2025 Valedictory RCTs in the Social Sciences Conference

NELI scale-up (wave 2) impact evaluation

Andrew Smith, Senior Evaluation Analyst, NFER

20th May 2025

NFER Project team: Jack Worth, Aarti Sahasranaman, Ruth Staunton, Kathryn Hurd, Jishi Jose, Max Falinski

Nuffield Early Language Intervention

:: NFER
National Foundation for
Educational Research
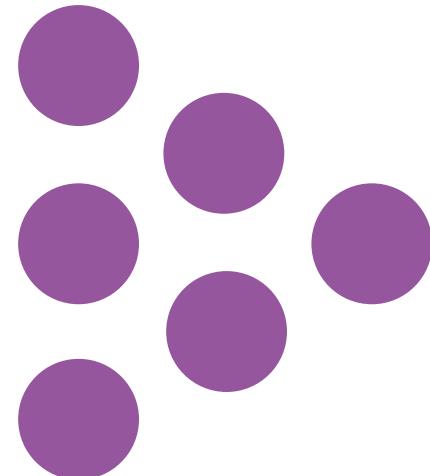
- Designed for children in Reception (4-5 years)

- Aims to develop children's vocabulary, listening and narrative skills

- Member of school staff (usually a Teaching Assistant) delivers:

  - 3 x 30 minute small **group sessions** each week to a targeted group of around 3-6 pupils for 20 weeks.

  - 2 x 15 minute **1:1 sessions** per week.

NELI: LanguageScreen

NELI: LanguageScreen

Hogwart School, OX1 1XO

Press Esc to exit full screen

Reception, NELI

First Names	Surname	DOB	Age (Months)	Year	Class	EV	RV	SR	LC	Score	Percentile	Description	Alert
						115	121	121	134	134	99	Much Above Average	🟢
						124	134	134	118	134	99	Much Above Average	🟢
						126	121	134	126	134	99	Much Above Average	🟢
						134	118	120	121	128	97	Much Above Average	🟢
						134	114	128	110	126	96	Much Above Average	🟢
						134	103	128	112	122	93	Much Above Average	🟢
						122	121	112	113	121	92	Much Above Average	🟢
						121	112	134	86	120	91	Much Above Average	🟢
						111	118	128	109	118	89	Much Above Average	🟢
						111	100	113	118	113	81	High Average	🟢
						105	112	123	98	112	79	High Average	🟢
						108	114	103	104	108	72	Average	🟢
						93	118	120	93	106	66	Average	🟢
						111	105	121	88	106	67	Average	🟢
						96	107	116	101	105	64	Average	🟢
						118	102	108	92	104	63	Average	🟢
						93	100	116	83	99	49	Average	🟢
						96	111	93	83	96	42	Average	🟢
						96	87	103	104	96	42	Average	🟢
						100	102	95	77	95	37	Average	🟢
						93	100	96	83	94	35	Average	🟢
						88	87	90	97	88	22	Low Average	🟠
						78	85	93	90	83	14	Below Average	🟠
						76	86	95	77	82	12	Below Average	🟠

Key: **EV** Expressive Vocabulary **RV** Receptive Vocabulary **SR** Sentence Repetition **LC** Listening Comprehension (all subtest scores are standard scores)
Score LanguageScreen standard score (average = 100) **Percentile** LanguageScreen percentile rank
Alert: 🟢 **Green** No concerns 🟠 **Amber** Slight concerns 🔴 **Red** Clear concerns

NELI: impact evaluations

- **Pilot study (Fricke et al., 2013)**

  - Two-arm RCT (pupil-level randomisation, 15 schools, 180 pupils)

    - 30 week programme: positive effect in intervention group across a range of measures; immediately post-intervention, and at 6 months.

- **Efficacy trial (Sibieta et al., 2016; commissioned by EEF)**

  - Three-arm RCT (pupil-level randomisation, 34 schools, 394 pupils)

    - 30 week programme: ES (language skills) **0.27** (0.07, 0.46), 4 months additional progress.

    - 20 week program: ES (language skills) **0.16** (-0.02, 0.34), 2 months additional progress.

NELI: impact evaluations

- **Effectiveness trial (Dimova et al., 2020; commissioned by EEF)**

  - Two-arm RCT (school-level randomisation, 193 schools, 1,156 pupils)

    - Primary outcome (language skills): ES **0.26** (0.17, 0.35), 3 months additional progress.

    - Secondary outcome (LanguageScreen): ES **0.36** (0.22, 0.47).

- **Scale-up (wave 2) impact evaluation (Smith et al., 2023; commissioned by EEF)**

  - Primary outcome (LanguageScreen)

  - Separate implementation and Process Evaluation (Disley et al., 2023; RAND Europe; commissioned by EEF)

NELI scale-up: waves 1 and 2

- **Wave 1 (2020/21):** 6,672 schools, £9 million DfE funding

- **Wave 2 (2021/22):** 4,422 schools registered for NELI, £8 million DfE funding

  - Targeting of schools: i) with a higher proportion of FSM-eligible pupils; ii) schools in local authorities in the bottom third of Ofsted ratings; iii) schools in Opportunity Areas (OAs)

- **Subsequent waves: 2022/23, 2023/24**

Timeline and Methodology

Scale-up (wave 2) impact evaluation: high-level timeline

NFER
National Foundation for Educational Research

- **2021**

  - **May - October**, recruitment (*to programme*) of NELI wave 2 schools (DfE, OUP).

  - **September – December**, online training for school staff (OxEd).

- **2022**

  - **January**, intervention delivery scheduled to begin in schools (*baseline testing*)

  - **March**, NFER appointed to undertake scale-up (wave 2 impact evaluation).

  - **June**, NFER approached schools - *recruitment to evaluation*.

  - **July 2022,** intervention delivery scheduled to be completed (*endline testing*).

  - **December 2022**, data for final analysis matched and available in NPD.

- **2023**

  - **January - March**, data analysis and reporting

Scale-up impact evaluation: overview

:::NFER
National Foundation for
Educational Research

- Aimed to assess the impact of NELI when delivered at scale in real-world conditions

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

  - *all pupils (RQ1);*
  - *FSM-eligible pupils (RQ2 - subsample);*
  - *EAL pupils (RQ3 - subsample)?*

- *How does the impact of NELI on pupils' oral language outcomes vary by:*

  - *dosage (RQ 4 - exploratory);*
  - *training fidelity (RQ 5 - exploratory)?*

Selecting pupils for NELI

- All pupils thought to be able to potentially benefit; no specific exclusion criteria.

- OxEd recommended that schools use LanguageScreen to:

  – baseline assess pupils in each class;

  – rank them;

  – select the 3-6 pupils with the lowest scores in each class for NELI.

  - Selection mechanism suggested **Regression Discontinuity** as a potential method.

Regression Discontinuity

Regression Discontinuity

- Binned Control Observations
- Binned Treatment Observations
- Local linear fit

$E[Y(1)|X]$

$\hat{\mu}_+$

μ_+

$\hat{\mu}_-$

μ_-

$E[Y(1)|X], E[Y(0)|X]$

$E[Y(0)|X]$

Cutoff

$c-h$ c $c+h$
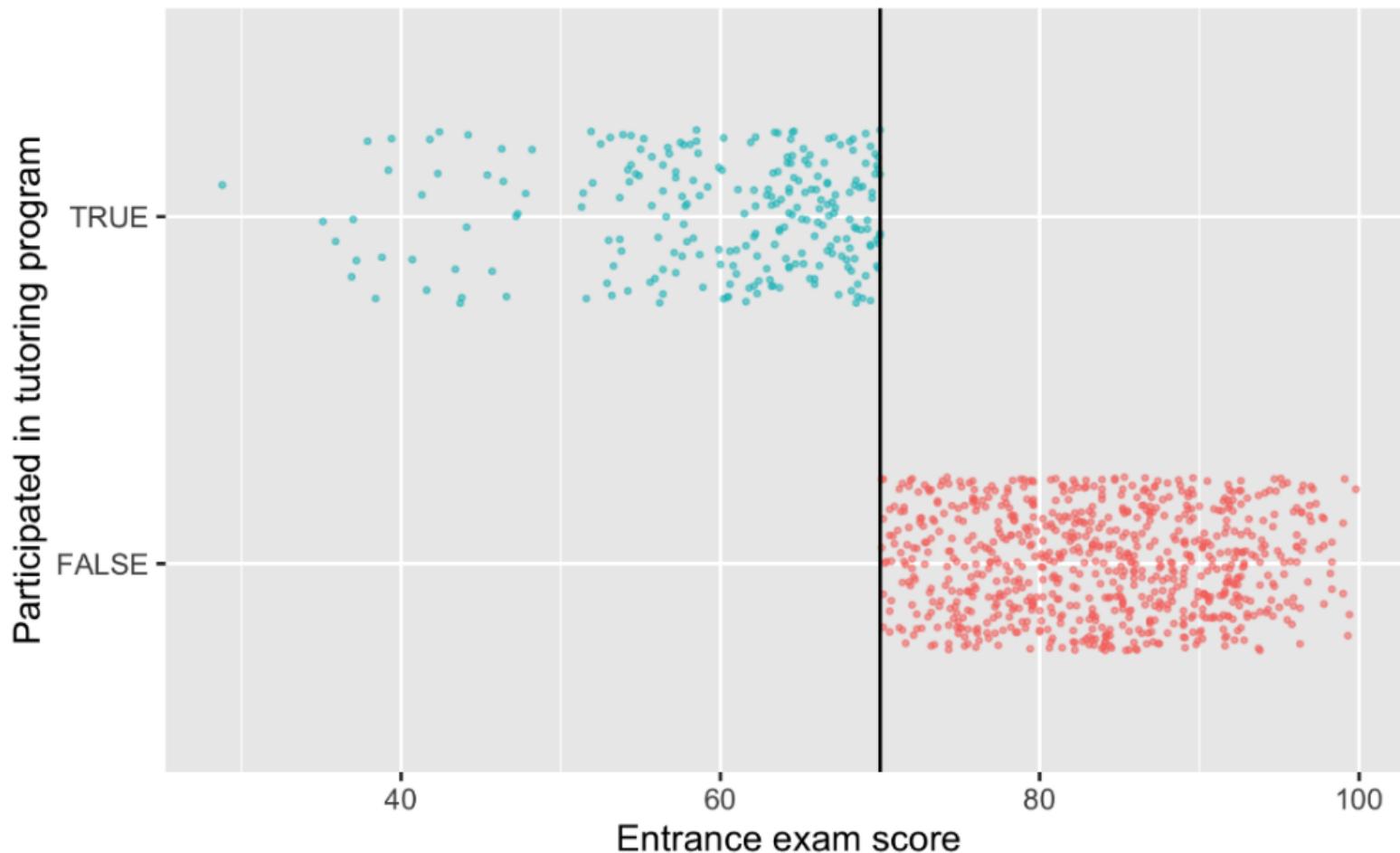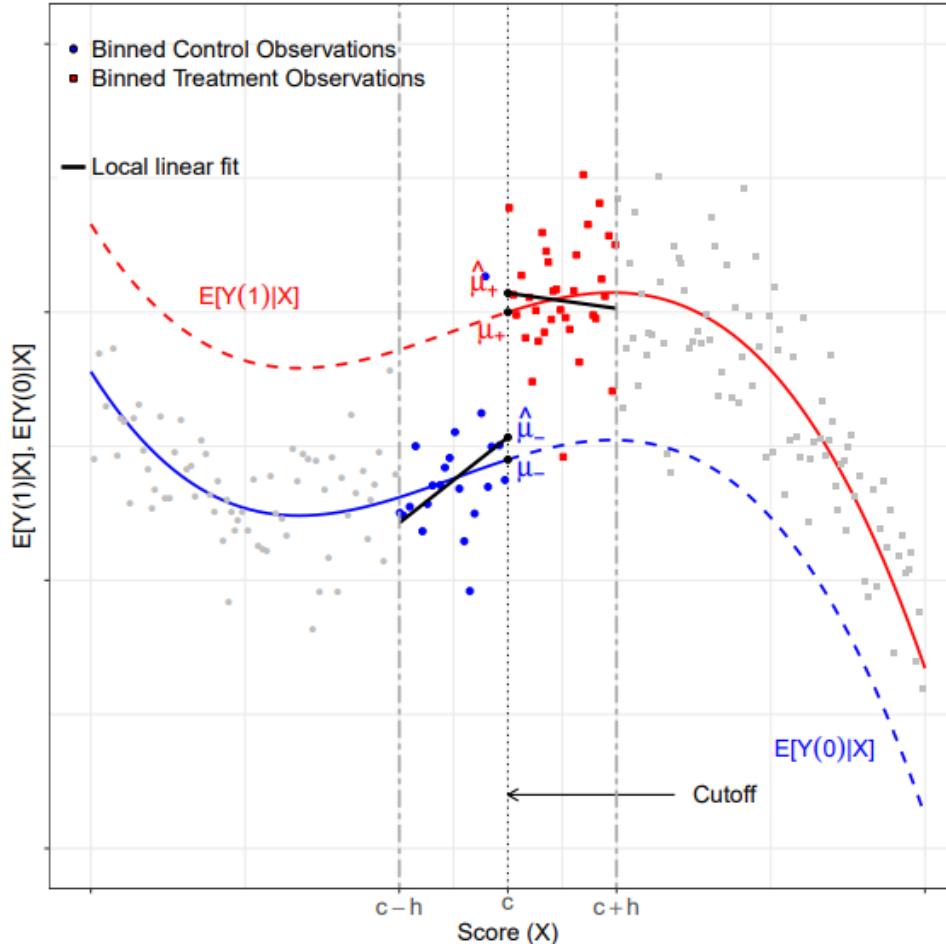
Score (X)

Cattaneo et al., 2019, p. 43

Selecting pupils for NELI

NFER
National Foundation for
Educational Research

**You have indicated that you selected pupils to participate in NELI at least partly based on factors other than the LanguageScreen score. Please select the other factors that were considered when selecting reception pupils for NELI from the list below**

Factor	Percentage
Behavioural factors (e.g. pupils' attention span and concentration, listening skills, interactions with other pupils, confidence)	55%
Assessments made by pupils' teacher, teaching assistant or other school staff member (e.g. Speech and Language Therapist)	53%
Pupils' Special Educational Needs (SEN)	51%
Other indicators of pupils' speech and language skills	46%
Whether pupil speaks English as an Additional Language (EAL)	41%
Pupils' participation in other speech/language programmes	31%
Pupils' performance in other subject areas	17%
Pupils' attendance record	12%
Other (please specify below):	3%
Don't know	2%

source: reproduced from Disley et al., 2023, p.42
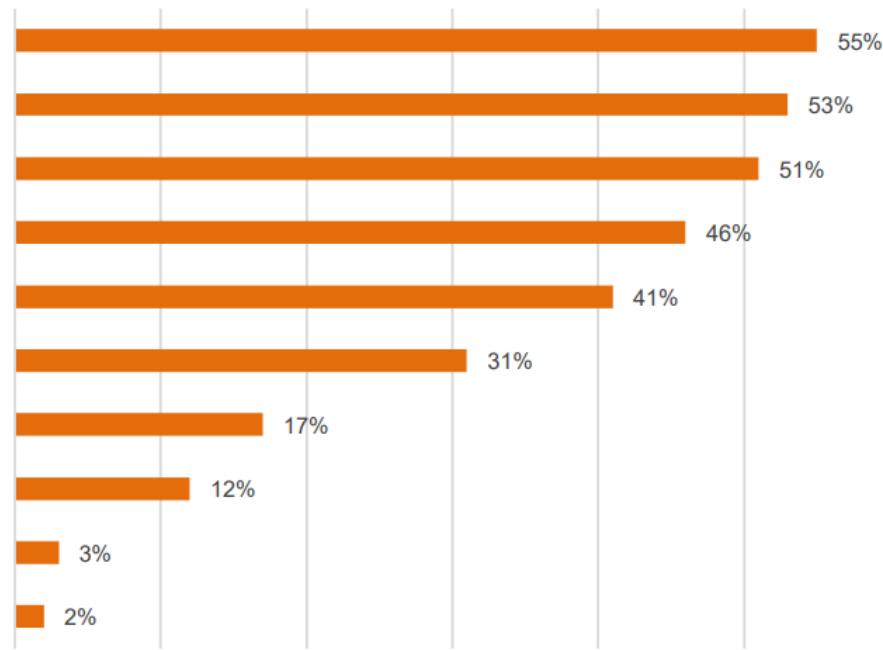
Scale-up (wave 2) data following cutoff modelling
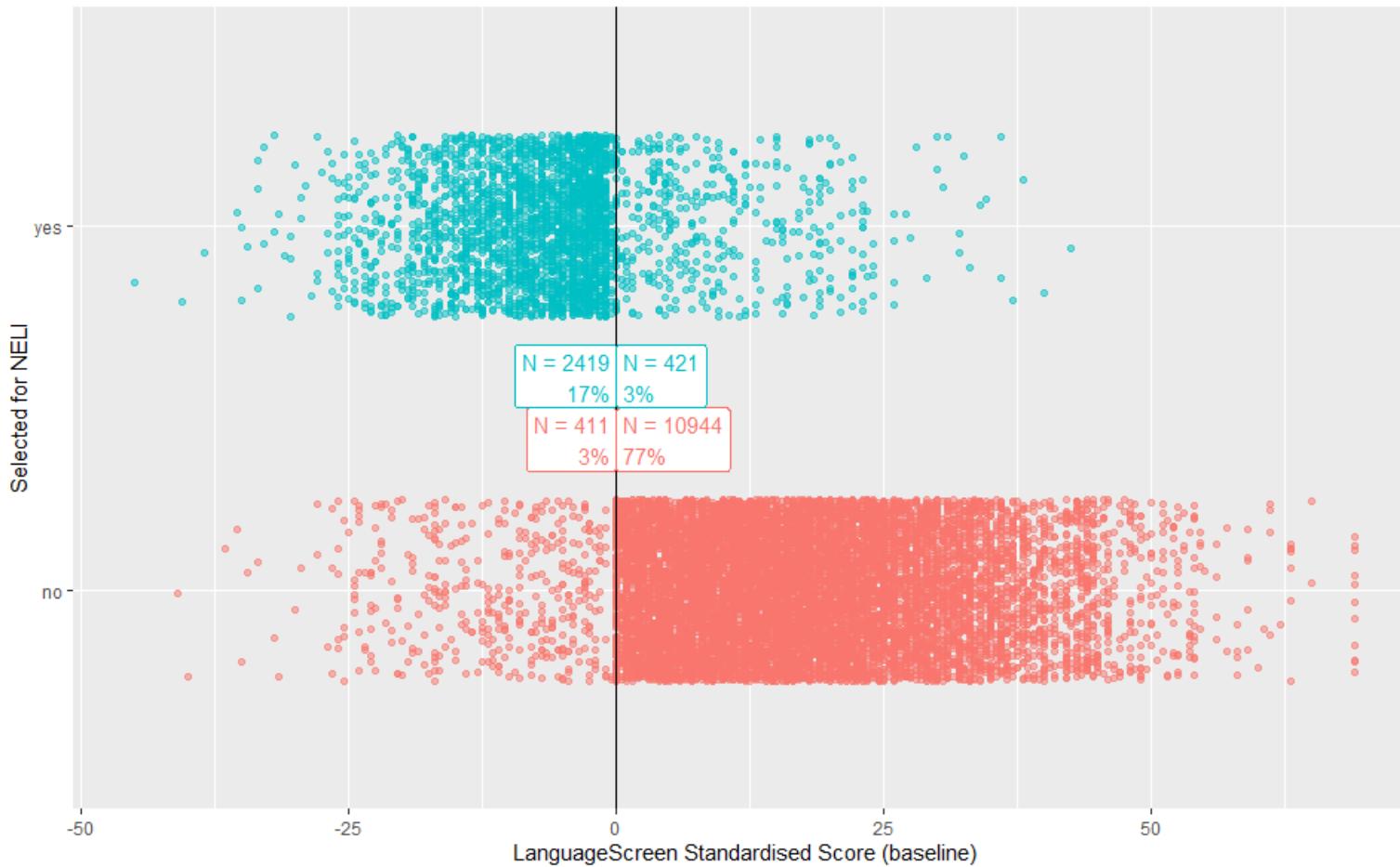
Analysis and Main Results

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)

Analysis

:·NFER
National Foundation for
Educational Research

- **Models:** unadjusted; covariate adjusted model (pre-test); covariate adjusted (pre-test plus other pupil-level covariates)

- **Falsification checks:**

  - Density of the running variable.

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

  - Treatment effects at placebo cutoffs.

  - Sensitivity to bandwidth choices.

  - Test whether treated and comparison units close to the cutoff are dissimilar (i.e. based on characteristics gender, FSM-eligibility, EAL, SEN).

  - Estimates excluding classes where treatment assignment is fuzzy (i.e. running variable does not uniquely determine treatment).

RQ 1. What is the impact of NELI when delivered at scale on pupils' oral language outcomes, as measured by LanguageScreen?
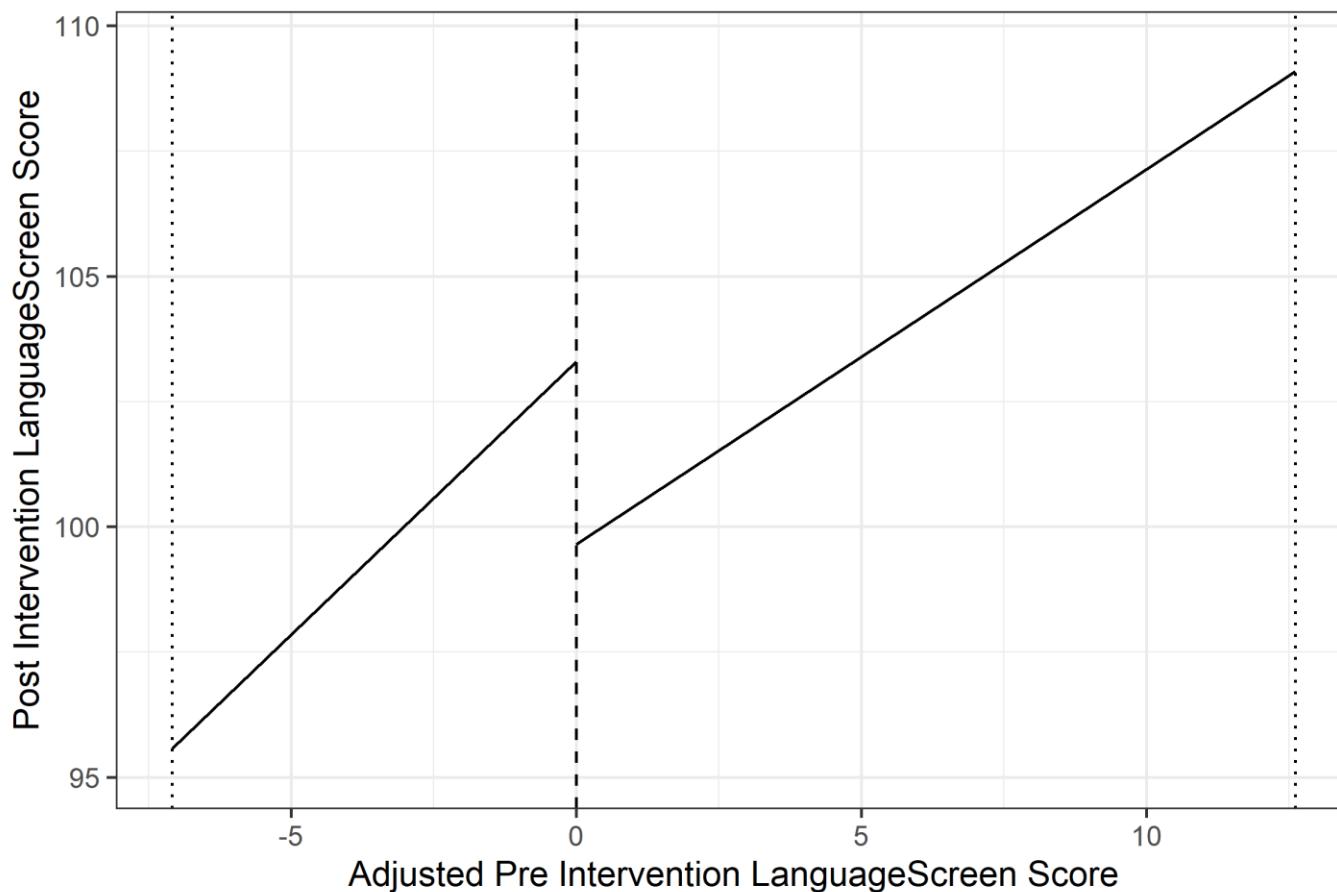
- Hedges' *g* **0.297** (95% CI 0.120, 0.474; p = <0.001)
- Pupils within bandwidth (intervention = 1,269, comparison = 3,207)

- Hedges' *g* **0.569** (95% CI 0.142, 0.997; p = 0.009)
- Pupils within bandwidth (intervention = 215, comparison = 473)

- Hedges' *g* **0.294** (95% CI -0.003, 0.623; p = 0.079)
- Pupils within bandwidth (intervention = 276, comparison = 577)
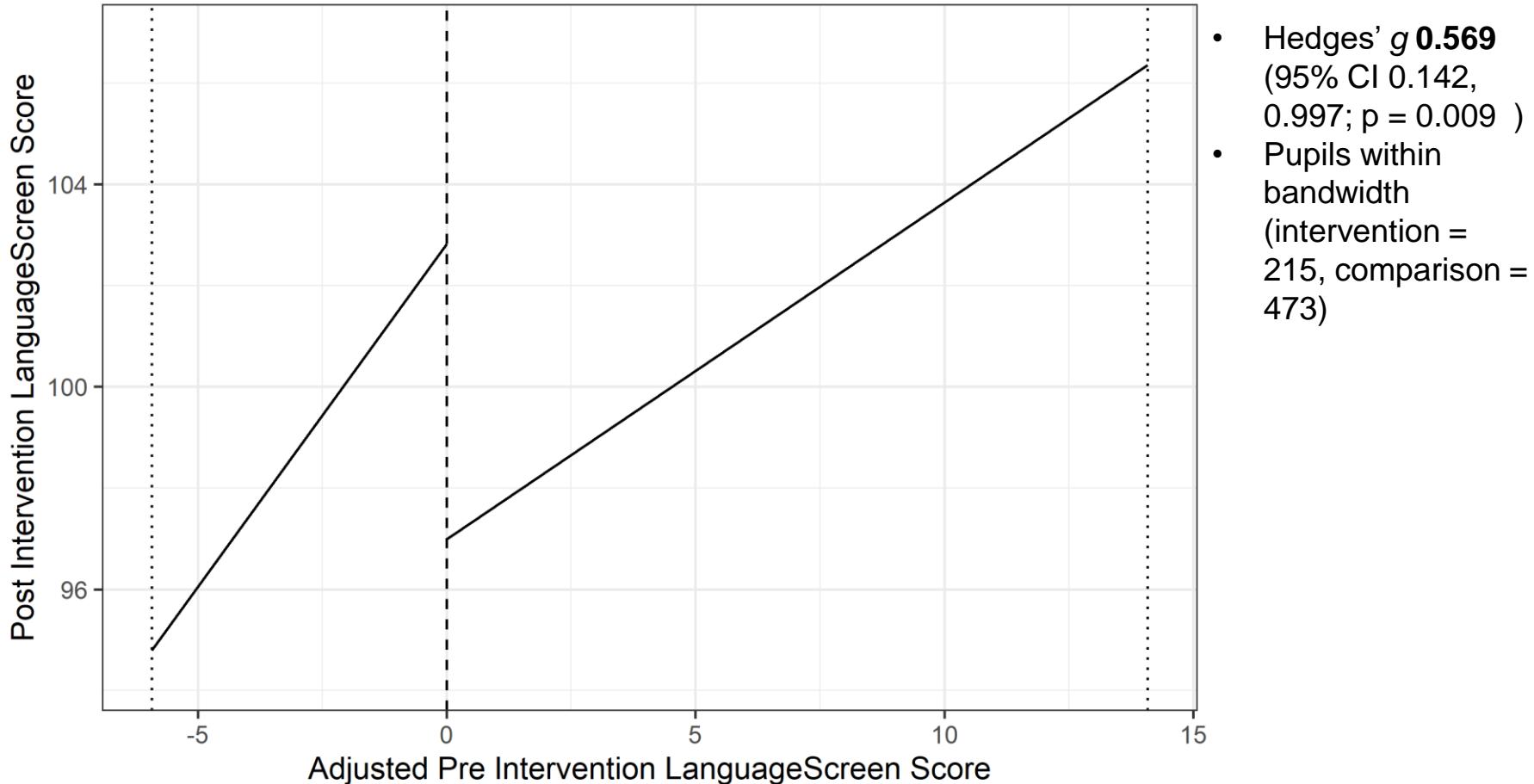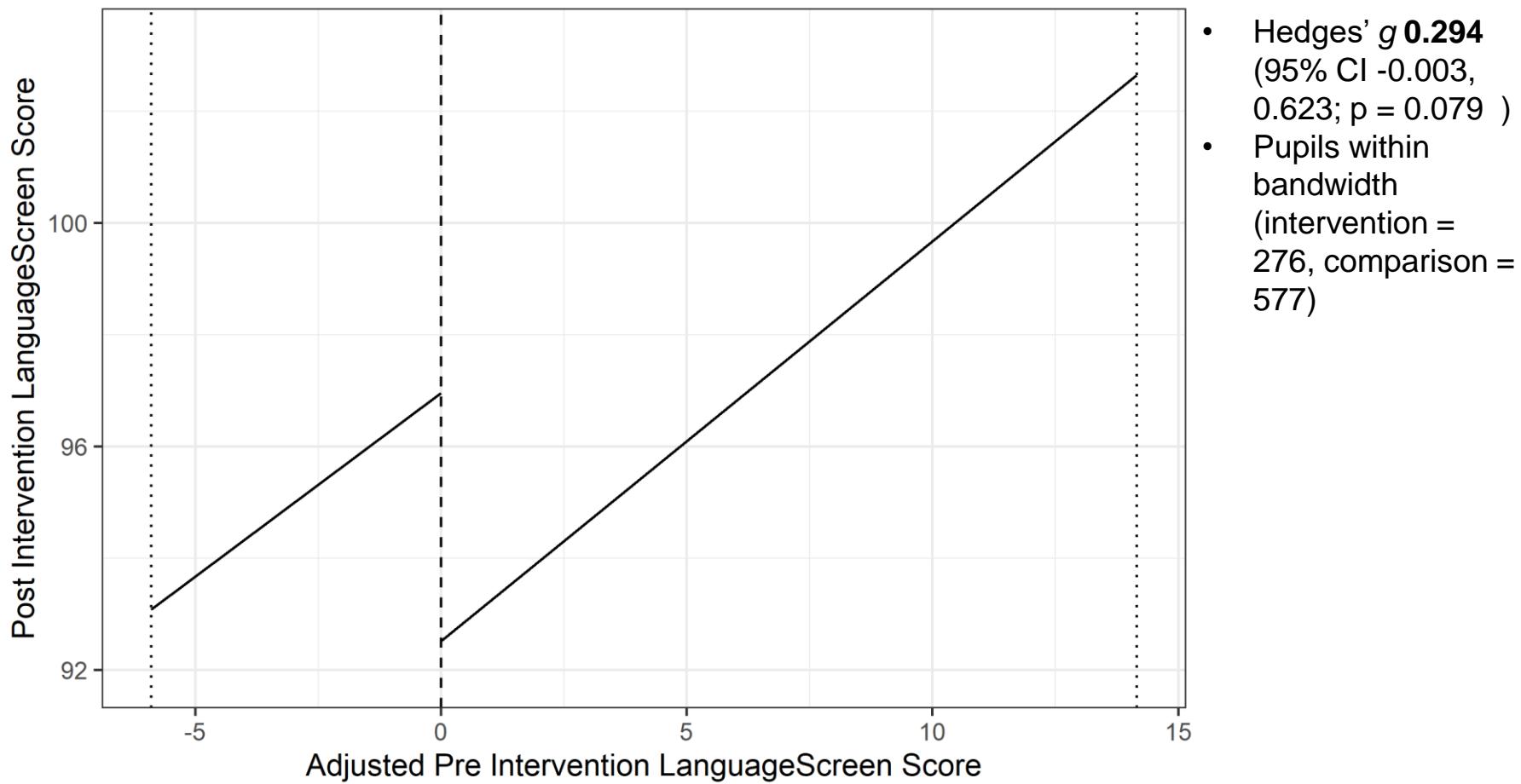
Conclusions and Limitations

Conclusions

- Evaluation schools and those in final analytical dataset representative of NELI wave 2 registered schools

- Positive impact on the development of oral language skills for all pupils (+4 months) and FSM pupils (+7 months) selected for NELI.

- Indication that pupils in receipt of more weekly group sessions made greater progress than pupils in receipt of fewer weekly group sessions.

- Indication of larger effects for pupils in schools in which more than 50% of TAs had attended at least one training session.

- The scale-up impact evaluation findings provide some support for the NELI logic model (short-term outcomes)

Limitations

- RD estimand (using noncumculative normative cutoffs)

- The effects of pupil characteristics, and missing data

- Use of non-independent (Wolf and Harbatkin, 2002) outcome (proximal, short-term)

Thank you

Questions and comments
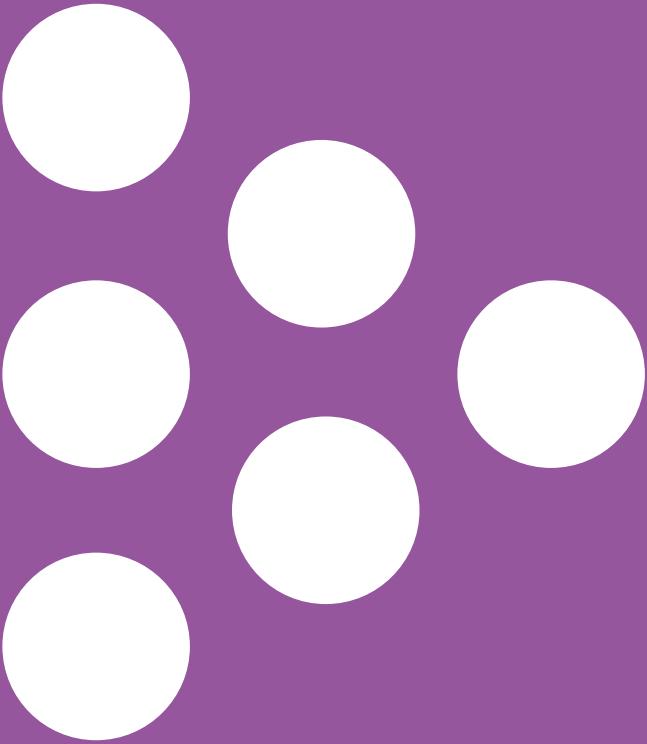
Evidence for excellence in education