five years old) with poor oral language skills in state-funded primary schools that signed up to receive the NELI programme in the 2021/22 academic year were eligible. Participating schools were advised to screen *all* reception pupils using LanguageScreen, a short tablet-based standardised assessment to identify pupils with the weakest language skills in their cohort. LanguageScreen has been shown to identify language difficulties in children as young as three and a half up to eight years and can be used to identify suitable pupils for interventions such as NELI.[1] Although schools were encouraged to screen pupils using LanguageScreen, this was not mandatory and teachers or TAs could use their judgement to select pupils to receive NELI.

Recruitment was on a first-come-first-served basis with particular targeting of schools with (a) a higher proportion of free school meal (FSM) eligible pupils, (b) schools in local authorities in the bottom third of Ofsted ratings, and (c) schools in Opportunity Areas; 4,422 primary schools in England signed up for wave two of NELI and over 50% of

---

[1] https://oxedandassessment.com/languagescreen/

schools met the priority targeting requirements described above. Assuming 26.6 pupils per class, 1.5 forms per school,[2] and three to six NELI pupils per class, we estimate that up to 171,570 reception pupils were screened and between 19,350 and 38,700 pupils received the programme. The sample size for the scale-up impact evaluation was 548 schools and 19,212 pupils, consisting of schools agreeing to take part in the evaluation and assessing their reception pupils at baseline using LanguageScreen. Of these 548 schools, only 435 provided the NELI indicator for their pupils (by recording via the OxEd and Assessment Ltd (OxEd) school account which pupils had received the intervention). Pupil data from one school could not be matched to NPD data (in order to include additional pupil characteristics such as FSM and EAL status), meaning that in total data for 114 schools needed to be excluded from the analysis.

## What (Materials)

Participating schools received one NELI 'kit' per reception class. The materials were published by Oxford University Press and included two teacher handbooks with detailed lesson plans, two sets of A4 flashcards, and one Ted puppet to support session delivery.

## What (Procedures, activities and/or processes)

*Training model*

An online training programme developed by the University of Oxford was offered to nominated staff members—typically a teaching assistant (TA), reception class teacher, and the school's NELI lead—at participating schools. The training was delivered by OxEd via the FutureLearn platform. The core training content to be completed by the staff member delivering the programme (usually the TA) consisted of:

- three linked online courses expected to take ten to twelve hours to complete in a self-paced manner;
  - two courses to be completed before NELI delivery begins; and
  - a third short course, which could be completed halfway through programme delivery.

Course 1 covered the fundamentals of understanding and supporting language development and an overview of the NELI programme. Course 2 covered the details of delivering the programme in schools, including the screening process to identify eligible pupils, and Course 3 covered training on teaching letter sounds and phonological awareness. The reception class teacher was only expected to complete the first of the three linked courses although they had access to the remaining two courses should they wish to complete the training. The NELI lead (if different from the reception class teacher) also had the option to access the training if they wanted to familiarise themselves with the programme. Staff members also received ongoing online support throughout the course of the programme via the NELI delivery support hub.

All participating schools also received access to the LanguageScreen assessment tool. LanguageScreen has four subtests: expressive vocabulary, receptive vocabulary, listening comprehension, and sentence repetition and takes about ten minutes per pupil to administer. Schools were encouraged to use LanguageScreen to identify the appropriate pupils to receive the NELI programme. This typically included three to six pupils in each participating reception class with the lowest LanguageScreen scores. Schools were also provided support in the form of chat, email, and in-person support by OxEd to address any queries relating to the screening process.

## Who (Intervention providers / implementers)

NELI was developed by researchers led by Professors Charles Hulme and Maggie Snowling (now at the University of Oxford) including Silke Fricke and Claudine Bowyer-Crane (now at the University of Sheffield). It was funded by the Nuffield Foundation. The delivery of wave two of NELI is managed by a consortium of delivery partners led by Nuffield Foundation Education Ltd, a special purpose vehicle of the Nuffield Foundation comprising two agencies:

---

[2] 26.6 pupils per class—https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics; 1.5 forms per school—NELI scale-up Y2 impact evaluation ITT, February 2022.

- OxEd and Assessment Ltd—intervention developer

  OxEd is the developer of the LanguageScreen digital application, the developer of the online training and delivery support model hosted by it on the FutureLearn platform, the provider of all support for intervention delivery (chat, email and in-person support for schools for both screening and programme delivery) and of weekly reporting to all delivery partners. Delivery support included employing and supervising the NELI mentor team for the online training and delivery support hub. If any information was not available during the training sessions or via the delivery support hub, schools could contact NELI mentors via the interactive course or hub discussion functionality. NELI mentors are speech and language professionals who offer support to schools during training and delivery of the NELI programme. The team included several members contracted via Elklan, one of the delivery partners in wave one of the NELI scale-up.

- Oxford University Press—OUP publishes and distributes all materials that form part of the NELI kit provided to participating schools and provided targeted outreach during school recruitment.

Recruitment of schools for the scale-up impact evaluation was conducted by the NFER.

**How (model of delivery)**

The TA or early years educator delivering the programme was expected to conduct three 30-minute small group sessions consisting of three to six eligible pupils (in reception, aged four to five and with poor oral language skills) and two 15 minute individual sessions per week for each targeted child. While NELI sessions were delivered during normal classroom hours, pupils selected to receive the programme were taken out of classes. School staff were responsible for identifying which classes pupils would miss in order to participate in NELI sessions. The group and individual sessions were to be scheduled on different days.

**Where (Location of the intervention)**

There were no geographical limitations (within England) on recruitment of schools for wave two of NELI. Training was delivered online and could be completed at a time of the trainee's choice from any location. Within each participating school, it was recommended that the programme be delivered in a quiet area such as a classroom, library, staff room, or dining room.

**When and how much (Duration and dosage of the intervention)**

Recruitment of wave two schools took place from May to end of October 2021. Online training of staff was completed in the autumn term (September to December) of 2021. Following completion of training, intervention delivery was scheduled to begin in schools in January 2022 and to be completed by end of the summer term of 2022.

The 20-week intervention was divided into two blocks of ten weeks. In the first ten, the intervention focused on vocabulary, active listening, and narrative skills and—in the final ten weeks—also included an additional session element (three minutes only) focusing on letter sounds and phonological awareness as foundations of early literacy skills. There was no minimum recommended dosage but participating schools were encouraged to deliver all 20 weeks of the intervention.

We are not aware of any other interventions taking place at scale during the period of NELI delivery covered by this evaluation, although schools may have put additional support in place for specific pupils as part of their standard practice.

**Tailoring (Adaptation of the intervention)**

Tailoring was built into the programme through the bi-weekly individual sessions; staff were encouraged to refer to their notes and observations and to tailor these sessions to the needs of each pupil.

**Costs**

There was no cost to schools to receive the NELI programme as part of the scale-up: DfE funding covered all programme costs for participating schools.

Schools participating in the impact evaluation received up to £250 for completing the following evaluation activities:

- carrying out endpoint LanguageScreen assessments for the vast majority of reception pupils who received a baseline LanguageScreen assessment;

- providing NELI indicator data indicating which pupils received NELI; and

- participating in a final TeachNELI delivery survey administered by the Nuffield Foundation.

Regarding the latter: three delivery surveys were administered in February 2022, March 2022, and June 2022 to understand how NELI implementation was progressing in schools. The data collected included LanguageScreen use, number of pupils receiving NELI in schools, number of NELI groups in schools, number of group and individual sessions delivered, and approximate duration of sessions. Analysis of the delivery survey data is presented in Disley et al., 2023b.

## Evaluation objectives

This evaluation was undertaken to understand the impact of NELI, when delivered at scale, on children's early language outcomes. The research questions all focused on one primary outcome: pupils' oral language outcomes measured by the LanguageScreen standardised score post-intervention.

**RQ1**  What is the impact of NELI when delivered at national scale on pupils' oral language outcomes, as measured by LanguageScreen?

**RQ2**  What is the impact of NELI when delivered at national scale on FSM (everFSM) pupils' oral language outcomes, as measured by LanguageScreen?

**RQ3**  What is the impact of NELI when delivered at national scale on EAL pupils' oral language outcomes, as measured by LanguageScreen?

**RQ4**  How does the impact of NELI on pupils' oral language outcomes vary by dosage?

**RQ5**  How does the impact of NELI on pupils' oral language outcomes vary by training fidelity?

**Ethics and evaluation registration**

The evaluation was conducted in accordance with the NFER Code of Practice. All of the 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 Marketing Research Association), and SRA (the Social Research Association), among others. The NFER is committed to the highest ethical standards in all of its activities and ethical considerations are embedded in its detailed quality assurance processes.

When schools signed up to participate in wave two of the NELI roll-out, they were asked if they would like to hear more about the impact evaluation at a later date. Schools that had indicated their willingness to do so were contacted by the NFER with a memorandum of understanding (MoU) detailing the responsibilities of schools and the evaluator (see Appendix H for the MoU). Each school's headteacher gave permission for the school's participation in the impact evaluation by signing the MoU. Parents of all reception pupils in schools participating in the impact evaluation were sent a letter explaining the evaluation and given the opportunity to withdraw their child from data processing. Schools

notified OxEd, the delivery partner, of any pupil withdrawals. OxEd ensured that data for withdrawn pupils was not shared with the NFER. This QED was registered on 2 February 2023 with OSF Registries.[3]

## Data protection

**Data protection statement and GDPR compliance**

The NFER is registered with the Information Commissioner's Office for all of its research and other activities. It ensures that all projects comply with the seven principles of data protection legislation—the UK General Data Protection Regulation (GDPR) and the Data Protection Act, 2018. The NFER is ISO/IEC 27001 certified (GB17/872763) and holds Cyber Essentials Plus (IASEM-CEP-004922). It maintains a full Information Security Management Strategy (ISMS) including a Data Security Policy with which all staff are required to comply.

To carry out this evaluation it was necessary to use and share personal data about pupils—both those who received the NELI programme and those who did not—as well as key staff members including the school's headteacher, NELI lead, and bursar or equivalent. Additionally, we also received special category data (that is, ethnicity and special education needs) for pupils matched from the NPD. The NFER also received pseudonymised pupil data ('initial pupil-level data') in August 2022 from OxEd for the purpose of preliminary analyses to assess the viability of the FRD approach and to identify schools that met the requirements for incentive payments. All data sharing between the NFER and OxEd was carried out via a secure portal. For further information, please see the NELI impact evaluation privacy notice (see Appendix I).[4]

**Legal bases**

The legal basis for processing the personal data accessed and generated by this evaluation is covered by GDPR Article 6 (1) (f), which provides a justification when 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

The legal basis for processing pupils' special personal data is covered by GDPR Article 9 (2) (j), which provides a justification when 'processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) (as supplemented by section 19 of the 2018 Act) based on domestic law which shall be proportionate to the aim pursued, respect the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject'.

We do not believe this processing has caused damage or distress to the data subjects.

**Linking to the National Pupil Database and use of the Secure Research Service (SRS)**

OxEd securely submitted pupil-level data including pupil identifiers to the NPD team to be matched to the pupil data held in the NPD. The NFER also securely submitted school-level data to the NPD team. The NPD team created a composite dataset that included pupil-level and school-level data and matched pupil data from the NPD. The team also replaced pupil identifiers with the Pupil Matching Reference and removed any direct school identifiers. The NFER was only able to access this data within the SRS and any outputs were checked to ensure that no pupils can be identified. The project met the Office of National Statistics' 'five safes'.

**Rights and retention periods**

Parents could withdraw their child from the evaluation or from their data being processed. Were pupils withdrawn from the programme or evaluation, the NFER still used the evaluation data that the school had provided up to that point and

---

[3] Registration DOI: https://doi.org/10.17605/OSF.IO/5M8JF

[4] Also available at: https://www.nfer.ac.uk/media/4945/nelv_school_information_sheet.pdf

linked it to the NPD unless the parent indicated otherwise. If at any time parents wished to withdraw their child's data or have errors corrected in it, contact details were provided in the Privacy Notices for whom to contact about this.

As noted in the grant agreement, three months after the publication of this evaluation report, all of the pseudonymised matched data will be added to the EEF archive, which is managed by FFT on behalf of the EEF and hosted by the ONS. This will enable the EEF and other research teams to use the pseudonymised data as part of subsequent research through the ONS Approved Researcher Scheme, including analysing long term outcomes through the National Pupil Database. This data may also be linked to other research datasets for the purpose of educational research.

The NFER will securely delete any personal data relating to the evaluation one year after the publication of this final report.

**Data controller and processing roles**

The DfE is the data controller and makes decisions about how personal data is used in the evaluation. The EEF and OxEd and Assessment Ltd are the data processors and the NFER is the data sub-processor.

# Project team

| Name | Affiliation | Roles and responsibilities |
|---|---|---|
| Jack Worth | The NFER | Project Director—responsible for overall quality of project delivery |
| Aarti Sahasranaman | The NFER | Project Leader—responsible for day-to-day management of the project |
| Andrew Smith | The NFER | Design Lead—responsible for design of impact evaluation and QA of analysis |
| Kathryn Hurd | The NFER | Operations Lead—responsible for recruitment of NELI schools for impact evaluation, contacting schools to coordinate data collection, delivering incentive strategy |
| Jishi Jose | The NFER | Project Manager—responsible for overseeing day to day running of operations |
| Max Falinski | The NFER | Researcher—responsible for school communications and administering incentives |
| Ruth Staunton | The NFER | Study statistician—responsible for preliminary and main analyses |
| Ishbelle Norris | The NFER | Data preparation and QA of statistical analysis |
| Charles Hulme | OxEd | CEO of OxEd and intervention developer, led wave two NELI roll-out delivery |
| Gillian West | OxEd | Director of the wave two NELI evaluation at OxEd and responsible for data extraction strategy |
| Joe Lowe | OxEd | Evaluation data extraction and data sharing |
| Sarah Hearne | OxEd | Project manager—responsible for day to day management of wave two NELI evaluation at OxEd |

# Methods

## Evaluation design

*Table 3: Evaluation design*

| Design | | Fuzzy Regression Discontinuity |
|---|---|---|
| Unit of analysis | | Pupils in reception classes |
| Number of units to be included in analysis (intervention, comparison) | | 10,759 (2,329 intervention, 8,430 comparison)* |
| Primary outcome | Variable | Oral language skills |
| | Measure (instrument, scale, source) | LanguageScreen standardised score (endline measurement) |
| Secondary outcome(s) | Variable(s) | n/a |
| | Measure(s) (instrument, scale, source) | n/a |
| Baseline for primary outcome | Variable | Oral language skills |
| | Measure (instrument, scale, source) | LanguageScreen standardised score (baseline measurement) |
| Baseline for secondary outcome(s) | Variable | n/a |
| | Measure (instrument, scale, source) | n/a |

* These numbers indicate pupils for whom LanguageScreen endline assessments have been completed and who also have their NELI indicator status completed. Of the 19,212 pupils in the final pupil-level dataset (indicated by the corresponding study plan; Worth et al., 2022), 15,570 remained after schools (n = 113) that had not completed the NELI indicator of their pupils were removed (in addition to data for one school which could not be matched to the NPD). Of these, 4,811 were not retested at endline or were excluded for other reasons (see Figure 4, Participant Flow Diagram), leaving 10,759 for analysis.

This evaluation was designed to estimate the impact of the NELI wave two scale-up (in 2021/2022) on pupils' oral language skills. As implementation of the second wave of the programme had already begun when the evaluation was commissioned, and one of the objectives of the evaluation was to understand the impact of NELI delivered at scale under real-world conditions, the evaluation did not lend itself to a randomised controlled trial. We therefore adopted a quasi-experimental approach using FRD—this was contingent upon the way in which pupils were selected by schools to receive NELI (further details about the selection mechanism and the implied design choice follow in the two subsequent sections, Participant Selection and Fuzzy Regression Discontinuity). Attainment outcomes for pupils selected for the programme were compared with outcomes for a counterfactual group of pupils who were not selected (and who therefore received usual teaching). All schools enrolled in NELI wave two were asked to consent to receiving information about a potential impact evaluation and those that did so were invited to take part in the evaluation; the number of pupils included in the analysis was determined by those schools that agreed to do so and returned data.

LanguageScreen is a tablet-based standardised assessment (see Outcome Measures section below) provided to schools new to NELI as part of the DfE-funded offer and hence was already being used as part of the wave two implementation to undertake baseline assessments. Given the large volume of data being collected and the evidence for the psychometric properties of LanguageScreen (for example, West et al., 2021; West et al., 2022, Hulme et al., 2023.), we chose the LanguageScreen standardised score—measuring pupils' oral language skills—as the primary outcome. This is the same measure that was used (at baseline, before group and individual sessions began) as a criterion to select pupils for NELI. We decided against further primary data collection in line with the EEF's preference that schools should not be further burdened by requiring additional data collection and to ensure that the evaluation could be completed within the agreed budget and timeline. We considered using national assessment data (EYFSP)

as a secondary outcome but ultimately chose not to include this measure due to the timing of the assessments (which may have been early in the summer term, before a majority of NELI group sessions had been delivered) and its coarse nature—emerging' (1), 'expected' (2), and 'not assessed'. The evaluation therefore proceeded without a secondary outcome.

## Participant selection

**School participation in the evaluation**

All 4,422 NELI wave two enrolled schools were eligible to take part in the evaluation and at the point of enrolling (prior to the commissioning of evaluators) schools were asked whether they would be willing to be contacted about an evaluation: 2,029 schools responded affirmatively and had tested their pupils at baseline using LanguageScreen. These schools were subsequently contacted by the NFER with details of the project, reasons for taking part, and the privacy notice. To take part, headteachers needed to sign an MoU to commit to the impact evaluation and confirm they were happy for the NFER to receive their pupils' LanguageScreen data. The MoU was accessed by logging on to the NFER secure school portal using the details provided in the invitation email. In the MoU, schools were asked to share the NELI lead's contact details and school-level details required for the impact evaluation. Schools that completed their MoU were sent a confirmation email and asked to share the parent letter we had provided with parents of all pupils in reception class(es) and then to share any evaluation withdrawal requests with OxEd on support@teachneli.org. Schools were also eligible to receive incentives of up to £250 for (1) completing LanguageScreen assessments for all pupils who were initially assessed ahead of the delivery of the programme (that is, both pupils who received NELI and those who did not), (2) indicating which pupils have received the NELI programme in the school's LanguageScreen account, and (3) completing the final TeachNELI delivery survey sent by Nuffield Foundation Education Ltd.

Five hundred and forty-eight schools signed the MoU and opted to take part in the evaluation, agreeing to retest all pupils (whether or not they had received the NELI intervention) in the summer of 2022, and to supply this alongside data indicating which pupils received the intervention (the NELI indicator). Of these, 113 schools could not be included in the analysis as they did not complete the NELI indicator, one school which could not be matched to the NPD was excluded, and additional pupils and classes were excluded for other reasons (see Figure 4, Participant Flow Diagram) to give a final analytical sample of 356 schools.

**Pupil selection**

Schools selected pupils to receive the intervention on the basis of the LanguageScreen baseline assessment scores and other criteria described below. NELI was understood to be of potential benefit to all children and there were no criteria specifying ineligibility for the intervention.[5] The developers suggested that schools should use LanguageScreen to undertake a baseline assessment of pupils' oral language abilities and use this data to rank pupils within each class, selecting the three to six pupils with the lowest scores in each class for NELI. Doing so would imply a single criterion (an integer variable, LanguageScreen baseline standardised score) for selecting pupils.

However, early findings from the IPE suggested that only 33% of school staff (of 181 surveyed in December 2021) selected pupils based on LanguageScreen scores alone: 66% selected pupils based on other factors *in addition* to LanguageScreen (these findings are supported by our observation that LanguageScreen baseline scores were not uniquely determinant of receipt of NELI; see Figure 2 and accompanying discussion).[6] One hundred and twenty-one members of staff provided further information about these additional factors (Figure 1), which were not known by the impact evaluation team for individual pupils. As Figure 1 indicates, these comprise a range of objective and subjective criteria, with some not being recorded in administrative data.

---

[5] https://www.teachneli.org/faqs/
[6] Only one member of staff reported that their school did not use LanguageScreen assessments in selecting pupils to participate in NELI (Disley et al., 2023, p.43).

*Figure 1: IPE data describing criteria used by respondents to select pupils for NELI*

Reproduced from Disley et al., 2023, p.42.
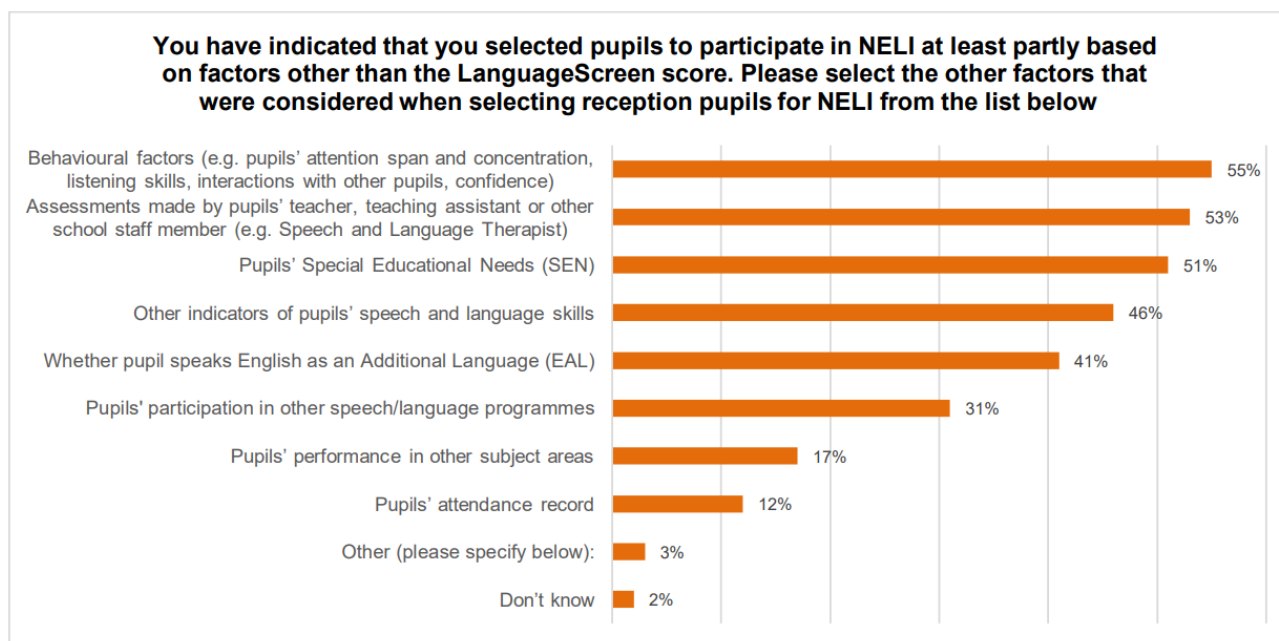


## Fuzzy Regression Discontinuity

Our evaluation was based on a Fuzzy Regression Discontinuity design with noncumulative normative cutoffs. The recommendation that schools use LanguageScreen baseline assessment scores to effectively rank and select pupils to receive NELI implied a regression discontinuity as the basis for estimating treatment effects. However, LanguageScreen baseline scores were not absolutely deterministic of treatment assignment as there were other factors teachers considered in addition (as per Figure 1 above). Therefore, the cutoff signified a probability of receiving treatment rather than determining it absolutely – units either side of the cutoff may have been treated or untreated as additional factors also played a part in the allocation of some units to treatment; in such scenarios these may be either known or unknown to an evaluator and may be unobservable in the evaluation data (Van der Klaauw, 1997).

This is illustrated in Figure 2 using the initial pupil-level data we received in August 2022 (n = 14,272 pupils),[7] with class cutoffs normalised to 0 and pooled data.[8] The cutoff varied across classes due to the use of ranking and the relative difference in abilities across schools (that is, schools with pupils who, on average, are of higher ability will have a higher cutoff than those with lower than average ability children). In the context of regression discontinuity, this can be characterised as noncumulative normative cutoffs, whereby individual units' running variable scores are recalculated relative to the group (in this case, class) cutoff so that data from all groups can be pooled (see Preliminary Analysis for more information). Figure 2 illustrates that the majority of pupils selected for NELI scored below the class cutoff, while the majority who were not selected scored above the class cutoff (3% of each group scored relative to their class cutoff on the opposite side to that indicated by their ultimate treatment assignment). Overall, this visual analysis suggested that LanguageScreen baseline scores played a large part in determining treatment assignment. However, it was necessary for us to take additional steps to confirm the viability of the implied *FRD* design, further details of which can also be found in Preliminary Analysis.

---

[7] The NFER received pseudonymised pupil data from OxEd for the purpose of preliminary analyses to assess the viability of the FRD approach and to identify schools that have met the requirements for incentive payments.

[8] Cutoffs were normalised by subtracting the modelled (non-zero) cutoff from pupil baseline scores. For more information see section Statistical Analysis: Preliminary Analysis to Assess Viability of FRD Design (page 20).

*Figure 2. Fuzzy regression discontinuity (using NELI initial pupil-level data received in August 2022)*



## Outcome measures

### Baseline and primary outcome measure

The evaluation included one primary outcome, a standardised score measuring pupils' oral language skills measured in the summer of 2022 (endline), around 20 weeks after the expected start of the delivery of the intervention to pupils. This was assessed using the LanguageScreen application, having previously been assessed before the commencement of delivery of the intervention to pupils (baseline). The choice of LanguageScreen standardised score as an outcome aligns with the intention of the intervention— 'to improve the spoken language ability of young children with relatively poor spoken language skills' (Disley et al., 2023, p.11—and the outcome of improved language ability identified by the NELI wave two scale-up logic model (ibid, p.9).[9] Furthermore, the timing of the outcome measurement is congruent with the proximal nature of the theorised outcome to the intervention.

The LanguageScreen application is a screening assessment administered by TAs which takes around ten minutes to complete and is scored automatically. It comprises four subtests (West et al., 2022):

- Expressive Vocabulary—naming 24 pictures;

- Receptive Vocabulary—matching each of 31 spoken words to one of four pictures;

- Sentence Repetition—repeating each of 12 sentences verbatim); and

- Listening Comprehension—answering 12 questions about three spoken stories that tap literal and inferential comprehension.

The LanguageScreen score is derived from the four subtests as a latent variable (West et al., 2021), with the standardised score being based on a sample of 348,944 children that was used for standardisation (Hume et al, 2023,

---

[9] The LanguageScreen standardised score was chosen as a screening measure and secondary outcome in the effectiveness trial; the primary outcome was 'language skills', a latent variable created from four individually administered language tests (Dimova et al., 2020, p.13).

2022).[10] West et al. (2021) reported that LanguageScreen reliability was high in the effectiveness RCT (pre-test screening Cronbach's alpha = 0.84) with good concurrent validity, while the standardisation paper, Hulme et al. (submitted), reported that the LanguageScreen total score has excellent reliability (Cronbach's alpha = 0.92; Person Separation Reliability = 0.94).

The standardised score used as the outcome measure in the evaluation therefore used data directly from the LanguageScreen system in its complete format and without adaptation by the evaluation team.[11] The data was collected by teachers and other school staff during the course of the school day, with scoring occurring automatically within the LanguageScreen system. Data was collected on tablet devices and uploaded to OxEd's database of LanguageScreen data.

**Secondary outcomes**

The evaluation did not include any secondary outcomes.

## Sample size

Schools were recruited to the impact evaluation from the 4,422 schools taking part in NELI wave two that had expressed an interest in participating. As schools were incentivised to take part in the evaluation, it was necessary to calculate a sample size to determine the number of schools required—in order to make decisions about the viability of the evaluation and to manage financial resources for incentives. At this stage, there was no data available for undertaking this calculation and a number of factors associated with the final dataset were unknown. The evaluation team therefore carried out initial sample size calculations for a prospective FRD study without data.[12] These were based on Deke and Dragoset (2012) and McKenzie (2022). Using the former we estimated the number of schools required to achieve different Minimal Detectable Effect Sizes (MDESs). These estimates were based on the assumption of 26.6 pupils per class with an average of 1.5 classes per school and a pre-post test correlation of 0.5.[13] They were also based on three different Regression Discontinuity Design Effects, that is, the number of times to multiply the equivalent RCT sample size for an RD design (9, 14, or 17; Deke and Dragoset, 2012). Using McKenzie's (2022) guidance, this number was then adjusted based on three different levels of assumed fuzziness under the anticipated FRD design (for example, a probability of 0.5 instead of 1 in the chance of being treated below the cutoff— repeated for 0.7 and 0.9). These methods assume some properties about the data but do not include others which are unknown in the absence of data. For example, we did not know (i) the distribution of scores on the running variable (LanguageScreen standard score at baseline), (ii) the location of the cutoff, or (iii) the proportion of treated and comparison pupils in the sample (although for the latter we assumed 23:77 based on three to six pupils per class receiving NELI). Following discussion with the EEF and OxEd we selected a number of schools to recruit based on a target MDES of 0.2. These calculations are presented in Table 4 (in column two, 'initial calculation without data') and are included in Appendix C.

Following the recruitment of schools to the evaluation and baseline (and some endline) testing of pupils, the developers were able to share a preliminary dataset with the evaluators so that preliminary analysis (including further sample size

---

[10] Standardisation was categorised into six-month age bands (48–53 months, n = 55,306; 54–59 months, n = 157,642; 60–65 months, n = 115,400). Data was analysed using Rasch modelling and provided a good fit to the model (RMSEA = 0.03, SRMSR = 0.09).

[11] The standardised score used as the baseline measure was adjusted to a class-level cutoff, as described in the Preliminary Analysis subsection of the Statistical Analysis section in this report. Using a standardised score as an outcome measure for an RCT may not be optimal (for example, because the standardisation process can lead to floor and ceiling effects). In this case it was necessary due to the way the measure is constructed (that is, derived from the subtests).

[12] In this and all subsequent power calculations we worked on the basis of specifying a target sample size for analysis of all pupils, rather than the FSM subgroup.

[13] 26.6 pupils per class (https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics); 1.5 forms per school (NELI scale-up Y2 impact evaluation ITT, Feb 2022); pre-post-test correlation as per Deke and Dragoset (2012).

calculations) could be carried out with the aim of confirming the viability of the proposed analysis using an FRD design. This dataset contained data for 14,272 pupils, of which 3,569 did not have outcome data (see Missing Data Analysis and Attrition). After cleaning, the preliminary dataset contained 10,703 pupils with pre- and post-intervention assessment data and treatment status indicated. This second sample size calculation ('calculation with preliminary pupil-level data', Table 4 below) was undertaken using the `rdsampsi` function in the `rdpower` R package (Cattaneo et al., 2019). Data variability and fuzziness were taken from the preliminary data. Two mean square error bandwidth selectors (above and below cutoff) were selected as the bandwidth selection procedure (Cattaneo et al., 2019, p.47). Mass points (see Preliminary Analysis section later) were addressed by requiring that initial bandwidths contain at least ten unique values. Pre-intervention LanguageScreen score was included as a baseline covariate. The adjustment of outcome scores to a class-level cutoff resulted in all classes having a mean of zero after adjustment. Since we expected no variability in mean scores across classes, a multilevel model approach was not implemented.[14] While the distribution of residuals in the final dataset was not known at this stage, the preliminary data suggested that heteroskedasticity might be expected. Therefore, a heteroskedasticity-robust plug-in residuals variance estimator was selected as the variance-covariance estimator. Sample size calculations were at a pupil level and class numbers in Table 4 are based on average cluster size. This sample gave an MDES of 0.22.

*Table 4: Sample size calculations*

| | | Initial calculation (without data) | Calculation with preliminary pupil-level data (n = 10,703)* | | Calculation with analysed pupil-level data (n = 10,759) | |
|---|---|---|---|---|---|---|
| | | OVERALL | OVERALL | FSM | OVERALL | FSM |
| Minimum Detectable Effect Size (MDES) | | 0.2 | 0.22 | 0.46 | 0.21 | 0.55 |
| Pre-test/ post-test correlations | Level 1 (pupil) | 0.5 | 0.66 | 0.66 | 0.66 | 0.66 |
| | Level 2 (class) | - | - | - | - | - |
| | Level 3 (school) | - | - | - | - | - |
| Intracluster correlations (ICCs) | Level 2 (class) | - | - | - | - | - |
| | Level 3 (school) | - | - | - | - | - |
| Alpha | | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| One-sided or two-sided? | | Two-sided | Two-sided | Two-sided | Two-sided | Two-sided |
| Average cluster size | | 27 | 19 | 4** | 21 | 6 |
| Number of classes | Intervention | - | - | - | - | - |
| | Comparison | - | - | - | - | - |
| | Total | 702 | 561 | 561 | 510 | 220*** |

---

[14] Data was also pooled following normalisation; our analysis did not explore treatment effects at specific cutoff values (i.e. as a proxy for site-by-treatment effects)

| | | | | | | |
|---|---|---|---|---|---|---|
| Number of pupils | Intervention | 3,848 | 2,589 | 487 | 2,329 | 424 |
| | Comparison | 14,833 | 8,114 | 1,526 | 8,430 | 885 |
| | Total | 18,681 | 10,703 | 2,013 | 10,759 | 1,309 |

\* Errors in Table 2 of the study plan (Worth et al., 2022)—sample size calculations—corrected here. These were an incorrect row total and pre/post correlation.
\** Assuming 18.81% of pupils are eligible for FSM (everFSM). Based on state funded primary school reception pupils in 2021/2022 from https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics.
\*** After subsetting the data to only include FSM pupils, classes with all or no pupils receiving the intervention were excluded. Some classes were lost because they had no FSM pupils and there was an additional loss of classes where all or no FSM pupils received the intervention.

Following receipt of the preliminary pupil-level data, schools were given more time to supply the required data for the evaluation and further data collection (of the NELI indicator status only) took place in schools in September 2022. This additional data collection marginally increased the number of pupils with both pre- and post-intervention assessment data and provided the final pupil-level dataset for analysis. Prior to analysis, the exclusion of pupils in year groups other than reception and in classes where all or no pupils received the intervention reduced the total number to 10,759, which gave an MDES of 0.21 ('calculation with analysed pupil-level data' in Table 4).

## Data sources

Details of the variables used can be found in Appendix D. OxEd supplied pupil-level data from the LanguageScreen system to the Department for Education for matching with National Pupil Database pupil data, school-level characteristics, and school-level dosage and fidelity data, shared by Nuffield Foundation Education Ltd and OxEd respectively. The matched dataset was made available to the evaluation team for analysis using the Secure Research Service. In addition, OxEd supplied pseudonymised preliminary pupil-level data (as of August 2002) to the evaluators to allow for early preliminary analysis to take place to assess the viability of the intended research design. The following datasets were combined for the analysis:

- LanguageScreen data including the baseline and primary outcome measures described above, school and class identifiers, and pseudonymised pupil data including unique identifiers and characteristics (for example, date-of-birth and EAL);

- NPD data included variables for each pupil included in the LanguageScreen data, contributing further characteristics such as FSM eligibility and Special Educational Needs) for balance checks across treated and untreated groups and for covariate adjustment;

- school-level characteristics derived from the DfE's publicly available Get Information about Schools dataset; school-level characteristics were used for our descriptive analysis of the sample; and

- school-level dosage and fidelity data derived from Nuffield Foundation Education Ltd's delivery surveys and OxEd's school staff training hosted on the FutureLearn platform.

## Statistical analysis

**Preliminary analysis to assess viability of FRD design—using initial pupil-level data available August 2022**

We carried out preliminary analysis using initial pupil-level data available to us in August 2022 (n = 10,703) with the primary objective of determining the viability of the FRD design suggested by the data and selection of pupils to receive NELI, and also to produce more accurate sample size calculations than those estimated without data (the first column of Table 4). This analysis indicated that our planned approach to the evaluation was appropriate and we repeated parts of this preliminary analysis a second time with the final pupil-level dataset before proceeding with the primary and subsequent analyses detailed below. This step therefore helped to shape final analytic choices, in particular by testing some of the RD identifying assumptions and modelling noncumulative multiple cutoffs in the data. Although outcome

data were included in both the initial and final pupil-level datasets, we did not use outcome data to shape analytic choices (other than using it to check for evidence of a discontinuity—point three below).

Preliminary analysis comprised the following steps. (Further details of the preliminary analysis carried out using initial pupil-level data (August 2022) can be found in in Appendix E.)

1. The first step involved describing the characteristics of the school sample—by comparing them with all NELI wave two and all English primary schools (with reception classes).

2. LanguageScreen baseline scores were analysed to model cutoffs by class, including calculating pupil baseline scores relative to the class cutoff. We selected the cutoff for each class that minimised the number of non-compliant pupils, normalising the cutoff to be zero across all classes by subtracting the cutoff from the scores and allowing for fuzziness in the analytical approach. The method for selection of a cutoff for each class was:

    a. set the cutoff to be the minimum baseline LanguageScreen score for the class and calculate the proportion of pupils correctly classified as intervention or comparison;

    b. repeat using each integer between the minimum and maximum baseline LanguageScreen score for the class as the cutoff; and

    c. take the median of the assessed cutoffs which produce the maximum classification accuracy.

3. It was necessary to model class cutoffs in this manner as they were implicit (not specified by teachers and unknown to us). The modelling approach was pre-specified in the evaluation study plan and was data-driven rather than allowing researcher discretion to choose cutoffs. Our finding that 35% of schools appeared to have a sharp cutoff (point five of Additional Analyses and Robustness Checks) was similar to that described by RAND Europe's IPE; on this basis we conclude that our approach was appropriate and did not significantly over- or under-estimate the cutoffs.

4. Graphical analysis was undertaken to confirm the validity of the proposed FRD design by checking for treatment assignment on either side of the cutoff and plotting the outcome against the cutoff to visually inspect for evidence of a discontinuity. Graphical analysis did not provide obvious evidence of a discontinuity at the normalised cutoff of zero but did not provide evidence which would preclude an RD design. We determined it to be appropriate to proceed with FRD analysis nonetheless, given the developers' recommendation about the use of LanguageScreen to select pupils for NELI (and the cutoff implied thereby). A visual inspection of a graph of treatment assignment (Figure 10, Appendix E) suggested a high probability of receiving treatment as a function of the LanguageScreen cutoff and, therefore, that analysis under FRD may be appropriate. This was confirmed in the following step.

5. Step five involved estimating the probability of receiving treatment as a function of the LanguageScreen cutoff. As FRD can be understood as an instrumental variables model (with the cutoff functioning as an instrument which affects treatment probability) it is important to check that the instrument is not weak, that is, that there is a substantial difference in treatment assignment either side of the cutoff. The What Works Clearinghouse Handbook (What Works Clearinghouse, 2022) indicates that study authors must run the first-stage regression of the participation indicator on the forcing variable, and the indicator for being above or below the cutoff, and provide either the F statistic from this regression, and that 'an F statistic of 16 will be used as the interim criterion for assessing instrument strength'. In the preliminary analysis, comparing the intervention received to whether the pre-intervention LanguageScreen score was above or below zero, we determined that the percentage of 'non-compliers' was 6% in total (for both those selected and not selected for NELI; Figure 2). When running the first-stage regression of the participation indicator on the forcing variable, and the indicator for being above or below the cutoff, the F statistic for the forcing variable was 85.85 and the F statistic for the indicator for being above or below the cutoff was 164.89. The low percentage of non-compliers and high F statistics suggest that the fuzziness in our data set is not extreme, so a FRD analysis is acceptable.

6. The number of mass points in the data was determined (due to non-continuous data on the running variable) and an assessment was made as to whether this precluded adopting a continuity-based RD approach (as per Cattaneo et al., 2020, pp.60–62). We found that the 10,703 observations (with non-

missing endline LanguageScreen scores) took 191 unique values and therefore considered the number of mass points to be sufficiently large for a continuity-based approach.

7. Power calculations were made to estimate the MDES with much greater precision than our initial calculations without data had allowed (see Sample Size Calculations). The initial calculations were not able to incorporate some of the properties unique to the dataset, such as the number of observations on each side of the cutoff and the degree to which the cutoff determines treatment (that is, fuzziness). Final power calculations were undertaken using the `rdpower` package in R, based on:

- outcome data (LanguageScreen endline assessment) for the final pupil sample;

- the modelled cutoffs (by class) with pupil running variable scores relative to these;

- known treatment assignment;

- treatment probability inherent in the data;

- covariates, including pre-test scores;

- bandwidth selection using two different mean squared error optimal (MSE-optimal) bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator (see note on bandwidth selection below); and

- heteroskedasticity-robust plug-in residuals variance estimator without weights.

Power calculations based on preliminary data estimated an MDES of 0.22 (rather than 0.20 as originally planned). This reduced to 0.21 in the final analytical sample.

**Primary analysis—using the final sample for analysis (Figure 4)**

Primary analysis used the non-parametric, continuity-based approach to RD, 'the most commonly employed in practice' (Cattaneo et al., 2019, p.5), which relies on different assumptions than the alternative local randomisation approach (although we used the latter as a robustness test of our primary analysis). As pupils' scores on the running variable (LanguageScreen standard score at baseline) were calculated relative to a cutoff normalised at zero during preliminary analysis, the primary analysis was of noncumulative multiple cutoffs in the data (Cattaneo et al., 2016). Hence cutoffs were aligned across classes to mitigate clustering effects. Impacts were estimated using a suite of R packages written specifically for RD analysis within the continuity-based approach,[15] including:

- `rdbwselect` for data-driven bandwidth selection methods;

- `rdrobust` for local polynomial point estimation and inference; and

- `rdplot` for graphical RD analysis.

RD estimation under continuity-based assumptions involves fitting two weighted least squares regressions (one each side of the cutoff) and calculating a point estimate, $\hat{\mu}+ - \hat{\mu}-$ (as illustrated by Figure 3). To do so, evaluators may choose the bandwidth within which to make the estimate. To avoid the introduction of evaluator bias (for example, testing different bandwidths until a preferred impact estimate is obtained), we allowed the appropriate bandwidth to be determined in a data-driven manner (MSE-optimal) for the primary analysis and subsequently conducted sensitivity tests using different bandwidths. The MSE-optimal method selects a bandwidth which minimises the mean squared error of the local polynomial RD point estimator, thus optimising the bias-variance trade-off (Cattaneo et al., 2019, p.47; Lee and Lemieux, 2010).

---

[15] Further details can be found at https://rdpackages.github.io/

*Figure 3: RD estimation—continuity-based approach (Cattaneo et al., 2019, p. 43)*
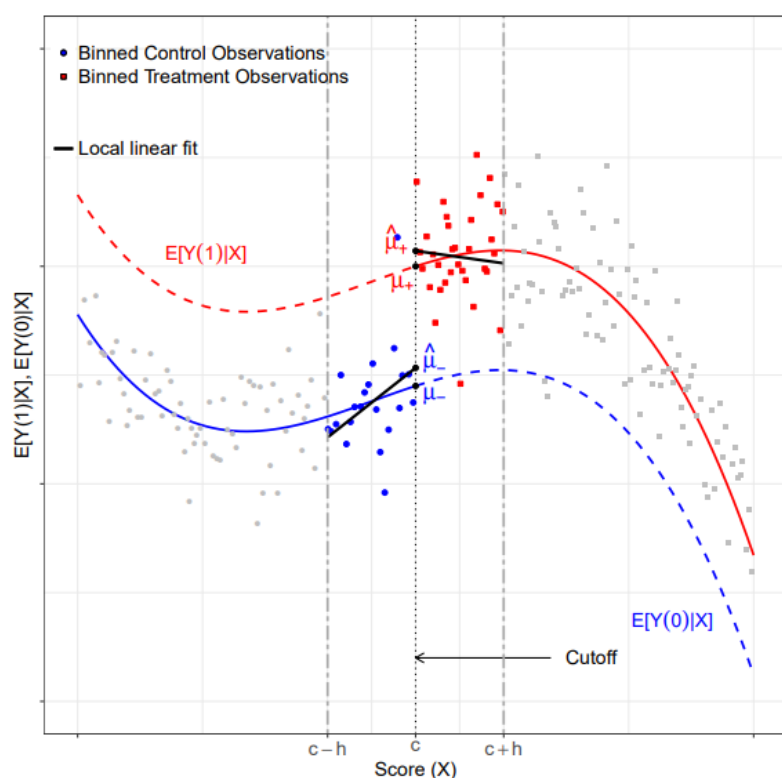


Primary analysis was undertaken using three FRD models (Cattaneo et al., 2021):

1. an unadjusted model—LanguageScreen standardised score (endline) regressed on pupil's score relative to cutoff, weighted for each observation, using cluster-robust standard errors;

2. a covariate adjusted model (pre-test)—as Model 1, including also LanguageScreen standardised score (baseline) as pre-test covariate; and

3. a covariate adjusted model (pre-test plus other pupil-level covariates)—as Model 2, including also gender, month of birth, FSM status, EAL status, and SEN status.

Models 2 and 3 required further balance checks using the pupils included within the bandwidth and allowed us to determine the effects of including the covariates progressively. All findings are based on Model 3 (covariate-adjusted).

**Missing data analysis**

We note the possibility of false discoveries in regression discontinuities should the missing data mechanism be related to the LanguageScreen endline assessment outcome (De la Cuesta and Imai, 2016). Our assumptions about missingness were based on our understanding of schools' use of LanguageScreen and their evaluation activity. We anticipated that the matched (with NPD) data would be complete with the exception of NELI indicator status or LanguageScreen endpoint assessment for a minority of pupils. Regarding the latter, we assumed missingness to be determined by school and teacher decisions about whether to retest pupils' post-intervention and, in some instances, which groups of pupils (intervention or non-intervention) to retest. Furthermore, schools were encouraged to record via their LanguageScreen accounts which pupils were receiving NELI as a requirement of the evaluation, although not all schools ultimately completed this step.

To investigate our assumption, we explored the nature and extent of missing data. We checked for entire schools that had missing data on NELI indicator and LanguageScreen endpoint assessment variables and compared their characteristics with schools with complete data. We modelled missingness, firstly, at the school level by means of a logistic regression where the outcome (for all pupils in the school) was either 'missing NELI indicator' or 'missing LanguageScreen endpoint assessment'. We also modelled missing data at the pupil level using two-level (pupil and

22

school) logistic regressions with the same outcomes. We included all pupil-level covariates specified in our analysis (LanguageScreen baseline score, EAL, SEN etc.) in addition to school-level covariates.

**Subgroup analyses**

We explored treatment effect heterogeneity by repeating our primary analysis for FSM-eligible pupils (using EVERFSM_6_P_[term][yy])[16] and EAL pupils. As these are relatively straightforward binary variables, we used indicator variables for each category to subset the data, then reapplied the data cleaning step where all classes with only intervention or only comparison pupils were removed. We then proceeded to estimate treatment effects for the subset groups—FSM and EAL—using the covariate adjusted model described in the primary analysis methods.

**Exploratory analysis**

We explored treatment effect heterogeneity by dosage (number of group sessions delivered) and training fidelity using a similar approach to our primary analysis (that is, covariate adjusted). This data was derived from delivery partner school surveys (completed in July 2022) and school staff training data supplied by OxEd, respectively. As surveys were returned by 344 schools, we considered this analysis to be exploratory due to the incomplete data (in terms of the sample used for primary analysis) and also the fact that schools returning surveys may not be representative of the entire sample.

Our definition of 'dosage' was based solely on the frequency of group sessions, which was shown to be prioritised by schools over individual sessions in the IPE (Disley et al., 2023b). Dosage was defined as the number of group sessions multiplied by the average session length in minutes, that is, the expected total number of minutes delivered in group sessions. Schools were split into dosage quartiles and the primary analysis model was applied to each quartile subset. We used school-level data on the number of group sessions delivered from the delivery partner survey as pupil-level dosage data was not available. Additionally, data for individual sessions was not included in the dosage metric as the delivery partner survey did not collect data on the number of individual sessions delivered.

Fidelity was defined using the percentage of TAs attending training sessions. Schools in our sample trained between zero and 11 TAs with a median of four. If more than 81% of TAs attended a session a score of two was given; if 50%–80% of TAs attended a session a score of one was given, and if fewer than 50% a score of zero was given. Three sessions were recorded in the dataset so schools could have fidelity scores of zero to six inclusive. Given the limited possible values for fidelity score, quartiles were not calculated but schools were split into the subsets of the seven possible integer values and the primary analysis model was applied to each subset. These definitions were derived from the prior effectiveness evaluation report (Dimova et. al., 2020).

**Additional analyses and robustness checks**

We tested the robustness of our findings to assumptions implied by our use of the continuity-based approach to model the regression discontinuity and their sensitivity to other analytical choices. This involved checking:

- the density of the running variable, to check whether the number of observations just below the cutoff is greatly different to the number of observations just above the cutoff—generally in RD designs this would suggest some manipulation (that is, that units could manipulate the score they receive on the running variable to influence treatment status) and would potentially undermine some of the assumptions on which RD is based. Whilst we did not expect pupils to be able to do this directly in the NELI wave two scale-up, they may have been assigned scores just above or below the cutoff in order to affect their treatment status, particularly as teachers used criteria in addition to LanguageScreen scores to select pupils for NELI (Figure 1). We used the rddensity package which provides manipulation tests of density discontinuity based on local polynomial density estimation methods (Cattaneo et al., 2018);

---

[16] In line with EEF statistical analysis guidance. Due to the age of the pupils this was assumed to be the same as FSMeligible_[term][yy].

- the exclusion of observations near the cutoff to understand the reliance by estimators on those units; this step is also relevant when there is evidence of manipulation (point 1 above);

- treatment effects at placebo cutoffs—we checked for treatment effects at cutoff values other than the actual cutoff (that is, 0): RD analysis is based on the assumption that regression functions are continuous at points other than the cutoff, that is, that there should be no discontinuities away from the cutoff;

- sensitivity to bandwidth choices—by varying bandwidth choice from that selected by MSE-optimal methods, we effectively added pupils to, or removed them from, the area of bandwidth within which the estimations are made adding pupils by increasing the bandwidth has the effect of decreasing variance (and hence, confidence intervals) but increasing bias. We investigated our findings under alternative bandwidth choices;

- estimations excluding classes where treatment assignment is fuzzy—that is, the running variable does not uniquely determine treatment. This allowed us to effectively analyse a subset of data as a Sharp RD: in the initial preliminary analysis, 38% of classes had a sharp cut off;

- testing the alternative RD framework (that is, the local randomisation approach)—by doing this we tested some of the fundamental assumptions of the preferred approach, for example, the continuity of potential outcomes near the cutoff; and

- testing whether treated and comparison units close to the cutoff are similar—predetermined covariates were fit as the outcome in an RD model to test whether systematic differences existed between units just above and below the cutoff (Cattaneo et. al. 2019).

**Estimation of effect sizes**

We calculated effect sizes (Hedges' g; Hedges, 1981) for primary and subgroup analyses, dividing the RD model coefficients by the pooled standard deviation ($s_p$ as calculated using the equation below). This standard deviation was calculated across all analysed pupils rather than the narrower sample within the bandwidth to avoid inflation of the effect size as it was expected that the bandwidth sample would be less variable than the total sample. Confidence intervals for coefficients were also converted to the Hedges' g scale using the pooled standard deviation.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 - 1 + n_2 - 1}}$$

# Timeline

*Table 5: Timeline*

| Dates | Activity | Staff responsible/ leading |
|---|---|---|
| March–May 2022 | Project set-up, development of recruitment materials, grant agreement sign off. | The NFER led but all staff involved |
| June 2022 | Receive list of schools that want to be contacted for impact evaluation from OxEd. Approach schools to bring them on board. Remind participating schools to complete endpoint LanguageScreen assessments of all reception pupils and NELI indicator. Receive weekly school-level data update from OxEd and monitor completion of assessments and NELI indicator data. | The NFER OxEd provided school list and weekly school-level data |
| July 2022 | Monitor school-level data to assess completion of assessments and NELI indicator data. Send tailored communications to schools reminding them to complete missing assessments/NELI indicator data. The NFER and OxEd agree on data spec for pupil-level data for preliminary analysis. | The NFER OxEd provided weekly school-level data |

| | Final delivery survey sent to schools by Nuffield Foundation Education Ltd. | |
|---|---|---|
| August–September 2022 | Receive pupil-level data for preliminary analysis to assess viability of FRD and to identify schools eligible to receive incentives.<br>QA of pupil-level data.<br>Additional data collection for NELI indicator data in schools.<br>The NFER and OxEd agree on data spec for pupil-level data for main analysis.<br>Submit NPD data request. | The NFER and OxEd extracted and shared pupil-level LanguageScreen data |
| October 2022 | Submit draft of study plan.<br>OxEd shares final de-identified pupil-level dataset with the NFER.<br>OxEd submits final pupil-level dataset to the DfE for matching of NPD variables.<br>The NFER submits school-level dosage metric to the DfE to combine with pupil-level data. | The NFER<br>OxEd |
| November–December 2022 | Finalise and publish study plan. | The NFER |
| December 2022–March 2023 | Data linking with NPD.<br>Complete preliminary analysis .<br>Main analysis.<br>Start report writing. | The NFER |
| April 2023 | Submit first draft of report. | The NFER |
| May–July 2023 | EEF / peer / developer reviews, the NFER edits and finalises report. | The NFER |
| September 2023 | Publish final report. | The NFER and the EEF |

# Impact evaluation results

## Participant flow including losses and exclusions

Figure 4 provides the details of the flow of participants through the evaluation to produce the final sample for analysis (prior to bandwidth selection for specific analyses: see Outcomes and Analysis section tables for numbers within bandwidths). Schools registered for the NELI wave two scale-up (n = 4,422) were contacted during the wave two recruitment process (prior to the evaluation team being commissioned) and were asked whether they consented to being contacted again about a potential impact evaluation. Approximately 75% of registered schools gave consent[17] and, of these, pupil-level data was available for 2,029 schools that completed baseline testing for the majority of their reception cohort using LanguageScreen. We therefore invited the 2,029 schools that had agreed to be contacted and had completed baseline testing to take part in the evaluation. Schools that were invited to take part in the evaluation were asked to do the following:

- sign an MoU committing to participating in the project;

- share a parent letter with parents of all pupils in reception class(es);

- indicate which pupils are receiving NELI on the LanguageScreen app;

- complete a final LanguageScreen assessment at the end of the 20-week programme or at the end of the summer term for all pupils who were initially assessed ahead of delivering the programme (that is, both pupils who received NELI and those who did not); and

- complete the final TeachNELI delivery survey.

Five hundred and forty-eight schools signed up to the evaluation (19,212 pupils in 823 classes). One school could not be matched with NPD records and 113 schools did not indicate which pupils received NELI on the LanguageScreen website. This meant that we were unable to use the pupil data (n = 3,613) from these schools as we were not able to categorise the pupils as treatment or comparison. This left 15,570 pupils (in 434 schools) for whom treatment status was known. Of the remaining 3,056 pupils who were indicated as having received NELI, 301 did not have a final LanguageScreen assessment (compared with 3,919 of 12,514 pupils who were untreated) and were therefore not included in the analytical sample. Pupil dates of birth and year group recorded in the NPD indicated that 22 treated pupils and three untreated pupils were not in reception classes (being older) and these pupils were similarly not included in the final sample for analysis. Finally, there were a number of classes where, after the above exclusions had been applied, either all pupils, or no pupils, were indicated as having received the intervention. We were cautious about including these for two reasons. Firstly, although schools were only asked to participate in the evaluation if they had completed baseline testing for all children, the presence of classes in the dataset where there was no comparison group suggested that this had not in fact been done. Furthermore, in some classes, endline LanguageScreen assessments were only returned for the pupils who had received NELI. Since this step of exclusion took place after excluding pupils missing endline assessment data, these classes appeared to have no comparison group and were excluded. To include classes with no comparison group in the pooled analysis may have biased findings. We therefore removed data for 87 classes where all pupils were indicated has having received the intervention (404 treated pupils) and 12 classes where no pupils were indicated as having not received the intervention (162 comparison pupils). These exclusions (from the 15,570 pupils in 434 schools) were classified as attrition; further detail can be found in Table 6.

Our final sample for analysis therefore comprised both baseline and endline data for 10,759 pupils in 510 classes (356 schools): 2,329 pupils received NELI while the remaining 8,430 pupils did not, likely receiving teaching as usual. Table 24 (Appendix J) describes the characteristics of English primary schools (n = 16,784), the 4,422 schools registered for the NELI wave two scale-up, and schools in our final analytical sample (n = 356). It also describes the characteristics of pupils in the 434 schools recruited to the evaluation and which provided NELI indicator data. Overall, the data

---

suggests that schools in the final analytical sample were broadly representative of both those in the NELI wave two scale-up schools and English primary schools in general.

*Figure 4: Participant flow diagram*

**NELI Wave 2**



Registered 2021/2022 (school n=4,422)

Did not agree to be contacted about evaluation and/or did not complete baseline assessments (school n=2,393)

Agreed to be contacted about evaluation and completed baseline assessments (school n=2,029)

Did not agree to be take part in the evaluation (school n=1,481)

**Schools recruited to evaluation**

School n=548; class n=823; pupil n=19,212

- Lost to NPD match – school n=1; class n=1; pupil n=29
- Did not complete NELI indicator – school n=113; class n=169; pupil n=3,613

School n=434; class n=653; pupil n=15,570

**NELI**
**Pupil n = 3,056**
(in 434 schools / 633 classes)

**Comparison**
**Pupil n = 12,514**
(in 427 schools / 643 classes)

- No endline record – pupil n=301 (in 22 schools / 33 classes)
- Year group not R – pupil n=22 (in 3 schools / 3 classes)
- In a class where all pupils indicated as receiving intervention – pupil n=404 (in 53 schools / 87 classes)

- No endline record –pupil n=3919 (in 68 schools / 120 classes)
- Year group not R –pupil n=3 (in 1 school / 1 class)
- In a class where no pupils indicated as receiving intervention –pupil n=162 (in 2 schools / 12 classes)

**Final sample for analysis\***

**Pupil n=10,759** (in 356 schools / 510 classes)

**NELI**
**Pupil n = 2,329**
(in 356 schools / 510 classes)

**Comparison**
**Pupil n = 8,430**
(in 356 schools / 510 classes)

\*Prior to bandwidth selection for specific analysis (see 'Outcomes and analysis' section tables for numbers within bandwidths)

## Missing data analysis and attrition

We ran missing data models on schools recruited to the evaluation (n = 548) and matched to the NPD. In the school-level missing data model, we found no evidence that school characteristics were related to likelihood of missingness on NELI indicators or LanguageScreen endline assessments. School governance type, region, urban or rural, and Ofsted rating did not have a significant impact on the probability of a school having data on these variables.

In the pupil-level analysis, we found evidence that pre-intervention LanguageScreen score, and FSM status were related to likelihood of missingness. Higher pre-intervention LanguageScreen score (likely to indicate pupils who did not receive NELI) was significantly associated with higher probability of a pupil having missing NELI indicator or missing endline assessment (odds ratio (95% CI): 1.037 (1.032, 1.041); p-value ≤ 0.001). Pupils eligible for FSM were significantly associated with higher probability of a pupil having missing NELI indicator or missing endline assessment (odds ratio (95% CI): 1.336 (1.155, 1.546); p-value ≤ 0.001). Gender, EAL, and SEN did not have a significant impact on the probability of a pupil having missing NELI indicators or missing endline assessments (Appendix J).

This suggests that schools may have systematically not retested pupils with certain characteristics, and that these pupils may have been more likely to have been those that did not receive NELI. However, we are not aware of any other evidence which would suggest deliberate systematic retesting of specific groups of pupils, and to do this would have been contrary to all communications about the scale-up and evaluation (for example, an email from OxEd to evaluation schools in June 2022 advised them to retest all pupils).

We further explored these pattens of missingness in the 434 schools which agreed to take part in the evaluation and completed the NELI indicator for pupils who had been tested at baseline. Table 6 characterises this missingness as pupil-level attrition. Of the 15,570 pupils recruited and with known treatment status, 69% were included in the analytical sample. Reasons for attrition are outlined in the previous section, Participant Flow, and include:

- pupils without endline assessments (27.1%);
- pupils with dates of birth out of range for reception pupils (0.2%); and
- classes where all or none of the pupils were indicated as having received the intervention (3.6%).

*Table 6: Pupil-level attrition from the evaluation (primary outcome)*

|  |  | Intervention | Comparison | Total |
|---|---|---|---|---|
| Number of pupils | Recruited (with NELI indicator) | 3,056 | 12,514 | 15,570 |
|  | Analysed | 2,329 | 8,430 | 10,759 |
| Pupil attrition (from recruited to analysis) | Number | 727 | 4,084 | 4,811 |
|  | Percentage | 24% | 33% | 31% |

Prompted by our pupil-level missing data analysis (above), we considered the possibility that only certain groups (for example, intervention pupils) or only those who staff perceived as benefiting from the intervention (or otherwise making greater than average progress) were retested at endline. Attrition among intervention pupils due to missing endline testing was 9.8%, compared with 31.3% for comparison pupils. This suggests that schools may have been making greater efforts to retest those pupils who participated in the intervention, leading to a larger proportion of intervention pupils in the analytical sample relative to the distribution in all schools participating in the evaluation and having completed the NELI indicator. Although there was a greater proportion of attrition for comparison pupils compared to intervention pupils due to missing endline data, the opposite was found when looking at attrition due to classes indicated as comprising all or no intervention pupils. Attrition for intervention pupils was higher in this case (13.2% compared to 1.3% for comparison pupils).

Taking evidence about these two sources of attrition together suggests that schools may have been prioritising testing and data submission for intervention pupils at both baseline and endline. The opposite direction of the attrition across both sources has the effect of cancelling out some of the larger difference, resulting in final overall attrition of 24% for treated pupils and 33% for untreated pupils.

We considered whether the differential missingness on outcomes was likely to introduce bias into our findings. Attrition in RD designs may be viewed somewhat differently to that in an RCT as impacts are estimated within a data-driven bandwidth (using post-attrition data) thus mitigating against any potential attrition bias. Furthermore, comparing the baseline characteristics of pupils (including pupils with missing outcomes, Table 24, Appendix J) to the characteristics of pupils in our final analytical sample (Table 8) suggests a high degree of similarity between the intervention pupils with and without missing outcomes (and also for the comparison pupils with and without missing outcomes). This therefore suggests that the final analytical sample is acceptably representative of pupils recruited to the evaluation and we did not conclude that the missing outcome data introduced a problematic degree of uncertainty into our findings.

## Outcomes and analysis

**Primary analysis**

**RQ1** What is the impact of NELI when delivered at national scale on pupils' oral language outcomes, as measured by LanguageScreen?

Table 7 and Figure 5 describe our primary analysis, which included all pupils in receipt of NELI in the analytical sample, and a comparison group of all pupils in the sample who did not receive the intervention. As described in the Methods section, primary analysis was undertaken using three FRD models. The results reported in Table 7 and Figure 5 are from the covariate adjusted model (Model 3, including LanguageScreen baseline as a pre-test measure plus other pupil-level covariates: gender, month of birth, FSM status, EAL status, and SEN status).[18] RD estimation selects units within a bandwidth each side of the cutoff as the basis for estimation. Our primary analysis detailed in Table 7 is based on data-driven bandwidth selection (that is, MSE-optimal) as per our proposed methodology, and we also allow for different bandwidth choices as a sensitivity analysis (Table 14). Estimation in this example is therefore based on 1,147 treated pupils and 3,329 comparison pupils from the analytical sample comprising 10,759 pupils (41.6%). Covariate balance for all pupils within the bandwidth is detailed in the table which follows (Table 8)—imbalance in certain characteristics (such as pupils with SEND or EAL) was considered when assigning the security rating (see also Table 17).

For all pupils within the bandwidth, we found a statistically significant ($p < 0.001$) effect of the NELI intervention on pupils' oral language skills measured by LanguageScreen of 0.297 (CI: 0.120, 0.474), which equates to four additional months' progress.

*Table 7: Primary analysis*

| Outcome | N class | Intervention group | | Comparison group | | Effect size | | |
| | | N pupil total | N pupil in bandwidth | N pupil total | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
|---|---|---|---|---|---|---|---|---|
| LanguageScreen Score | 510 | 2329 | 1269 | 8430 | 3207 | 4476 (1269; 3207) | 0.297 (0.120, 0.474) | <0.001 |

---

[18] The results from Models 1 and 2 are shown in Appendix J.

*Figure 5: Primary analysis model fit*
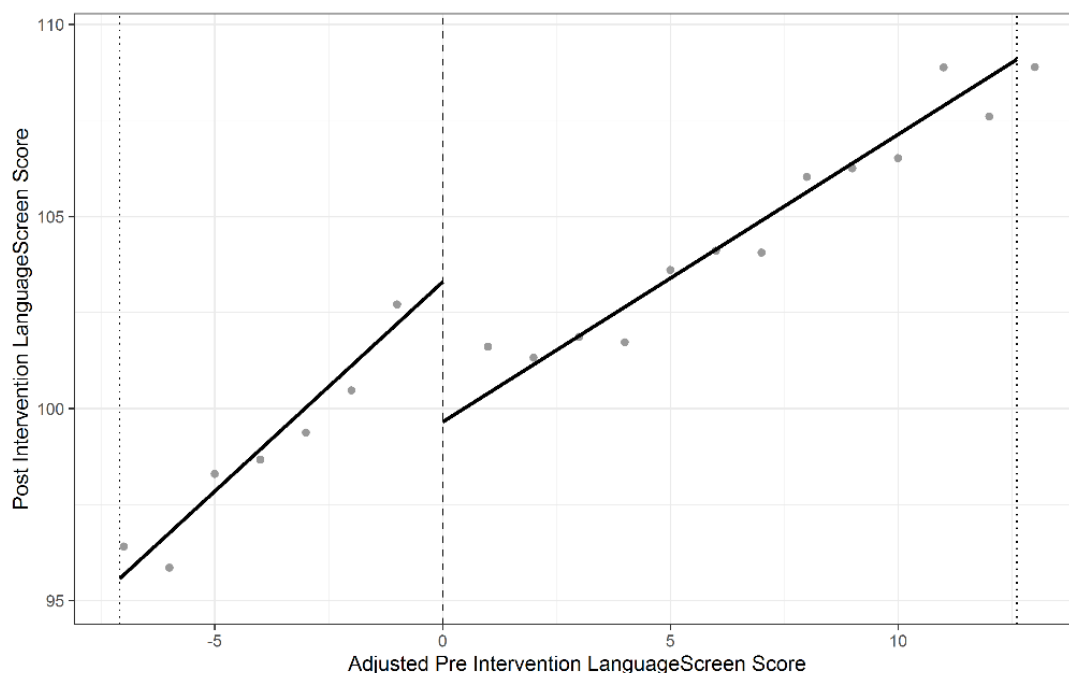


Cutoff represented by vertical dashed line. Bandwidth limits represented by vertical dotted lines. Data points represent the mean post intervention LanguageScreen score for an interval of one on the adjusted pre-intervention LanguageScreen score axis (±0.5 from where the point is shown). This is to avoid disclosive data points that represent fewer than ten individuals.

*Table 8: Covariate balance for pupils included in the primary analysis model*

| School level (categorical) | All pupils entering into the primary analysis model | | | | Pupils within the MSE-optimal bandwidth | | | |
|---|---|---|---|---|---|---|---|---|
| | Intervention group | | Comparison group | | Intervention group | | Comparison group | |
| | n/N | % | n/N | % | n/N | % | n/N | % |
| N | 2329 | | 8430 | | 1269 | | 3207 | |
| **Region** | | | | | | | | |
| East Midlands | 163/2329 | 7 | 575/8430 | 6.8 | 94/1269 | 7.4 | 221/3207 | 6.9 |
| East of England | 238/2329 | 10.2 | 929/8430 | 11 | 136/1269 | 10.7 | 382/3207 | 11.9 |
| London | 408/2329 | 17.5 | 1332/8430 | 15.8 | 222/1269 | 17.5 | 492/3207 | 15.3 |
| North East | 83/2329 | 3.6 | 401/8430 | 4.8 | 54/1269 | 4.3 | 138/3207 | 4.3 |
| North West | 351/2329 | 15.1 | 1011/8430 | 12 | 185/1269 | 14.6 | 410/3207 | 12.8 |
| South East | 383/2329 | 16.4 | 1618/8430 | 19.2 | 200/1269 | 15.8 | 630/3207 | 19.6 |
| South West | 249/2329 | 10.7 | 909/8430 | 10.8 | 128/1269 | 10.1 | 353/3207 | 11 |
| West Midlands | 243/2329 | 10.4 | 781/8430 | 9.3 | 131/1269 | 10.3 | 301/3207 | 9.4 |

| Yorkshire and the Humber | 211/2329 | 9.1 | 874/8430 | 10.4 | 119/1269 | 9.4 | 280/3207 | 8.7 |
|---|---|---|---|---|---|---|---|---|
| **Rural or urban** | | | | | | | | |
| Rural | 514/2329 | 22.1 | 1747/8430 | 20.7 | 286/1269 | 22.5 | 720/3207 | 22.5 |
| Urban | 1815/2329 | 77.9 | 6683/8430 | 79.3 | 983/1269 | 77.5 | 2487/3207 | 77.5 |
| **Ofsted rating** | | | | | | | | |
| Outstanding | 350/2329 | 15 | 1393/8430 | 16.5 | 181/1269 | 14.3 | 462/3207 | 14.4 |
| Good | 1640/2329 | 70.4 | 5598/8430 | 66.4 | 902/1269 | 71.1 | 2240/3207 | 69.8 |
| Requires Improvement | 138/2329 | 5.9 | 547/8430 | 6.5 | 73/1269 | 5.8 | 191/3207 | 6 |
| Missing Ofsted Rating | 201/2329 | 8.6 | 892/8430 | 10.6 | 113/1269 | 8.9 | 314/3207 | 9.8 |
| Pupil level (categorical) | n/N | Percentage | n/N | Percentage | | | | |
| **Gender** | | | | | | | | |
| Female | 1050/2329 | 45.1 | 4167/8430 | 49.4 | 593/1269 | 46.7 | 1541/3207 | 48.1 |
| Male | 1279/2329 | 54.9 | 4263/8430 | 50.6 | 676/1269 | 53.3 | 1666/3207 | 51.9 |
| **FSM-eligibility status** | | | | | | | | |
| Non FSM | 1818/2329 | 78.1 | 7204/8430 | 85.5 | 987/1269 | 77.8 | 2611/3207 | 81.4 |
| FSM | >501/2329 | >21.5 | >1216/8430 | >14.4 | >272/1269 | >21.4 | >586/3207 | >18.3 |
| FSM Missing | <10/2329 | <0.4 | <10/8430 | <0.1 | <10/1269 | <0.8 | <10/3207 | <0.3 |
| **EAL status** | | | | | | | | |
| Non EAL | 1455/2329 | 62.5 | 7229/8430 | 85.8 | 894/1269 | 70.4 | 2578/3207 | 80.4 |
| EAL | 874/2329 | 37.5 | 1201/8430 | 14.2 | 375/1269 | 29.6 | 629/3207 | 19.6 |
| **SEN status** | | | | | | | | |
| Non SEN | 1957/2329 | 84 | 7862/8430 | 93.3 | 1089/1269 | 85.8 | 2935/3207 | 91.5 |
| SEN | 372/2329 | 16 | 568/8430 | 6.7 | 180/1269 | 14.2 | 272/3207 | 8.5 |

**Subgroup analyses**

**RQ2** What is the impact of NELI when delivered at national scale on FSM (everFSM) pupils' oral language outcomes, as measured by LanguageScreen?

**RQ3** What is the impact of NELI when delivered at national scale on EAL pupils' oral language outcomes, as measured by LanguageScreen?

We conducted subgroup analyses by subsetting the data to create two datasets derived from our analytical sample, one for pupils eligible for free school meals (using everFSM6) and one for pupils identified as having English as an additional language. After subsetting the data to include only the relevant pupils (FSM and EAL separately), classes with all or no pupils receiving the intervention were excluded. Some classes were lost because they had, for example, no FSM pupils, and there was an additional loss of classes where all or no FSM pupils received the intervention. Each analysis used Model 3 (pre-test measure and pupil-level covariates) as per our primary analysis and MSE-optimal data-driven bandwidths to select pupils to include in the analysis.

Analysis for the FSM subgroup included 688 pupils from 220 classes within the bandwidth and found a statistically significant effect ($p$ = 0.009) of participation in NELI of 0.569 (CI: 0.142, 0.997). This equates to seven additional months' progress.

Analysis for the EAL subgroup included 853 pupils from 219 classes within the bandwidth. Although we did not estimate the power of the evaluation for this subgroup, we believe that it was not powered to detect an effect of the magnitude observed (0.294; CI: -0.003, 0.623) and this was not statistically significant (p = 0.079). Nevertheless, this effect size is positive and of a similar magnitude to the findings of our preliminary analysis, thus demonstrating some comparability.

*Table 9: Subgroup analyses*

| Outcome | N Class | Intervention group | | Comparison group | | Effect size | | |
|---|---|---|---|---|---|---|---|---|
| | | N pupil total | N pupil in bandwidth | N pupil total | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| LanguageScreen score for FSM subset | 220 | 424 | 215 | 885 | 473 | 688 (215; 473) | 0.569 (0.142, 0.997) | 0.009 |
| LanguageScreen score for EAL subset | 219 | 699 | 276 | 997 | 577 | 853 (276; 577) | 0.294 (-0.003, 0.623) | 0.079 |

*Figure 6: FSM subgroup analysis model fit*
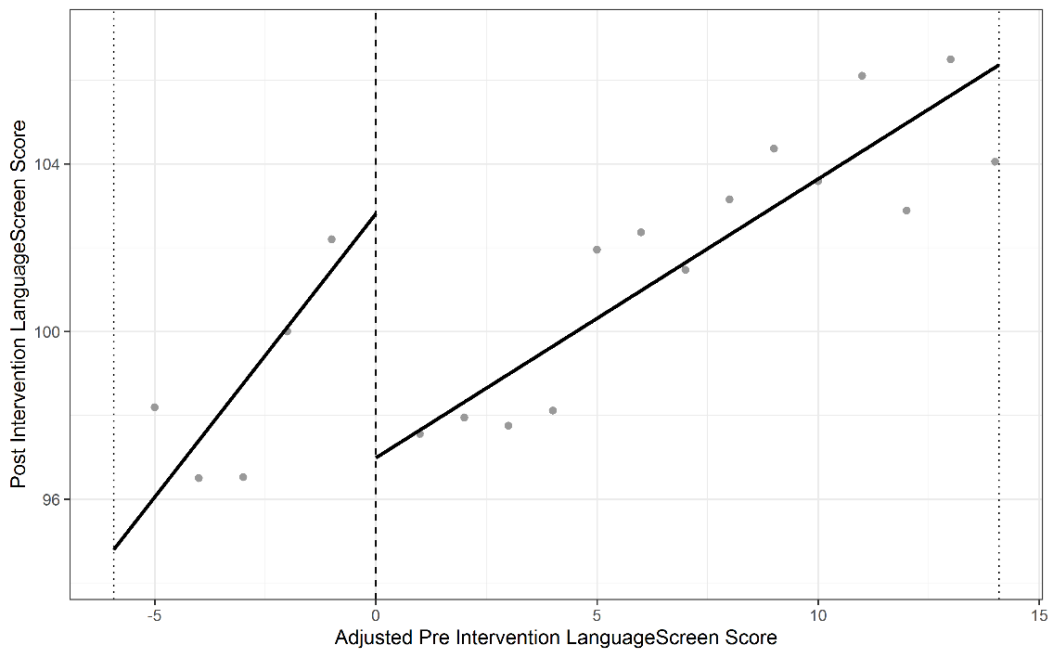


Cutoff represented by vertical dashed line. Bandwidth limits represented by vertical dotted lines. Data points represent the mean post intervention LanguageScreen score for an interval of one on the adjusted pre-intervention LanguageScreen score axis (±0.5 from where the point is shown). This is to avoid disclosive data points that represent fewer than ten individuals.

*Figure 7: EAL subgroup analysis model fit*
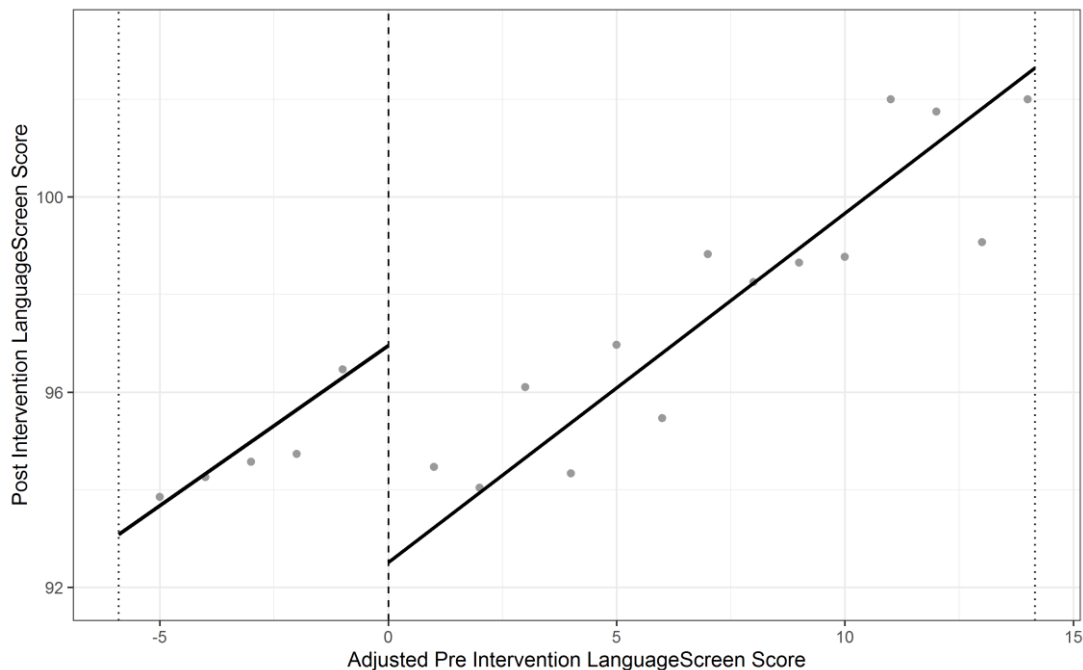


Cutoff represented by vertical dashed line. Bandwidth limits represented by vertical dotted lines. Data points represent the mean post intervention LanguageScreen score for an interval of one on the adjusted pre-intervention LanguageScreen score axis (±0.5 from where the point is shown). This is to avoid disclosive data points that represent fewer than ten individuals.

## Exploratory analyses

**RQ4**  How does the impact of NELI on pupils' oral language outcomes vary by dosage?

Figure 8 shows a histogram of the dosages for all schools in the dosage analyses. Table 10 gives details of the four dosage analyses which estimated treatment effects for each dosage quartile (derived from IPE survey responses) as

per the primary analysis model (Model 3).[19] We observed larger (and statistically significant) effects where schools reported dosages in the third and fourth quartiles compared with schools with reported dosages in the first and second quartiles, although effect sizes have not been statistically compared between quartiles.

*Figure 8: Histogram of the total minutes of group sessions for schools in the dosage analyses*
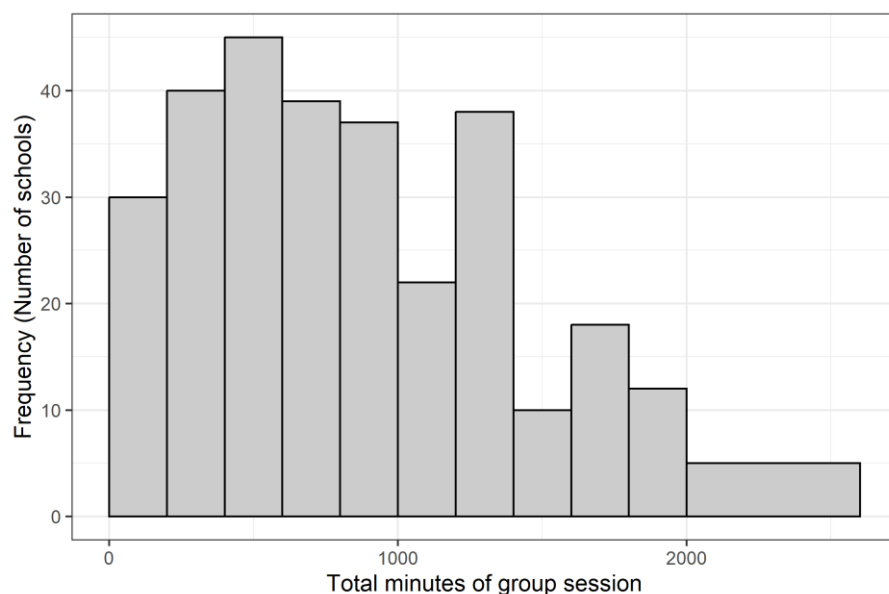


*Table 10: Dosage analysis*

| Outcome: LanguageScreen score | N class | Intervention group | | Comparison group | | Effect size | | |
|---|---|---|---|---|---|---|---|---|
| | | N pupil total | N pupil in bandwidth | N pupil total | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| Dosage Q1 (17.5 to 410.6 total minutes of group sessions) | 100 | 441 | 271 | 1491 | 564 | 835 (271; 564) | 0.119 (-0.194, 0.432) | 0.456 |
| Dosage Q2 (410.6 to 797.5 total minutes of group sessions) | 99 | 430 | 248 | 1550 | 837 | 1085 (248; 837) | 0.254 (-0.109, 0.617) | 0.170 |
| Dosage Q3 (797.5 to 1286.9 total minutes of group sessions) | 107 | 506 | 248 | 1937 | 611 | 859 (248; 611) | 0.476 (0.003, 0.950) | 0.049 |
| Dosage Q4 (1286.9 to 2550 total minutes of group sessions) | 114 | 486 | 180 | 2049 | 1023 | 1203 (180; 1023) | 0.399 (0.043, 0.754) | 0.028 |

[19] The mean number of group sessions delivered by schools in the final sample for analysis was 33 (equivalent to 11 weeks of the intervention).

**RQ5** How does the impact of NELI on pupils' oral language outcomes vary by training fidelity?

Table 11 gives details of our analysis of training fidelity, which were based on data from OxEd's Futurelearn training platform. Positive effects of receiving the NELI intervention were observed for fidelity scores of one to six (these representing schools in which more than 50% of TAs had attended at least one training session) although not all of these were statistically significant and some may have been observed due to chance. Statistically significant effects of NELI were observed for fidelity scores of two and three suggesting that some training of TAs delivering NELI has a positive impact on treated pupils' outcomes. However, the effects of training on the outcomes of pupils receiving NELI were of a lesser magnitude as a greater number of TAs attended more sessions.

*Table 11: Fidelity analysis*

| Outcome: LanguageScreen score | N Class | Intervention group | | Comparison group | | Effect size | | |
|---|---|---|---|---|---|---|---|---|
| | | N pupil total | N pupil in bandwidth | N pupil total | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| Fidelity Score 0 | 53 | 192 | 89 | 965 | 402 | 491 (89; 402) | -0.427 (-1.020, 0.166) | 0.158 |
| Fidelity Score 1 | 67 | 330 | 195 | 1180 | 480 | 675 (195; 480) | 0.362 (-0.076, 0.800) | 0.105 |
| Fidelity Score 2 | 84 | 338 | 213 | 1336 | 554 | 767 (213; 554) | 0.409 (0.050, 0.768) | 0.026 |
| Fidelity Score 3 | 111 | 553 | 346 | 1804 | 774 | 1120 (346; 774) | 0.475 (0.080, 0.869) | 0.018 |
| Fidelity Score 4 | 87 | 381 | 246 | 1497 | 783 | 1029 (246; 783) | 0.148 (-0.231, 0.527) | 0.444 |
| Fidelity Score 5 | 40 | 201 | 123 | 625 | 293 | 416 (123; 293) | 0.226 (-0.267, 0.719) | 0.368 |
| Fidelity Score 6 | 61 | 304 | 175 | 914 | 452 | 627 (175; 452) | 0.116 (-0.271, 0.502) | 0.557 |

**Additional analyses and robustness checks**

As outlined in our study plan (and detailed in the Methods section of this evaluation report), we completed a number of additional analyses to check the assumptions and choices on which our primary analysis was based, including checks specific to the FRD design.

*1. Density of the running variable*

We used the manipulation test of density discontinuity in the `rddensity` package (Cattaneo et al., 2018). The *p*-value for this test was 0.425 suggesting no evidence of manipulation of units close to the cutoff, that is, that pupils were not assigned scores just above or below the cutoff in order to affect their treatment status. The graphical analysis in the preliminary analysis (Appendix E) supports the conclusion that the density of the running variable does not display discontinuity.

*2. Exclusion of observations near the cutoff to understand the reliance by estimators on those units*

Inference in RD designs is based on units within a bandwidth either side of the cutoff. Units closest to the cutoff may be the most influential when fitting local polynomials (Cattaneo et al., 2020, p.104) so excluding them is a test of their influence on estimates. Furthermore, these may be the units which appear on one side of the cutoff due to manipulation (although as stated in point one above, we observed no evidence of manipulation in our data).

Table 12 shows our estimate for all pupils (as per primary analysis Model 3 above) and the same model re-run to exclude the 5% of units closest to the cutoff within each bandwidth, and again to exclude 10% of the same. Doing so does not meaningfully impact the results of the primary analysis: effect sizes are in the same direction, of similar magnitude (0.297 to 0.357), and still highly significant ($p < 0.005$).

*Table 12: Analysis excluding units close to the cutoff*

| Outcome: LanguageScreen score | Intervention group<br>N pupil in bandwidth | Comparison group<br>N pupil in bandwidth | Effect size | | |
|---|---|---|---|---|---|
| | | | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| All pupils | 1269 | 3207 | 4476 (1269; 3207) | 0.297 (0.120, 0.474) | <0.001 |
| Removing 5% closest to the cutoff within each bandwidth | 1193 | 3060 | 4253 (1193; 3060) | 0.357 (0.159, 0.554) | <0.001 |
| Removing 10% closest to the cutoff within each bandwidth | 1124 | 2904 | 4028 (1124; 2904) | 0.330 (0.115, 0.545) | 0.003 |

*3. Treatment effects at placebo cutoffs*

RD analysis is based on the assumption that regression functions are continuous (that is, discontinuities should not be observed) at points other than the cutoff as treatment occurs at the cutoff (albeit this assumption is strongest for Sharp RDs). We therefore checked for treatment effects at four placebo cutoffs, two on either side of the actual cutoff (0).

Table 13 below illustrates these: we observed no statistically significant discontinuities at any of the placebo cutoffs; this provided support for our assumption of continuous regression functions.

*Table 13: Analysis at cutoffs other than 0*

| Outcome: LanguageScreen score | Intervention group | Comparison group | Effect size | | |
|---|---|---|---|---|---|
| | N pupil in bandwidth | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| Cutoff = -8 | 449 | 1822 | 2271 (449; 1882) | -2.71 (-16.1, 10.6) | 0.691 |
| Cutoff = -4 | 148 | 1562 | 1710 (148; 1562) | -0.815 (-29.0, 27.4) | 0.955 |
| Cutoff = 4 | 3872 | 367 | 4239 (3872; 367) | 0.400 (-0.946, 1.750) | 0.560 |
| Cutoff = 8 | 4611 | 220 | 4831 (4611; 220) | 0.157 (-6.75, 7.07) | 0.964 |

*4. Sensitivity to bandwidth choices*

For our primary, subgroup, and exploratory analyses we used data-driven bandwidth selection. One of the reasons for this was to avoid introducing any evaluator bias into the choice of bandwidth (the size of which could influence estimates). However, in order to determine the sensitivity of estimates to bandwidth choice we re-ran our primary analysis under alternative bandwidth assumptions. Table 14 shows the effects of varying the bandwidth from that selected by MSE-optimal methods (effectively adding pupils to, or removing them from, the area of bandwidth within which the estimations are made). The primary analysis using the same bandwidth previously reported is given for context ('100% bandwidth') along with two variations, bandwidths half and double the size of the primary analysis MSE-optimal bandwidth.

Estimates using the alternative bandwidths remain statistically significant. As anticipated, including more pupils in the bandwidth gives a narrower confidence interval due to an increase in precision and in both the estimates are of a similar magnitude (0.387, 0.294) to the original (0.297).

*Table 14: Analysis of alternative bandwidths*

| Outcome: LanguageScreen score | Intervention group | | Comparison group | | Effect size | | |
|---|---|---|---|---|---|---|---|
| | N pupil in bandwidth | Bandwidth limit | N pupil in bandwidth | Bandwidth limit | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| 50% bandwidth | 742 | -3.548 | 1459 | 6.303 | 2201 (742; 1459) | 0.387 (0.0885, 0.685) | 0.011 |
| 100% bandwidth (i.e., primary analysis) | 1269 | -7.095 | 3207 | 12.605 | 4476 (1269; 3207) | 0.297 (0.120, 0.474) | <0.001 |
| 200% bandwidth | 1958 | -14.210 | 6272 | 25.210 | 8230 (1958; 6272) | 0.294 (0.179, 0.410) | <0.001 |

*5. Estimations excluding classes where treatment assignment is fuzzy (that is, running variable does not uniquely determine treatment)*

We did not know the basis on which pupils were selected to receive NELI in a particular school or class, although it was recommended that schools use LanguageScreen baseline scores to rank and select pupils. Cutoffs on the running variable (LanguageScreen baseline) were therefore not explicitly available to us and varied across classes due to differences in class ability levels. We therefore modelled cutoffs by class using an iterative approach to maximise classification accuracy (preliminary analysis point two). We found that 35% of classes appeared to have a sharp cutoff, that is, appeared to use LanguageScreen baseline scores alone as the basis for selecting pupils. This finding is similar to that of RAND's IPE which found that 33% of school staff (of 181 surveyed in December 2021: Disley et al., 2023) reported selecting pupils for NELI based on LanguageScreen baseline scores alone.

Table 15 repeats our primary analysis (Model 3) using only data from pupils in the classes where our modelling suggested a sharp cutoff. The effect size in this subset is in the same direction as the primary analysis but of smaller magnitude and not statistically significant, the latter likely being due to a combination of the reduced sample size and the smaller effect. Covariate balance within the bandwidth is detailed in Table 25 (Appendix J).

*Table 15: Analysis of sharp subset*

| Outcome | N Class | Intervention group | | Comparison group | | Effect size | | |
|---|---|---|---|---|---|---|---|---|
| | | N pupils total | N pupils in bandwidth | N pupils total | N pupils in bandwidth | N pupils in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| LanguageScreen score (excluding classes with a fuzzy cutoff) | 180 | 704 | 398 | 2875 | 1012 | 1410 (398; 1012) | 0.161 (-0.0620, 0.383) | 0.157 |

*6. Testing the alternative RD framework (the local randomisation approach)*

All RD analyses in this report have been conducted under the continuity-based approach. To test assumptions of this preferred approach—for example, the continuity of potential outcomes near the cutoff—we also applied the alternative, local randomisation approach to the primary analysis. Using the `rdlocrand` R package, the window with the most plausible local randomisation assumptions is given as -1 to 1 with 155 observations below the cutoff and 330 observations above the cutoff. This is in contrast to the primary analysis bandwidth of -7.10 to 12.6 with 1,269 observations below the cutoff and 3,207 observations above it. Using the suggested bandwidth of -1 to 1, randomisation inference gave a non-significant effect size of 0.146 as shown in Table 16. The magnitude of this effect is not as large as the primary analysis continuity-based approach effect size, but the direction of the effect is consistent and the reduced sample size could explain the lack of statistical significance.

*Table 16: Analysis under local randomisation approach*

| Outcome | Intervention group | Comparison group | Effect size | | |
|---|---|---|---|---|---|
| | N pupil in bandwidth | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| LanguageScreen Score (local randomisation approach) | 199 | 286 | 485 (199; 286) | 0.146 (-0.00456, 0.372) | 0.116 |

*7. Test whether treated and comparison units close to the cutoff are similar*

Predetermined covariates were fit as the outcome in an FRD model to test whether systematic differences existed between intervention and comparison units irrespective of bandwidths applied in estimation models (Table 17). Any differences may suggest that something other than the intervention was contributing to the observed outcomes. Two of the covariates (gender and SEN) showed a significant difference between the intervention and comparison group, with small magnitude effects in both cases. For gender, there is a significantly higher likelihood of being male in the intervention group compared to in the comparison group. For SEN, there is a significantly higher likelihood of being SEN in the intervention group compared to in the comparison group.

However, we expected some differences between the intervention and comparison groups given the data describing the way in which schools selected pupils for NELI (Figure 1). This data suggests that treatment assignment was determined in a majority of schools using characteristics such as these, in addition to LanguageScreen baseline scores as the basis for decisions about which pupils should receive NELI. This is in line with the characteristics of pupils in the sample of schools recruited to the evaluation and completing NELI indicator data (Table 24), which highlighted that treated pupils were more likely to be male, FSM-eligible, or indicated as EAL or SEN in school census data.

Further data concerning covariate balance can be found in Table 8, which displays the covariate balance for pupils included in the primary analysis model and that for pupils within the MSE-optimal bandwidth. Both show similar characteristics to those of pupils recruited to the sample (that is, all pupils, including those with missing outcome data): treated pupils were more likely to be male, FSM-eligible, or indicated as EAL or SEN, although the percentage difference for gender was smaller than for the other three characteristics.

Although both the falsification check analysis and balance among analysed pupils within the bandwidth suggest some imbalance between the two groups, we had planned to include covariates in our analysis in a progressive manner in order to determine their contribution to estimates of the treatment effect. These models are described in the Primary Analysis section of this report, and outputs from these are included in Appendix J (Table 23). We found that when the covariates described above were included in the model (and also including month of birth), the estimated treatment effect increased from 0.269 (Model 2: LanguageScreen baseline score only entered as a covariate) to 0.297 (Model 3: covariates above plus month of birth entered). It may therefore be the case that imbalances in some covariates such as SEN and EAL in particular were downwardly biasing the impact estimate observed in Model 2, and that their inclusion effectively corrects for this imbalance.

*Table 17: Falsification check analysis with covariates as outcome measures*

| Outcome | Intervention group | Comparison group | Effect size | | |
|---|---|---|---|---|---|
| | N pupil in bandwidth | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| Gender | 955 | 1267 | 2222 (955; 1267) | -0.133 (-0.263, -0.00370) | 0.0438 |
| FSM | 877 | 1143 | 2020 (877; 1143) | 0.0218 (-0.0898, 0.133) | 0.702 |
| EAL | 793 | 1029 | 1822 (793; 1029) | 0.0480 (-0.0943, 0.190) | 0.508 |
| SEN | 877 | 1143 | 2020 (877; 1143) | -0.129 (-0.245, -0.0145) | 0.0272 |

# Conclusion

*Table 18: Key conclusions*

| Key conclusions |
| --- |
| Pupils who received the NELI programme made the equivalent of four additional months' progress in language skills, on average, compared to pupils who did not receive NELI. This result has a moderate to high security rating. |
| Subgroup analysis found pupils eligible for free school meals (FSM) who received the NELI programme made an additional seven months' progress in language skills, on average, compared to pupils eligible for FSM who did not receive the programme. |
| Subgroup analysis found pupils with English as an additional language (EAL) who received the NELI programme made an additional four months' progress in language skills, on average, compared to pupils with EAL who did not. However, the sample of pupils for this subgroup was small and potentially not sufficient to confidently interpret the level of impact. |
| Exploratory analysis highlighted that the effect of receiving NELI was greater for pupils whose TA delivered more of the programme's group sessions compared with pupils whose TAs delivered fewer group sessions. |
| Exploratory analysis found that the effect of receiving NELI was greater for pupils in schools where more than 50% of TAs had attended at least one training session compared to pupils in schools where fewer than 50% of TAs had attended between zero and three training sessions. |

## Impact evaluation and IPE integration

### Interpretation

This evaluation set out to assess the impact of wave two of the scale-up of NELI, a 20-week oral language intervention for reception pupils. While previous pilot, efficacy, and effectiveness trials have demonstrated the positive impact of NELI on pupils' oral language skills, this is the first evaluation of the impact of NELI when delivered at national scale. Whereas the efficacy and effectiveness trials used randomised controlled trials to evaluate the impact of NELI, this quasi-experimental evaluation used a Fuzzy Regression Discontinuity design that leveraged the treatment assignment rule created by the LanguageScreen cutoff score. The majority of schools also applied other selection criteria in addition to pupils' LanguageScreen scores at baseline, thereby implying a 'fuzzy' rather than 'sharp' RD design.

The schools recruited to the evaluation and those ultimately included in the analysis were representative of NELI wave two scale-up schools. NELI had a positive impact on reception pupils' oral language skills when delivered at national scale. Care must be taken when interpreting the effects of RD designs (see Limitations, below), but nevertheless we found that pupils receiving NELI made an additional four months' progress in oral language skills compared to pupils who did not receive NELI. NELI also had a positive impact on FSM-eligible pupils: FSM pupils receiving NELI made an additional seven months' progress in oral language skills compared to FSM pupils who did not receive NELI. Given that the attainment gap at the end of reception between pupils from disadvantaged backgrounds and their more affluent peers is 4.6 months (Hutchinson, Reader and Akhal, 2020), NELI's positive impact on the oral language skills of FSM -eligible pupils has the potential to close the 'language gap' for these pupils. This is of particular importance as the NELI scale-up was part of the government's Covid-19 recovery efforts, with priority given to schools with a high proportion of disadvantaged pupils. Although NELI was shown to have a positive impact on the oral language skills of pupils with EAL who received NELI compared to pupils with EAL who did not receive NELI, there is more uncertainty around this finding as it was not observed to be statistically significant. This, however, may be due to characteristics of the data (for example, the size of the sample) and nevertheless the effect size observed would be of practical educational significance if indeed it is a true estimate of the actual effect for EAL pupils. Findings from the subset of schools which returned survey data indicating the number of group sessions delivered also suggest that pupils whose TAs delivered more weekly group sessions made greater progress than pupils whose TAs delivered fewer weekly group sessions.

This evaluation took place following a pilot study (Fricke et al., 2013), efficacy trial (Sibieta et al., 2016), and an effectiveness trial (Dimova et al., 2020) and thus represents the final stage in the evaluation pipeline for an intervention.

The effectiveness trial estimated an additional three months' progress for pupils receiving the NELI intervention, using a primary outcome of language skills based on a composite measure. In that evaluation, oral language skills measured by LanguageScreen was specified as a secondary outcome, with an effect size of 0.358 (CI: 0.22, 0.47; $p < 0.05$) being estimated for all pupils (equivalent to four months' additional progress). This is, therefore, comparable to the estimate made by this evaluation, although we estimated the effect size to be slightly lower (0.297). This is perhaps to be expected: treatment effect estimates from QEDs are often lower than those of RCT evaluations of the same intervention, and treatment effects when interventions are delivered at scale are often observed to be lower than in evaluations at earlier stages of the evaluation pipeline. This may be due to the intervention being delivered in increasingly less-than-ideal conditions as the scale of delivery increases (Cheung and Slavin, 2016).

The EEF had commissioned an independent IPE of wave one of the scale-up of NELI by RAND Europe to understand better the scale-up process in the context of the disruption caused by Covid-19. When additional funding was provided by the DfE for wave two of the scale-up, this evaluation was subsequently extended to cover the IPE of wave two. The wave two IPE focuses on the delivery of NELI in the 2021/22 academic year, including the sustainability of the intervention and wider lessons for the implementation of education interventions. Specifically, the research questions focused on five topic areas including school recruitment and reach, fidelity to intervention delivery, perceived impact of NELI on pupils and teachers/TAs, transition of the scale-up approach from wave one to wave two, and sustainability of NELI after the end of the funding period. Two reports detailing the IPE findings for each wave have now been published (Disley et al., 2023a; Disley et al., 2023b).

The IPE noted that there were deviations from the intended intervention delivery model during wave two. According to the intended delivery model, NELI was to be delivered over 20 weeks in the form of three 30-minute group sessions and two 15-minute one to one sessions per week. Around half of the wave two schools surveyed did not complete 20 weeks of intervention delivery due to a lack of staff time and capacity: data from delivery partner surveys revealed that, on average, schools had delivered about 30 group sessions (about ten weeks of NELI delivery) by the end of the delivery period for wave two schools, although there was wide variation in the number of group sessions delivered. This wide delivery variation was also observed in the previous effectiveness trial (Dimova et al., 2020). Covid-19 continued to present a barrier to delivery as staff and pupil absences derailed the progress of intervention delivery. Around a third of schools surveyed reported not having delivered any one to one sessions. Individual sessions were described as being resource intensive especially when staff did not always perceive a need for them, leading schools to prioritise group sessions.

Despite issues with the fidelity of intervention delivery, surveys of TAs and teachers on the perceived impact of NELI on pupils were largely positive and consistent with our impact findings (Disley et al., 2023b). The majority of teachers and TAs surveyed perceived that NELI had improved the language skills and confidence of recipient pupils. Staff from case study schools also commented on the positive impact of NELI, particularly on pupils' speaking and narration skills and vocabulary. The perceived positive impact was greater among schools that had delivered the intervention as intended, whether in terms of delivery of the full 20 weeks or in terms of delivery of both group and individual sessions. Our exploratory analyses examining the impact of the dosage of NELI on pupils' language outcomes is consistent with these findings. NELI pupils in schools that delivered more group sessions had better language outcomes than NELI pupils in schools that delivered fewer group sessions. This was also consistent with the compliance analysis carried out in the NELI effectiveness trial (Dimova et al., 2020). Our impact findings, in concert with those of the previous effectiveness trial, suggest that NELI can have a positive and significant impact on pupils' language outcomes even when the intervention is not implemented with great fidelity. Given that most schools that were surveyed in the IPE felt that delivering NELI was very time intensive and this was one of the barriers to implementation, there might be scope for the development and testing of a shorter or less intensive version of NELI that might have a positive impact on pupils' language outcomes.

Our exploratory analysis of the impact of training fidelity on pupils' language outcomes suggests that the effect of receiving NELI was greater for pupils in schools in which more than 50% of TAs had attended at least one training session, compared with pupils in schools where fewer than 50% of TAs had attended between zero and three training sessions. It is important to highlight that the TA training delivered as part of this national scale-up was fully online, in contrast to the mandatory face to face TA training delivered in the previous trials. While the delivery partners, OxEd, had started developing an online asynchronous training model prior to the waves one and two national scale-up, this fully online training model was first implemented in the context of both waves. The positive impact of NELI observed

in this evaluation of the wave two national scale-up suggests that the online training model may indeed be efficacious. However, further rigorous evaluation will be required to definitively understand the contribution of the mode of training delivery to the impact of NELI.

In exploring the sustainability of NELI in wave one and wave two schools after the end of the funding period, the IPE found that the majority of wave one schools continued to offer NELI to their next cohort of reception pupils in 2021/2022, demonstrating the appetite for NELI even in the current challenging context. Most wave one and wave two schools are open to delivering NELI in 2022/2023 and beyond, but many are undecided because of pressures on time and lack of clarity over the needs of the next cohort. In order to support wave one and two schools with continued delivery of NELI, OxEd and the DfE have reached an agreement that will allow these schools to access the NELI training and LanguageScreen app free of charge. The agreement also allows for a small number of new schools in priority areas to sign up for NELI. The DfE has recently announced that funding for NELI will be extended to the 2023/2024 academic year.

**Evidence to support the logic model**

As outlined in the introduction to this report and elsewhere, there is a significant body of evidence of the positive impact of NELI on pupils' oral language skills. The effectiveness trial evaluated by Dimova et al. (2020) included an extensive IPE, the results of which supported the original logic model of the intervention. For the purpose of this scale-up, RAND Europe, in collaboration with the intervention developers and the EEF, drafted a logic model setting out the inputs, activities, outputs, and outcomes for the delivery of NELI at scale. This logic model, that was developed while delivery was ongoing, was a description of how NELI was delivered at scale (Disley et al., 2020). Therefore, rather than testing the theory underlying the NELI intervention, the IPE of the wave two scale-up focused on research questions related to the delivery of NELI at scale. The IPE noted deviations from the original intervention delivery model, which we have discussed above in the context of our findings.

## Limitations and lessons learned

This impact evaluation of wave two of the NELI scale-up analysed data from nearly 11,000 children from 356 schools with reception classes from across England. It therefore represents the largest evaluation of the NELI intervention to date and the first of the intervention delivered at scale. Despite its strengths, the impact evaluation was subject to a number of limitations.

Using an RD design for evaluation may control for some unobservable characteristics by estimating a treatment effect using units within a bandwidth either side of a cutoff. In the absence of randomisation, this therefore provides a good basis for robust causal inference. However, RD treatment effects are estimated local to the cutoff and hence cannot be assumed to have external validity to units with scores far from the cutoff (either within or outside of the bandwidth). Cattaneo and colleagues (2019) note that additional assumptions would be required in order to provide further assurance about the external validity of RD estimates, and that this is a topic of active research. In addition, our analysis was based on an FRD, estimating the effect of treatment near the cutoff for compliers (that is, a subset of those assigned to each group, treatment and comparison) rather than using the data of all who were assigned to either condition. Given that there were non-compliers in both groups we cannot be certain as to the net effect on our impact estimates, other than to conclude that the FRD parameter is associated with some imprecision. Finally, the presence of noncumulative normative cutoffs and data pooling may indicate additional heterogeneity as the pooled estimand is effectively a weighted average of the average effects of treatment at each cutoff (Cattaneo et al., 2016). As we were principally interested in understanding the effectiveness of the intervention delivered at scale (rather than in specific schools) we chose not to explore any potential heterogeneity resulting from pooling the data. This heterogeneity may, however, partially offset some of the limitations in external validity previously identified as the cutoffs around which treatment effects were estimated varied across schools. In summary, we cannot assert that our estimates of treatment effects apply to all treated pupils, but we note, however, that they are of a similar magnitude to the average treatment effect estimated by the effectiveness trial.

A further limitation regarding the design was the imbalance between treated and comparison groups on key characteristics, which may be sources of selection bias. Although our analyses included these covariates in our analytical models, we cannot be sure that selection bias did not influence our estimates to some degree. We also know

that to some extent schools selected pupils to receive NELI based on unobservable characteristics and that it would, therefore, be very difficult to choose a research design or analytical strategy to address this (in the absence of randomisation). To some extent, however, an FRD can mitigate the issue of selection on unobservables by estimating treatment effects for units which complied with their treatment assignment based on the running variable.

It may be the case that school decisions about which pupils to retest at endline introduced some bias into the sample— for example, some teachers may have chosen not to retest pupils who had made less than average perceived progress. However, we have no evidence to indicate that this did in fact happen and it would have been contrary to all communications to schools about which pupils to retest at endline. Where schools had returned data suggesting that all pupils in a class were either in the intervention or control group, we dropped these classes from our analysis, thus addressing class-level missingness. We found pupil-level outcome missingness to be related to pupil characteristics and LanguageScreen baseline score but observed that the final analytical sample of intervention and comparison units was similar to the sample which included pupils with missing outcomes. Hence, we did not conclude that there was any evidence suggesting that missing data clearly undermined our findings.

Missing data also meant the analytical sample was smaller than might have otherwise been possible. Schools received incentives for their participation, for retesting pupils at endline, and for the return of data, and they also received frequent communication with regard to the importance of their role in the evaluation. While this was, in fact, the minimal data collection requirement which we could as evaluators have placed on schools, we recognise that this was likely to have been difficult for many schools still affected by staffing challenges associated with the Covid-19 pandemic, in addition to other priorities in the summer term. The reduction in sample size therefore inevitably reduced the power of the evaluation while still achieving an MDES of 0.21.

A further limitation of the evaluation, also related to the size of the sample, is that it was not sufficiently powered to detect treatment effects of the expected magnitude in the FSM and EAL subgroups. This was anticipated when the evaluation was designed, and to achieve a much larger sample of pupils designated as being members of these subgroups would have required more time and resources (for example, for incentivising schools) than this evaluation was able to access. Nevertheless, the policy context of the intervention was one in which disadvantaged pupils were understood to be disproportionately affected and therefore schools with higher-than-average proportions of FSM-eligible pupils were targeted in wave one. It may therefore have been appropriate to place more of a priority on this subgroup when initially designing and resourcing the evaluation of the wave two scale-up.

The evaluation focused on one outcome, a decision which was shaped by some of the constraints and preferences— such as timeline, lack of usable secondary data, and the preference not to collect further primary data from schools— discussed during the inception of the project. However, the wave two scale-up logic model hypothesised a number of outcomes for pupils, teachers, and schools. While not all of these would readily lend themselves to a quantitative impact evaluation, it may have been possible to design an evaluation which addressed other pupil outcomes, in particular long-term language development. The outcome chosen by this evaluation was appropriate to measure short-term language but does not provide immediate evidence about the impact of the intervention on pupils' broader English language outcomes measured by national assessments.

The impact evaluation was also limited in terms of its integration with the IPE, which needed to be commissioned prior to the start of the impact evaluation. This limited opportunities to understand some aspects of the evaluation in greater detail, and on a school-by-school basis. For example, greater data collection about the criteria used by schools to select pupils to receive NELI may have highlighted additional methodological options and may have negated the need to make inferences about which schools used LanguageScreen alone. Furthermore, there may have been an opportunity to collect more detailed and comprehensive data about pupil receipt of NELI (for example, pupil-level dosage data and the number of sessions attended—group and individual) which would have allowed for more detailed analysis of dosage effects.

## Future research and publications

This evaluation addressed one of the outcomes theorised by the NELI wave two scale-up logic model (Disley et al., 2023, p.9). Future research may focus on the other pupil-level outcome identified by the logic model, long-term reading comprehension, and may also take the opportunity to explore the effect of the intervention on attainment in English

measured by national assessments. Further long-term data analysis building on the effectiveness trial has recently been undertaken (Groom, Brown and Lymperis, 2023) but there is the opportunity for long-term impacts of the scale-up to be understood. Data from this evaluation will be deposited in the EEF's archive, linking it to NPD national assessment data (for example, KS1, KS2) thus providing the basis for subsequent analysis of the NELI wave two scale-up without the need for additional data collection.

Future research may also place a greater focus on subgroups of interest (FSM and EAL), which are believed to have been disproportionately impacted by the disruption caused by the Covid-19 pandemic. Indeed, this impact is one of the motivating factors behind large-scale early interventions such as NELI and although the present evaluation sought to understand impact for these groups it was not possible to power this evaluation with these subgroups specifically in mind. Therefore, a future study may be able to enrol a larger number of schools returning complete data to the evaluation in order to achieve required sample sizes for these groups.

While the current evaluation took advantage of the IPE data being collected by RAND to understand variability in levels of programme delivery across schools, there are opportunities to collect more granular and detailed quantitative data about the implementation of NELI in order to investigate the moderators of treatment effects. For example, understanding the delivery of group and individual sessions at scale may provide insight into the relative contribution of each to intervention effectiveness, and analysis of data from shorter or less intensive versions of the intervention may potentially highlight how schools can prioritise resources where necessary.

A future evaluation of NELI delivered at scale may also seek to understand better the selection criteria which staff are using to select pupils, on a pupil-by-pupil basis. This data would allow for more precise modelling, for example, by clearly being able to identify the classes in which only LanguageScreen was used to do this.

This evaluation provided data about sample size calculations for an FRD of NELI delivered at scale. This could inform a future RD-based evaluation specifically (whereas some of these data, such as the degree of fuzziness, were unknown to us when initially designing this evaluation) and may also be useful for sample size calculations under other designs. While other designs may be considered and may confer some advantages over RD (for example, the ability to estimate an average treatment effect) nevertheless they are still faced with the challenge of pupil-level selection bias. Finally, the fact that the probability of receiving treatment as a function of the LanguageScreen cutoff was relatively strong may encourage future evaluations using RD designs where RCTs may not be feasible or desirable, but it is necessary for the evaluation design to address the problem of (at least partial) selection on unobservables.

# References

Bowyer-Crane, C., Bonetti, S., Compton, S., Nielson, D., D'Apice, K. and Tracey, L. (2021) 'Impact of Covid19 on School Starters: Interim Briefing 1. Parent and School Concerns About Children Starting School': https://educationendowmentfoundation.org.uk/public/files/Impact_of_Covid19_on_School_Starters_-_Interim_Briefing_1_-_April_2021_-_Final.pdf

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (2019) *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge: Cambridge University Press. https://www.cambridge.org/core/elements/practical-introduction-to-regression-discontinuity-designs/F04907129D5C1B823E3DB19C31CAB905

Cattaneo, M. D., Jansson, M. and Ma, X. (2018) 'Manipulation Testing Based on Density Discontinuity', *Stata Journal*, 18 (1), pp. 234–261. https://doi.org/10.1177/1536867X1801800115

Cattaneo, M. D., Keele, L. and Titiunik, R. (2021) 'Covariate Adjustment in Regression Discontinuity Designs: http://arxiv.org/abs/2110.08410

Cattaneo, M. D., Keele, L., Titiunik, R. and Vazquez-Bare, G. (2016) 'Interpreting Regression Discontinuity Designs with Multiple Cutoffs', *Journal of Politics*, 78 (4), pp. 1229–248. https://doi.org/10.1086/686802

Cattaneo, M. D. and Titiunik, R. (2022) 'Regression Discontinuity Designs', *Annual Review of Economics*, 14 (1), pp. 821–51. https://mdcattaneo.github.io/papers/Cattaneo-Titiunik_2022_ARE.pdf (Accessed: 28 June 2023).

Cattaneo, M. D., Titiunik, R. and Vazquez-Bare, G. (2019) 'Power Calculations for Regression-Discontinuity Designs', *Stata Journal*, 19 (1), pp. 210–245. https://doi.org/10.1177/1536867X19830919

Cheung, A. C. K. and Slavin, R. E. (2016) 'How Methodological Features Affect Effect Sizes in Education', *Educational Researcher*, 45 (5), pp. 283–292. https://doi.org/10.3102/0013189X16656615.

Deke, J. and Dragoset, L. (2012) 'Statistical Power for Regression Discontinuity Designs in Education: Empirical Estimates of Design Effects Relative to Randomized Controlled Trials': https://files.eric.ed.gov/fulltext/ED533141.pdf

DfE and Ford, V. (2021) 'Every School with Reception Class Offered Early Language Training', GOV.UK: https://www.gov.uk/government/news/every-school-with-reception-class-offered-early-language-training

Dimova, S., Illie, S., Rosa Brown, E., Broeks, M., Culora, A. and Sutherland, A. (2020) 'The Nuffield Early Language Intervention. Evaluation Report': https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Nuffield_Early_Language_Intervention.pdf

Disley, E., Nightingale, M., Amodeo, G., Haider, A., Culora, A., Dimova, S., Gilder, L. and Virdee, M. (2023) 'The Nuffield Early Language Intervention Scale-Up: Evaluation Report—Year 1': https://d2tic4wvo1iusb.cloudfront.net/documents/projects/NELI-scale-up-Year-1-report.pdf?v=1679502241

Disley, E., Nightingale, M., Haider, A. and Amodeo, G. (2023) 'The Nuffield Early Language Intervention Scale-Up: Evaluation Report—Year 2': https://d2tic4wvo1iusb.cloudfront.net/documents/projects/NELI-Scale-Up-Year-2-report.pdf

Duff, F., Reen, G., Plunkett, K. and Nation, K. (2015) 'Do Infant Vocabulary Skills Predict School-Age Language and Literacy Outcomes?', *Child Psychology and Psychiatry*, 56 (8), pp. 848–856. https://doi.org/10.1111/jcpp.12378

EEF (2023) 'Pipeline of EEF trials', London: Education Endowment Foundation: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/process-and-people/pipeline-of-eef-trials

Feinstein, L. and Duckworth, K. (2006) 'Development in the Early Years: Its Importance for School Performance and Adult Outcomes': https://discovery.ucl.ac.uk/id/eprint/10005970/1/Feinstein2006Development.pdf

Fernald, A., Marchman, V. A. and Weisleder, A. (2012) 'SES Differences in Language Processing Skill and Vocabulary Are Evident at 18 Months', *Developmental Science*, 16 (2), pp. 234–248. https://doi.org/10.1111/desc.12019

Francis, B. (2022) EEF blog: 'Understanding the Impact of COVID on Learning' (18 May): https://educationendowmentfoundation.org.uk/news/eef-blog-understanding-the-impact-of-covid-on-learning

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C. and Snowling, M. J. (2012) 'Efficacy of Language Intervention in the Early Years', *Child Psychology and Psychiatry*, 54 (3), pp. 280–290. https://doi.org/10.1111/jcpp.12010

Groom, M., Rosa Brown, E. and Lymperis, L. (2023) 'The Nuffield Early Language Intervention: Addendum Report': https://d2tic4wvo1iusb.cloudfront.net/documents/projects/EEF_NELI_addendum_report_redacted.pdf?v=1680082686

Hainmueller, J. (2012) 'Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies', *Political Analysis*, 20 (1), pp. 25–46. https://doi.org/10.1093/pan/mpr025.

Hedges, L. V. (1981) 'Distribution Theory for Glass's Estimator of Effect Size and Related Estimators', *Educational Statistics*, 6 (2), pp. 107–128. https://doi.org/10.2307/1164588

Heiss, A. (2020) 'Fuzzy Regression Discontinuity': https://evalf20.classes.andrewheiss.com/example/rdd-fuzzy/

Howson, J. (2022) 'The Labour Market for Teachers in England January to July 2022: A Period of Unprecedented Turmoil': https://www.teachvac.co.uk/misc_public/Labour%20Market%20Report%20-%20January%20to%20July%202022.pdf

Hulme, C. and Snowling, M. J. (2015) 'Learning to Read: What We Know and What We Need to Understand Better', *Child Development Perspectives*, 7 (1), pp. 1–5: https://doi.org/10.1111/cdep.12005

Hulme, C., McGrane, J., Duta, M., West, G., Cripps, D., Dasguta, A., Hearne, S., Gardner, R. and Snowling, M. (submitted) 'LanguageScreen: The Development and Standardization of an Automated Language Assessment App'.

Hutchinson, J., Reader, M. and Akhal, A. (2020) 'Education in England: Annual Report 2020', Education Policy Institute: https://epi.org.uk/wp-content/uploads/2020/09/EPI_2020_Annual_Report_.pdf

Inside Government (2020) 'The Impact of COVID-19 for Pupils Using English as an Additional Language' (blog): https://blog.insidegovernment.co.uk/schools/covid-19-impact-for-eal-pupils

Law, J., Todd, L., Clark, J., Mroz, M. and Carr, J. (no date) 'Early Language Delays in the UK': https://resourcecentre.savethechildren.net/document/early-language-delays-uk/

Lee, D. S. and Lemieux, T. (2010) 'Regression Discontinuity Designs in Economics', *Economic Literature*, 48 (2), pp. 281–355. https://doi.org/10.1257/jel.48.2.281

McKensie, D. (2016) 'Power Calculations for Regression Discontinuity Evaluations: Part 1' (World Bank blogs, 6 September): https://blogs.worldbank.org/impactevaluations/power-calculations-regression-discontinuity-evaluations-part-1

Roulstone, S., Law, J., Rush, R., Clegg, J. and Peters, T. (2011) 'Investigating the Role of Language in Children's Early Educational Outcomes' Research Report DFE-RR134: https://doi.org/10.1037/e603032011-001

Scarborough, H. S., Neuman, S. and Dickinson, D. (2009) 'Connecting Early Language and Literacy to Later Reading (Dis)Abilities: Evidence, Theory, and Practice', in Fletcher-Campbell, F., Soler, J. and Reid, G. (eds) *Approaching Difficulties in Literacy Development: Assessment, Pedagogy and Programmes*, London: Sage, pp. 23–38.

Sibieta, L., Kotecha, M. and Skipp, A. (2016) 'Nuffield Early Language Intervention: Evaluation Report and Executive Summary': https://d2tic4wvo1iusb.cloudfront.net/documents/projects/EEF_Project_Report_Nuffield_Early_Language_Intervention.pdf?v=1680172777

Tracey, L., Bowyer-Crane, C., Bonetti, S., Nielson, D., D'Apice, K. and Compton, S. (2022) 'The Impact of the Covid-19 Pandemic on Children's Socio-Emotional Wellbeing and Attainment During the Reception Year': https://d2tic4wvo1iusb.cloudfront.net/documents/projects/EEF-School-Starters.pdf?v=1680772078

Van der Klau, W. (1997) 'A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrolment': https://core.ac.uk/download/pdf/6636941.pdf

West, G. (2022) 'LanguageScreen Technical Detail' (e-mail, personal communication, 17 November 2022).

West, G., Lervåg, A., Snowling, M. J., Buchanan-Worster, E., Duta, M. and Hulme, C. (2022) 'Early Language Intervention Improves Behavioural Adjustment in School: Evidence from a Cluster Randomized Trial', *School Psychology*, 92, pp. 334–345. https://doi.org/10.1016/j.jsp.2022.04.006

West, G., Snowling, M. J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., McLachlan, H. and Hulme, C. (2021) 'Early Language Screening and Intervention Can Be Delivered Successfully at Scale: Evidence from a Cluster Randomized Controlled Trial', *Child Psychology and Psychiatry*, 62 (12), pp. 1425–434. https://doi.org/10.1111/jcpp.13415

What Works Clearinghouse (2022) *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0*: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf

Whitehurst, G. J. and Lonigan, C. J. (1998) 'Child Development and Emergent Literacy', *Child Development*, 69 (3), pp. 848–872. https://doi.org/10.2307/1132208

Worth, J., Smith, A., Sahasranaman, A. and Staunton, R. (2022) 'Impact Evaluation of Nuffield Early Language Intervention (NELI) Wave 2': https://d2tic4wvo1iusb.cloudfront.net/documents/projects/Impact-Evaluation-of-NELI_Study-Plan_Final_2012022.pdf?v=1686567649

# Appendix A: Security classification of trial findings

**OUTCOME:** *Oral language (LanguageScreen standardised score)*

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

| Threats to validity | Risk rating | Comments |
|---|---|---|
| **Threat 1: Confounding** | Moderate | Some evidence of imbalance in characteristics (e.g., pupils with SEND, EAL) within the selected bandwidth, including from placebo tests. |
| **Threat 2: Concurrent Interventions** | Low | No substantial evidence of risks |
| **Threat 3: Experimental effects** | Low | No substantial evidence of risks |
| **Threat 4: Implementation fidelity** | Low | Some concerns about fidelity were expressed, but there is a consistency with previous EEF trials of this intervention in terms of this not necessarily being too detrimental to effectiveness. Effect sizes were larger in schools reporting higher dosage of NELI delivery. |
| **Threat 5: Missing Data** | Moderate | Attrition is above 30%, but reasoning for this not being systematic in a way that would imply bias is plausible, and sample composition evidence either side of the attrition is also encouraging. |
| **Threat 6: Measurement of Outcomes** | Low | Validated outcome measure with no evidence of floor or ceiling effects. |
| **Threat 7: Selective reporting** | Low | Well reported against planned analyses. |

- **Initial padlock score:** 3 Padlocks – Regression discontinuity design, powered to 0.21, attrition at 31%. Note that although attrition falls just above the 30% threshold for 3 padlocks, given the other strengths of the study and the greater relevance of attrition for external rather than internal validity in the context of the regression discontinuity design, dropping an additional padlock was not considered to be merited.
- **Reason for adjustment for threats to validity**: 0 Padlocks – Some concerns about confounding, but at a magnitude where they are likely handled by the covariate adjustment strategy employed.
- **Final padlock score:** Initial score adjusted for threats to validity = 3 Padlocks

# Appendix B: Effect size estimation

*Table 19: Effect size estimation*

| Outcome | Difference at the cutoff (95% CI) | Intervention group | | Comparison group | | Pooled standard deviation |
|---|---|---|---|---|---|---|
| | | n | Variance | n | Variance | |
| Primary analysis | 3.92 (1.59, 6.26) | 2329 | 175 | 8430 | 174 | 13.2 |
| Subgroup analysis - FSM | 7.37 (1.84, 12.9) | 424 | 159 | 885 | 171 | 12.9 |
| Subgroup analysis - EAL | 3.71 (-0.436, 7.86) | 699 | 140 | 997 | 173 | 12.6 |
| Exploratory analysis – sharp subset | 1.44 (-1.69, 4.57) | 704 | 163 | 2875 | 139 | 12.0 |

# Appendix C: Initial sample size calculations for school recruitment (without data)

*Table 20: Initial sample size calculations for school recruitment (without data)*

|  | Pupils | | | | Schools | | | |
|---|---|---|---|---|---|---|---|---|
|  | MDES | | | | MDES | | | |
|  | 0.200 | 0.230 | 0.280 | 0.385 | 0.200 | 0.230 | 0.280 | 0.385 |
| **RCT** | 555 | 420 | 285 | 152 | 14 | 11 | 7 | 4 |
|  |  |  |  |  |  |  |  |  |
| **SRD, RDDE=9** | 4995 | 3780 | 2565 | 1368 | 125 | 95 | 64 | 34 |
| **SRD, RDDE=14** | 7770 | 5880 | 3990 | 2128 | 195 | 147 | 100 | 53 |
| **SRD, RDDE=17** | 9435 | 7140 | 4845 | 2584 | 236 | 179 | 121 | 65 |
|  |  |  |  |  |  |  |  |  |
| **FRD, 0.9, RDDE=9** | 6144 | 4649 | 3155 | 1683 | 154 | 117 | 79 | 42 |
| **FRD, 0.9, RDDE=14** | 9557 | 7232 | 4908 | 2617 | 240 | 181 | 123 | 66 |
| **FRD, 0.9, RDDE=17** | 11605 | 8782 | 5959 | 3178 | 291 | 220 | 149 | 80 |
|  |  |  |  |  |  |  |  |  |
| **FRD, 0.7, RDDE=9** | 9890 | 7484 | 5079 | 2709 | 248 | 188 | 127 | 68 |
| **FRD, 0.7, RDDE=14** | 15385 | 11642 | 7900 | 4213 | 386 | 292 | 198 | 106 |
| **FRD, 0.7, RDDE=17** | 18681 | 14137 | 9593 | 5116 | 468 | 354 | 240 | 128 |
|  |  |  |  |  |  |  |  |  |
| **FRD, 0.5, RDDE=9** | 18681 | 14137 | 9593 | 5116 | 468 | 354 | 240 | 128 |
| **FRD, 0.5, RDDE=14** | 29060 | 21991 | 14923 | 7959 | 728 | 551 | 374 | 199 |
| **FRD, 0.5, RDDE=17** | 35287 | 26704 | 18120 | 9664 | 884 | 669 | 454 | 242 |

# Appendix D: Datasets and variables

*Table 21: LanguageScreen data (from OxEd)*

| Variable name | Description |
| --- | --- |
| school_nfer | NFER school number |
| school_urn | URN |
| school_name | School Name |
| school_postcode | School Postcode |
| pupil_upn | UPN |
| pupil_meaningless_identifier | Pupil Meaningless Identifier |
| pupil_firstname | Pupil First Name |
| pupil_lastname | Pupil Last Name |
| class_name | School Class |
| pupil_date_of_birth | Date of Birth |
| pupil_gender | Gender |
| pupil_eal | EAL Status |
| intervention_received | Recieved NELI programme (yes/no) |
| approx_neli_completion_date | Completion date of NELI programme |
| pre_neli_assessment_date | Baseline Assessment Date |
| pre_neli_standard_score_ev | Baseline Expressive Vocabulary Score |
| pre_neli_standard_score_rv | Baseline Receptive Vocabulary Score |
| pre_neli_standard_score_lc | Baseline Language Comprehension Score |
| pre_neli_standard_score_sr | Baseline Sentence Repetition Score |
| pre_neli_standard_score_ls | Baseline Language Screen Standard Score |
| pre_neli_percentile_score_ls | Baseline Percentile Score |
| pre_neli_assessment_id | Baseline Assessment ID |
| pre_neli_assessment_age_months | Assessment age at baseline |
| post_neli_assessment_date | Endline Assessment Date |
| post_neli_standard_score_ev | Endline Expressive Vocabulary Score |
| post_neli_standard_score_rv | Endline Receptive Vocabulary Score |
| post_neli_standard_score_lc | Endline Language Comprehension Score |
| post_neli_standard_score_sr | Endline Sentence Repetition Score |
| post_neli_standard_score_ls | Endline Language Screen Standard Score |
| post_neli_percentile_score_ls | Endline Percentile Score |

| post_neli_assessment_id | Endline Assessment ID |
|---|---|
| post_neli_assessment_age_months | Assessment age at endline |
| time_between_assessments_days | Time between assessment in days |
| time_between_assessments_weeks | Time between assessment in weeks |

**NPD data:**

1. PupilMatchingRefAnonymous_[term][yy]
2. Gender_[term][yy]
3. YearOfBirth_[term][yy]
4. MonthOfBirth_[term][yy]
5. EthnicGroupMajor_[term][yy]
6. EVERFSM_6_P_[term][yy]
7. EVERFSM_ALL_[term][yy]
8. FSMeligible_[term][yy]
9. LanguageGroupMajor_[term][yy]
10. PartTime_[term][yy]
11. SENprovisionMajor_[term][yy]
12. PrimarySENtype_[term][yy

**School-level characteristics** (derived from DfE's publicly available Get Information about Schools dataset):

1. EstablishmentTypeGroup (name)
2. GOR (name)
3. UrbanRural (name)
4. OfstedRating (name)

**School-level dosage** data (derived from Nuffield Foundation Ltd's TeachNELI delivery surveys):

1. What was the session number of the last group session delivered?
2. What is your average group session length?

**School-level fidelity** data (derived from OxEd's Futurelearn training platform):

1. Course1Invited_TAs
2. Course1InProgr_TAs
3. Course1Complete_TAs
4. Course2Invited_TAs
5. Course2InProgr_TAs
6. Course2Complete_TAs
7. Course3Invited_TAs
8. Course3InProgr_TAs
9. Course3Complete_Tas

# Appendix E: Preliminary analysis (using initial pupil-level dataset, August 2022)

1. Characteristics of the school sample.

The characteristics of schools analysed as part of the preliminary analysis were similar to those included the final analysis sample (shown in Table 24, Appendix J).

2. Analysis of LanguageScreen baseline scores to model cutoffs by class.

The cutoff for each class was calculated as described in the preliminary analysis methods. A histogram of these cutoffs is shown in Figure 9 below.

*Figure 9 - Histogram of class level cutoffs in preliminary analysis (n Class =584)*
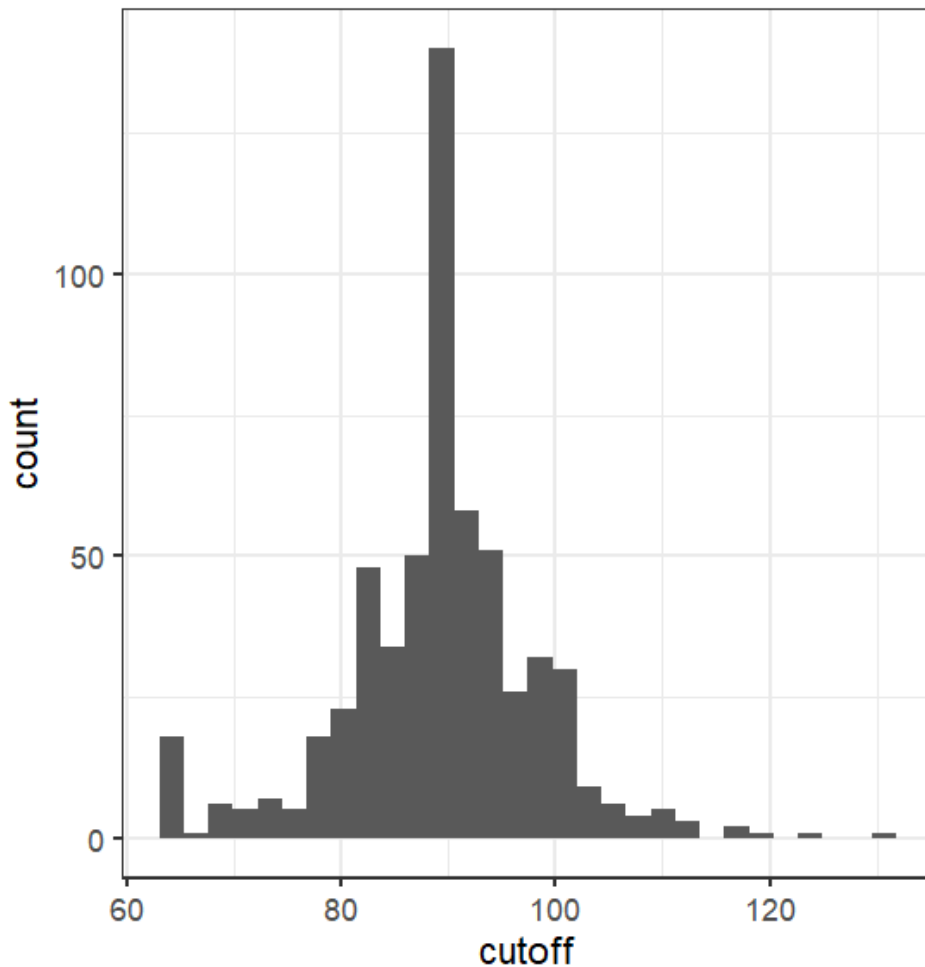


3. Graphical analysis to confirm the validity of the proposed FRD design by checking for treatment assignment either side of the cutoff and plotting the outcome against the cutoff to visually inspect for evidence of a discontinuity.

Figure 10 below shows the outcome LanguageScreen score against the cutoff adjusted pre intervention LanguageScreen score, close to the cutoff (±5). Although this graphical analysis does not provide obvious evidence of a discontinuity at the cutoff of 0, it also does not provide evidence which would preclude an FRD design. We determined it to be appropriate to proceed with FRD analysis nonetheless, given our understanding of the use of the LanguageScreen cutoff as the main criterion for selecting pupils for NELI.

*Figure 10 – Post Intervention LanguageScreen Score against Cutoff Adjusted Pre Intervention LanguageScreen Score*
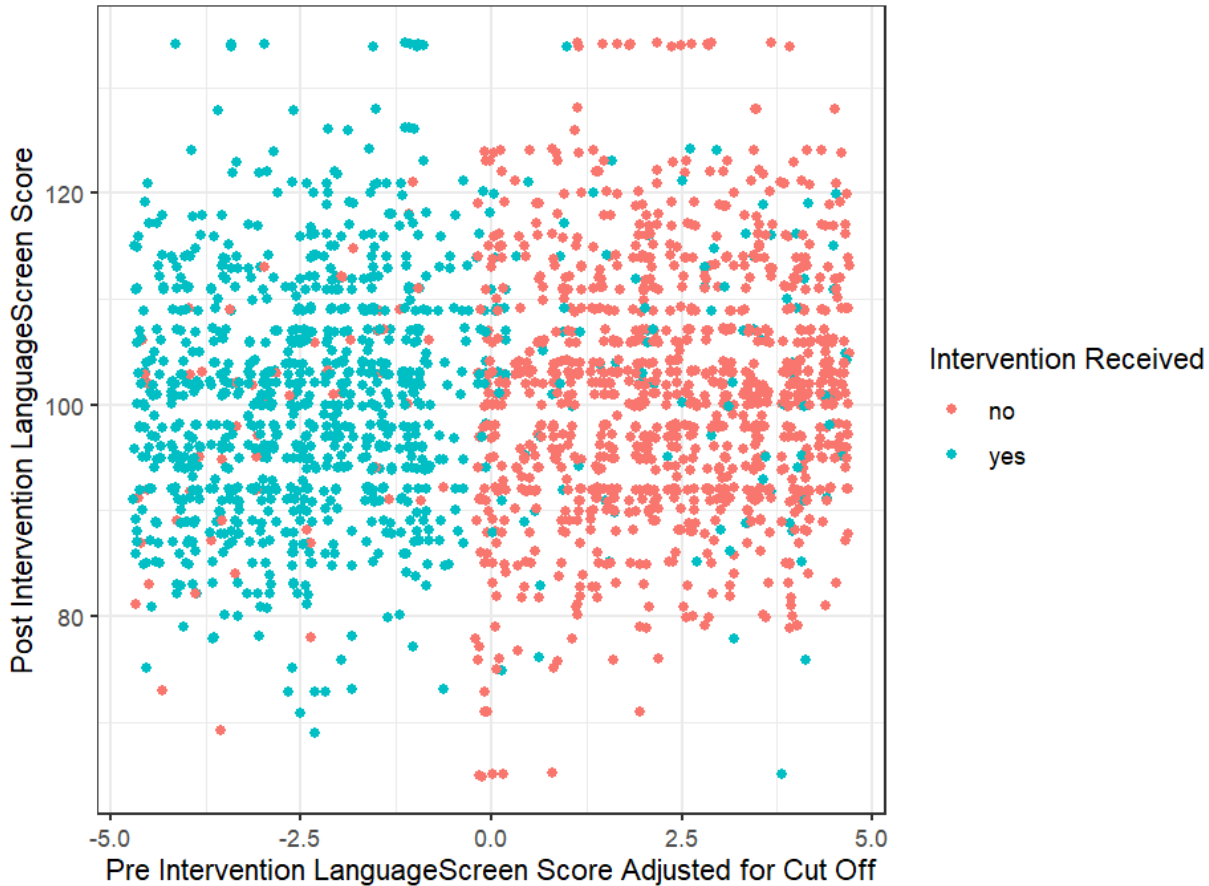


4. Estimating the probability of receiving treatment as a function of the LanguageScreen cutoff.

In the preliminary analysis, comparing the intervention received to whether the pre-intervention LanguageScreen score was above or below zero, we determined that the percentage of 'non-compliers' was 6% (Figure 2). When running the first-stage regression of the participation indicator on the forcing variable, and the indicator for being above or below the cutoff, the F statistic for the forcing variable was 85.85 and the F statistic for the indicator for being above or below the cutoff was 164.89. The low percentage of non-compliers and high F statistics suggest that the fuzziness in our data set is not extreme, so an FRD analysis is acceptable. Graphical illustration of the probability of intervention receipt against adjusted pre intervention LanguageScreen score (Figure 11 below) also supports this approach.

Figure 11 - Probability of receiving the intervention
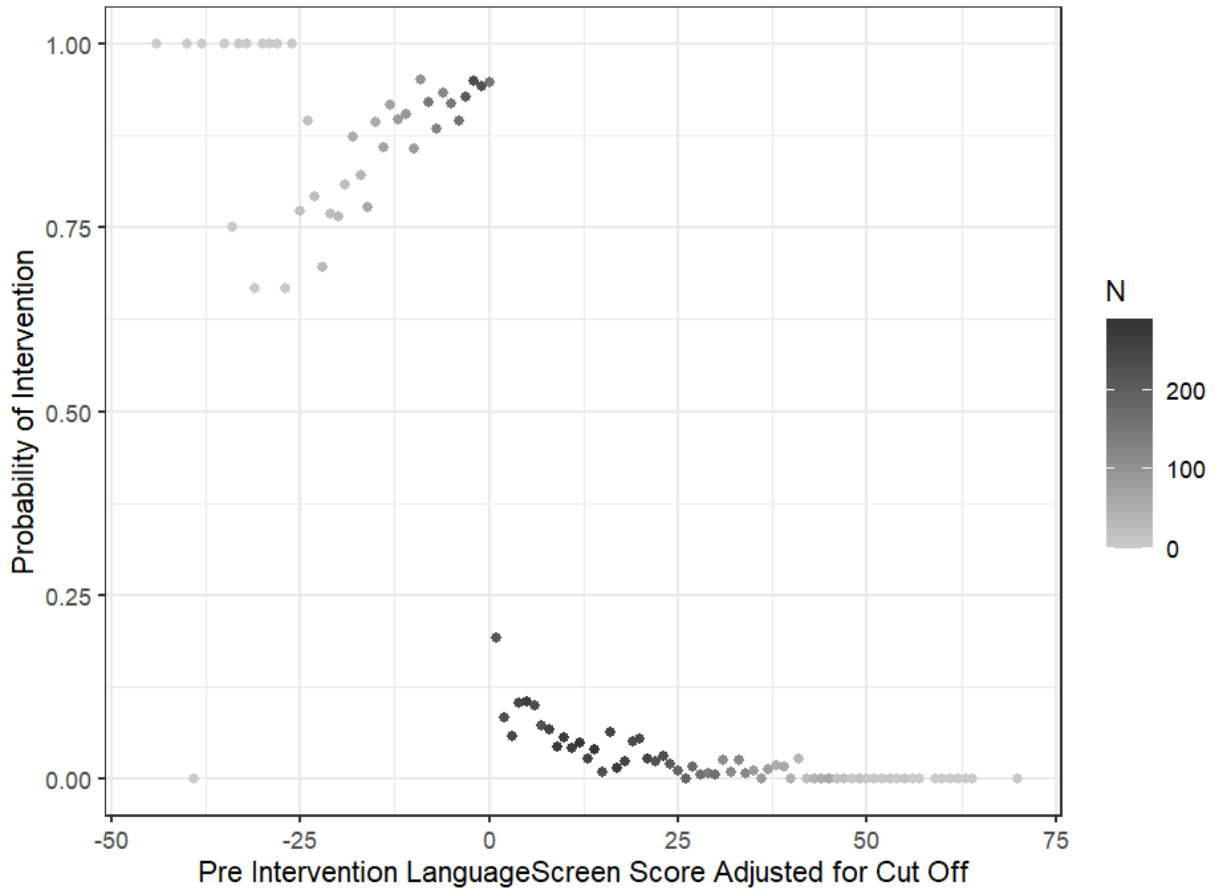


5. Determining the number of mass points in the data (due to non-continuous data on the running variable), and whether this precluded adopting a continuity-based RD approach (as per Cattaneo et al., 2020, pp.60-62).

We found that the 10,703 observations (with non-missing endline LanguageScreen scores) took 191 unique values and therefore the number of mass points was considered sufficiently large for a continuity-based approach.

6. Final power calculations to estimate the MDES with much greater precision than our initial calculations without data allowed.

Once preliminary pupil-level data had been collected, we were able to use this data to inform a second sample size calculation. This utilised the rdsampsi function in the *rdpower* R package (Cattaneo et al., 2019). Data variability and fuzziness was taken from the preliminary data. Two mean square error bandwidth selectors (above and below cut off) were selected as the bandwidth selection procedure. Mass points were addressed by requiring that initial bandwidths contain at least 10 unique values. Pre-intervention LanguageScreen score was included as a baseline covariate. Due to the normalisation of outcome scores to a class-level cutoff, we expect no variability in mean scores across classes so a multilevel model approach was not implemented. While the distribution of residuals in the final dataset was not known at the time of the preliminary analysis, the preliminary data suggested that heteroskedasticity might be expected. Therefore, a heteroskedasticity-robust plug-in residuals variance estimator was selected as the variance-covariance estimator. Sample size calculations were at a pupil level and class numbers were based on average cluster size. The results of this sample size calculation are shown in Table 4 in the main report. This part of the preliminary analysis was repeated for the final analysis sample (also shown in Table 4).

7.  Balance checks.

Covariate balance for pupils included in the preliminary analysis is shown in Table 22 below. Note that some of the covariates anticipated for the final analysis are only available after the match to the NPD data so are not reported here. This part of the preliminary analysis was repeated for the final analysis sample (Appendix J).

*Table 22: Covariate balance for pupils included in the preliminary analysis*

| School-level (categorical) | Intervention group | | Comparison group | |
|---|---|---|---|---|
| | n/N | Percentage | n/N | Percentage |
| N | 2589 | | 8114 | |
| Region | | | | |
| East Midlands | 183/2589 | 7.1 | 497/8114 | 6.1 |
| East of England | 268/2589 | 10.4 | 909/8114 | 11.2 |
| London | 522/2589 | 20.2 | 1330/8114 | 16.4 |
| North East | 97/2589 | 3.7 | 404/8114 | 5.0 |
| North West | 370/2589 | 14.3 | 990/8114 | 12.2 |
| South East | 409/2589 | 15.8 | 1606/8114 | 19.8 |
| South West | 246/2589 | 9.5 | 855/8114 | 10.5 |
| West Midlands | 250/2589 | 9.7 | 674/8114 | 8.3 |
| Yorkshire and the Humber | 244/2589 | 9.4 | 849/8114 | 10.5 |
| Rural or Urban | | | | |
| Rural | 533/2589 | 20.6 | 1703/8114 | 21 |
| Urban | 2056/2589 | 79.4 | 6411/8114 | 79 |
| Ofsted Rating | | | | |
| Outstanding | 393/2589 | 15.2 | 1379/8114 | 17 |
| Good | 1788/2589 | 69.1 | 5403/8114 | 66.6 |
| Requires Improvement | 178/2589 | 6.9 | 513/8114 | 6.3 |
| Inadequate | 9/2589 | 0.3 | 0/8114 | 0 |
| Missing Ofsted Rating | 221/2589 | 8.5 | 819/8114 | 10.1 |

8.  Missing data.

See 'Missing data analysis and attrition' section main report.

**Appendix F: School information sheet**



# Impact evaluation of the Nuffield Early Language Intervention (NELI)

## 1. Why are you receiving this Information Sheet?

You are receiving this Information Sheet because your school has agreed to receive information about the impact evaluation of the NELI Programme.

**We are inviting schools who are participating in the NELI programme to help us understand the impact of the programme on children's language skills by agreeing to participate in this independent impact evaluation by the National Foundation for Educational Research (NFER). Schools that help us in this way will specifically contribute to the evidence base about the impact of NELI and will receive a thank you payment of up to £250 for their contribution towards the impact evaluation.**

## 2. Which schools can take part in NELI impact evaluation?

Schools that meet all the following criteria can take part:

- Schools participating in NELI Programme in the 2021/2022 academic year
- Schools who completed the initial NELI LanguageScreen assessment for the vast majority of pupils in Reception class(es) before their NELI programme delivery started. (*The assessment may not have been completed for a very small number of pupils, for example, pupils who were absent on the day of the assessment. Schools to which this applies can still take part in the NELI impact evaluation.*)

## 3. What will schools need to do for NELI impact evaluation?

Timeline and outline of key activities required:

| | |
|---|---|
| May 2022 – July 2022 | Headteacher signs a memorandum of understanding (MoU) to commit to the impact evaluation and confirm they are happy for NFER to receive their pupils' LanguageScreen data. In the MoU, school will also need to share the NELI Lead's contact details and school level information required for the impact evaluation. |
| June 2022 – July 2022 | Share parent letter with all parents of pupils in Reception class(es) and share any evaluation withdrawal requests with Oxford Education and Assessment Ltd (OxEd and Assessment Ltd). |
| June 2022 – July 2022 | Complete a final LanguageScreen assessment at the end of this summer term for all pupils who were initially assessed ahead of delivering the programme (i.e., both pupils who received NELI and those who did not). <br><br> Ensure all LanguageScreen assessment data from the app has fully uploaded to the OxEd and Assessment Ltd server. <br><br> NELI Lead completes the final 3 minute TeachNELI Delivery survey which will be available from 4th to 22nd of July. |

## 4. Who decides what data will be shared?

The school's Headteacher will need to give permission for your school to sign up to being a NELI impact evaluation school.

**Parents/ guardians** have the right to withdraw their child from data processing for this impact evaluation. They can do so by using the withdrawal portion of the parent letter that will be provided by NFER and distributed by schools. After distributing the parent letter, schools should give parents one week to return any withdrawal slips. If the school receives any withdrawal slips, they should email OxEd and Assessment Ltd. on support@teachneli.org to confirm the number of withdrawals received. OxEd and Assessment will then provide schools with information on how to ensure children who withdrew from data processing for this impact evaluation are not included in the evaluation.

## 5.    What are the benefits for my school?

By becoming a NELI impact evaluation school, **your school will help to strengthen the evidence relating to the impact of the NELI programme on Reception pupils' language skills. It will help us understand the impact of the programme when delivered at national scale** and how best to implement the similar programmes in the future. The evaluation report will be publicly available on the Education Endowment Foundation's (EEF) website.

Schools that participate in the impact evaluation and complete the project requirements will receive a thank you payment of up to £250 for their contribution to the evaluation.

## 6.    How does my school sign up?

To sign up, please complete and submit the MoU we have provided using the secure portal details we have shared with you. Once we have received your MoU, we will be in touch with the NELI lead regarding the next steps for this evaluation.

## 7.    Who can I contact for more information?

If you have any queries on the impact evaluation please contact NFER on ImpactNELI@nfer.ac.uk

Further information on the evaluation and a link to the impact evaluation privacy notice can also be found at: https://nfer.ac.uk/for-schools/participate-in-research/impact-evaluation-of-the-nuffield-early-language-intervention/

## 8.    About the evaluation

The EEF has commissioned the National Foundation for Educational Research (NFER) to independently evaluate the impact of the NELI Programme during the second year of the delivery. We are now approaching schools that are taking part in the NELI Programme in 2021 – 2022 to invite them to participate in the impact evaluation.

Schools' contributions are vital for the successful impact evaluation of the NELI Programme, which in turn will play a pivotal role in strengthening the evidence on how to best implement similar programmes at scale in the future, and on the impact the intervention can have on pupils' learning outcomes.

## Appendix G: Parent opt-out letter



# Impact Evaluation of the Nuffield Early Language Intervention (NELI)
## Parent/Carer form for **Withdrawal** from data sharing

Dear Parent / Guardian,

We are writing to let you know that your child's school is participating in the Impact Evaluation of the Nuffield Early Language Intervention (NELI). The programme has been running in the school since October 2021 and is designed to support Reception pupils. NELI involves providing targeted small group and one-to-one support for children who would benefit from additional support with their language and early literacy skills. The initiative is funded by the Department of Education; the independent impact evaluation is being carried out by the National Foundation for Educational Research (NFER) and is funded by the Education Endowment Foundation (EEF).

As part of the evaluation your child's school has committed to share data with NFER as the independent evaluator. The Privacy Notice summarises what personal data will be shared and protected. We have robust procedures in place to make sure that we comply with the requirements of GDPR. No individual child or school will be identified in any of the reporting of this project.

If you are **happy for your child's data to be used for this evaluation**, **you do not need to return the reply slip**. However, if you would prefer your child's data not to be shared, stored and used for this project, please complete the form below and return it to your child's teacher within one week of receiving this letter. If you would like to withdraw your child's data from the evaluation at any subsequent stages, please inform your child's teacher.  If you have any queries please contact us via email at ImpactNELI@nfer.ac.uk

Yours sincerely,

Kathryn Hurd

Head of Survey Operations

National Foundation for Educational Research

**Evaluation of Nuffield Early Language Intervention: Withdrawal form from data sharing for parents**

You only need to complete this form if you **DO NOT** wish your child's data to be shared, stored, and used for this evaluation.

☐ I **DO NOT** give permission for information about my child (including name, date of birth, EAL status and assessment results) to be shared (and linked to information contained in the National Pupil Database) for use in the Impact Evaluation of the Nuffield Early Language Intervention

Name of Child: _____

Name of School: _____

School's Postcode _____

Please return this form to your child's class teacher **if you do not want your child's data to be shared.**

Thank you very much

**Confidential when completed**

## Appendix H: Memorandum of Understanding (Online Document)

# Memorandum of Understanding for the Impact Evaluation of the Nuffield Early Language Intervention (NELI)

The EEF has commissioned the National Foundation for Educational Research (NFER) to independently evaluate the impact of the NELI programme during the second year of the delivery. We are now approaching schools that are taking part in the NELI programme in 2021/22 to invite them to participate in the impact evaluation.

Schools' contributions are vital for the successful the impact evaluation of the NELI programme, which in turn will play a pivotal role in strengthening the evidence on how best to implement similar programmes at scale in the future, and on the impact the intervention can have on pupils' learning outcomes.

Schools that meet all the following criteria can take part:

- Schools participating in NELI programme in 2021/22 academic year
- Schools who completed the initial NELI LanguageScreen assessment for the vast majority of pupils in Reception class(es) before their NELI programme delivery started

Schools that sign this MoU agree to the following:

- NFER to receive your pupils' LanguageScreen data.
- Share parent letter with all parents of pupils in Reception class(es) and share any evaluation withdrawal requests with Oxford Education and Assessment Ltd (OxEd and Assessment Ltd).
- NELI Lead to indicate which pupils are receiving NELI on the LanguageScreen website by clicking on the pupil and then 'Record an intervention'.
- Complete a final LanguageScreen assessment at the end of the 20-week programme or at the end of this summer term for all pupils who were initially assessed ahead of delivering the programme (I.e., both pupils who received NELI and those who did not).
- Ensure all LanguageScreen data from the app has fully uploaded to the OxEd and Assessment Ltd server
- NELI Lead to complete the third 3 minute TeachNELI Delivery survey which will be available from end of June till July.

Further information on the evaluation and the link to the impact evaluation privacy notice can be found at:

https://www.nfer.ac.uk/for-schools/participate-in-research/impact-evaluation-of-the-nuffield-early-language-intervention/

☐ I confirm that my school is happy to participate in the impact evaluation of NELI programme and NFER to receive our pupils' LanguageScreen data.

### School Details

Please check that the details we have for you are correct:

|  | Are your details correct? | Please amend if not |
|---|---|---|
| School Name |  |  |
| Headteacher |  |  |
| Tel. No. |  |  |
| Fax No. |  |  |
| Email |  |  |

Please share name and contact details for the NELI Lead at your school.

Title:

Forename:

Surname:

Job title:

Email address:

Re-enter email address:

Telephone/mobile number:

In July we will be contacting your school to collect bank details to enable us to make payments directly to your school's bank. Please share contact details of the school bursar who will be able to share the details with us in order to make the thank you payment.

Forename:

Surname:

Job title:

Email address:

Re-enter email address:

Telephone/mobile number:

Please answer these questions about NELI programme delivery in your school:

What is the number of NELI groups running in your school that include Reception pupils?

Do any of your NELI groups include pupils who are from different Reception classes? For example, a two-form entry school may have three NELI groups with one group including pupils from both Reception classes. (Yes/No)

Do any of your NELI groups include pupils from both Reception and other year groups? (Yes/No)

Did your school complete the initial LanguageScreen assessment for the vast majority of pupils in Reception classes before your NELI groups started? The assessments may not have been completed for a very small number of pupils, for example, pupils who were absent on the day of the assessment. (Yes/No)

Did your school select pupils to receive NELI based ONLY on their initial LanguageScreen assessment scores? Please answer 'no' if you also considered other factors (e.g., EAL, SEN, behavioural). (Yes/No)

## Impact Evaluation of the Nuffield Early Language Intervention

### Confirm

Please check the details below and then click the **Submit** button to send your details to NFER. You can amend your details if required by clicking the **Previous** button. Please be aware that once you press submit, you will no longer be able to change this information online. You will need to contact us using the number to the right if you need to update the information you have provided.

| School Details |
| --- |
| School Name: |
| Headteacher: |
| Phone: |
| Fax: |
| Email: |
| Full Name: |
| Job title: |
| Email: |
| Phone: |
| What is the total number of NELI groups running in your school that include Reception pupils? |
| Full Name: |
| Job title: |
| Email: |
| Phone: |
| |

| | |
| --- | --- |
| Do any of your NELI groups include pupils who are from different Reception classes? For example, a two-form entry school may have three NELI groups with one group including pupils from both Reception classes. | Yes/No |
| Do any of your NELI groups include pupils from both Reception and other year groups? | Yes/No |
| Did your school complete the initial LanguageScreen assessment for the vast majority of pupils in Reception classes before your NELI groups started? The assessments may not have been completed for a very small number of pupils, for example, pupils who were absent on the day of the assessment. | Yes/No |
| Did your school select pupils to receive NELI based ONLY on their initial LanguageScreen assessment scores? Please answer 'no' if you also considered other factors (e.g., EAL, SEN, behavioural). | Yes/No |

<< Previous                    Submit

## Appendix I: Privacy Notice

# Privacy notice for schools participating in the impact evaluation of Nuffield Early Language Intervention (NELI)

## 1 Why are we collecting this data?

The Education Endowment Foundation (EEF) has commissioned the National Foundation for Educational Research (NFER) to carry out an independent impact evaluation of the Nuffield Early Language Intervention (NELI). The aim of this evaluation is to measure the impact of NELI on participating pupils' language skills.

NELI is a government-funded 20-week initiative. It aims to develop children's vocabulary, listening and narrative skills and in the last 10 weeks also involves work to develop phonological awareness and early letter-sound knowledge as foundations for early literacy skills. The programme is managed by the Nuffield Foundation Education Limited and their delivery partner Oxford Education and Assessment Ltd (OxEd and Assessment Ltd).  More information about the NELI programme can be found **here**.

This document outlines how personal data of the school staff and pupils will be collected and processed as part of the impact evaluation.

Note: NFER is only contacting the member of school staff who agreed to be contacted about the impact evaluation when signing up for the programme.

## 2 Who makes decisions about how personal data is used?

The Department for Education (DfE) as the data controller makes decisions about how personal data is used for this impact evaluation.  It has determined the means and purpose of the processing. EEF is the data processor and NFER is a sub-processor; it follows the instructions of DfE when processing personal data.

## 3 How is the use of personal data lawful?

For the use of your personal data to be lawful, the DfE as the data controller for this evaluation needs to ensure that one of the conditions in data protection legislation are met.  For the impact evaluation, the relevant condition is:

Article 6 (1) (e) UK GDPR to perform a public task

The statutory basis for these tasks are set out in:

- S.10 The Education Act 1996: The Secretary of State shall promote the education of the people of England and Wales.

The legal basis for processing pupils' special personal data is covered by:

GDPR Article 9 (2) (j) which states that 'processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) (as supplemented by section 19 of the 2018 Act) based on domestic law which shall be proportionate to the aim pursued, respect the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject'.

We do not believe this processing will cause damage or distress to the data subjects.  The outcomes of the evaluation will not result in the creation of measures or decisions being made about the data subjects.

# 4 How will personal data be obtained?

On signing up for the NELI programme in the 2021/22 academic year, schools were asked if they wanted to be contacted about being involved in the impact evaluation. Nuffield Foundation Education Limited and its delivery partner OxEd and Assessment Ltd have shared details of the staff contact at interested schools who consented to be contacted about the evaluation with NFER. Any additional personal data required by NFER will be collected directly from the staff involved when schools agree to participate.

Pupil data will also be analysed as part of the impact evaluation. NFER will not obtain this directly but will use the Secure Research Service (SRS) hosted by the Office for National Statistics (ONS).

# 5 What personal data is being collected by this project?

**<u>NELI Programme Signatory</u>**

OxEd and Assessment Ltd will share the following personal data with NFER:

- Name of the contact in the school who signed the MOU for the school to participate in NELI (typically the Headteacher),
- their job roles or job titles and
- their contact details such as email address and/or telephone number.

The NFER will use this data to contact each school, NFER will share an MoU which will need to be signed by the headteacher to confirm the school's participation in the impact evaluation. As part of the MoU, NFER will also collect some school-level data about NELI programme delivery in each school participating in the impact evaluation.

All schools participating in the evaluation will be asked by the NFER to provide the following details:

- NELI Lead name,
- NELI Lead job role or job title
- NELI Lead contact details such as email address and/or telephone number
- School Bursar name
- their job role or job title,
- their contact details such as email address and/or telephone number

The NFER will use the contact details of the NELI Lead to remind them to complete the final LanguageScreen assessment, NELI status indicator and the NELI delivery survey. School bursar contact details will be used to collect school's payment details to send out the thank you payment after the evaluation activities are completed.

**<u>Reception pupils</u>**

For schools that have signed up to take part in the impact evaluation, OxEd and Assessment Ltd will share the following data for reception pupils registered on the LanguageScreen app (except where parents withdrew their child from data processing for this impact evaluation) with the DfE's National Pupil Database (NPD) team.

The pupil personal data will include:

- Unique Pupil Identifier (UPN)
- First name
- Last name
- Date of birth
- Gender
- School class
- Whether the pupil is a learner of English as an Additional Language (EAL)
- Whether the pupil has been selected for the NELI programme
- Initial LanguageScreen assessment results

- Final LanguageScreen assessment results

DfE's NPD[20] will match above pupil data and will add the following NPD variables to the dataset:

- Whether the pupil is a learner of English as an Additional Language (EAL) – collected again in case these data in LanguageScreen are missing
- Special Educational Needs provision
- Nature of pupils' primary special educational need
- Child's ethnic code
- Whether the pupil is part-time or not
- Free School Meals eligibility
- Early Years Foundation Stage Profile assessment data

The NPD team will replace pupil names and UPN with a pseudonym (such as a reference number). This process is known as pseudonymisation[2]. NFER researchers can only access and analyse this dataset within the ONS SRS and any output will be checked to ensure that no pupils can be identified from the analysis.

Above pupil data will be matched to data about school characteristics and school-level measures that NFER derives based on the number of NELI staff attending the training and the number of NELI sessions delivered in a school. NFER will create a school-level dataset and will upload it on SRS for analysis. This is derived from programme delivery surveys fielded by Nuffield Foundation Education Limited and training data held within OxEd and Assessment Ltd's FutureLearn platform.

NFER needs to check that its approach to the analysis will answer the research questions. It will therefore be necessary to test this on subset of pupil data detailed above. OxEd will directly provide NFER with a pseudonymised dataset (containing gender, month and year of birth, EAL status and assessment data for each pupil, which is associated with class and school) through a secure data portal to carry out this preliminary analysis.

---

[20] The National Pupil Database (NPD) is a collection of data relating to education in England compiled by the Department for Education (DfE). The NPD is used by the DfE to inform policy and approved users can apply for extracts of it "for the purpose of promoting the education or wellbeing of children in England". If an application is approved, pupil data extracted from the tool is shared with the NPD team and matched to the requested datasets. The NPD team then send the matched set to the Office for National Statistics (ONS) (Secure Research Service) where NFER's approved researchers perform their analysis.

# 6    Who will personal data be shared with?

**No schools or individuals will be named in any report for this project.**

Your school already shares data directly with OxEd and Assessment Ltd as part of the delivery of the programme.  They will share these data with NFER and the NPD team at DfE as described in section 5 of this privacy notice.

After three months from the completion of the study, pseudonymised[21] pupil data will be added to the Education Endowment Foundation (EEF) archive. The EEF archive is hosted by the ONS and managed by the EEF archive manager. This will enable DfE and other research teams to use the pseudonymised data as part of subsequent research. The pseudonymised data may also be linked to other relevant datasets after archiving.

The staff personal data collected for the evaluation will not be archived and will be deleted once the evaluation activities have been completed.

# 7 Is personal data being transferred outside of the European Economic Areas (EEA)?

No personal data being processed as part of this evaluation is being transferred outside of the EEA.

# 8    How long will personal data be retained?

The NFER will securely delete any personal data relating to the evaluation one year after the publication of the final report, currently expected to be June 2023.

Pupil names and UPN will be deleted after the data is linked to the NPD, expected to be November 2022. The pseudonymised data set will be stored indefinitely in the EEF archive to enable researchers to track the impact of the programme on attainment at subsequent educational stages.

# 9    How is the security of data maintained?

NFER has measures in place to prevent personal data being accidentally lost, used or accessed in an unauthorised way, altered or disclosed. NFER will limit access to personal data to staff members who have a business need to see it.

Arrangements for protection of personal data processed for the evaluation are below.

NFER has been certified to ISO27001 (GB17/872763) the internal standard for information security and holds Cyber Essentials Plus (IASME-CEP-004922). NFER operates Microsoft Windows Operating Systems and industry standard enterprise software such as databases and email, all managed to recognised industry standards with a full patching regime. All NFER laptops and mobile storage devices are encrypted and accessed with PIN-codes and strong passwords.  Annual penetration tests are carried out by a CHECK-accredited supplier and remediation undertaken. We use a replicated disaster recovery service (RDRS) which allows the business to continue to operate in the event of failure. Any personal data which is shared with us is transferred using our secure portal and is encrypted in transit (HTTPS and TLS 1.2).

---

[21] Pseudonymisation is a technique that replaces or removes information (like names or other meaningful identifiers) in a data set that identifies an individual.

## 10    What rights do I have over my personal data?

Any school or individual can withdraw from their data being processed. DfE, EEF and the NFER appreciate schools' and staff's support in collecting this data since it is very important for the validity of the results. Should you withdraw from the evaluation, the DfE and NFER will still use the evaluation data that the school has provided up to that point and link it to NPD unless you indicate otherwise.

Under data protection legislation, individuals have the right:

- to request access to information that we hold about them (subject access request)
- to have their personal data rectified, if it is inaccurate or incomplete
- to request the deletion or removal of personal data where there is no compelling reason for its continued processing
- to restrict our processing of pupil's personal data (for example, permitting its storage but no further processing)
- to object to our processing
- not to be subject to decisions based purely on automated processing where it produces a legal or similarly significant effect on the pupil

If at any time you wish us to withdraw your data from the evaluation or correct errors in it, please contact ImpactNELI@nfer.ac.uk

DfE determines the purposes and means of processing personal data as part of this project. Please see the DfE's Personal Information Charter for further information and contact details for their Data Protection Officer.

## 11    Who can I contact about this project?

The NFER is responsible for the day-to-day management of this impact evaluation. Contact ImpactNELI@nfer.ac.uk with any queries.

If you have any questions about how we use your personal information, please contact DfE and quote 'Impact evaluation of NELI' as a reference.

If you want to contact the Data Protection Officer (DPO), please contact DfE and mark it 'for the attention of the DPO'. If you have a concern about the way this project processes personal data, we request that you raise your concern with the DfE in the first instance.  If you remain dissatisfied, you can contact the Information Commissioner's Office, the body responsible for enforcing data protection legislation in the UK, at https://ico.org.uk/concerns/.

## 12   Updates

We keep this privacy notice under review to make sure it is up to date and accurate.

# Appendix J: Additional analysis outputs

*Table 23 - Primary analysis models*

| Outcome | N Class | Intervention group | | Comparison group | | Effect size | | |
|---|---|---|---|---|---|---|---|---|
| | | N pupil total | N pupil in bandwidth | N pupil total | N pupil in bandwidth | N pupil in bandwidth (intervention; comparison) | Hedges g (95% CI) | p-value |
| LanguageScreen Score – Model 1, no covariates | 510 | 2329 | 1449 | 8430 | 3494 | 4943 (1449; 3494) | 0.384 (0.195, 0.573) | <0.001 |
| LanguageScreen Score – Model 2, baseline score only as a covariate | 510 | 2329 | 1263 | 8430 | 2955 | 4218 (1263; 2955) | 0.269 (0.0979, 0.439) | 0.002 |
| LanguageScreen Score – Model 3, all covariates i.e., primary analysis | 510 | 2329 | 1269 | 8430 | 3207 | 4476 (1269; 3207) | 0.297 (0.120, 0.474) | <0.001 |

Table 24: Baseline characteristics of groups as recruited (with NELI indicator)

| School-level (categorical) | % in population‡ (open primary schools) | % in sample (NELI wave 2 scale-up registered schools) | % in final sample for analysis | Comparison group (in schools recruited to the evaluation and which provided NELI indicator data – not final sample for analysis) | | Intervention group (in schools recruited to the evaluation and which provided NELI indicator data – not final sample for analysis) | |
|---|---|---|---|---|---|---|---|
| | | | | n/N | % | n/N | % |
| **N** | 16,784 schools | 4,422 schools | 356 schools | 12,514 pupils | | 3,056 pupils | |
| **School Governance** | | | | | | | |
| Academies | 39.0 | 35.3 | 31.5 | 4117/12514 | 32.9 | 994/3056 | 32.5 |
| Free Schools | 1.5 | 2.2 | 1.1 | >116*/12514 | >0.9* | >42*/3056 | >1.4* |
| LA Maintained Schools | 59.5 | 61.5 | 67.1 | 8271/12514 | 66.1 | 2010/3056 | 65.8 |
| Special Schools | 0 | 1.0 | 0.3 | <10*/12514 | <0.1* | <10*/3056 | <0.3* |
| **Region** | | | | | | | |
| East Midlands | 9.8 | 10 | 9.0 | 853/12514 | 6.8 | 226/3056 | 7.4 |
| East of England | 11.9 | 12.3 | 11.5 | 1357/12514 | 10.8 | 316/3056 | 10.3 |

| School-level (categorical) | % in population‡ (open primary schools) | % in sample (NELI wave 2 scale-up registered schools) | % in final sample for analysis | Comparison group (in schools recruited to the evaluation and which provided NELI indicator data – not final sample for analysis) | | Intervention group (in schools recruited to the evaluation and which provided NELI indicator data – not final sample for analysis) | |
|---|---|---|---|---|---|---|---|
| London | 10.7 | 9.2 | 12.1 | 2250/12514 | 18.0 | 576/3056 | 18.8 |
| | | | | n/N | % | n/N | % |
| North East | 5.1 | 5.1 | 4.5 | 523/12514 | 4.2 | 105/3056 | 3.4 |
| North West | 14.6 | 14.5 | 14.6 | 1652/12514 | 13.2 | 447/3056 | 14.6 |
| South East | 15.5 | 15.4 | 17.1 | 2290/12514 | 18.3 | 471/3056 | 15.4 |
| South West | 11.2 | 13.1 | 12.4 | 1171/12514 | 9.4 | 317/3056 | 10.4 |
| West Midlands | 10.6 | 10.3 | 8.4 | 1177/12514 | 9.4 | 302/3056 | 9.9 |
| Yorkshire and the Humber | 10.6 | 10.0 | 10.4 | 1241/12514 | 9.9 | 296/3056 | 9.7 |
| **Rural or Urban** | | | | | | | |
| Rural | 29.0 | 37.4 | 30.6 | 2339/12514 | 18.7 | 622/3056 | 20.4 |
| Urban | 71.0 | 62.6 | 69.4 | 10175/12514 | 81.3 | 2434/3056 | 79.6 |
| **Ofsted Rating** | | | | | | | |
| Outstanding | 11.2 | 12.9 | 13.8 | 2145/12514 | 17.1 | 475/3056 | 15.5 |
| Good | 68.4 | 68.3 | 68.3 | 8362/12514 | 66.8 | 2111/3056 | 69.1 |
| Requires Improvement | 6.3 | 5.9 | 6.7 | 670/12514 | 5.4 | 163/3056 | 5.3 |
| Inadequate | 0.4 | 0.9 | 0 | 20/12514 | 0.2 | 11/3056 | 0.4 |
| Missing Ofsted Rating | 13.7 | 12.1 | 11.2 | 1317/12514 | 10.5 | 296/3056 | 9.7 |

| School-level (categorical) | % in population‡ (open primary schools) | % in sample (NELI wave 2 scale-up registered schools) | % in final sample for analysis | Comparison group (in schools recruited to the evaluation and which provided NELI indicator data – not final sample for analysis) | | Intervention group (in schools recruited to the evaluation and which provided NELI indicator data – not final sample for analysis) | |
|---|---|---|---|---|---|---|---|
| | | | | n/N | % | n/N | % |
| **Pupil-level (categorical)** | | | | n/N | % | n/N | % |
| **Gender** | | | | | | | |
| Female | | | | 6198/12514 | 49.5 | 1346/3056 | 44.0 |
| Male | | | | 6316/12514 | 50.5 | 1710/3056 | 56.0 |
| **FSM-eligibility status** | | | | | | | |
| Non FSM | | | | 10573/12514 | 84.5 | 2383/3056 | 78.0 |
| FSM | | | | 1923/12514 | 15.4 | >663*/3056 | >21.7* |
| FSM Missing | | | | 18/12514 | 0.1 | <10*/3056 | <0.3* |
| **EAL Status** | | | | | | | |
| Non EAL | | | | 10518/12514 | 84.0 | 1884/3056 | 61.6 |
| EAL | | | | 1996/12514 | 16.0 | 1172/3056 | 38.4 |
| **SEN status** | | | | | | | |
| Non SEN | | | | 11600/12514 | 92.7 | 2536/3056 | 83.0 |
| SEN | | | | 914/12514 | 7.3 | 520/3056 | 17.0 |
| **Pupil-level (continuous)** | | | | n/N | Mean (SD) | n/N | Mean (SD) |
| Age in months in July 2022 | | | | 12514 | 64.6 (3.6) | 3056 | 64.4 (3.9) |

* Please note ONS statistical disclosure controls for DfE data prevent the reporting of cell counts lower than 10 at individual level. In this table, limits have been included to avoid statistical disclosure where pupil numbers are less than 10 for any level within a category.

‡ Calculated from a [Get Information About Schools](#) extract downloaded on 24/02/2023 and subset to include only schools where the Establishment Status (name) field was 'Open' or 'Open, but proposed to close' and the Phase of Education (name) field was 'Primary' or 'Middle deemed primary'.

*Table 25: Covariate balance for pupils included in the sharp subset analysis model described in point 5 of the Additional analyses and robustness checks*

| School-level (categorical) | All pupils entering into the primary analysis model | | | | Pupils within the MSE-optimal bandwidth | | | |
| | Intervention group | | Comparison group | | Intervention group | | Comparison group | |
| | n/N | % | n/N | % | n/N | % | n/N | % |
|---|---|---|---|---|---|---|---|---|
| N | 2329 | | 8430 | | 1539 | | 3111 | |
| Region | | | | | | | | |
| East Midlands | 163/2329 | 7 | 575/8430 | 6.8 | 111/1539 | 7.2 | 216/3111 | 6.9 |
| East of England | 238/2329 | 10.2 | 929/8430 | 11 | 167/1539 | 10.9 | 372/3111 | 12 |
| London | 408/2329 | 17.5 | 1332/8430 | 15.8 | 255/1539 | 16.6 | 482/3111 | 15.5 |
| North East | 83/2329 | 3.6 | 401/8430 | 4.8 | 66/1539 | 4.3 | 136/3111 | 4.4 |
| North West | 351/2329 | 15.1 | 1011/8430 | 12 | 222/1539 | 14.4 | 397/3111 | 12.8 |
| South East | 383/2329 | 16.4 | 1618/8430 | 19.2 | 254/1539 | 16.5 | 605/3111 | 19.4 |
| South West | 249/2329 | 10.7 | 909/8430 | 10.8 | 164/1539 | 10.7 | 344/3111 | 11.1 |
| West Midlands | 243/2329 | 10.4 | 781/8430 | 9.3 | 156/1539 | 10.1 | 286/3111 | 9.2 |
| Yorkshire and the Humber | 211/2329 | 9.1 | 874/8430 | 10.4 | 144/1539 | 9.4 | 273/3111 | 8.8 |
| Rural or Urban | | | | | | | | |
| Rural | 514/2329 | 22.1 | 1747/8430 | 20.7 | 356/1539 | 23.1 | 698/3111 | 22.4 |
| Urban | 1815/2329 | 77.9 | 6683/8430 | 79.3 | 1183/1539 | 76.9 | 2413/3111 | 77.6 |
| Ofsted Rating | | | | | | | | |
| Outstanding | 350/2329 | 15 | 1393/8430 | 16.5 | 219/1539 | 14.2 | 445/3111 | 14.3 |
| Good | 1640/2329 | 70.4 | 5598/8430 | 66.4 | 1094/1539 | 71.1 | 2180/3111 | 70.1 |
| Requires Improvement | 138/2329 | 5.9 | 547/8430 | 6.5 | 84/1539 | 5.5 | 183/3111 | 5.9 |

| School-level (categorical) | All pupils entering into the primary analysis model | | | | Pupils within the MSE-optimal bandwidth | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Intervention group | | Comparison group | | Intervention group | | Comparison group | |
| | n/N | % | n/N | % | n/N | % | n/N | % |
| Missing Ofsted Rating | 201/2329 | 8.6 | 892/8430 | 10.6 | 142/1539 | 9.2 | 303/3111 | 9.7 |
| Pupil-level (categorical) | n/N | Percentage | n/N | Percentage | | | | |
| Gender | | | | | | | | |
| Female | 1050/2329 | 45.1 | 4167/8430 | 49.4 | 704/1539 | 45.7 | 1476/3111 | 47.4 |
| Male | 1279/2329 | 54.9 | 4263/8430 | 50.6 | 835/1539 | 54.3 | 1635/3111 | 52.6 |
| FSM-eligibility status | | | | | | | | |
| Non FSM | 1818/2329 | 78.1 | 7204/8430 | 85.5 | 1200/1539 | 78 | 2527/3111 | 81.2 |
| FSM | >501/2329 | >21.5 | >1216/8430 | >14.4 | >329/1539 | >21.4 | >574/3111 | >18.5 |
| FSM Missing | <10/2329 | <0.4 | <10/8430 | <0.1 | <10/1539 | <0.6 | <10/3111 | <0.3 |
| EAL Status | | | | | | | | |
| Non EAL | 1455/2329 | 62.5 | 7229/8430 | 85.8 | 1073/1539 | 69.7 | 2485/3111 | 79.9 |
| EAL | 874/2329 | 37.5 | 1201/8430 | 14.2 | 466/1539 | 30.3 | 626/3111 | 20.1 |
| SEN status | | | | | | | | |
| Non SEN | 1957/2329 | 84 | 7862/8430 | 93.3 | 1306/1539 | 84.9 | 2828/3111 | 90.9 |
| SEN | 372/2329 | 16 | 568/8430 | 6.7 | 233/1539 | 15.1 | 283/3111 | 9.1 |

*Table 26 – Missing data analysis model*

| Variable | Odds ratio | *p*-value |
|---|---|---|
| Pre-intervention LanguageScreen Score | 1.037 (1.032, 1.041) | <0.001 |
| Gender | 1.043 (0.941, 1.157) | 0.947 |
| FSM | 1.336 (1.155, 1.546) | <0.001 |
| EAL | 1.130 (0.960, 1.330) | 0.554 |
| SEN | 1.198 (0.989, 1.453) | 0.303 |