The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry **:* (2024), pp **–** doi:10.1111/jcpp.13947

Oral language enrichment in preschool improves children's language skills: a cluster randomised controlled trial

Gillian West,[1] Arne Lervåg,[2] Julia M. H. Birchenough,[3] Caroline Korell,[3] Mariela Rios Diaz,[3] Mihaela Duta,[3] Denise Cripps,[3] Rachel Gardner,[3] Caroline Fairhurst,[4] and Charles Hulme[5]

[1]University College London, London, UK; [2]University of Oslo, Oslo, Norway; [3]University of Oxford, Oxford, UK; [4]University of York, York, UK; [5]Oxford Brookes University, Oxford, UK

Background: Oral language skills provide the foundation for formal education, yet many children enter school with language weaknesses. This study evaluated the efficacy of a new language enrichment programme, the Nuffield Early Language Intervention—Preschool (NELI Preschool), delivered to children in the year before they enter formal education. **Methods:** We conducted a preregistered cluster randomised controlled trial in 65 nursery schools in England (https://doi.org/10.1186/ISRCTN29838552). NELI Preschool consists of a 20-week whole-class language enrichment programme delivered by a teacher each day for 20 min. In addition, children with the weakest language skills in each class are allocated to receive additional targeted support delivered by classroom assistants (whole-class + targeted). The language skills of all children (n = 1,586) in participating classrooms were assessed using the LanguageScreen automated app (https://oxedandassessment.com/languagescreen/). Settings were then randomly allocated to an intervention or control group. The children with the weakest language in each class (whole-class + targeted children n = 438), along with four randomly selected children in each class allocated to the whole-class only programme (n = 288) were individually tested on a range of language measures. **Results:** Children receiving NELI Preschool made larger gains than children in the control group on an oral language latent variable (whole-class children d = .26; whole-class + targeted children d = .16). **Conclusions:** This study provides good evidence that whole-class intervention delivered in preschool can produce educationally significant improvements in children's language skills. The intervention is scaleable and relatively low cost. These findings have important implications for educational and social policy. **Keywords:** Language; RCT; education; preschool; intervention.

Introduction

Language skills are the foundation of virtually all aspects of education, including literacy (Hjetland, Brinchmann, Scherer, Hulme, & Melby-Lervåg, 2020; Hulme, Nash, Gooch, Lervåg, & Snowling, 2015; Snow, 2016) and numeracy development (Chow & Ekholm, 2019; Hornburg, Schmitt, & Purpura, 2018). They are also vital for psychosocial development (Norbury et al., 2016; van Agt, Verhoeven, van den Brink, & de Koning, 2011). Furthermore, it is well established that social class and language skills are strongly related; children from socio-economically disadvantaged backgrounds are much more likely to enter education with less-well-developed language skills (Guo & Harris, 2000; Hart & Risley, 1995; Roulstone, Law, Rush, Clegg, & Peters, 2011; Sampson, Sharkey, & Raudenbush, 2008; Sirin, 2005).

Language skills develop rapidly between the ages of 3–6 years making preschool an excellent time to intervene to support language development. Potentially, such interventions may also help to narrow the gap in language skills associated with social disadvantage. The newly developed Nuffield Early Language Intervention—Preschool (NELI Preschool) programme aims to do both these things, combining language enrichment for all children with additional targeted support for those with language needs. NELI Preschool is a fully scripted programme built around the principles of shared book reading and guided play, introducing children to richer and more structured language than they would typically encounter in spoken language (Nation, Dawson, & Hsiao, 2022). It is based around 20 pre-reading books, containing a mixture of traditional tales, contemporary stories and non-fiction. Shared reading and dialogical questioning introduces the stories to the children. The programme involves direct teaching of vocabulary knowledge and builds children's narrative skills, using tailored scaffolded support when children retell story elements in their own words. Activities lead children to produce more and more integrated and coherent language of their own, while building their confidence in speaking. A focus on vocabulary and narrative development linked to shared reading of stories is theoretically and empirically motivated. Multi-contextual teaching of the meanings of words from books is a highly effective strategy to boost receptive and expressive language skills (Beck, McKeown, & Kucan, 2013;

Conflict of interest statement: See Acknowledgements for full disclosures.

Biemiller & Boote, 2006; Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009). Narrative development, encompassing knowledge of story structure and practice in retelling stories supports the development of language skills integral to communication, from concept formation and the ordering of personal experiences, to syntactic development and comprehension, as well as preparing children for literacy instruction (Lever & Sénéchal, 2011; Nielsen, Friesen, & Fink, 2012; Stadler & Ward, 2005). A major focus of the programme is on getting children to produce language, which reflects the theoretical view that language production is particularly important for language learning (Macdonald, 2013). NELI Preschool also develops children's active listening skills, which in turn, supports their use of language for learning, as well as the development of pro-social behaviour suited to a classroom environment (West et al., 2022; Yew & O'Kearney, 2013).

Teacher training is an essential ingredient in successful language intervention (Egert, Fukkink, & Eckhardt, 2018; Siraj et al., 2023). The NELI Preschool programme includes comprehensive training for educators to ensure they learn not just how to deliver the intervention, but also about a range of evidence-based teaching strategies to support language development, both inside and outside of the programme. More broadly, the training seeks to ensure that school staff understand the importance of language skills for education and psycho-social development.

The combination of whole group language enrichment and additional targeted support for those with weak language skills in the NELI Preschool programme, is designed to narrow the gap between children entering school with poor language and their peers, while also ensuring that all children receive language enrichment. Children are selected to receive individualised targeted help, using the LanguageScreen app (https://oxedandassessment.com/languagescreen/), which is also used to monitor improvements in language skills. The NELI Preschool programme is designed to be suitable for both monolingual children and children with English as an Additional language (EAL children, i.e. those for whom English is not the language predominantly spoken in the home). Evidence shows that EAL children may benefit from early years oral interventions as much as their monolingual peers (West et al., 2021), and the scaffolded support in NELI Preschool is well-suited to children learning English as a second (or further) language (Bowles, Radford, & Bakopoulou, 2018).

Method

A cluster randomised controlled trial (RCT) was conducted in nursery schools in England. A process evaluation was also undertaken to assess programme acceptability, feasibility and delivery fidelity. Schools were randomly allocated to the treatment or waiting control group by an independent clinical trials unit. Only schools in the treatment group delivered the intervention, whilst schools in the control group continued to provide their usual curriculum. The language skills of children in both arms of the trial were assessed before and after intervention delivery. Study design, measures and analysis were pre-registered (https://doi.org/10.1186/ISRCTN29838552). To avoid bias, a number of critical analyses of the primary outcome were completed blind to group allocation by one of the authors (AL), who was supplied with an anonymised dataset with treatment group coded arbitrarily.

Participants

Ethical permission for the study was granted by the Research Ethics Committee of the University of Oxford. Head teachers gave consent to take part in the trial. Children in participating nursery classrooms were enrolled on an opt out basis. Sixty-five schools (n = 70 preschool classrooms) from 7 geographical areas in England (Buckinghamshire, Hertfordshire, Liverpool, London South East, London North West, Nottinghamshire, Oxfordshire) took part in the trial. All children aged 3–4 years, attending preschool 4+ days (or half days) a week were eligible for the trial. In total, 1,586 preschool children (773 boys, 48.7%) took part. Of these children, 543 (34.2%) were reported to be EAL.

Design

Prior to the intervention, LanguageScreen was used by school staff to assess the language skills of all eligible children in participating schools. The scores from LanguageScreen were used by the schools to identify the six children in each classroom with the weakest language skills to receive the additional targeted element of programme. Preschools were then randomised to group by an independent statistician at the York Trials Unit at the University of York, using stratified block randomisation with a block size of two, by geographical area, number of children assessed at baseline (dichotomised by median number of children attending), and mean LanguageScreen score (dichotomised at the median).

The children identified as having the weakest language skills in each class were then tested individually on a battery of standardised assessments by the research team ($t1$). In order to be able to evaluate the whole-class only element of the programme, four of the remaining children in each preschool class were randomly allocated to receive individual in-depth testing.

Following the completion of pre-testing, all preschool staff delivering the programme received online training. Delivery of the intervention took place between January and July 2022. Observation visits to all intervention schools were conducted between February and March 2022.

After the programme had been completed, all children were re-assessed ($t2$) with the LanguageScreen app and the battery of in-depth tests were re-administered to the children who had received them at $t1$. Figure 1 shows the timeline of assessments, training and intervention. Figure 2 shows the flow of participants through the trial.

Measures

Screening. All children were assessed at $t0$ and at $t2$ with LanguageScreen (https://oxedandassessment.com/languagescreen/). LanguageScreen is an App with four subtests: Expressive vocabulary (naming 24 pictures); Receptive vocabulary (matching a spoken word to one of four pictures, 23 items); Sentence repetition (repeating 14 sentences verbatim);

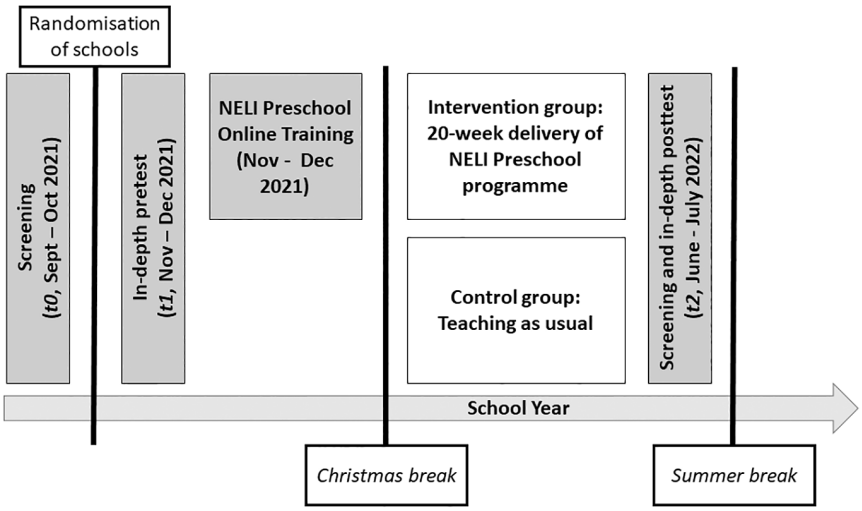

Figure 1 Timeline of the trial showing times of assessment, training and intervention.

and Listening comprehension (answering 16 literal and inferential questions about three short stories spoken by the app).

In-depth language assessments. Children receiving in-depth tests were assessed by the research team and speech and language therapists trained by the team to deliver and score the tests. Assessor training comprised a day of face-to-face training, practice assessments and a calibration of scoring before each testing period. The battery included two subtests of the Clinical Evaluation of Language Fundamentals (CELF) Preschool II UK (Semel, Wiig, & Secord, 2006). For the Expressive Vocabulary subtest, the child is shown a picture and asked to name it. For the Recalling Sentences subtest, the child is asked to repeat sentences, which increase in difficulty and length. Children were also assessed with the Renfrew Action Picture Test (Renfrew, 2003), which requires the child to answer a set question about each of 10 pictures. Answers are recorded verbatim and scored for information content and grammar. It should be noted that the CELF Expressive Vocabulary contained one word (pour) that was also a 'Special Word' taught in the programme.

Additional measures. Knowledge of vocabulary items directly taught in the intervention was assessed using twenty-nine pictures that had to be named. The measure used in analyses was the total number correct. Narrative skills were measured using a simplified version of the Renfrew Bus Story. The child heard a short story about a child failing to catch a bus and then being caught in the rain. They are then asked to re-tell the story while being shown three picture prompts representing the key elements of the story. The measure of narrative skill was the total number of words uttered by the child in recounting this story.

Children's behavioural adjustment in school was measured using the subscale from the Brief Early Skills and Support Index (BESSI; Hughes, Daly, Foley, White, & Devine, 2015). This teacher rating scale has 12 items, each rated on a four-point scale ranging from strongly agree to strongly disagree. Items were grouped into three subsets representing behavioural regulation, attention/hyperactivity and sociability, which were used to construct a behavioural adjustment factor for analysis.

Finally, self-regulation was assessed using the Head Toes Knees and Shoulders-Revised (HTKS-R, Gonzales et al., 2021), which is a widely used assessment of inhibitory control, cognitive flexibility and working memory. Children are taught pairs of behavioural commands (e.g. 'when I say touch your head, you touch your toes'). The task incudes four blocks of trials (59 items in total), which increase in complexity as the task progresses. Two points are included for correct responses and one point for self-corrections. The measure used in analyses was the total raw score.

NELI Preschool Programme

NELI Preschool is a 20-week scripted language programme that consists of a whole-class and a targeted component. The programme is designed to enrich children's vocabulary and develop their narrative and active listening skills and combines class-based language instruction for all children with additional targeted support in small group and individual sessions for children with weak language skills. The programme is built around 20 pre-reading books, with a new book being used each week. Whole class sessions are delivered every day for 15–20 min by the teacher. These daily sessions are displayed on a whiteboard or screen and engage the children with the book of the week and pursue related activities to support vocabulary learning and speech production. Each new book introduces four carefully selected words, which are at the centre of the vocabulary learning activities. Sessions are scripted with flexibility to adapt to the cohort's ability level. For more advanced children, additional optional activities provide the opportunity to extend their vocabulary further.

The children identified as needing additional targeted support take part in three small-group (10–15 min each) and one individual session (10 min) per week. These sessions focus on developing vocabulary and narrative skills by getting children to retell aspects of the stories used in the programme. There is an emphasis in these sessions on getting children to produce language, and for the staff member working with the child to expand on what children say to lead them towards more complex or better-formed utterances using "scaffolding". The two strands of the programme for children with typically developing language skills and those with language weaknesses are referred to in the current study as *whole-class* and *whole-class + targeted* strands respectively.

Educator training and support

Training for preschool staff delivering the programme used an online training course that took approximately 10–12 h to complete. Following a detailed introduction to children's oral language development, staff were introduced to the techniques used for shared book reading and teaching in the programme, as well as good practice in encouraging language production

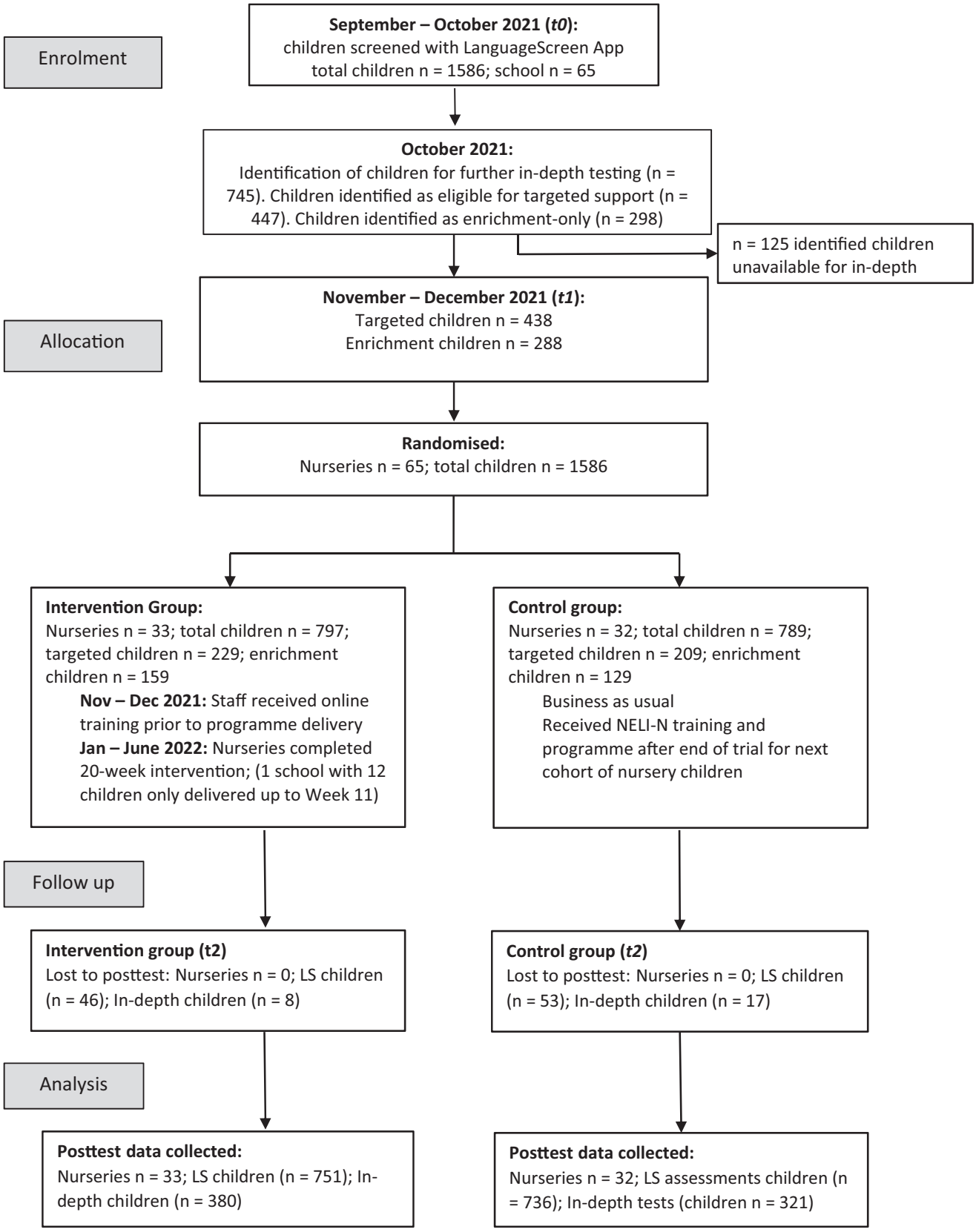

Figure 2 CONSORT diagram showing the flow of participants through the cluster RCT.

and active listening. The course was self-paced. The training included filmed sessions of programme delivery, enabling trainees to watch 'Best Practice' for each type of programme session, alongside additional expert commentary. At the end of each training step, learners are invited to reflect on, and share thoughts or questions, with other learners and course mentors, encouraging the development of a community of practice. Each section of the training finished with a quiz to enable participants to monitor their progress. Participants could revise content at any time during or after completing

training. On successful completion of the course (defined as completing >80% of the online training modules) and subject to passing a test (a score of >70% of answers correct) participants received a practitioner certificate. Practitioners were required to complete the training before they could begin to deliver the programme.

An online Support Hub provided a forum for schools to exchange ideas and feedback during delivery, and to continue to build their community of practice and to keep motivation for delivery high. Each week the Hub introduced the new book and a rolling programme of tips designed to enhance delivery of the programme. Educators were encouraged to share new activity ideas to enhance the programme even further on a secure platform (Padlet). Every 5 weeks, schools were sent a summary newsletter with news from the Hub and reminders of any trial-related administration tasks.

Results

All analyses were performed on an intention-to-treat basis. Analyses were conducted in Stata 14.0 (Stata Corp, College Station, Texas, USA) or Mplus 8.4 (Muthén & Muthén, 1998–2019) with Full Information Maximum Likelihood estimators to allow for missing data. Most analyses followed a pre-registered plan (https://doi.org/10.1186/ISRCTN29838552).

At screening (*t*0) 209 control and 229 intervention children were identified, based on their LanguageScreen scores, as eligible to receive the whole-class + targeted programme (mean LanguageScreen standard score at *t*0 = 79.03; *SD* = 9.49). In addition, 129 enrichment-only children were allocated to in-depth testing in the control group and 159 in the intervention group (mean LanguageScreen standard score at *t*0 = 98.18 (*SD* = 12.58). At posttest, 17 control children (8%) and 8 intervention children (3%) were lost to follow up (odds ratio = 0.36; 95% CI 0.14, 0.88); indicating that although attrition overall was low, it was slightly lower in the intervention group than in the control group. Critically, there were no significant differences at pretest between children who completed the study and those who dropped out in terms of gender χ^2 (1) = 0.15; p = .70), age (t = 0.47; p = .64; d = .03) or language factor scores derived from language tests (t = .44; p = .67; d = .15), indicating that attrition is unlikely to bias the estimates of effect sizes reported below.

Descriptive statistics for all measures at baseline and post-test for both the intervention and control groups are shown in Table 1. The groups were well equated on language skills at baseline.

Correlations between the standardised language assessments and LanguageScreen were high, ranging from r = .70 to .82 at pretest and r = .70 to .79 at posttest, evidencing its validity as a reliable screener of language skills for this young age group. The primary outcome was a language latent variable defined by loadings from the four LanguageScreen subtests (expressive vocabulary, receptive vocabulary, sentence repetition and listening comprehension) and the individually administered language tests (CELF recalling sentences subtest, CELF expressive vocabulary subtest, Renfrew Action Picture test [information and grammar]). This language latent variable was created for baseline (pretest) and posttest scores. Analyses are based on latent variable ANCOVA models implemented in an SEM framework. The pretest latent language variable was the covariate, and the posttest latent variable was the outcome measure. Errors for the language latent variable indicators were correlated to provide an adequate model fit. The effects of the intervention were measured by the y-standardised regression coefficient for a group dummy variable. The intra-class correlation coefficient for the latent variable was (ICC = .12). The effects of clustering within schools was accounted for by using robust (Huber-White) cluster standard errors.

Separate ANCOVA models were conducted to assess the effects of the language intervention programme on (a) typically developing children (the randomly selected four children in each class not identified as having poor language skills who received the whole-class only strand of the programme); and (b) children identified as having language difficulties (the six children in each class with the lowest language composite score who received the whole-class + targeted strands of the programme).

Confirmatory factor analyses and model modifications

Before estimating the effects of intervention, confirmatory factor analyses (CFAs) for the whole-class and whole-class + targeted children were estimated separately to check the dimensionality of the data. The baseline model was a two-factor CFA (the same factor at pretest and posttest), with all eight indicators reflecting the latent language construct at each time point; only correlations between the residuals of the same variable across time were estimated. As expected, based on earlier trials, this model did not fit the data well. Modification indices were then used to identify misspecifications relating to correlations between residuals that made sense from a theoretical point of view. Only misspecifications that were consistent across time were addressed, in order to achieve configural invariance (see Table S1 for details).

The two models (for the whole-class and whole-class + targeted strands, separately) were tested for factorial invariance across time. The configural models (factor loadings and intercepts freely estimated) were used as a baseline with which the more restricted models were compared. Only partial metric (equal factor loadings) invariance held for the whole-class + targeted children (3 factor loadings differed between the pretest and posttest latent language variables). For the whole-class model full metric and partial scalar (equal factor loadings and intercepts) invariance held (with three intercepts

Table 1 Mean raw scores (*SD*) for intervention and control group for outcome measures pre-intervention (*t*0, *t*1) and post-intervention (*t*2)

		Intervention Group (*n* = 388)				Control Group (*n* = 338)				
		Enrichment-only children		Enrichment + Targeted children		Enrichment-only children		Enrichment + Targeted children		
	Reliability	*N*	Mean (*SD*)	*N*	Mean (*SD*)	*N*	Mean (*SD*)	*N*	Mean (*SD*)	*d*[e]
Age (months)										
*t*0			43.30 (3.71)		43.57 (3.62)		43.06 (3.91)		43.50 (3.74)	
LanguageScreen subtests										
Expressive Vocabulary	.82[a]									.18; .02
*t*0 (24)		159	10.33 (4.05)	229	4.72 (3.66)	129	10.32 (4.12)	209	4.79 (3.44)	
*t*2 (24)		156	13.88 (4.25)	221	8.54 (4.43)	121	13.27 (3.89)	192	8.47 (4.23)	
Receptive Vocabulary	.75[a]									.22; .04
*t*0 (23)		159	14.39 (3.69)	229	9.23 (4.06)	129	13.80 (3.38)	209	8.97 (3.67)	
*t*2 (23)		156	16.95 (3.32)	221	13.48 (4.03)	121	15.92 (3.74)	192	13.01 (3.65)	
Sentence Repetition	.84[a]									.11; .02
*t*0 (14)		159	6.16 (3.39)	229	2.34 (2.41)	129	5.93 (3.57)	209	2.02 (2.12)	
*t*2 (14)		156	9.35 (2.86)	221	6.10 (3.60)	121	8.91 (3.55)	192	5.71 (3.31)	
Listening Comprehension	.81[a]									.35; .32
*t*0 (16)		159	4.70 (3.59)	229	1.20 (2.09)	129	4.29 (3.60)	209	1.09 (1.77)	
*t*2 (16)		156	9.35 (3.99)	221	5.68 (4.20)	121	7.70 (3.95)	192	4.26 (3.63)	
LanguageScreen Total	.82[b]								.27; .10	
*t*0 (77)		159	35.58 (12.01)	229	17.48 (9.65)	129	34.33 (11.95)	209	16.88 (8.47)	
*t*2 (77)		156	49.53 (11.91)	221	33.80 (13.46)	121	45.80 (12.19)	192	31.45 (12.04)	
In-depth assessments										
CELF EV	.82[a]									.20; .08
*t*1 (40)		159	13.25 (6.52)	227	6.14 (5.28)	129	13.20 (6.28)	206	6.30 (4.52)	
*t*2 (40)		156	19.26 (6.77)	224	10.92 (6.50)	124	17.98 (6.95)	196	10.46 (5.76)	
CELF RS	.90[a]									.08; .17
*t*1 (37)		159	9.58 (8.14)	227	3.56 (4.75)	129	10.34 (7.58)	204	3.54 (4.66)	
*t*2 (37)		156	15.76 (8.24)	221	7.88 (6.64)	124	15.83 (6.34)	195	6.87 (5.90)	
APT-Information	.86[a]									.28; .05
*t*1 (43)		159	18.18 (6.05)	227	10.75 (7.13)	128	19.51 (6.10)	205	11.56 (6.77)	
*t*2 (43)		156	24.93 (5.63)	223	17.65 (7.59)	123	24.02 (5.93)	196	17.80 (6.61)	
APT-Grammar	.85[a]									.37; .25
*t*1 (39)		159	14.54 (7.15)	227	6.79 (6.69)	128	15.61 (6.34)	205	7.05 (5.77)	
*t*2 (39)		156	22.06 (7.00)	223	13.84 (8.07)	123	20.26 (6.30)	196	12.15 (6.59)	
Taught Vocabulary	.95[a]									.46; .30
*t*1 (29)		159	18.44 (5.16)	225	11.12 (6.94)	128	18.46 (4.45)	205	11.50(6.12)	
*t*2 (29)		156	23.84 (3.37)	221	18.55 (6.61)	123	21.58 (3.72)	192	16.84 (5.62)	
Narrative	.44[b]									.67; .37
*t*1 (total words)		159	25.59 (16.46)	227	12.97 (13.40)	128	27.21 (16.16)	203	13.54 (11.54)	
*t*2 (total words)		155	42.32 (23.76)	221	30.04(21.70)	122	31.76 (16.40)	192	25.11 (17.77)	
HTKS-R	.71[c]									.31; .22
*t*1 (118)		159	27.91 (23.03)	229	13.56 (14.95)	128	25.01 (19.33)	196	13.62 (15.88)	
*t*2 (118)		156	54.97 (29.04)	221	31.22 (26.32)	124	45.08 (28.86)	192	26.79 (22.09)	
Teacher-rating										
Behavioural Adjustment[d]	.92[a]									−.07; .02
*t*1 (36)		159	12.25 (6.82)	223	14.41 (7.00)	129	12.50 (6.98)	209	14.81 (6.86)	
*t*2 (36)		155	9.52 (6.70)	221	11.28 (6.93)	126	10.37 (6.84)	201	11.63 (7.11)	

APT, Action Picture Test; EV, Expressive Vocabulary; HTKS-R, Heads Toes Knees and Shoulders – Revised; RS, Recalling Sentences. Maximum total score for each subtest given in brackets after each item.
[a]Cronbach's alpha calculated at pre-intervention (*t*0 or *t*1).
[b]Test–retest reliability (Pearson *r*).
[c]Cronbach's alpha calculated across subtotals of 4 parts of HTKS-R at t1.
[d]Lower scores denote better behavioural adjustment.
[e]Cohen's *d* effect size for whole-class and whole-class + targeted strands respectively, calculated as the difference in marginal means between control and intervention groups at posttest from ANCOVA models divided by the pooled standard deviation at pretest.

freely estimated; see Table S2 for details). These were the models used to calculate effect sizes. The lack of metric invariance in the model for the whole-class + targeted children probably reflects the fact that the different language measures show differential degrees of improvement as a result of intervention: the effect sizes for improvements on the different measures in Table S2, show that those for the whole-class + targeted group tend to be smaller and more variable than those for the whole-class group. This pattern indicates that the overall effect size for improvements on the language latent variable in the whole-class + targeted group needs to be interpreted with caution.

Effect size estimates for the pre-registered primary outcome (language)

The effect of the intervention was estimated with ANCOVA models, where the latent language variable at t2 was regressed on both the latent language variable at t1 and a dummy variable indicating group membership (intervention 1; control 0). In addition, language at t1 was correlated with the group-membership dummy, in order to take into account any possible difference between groups at baseline (no significant differences at baseline were found in any of the ANCOVA models). As can be seen from Figures 3 and 4, there was a significant effect from the intervention for both whole-class and whole-class + targeted children. Furthermore, there were no interactions between Language t1 and the intervention dummy for the whole-class-only children: $\beta = .113$, $p = .164$ or for the whole-class + targeted children: $\beta = -.017$, $p = .799$ (βs standardised on y only). The absence of these interactions indicates that the intervention did not vary in its effectiveness as a function of initial level of language ability.

Analyses of secondary preregistered outcome measures

The effects of intervention on two further pre-registered secondary outcomes were assessed using mixed effects ANCOVA models with intervention group dummy coded, posttest scores as the outcome, pretest scores as the covariate and school as a random effect, to account for clustering within schools. Both strands of the programme were associated with significant improvements in knowledge of vocabulary taught in the programme. The model for the whole-class strand showed a significant negative interaction between the covariate and group (the intervention had a larger effect for children with poorer scores at pretest). Plots of the interaction showed that the intervention group had significantly higher posttest scores throughout most of the distribution; at the mean of covariate there was relatively large advantage for the intervention group (difference in marginal means = 2.23; $\chi^2 = 25.58$; *df*. 1; $p < .001$; $d = 0.46$). For the whole-class + targeted strand there were parallel slopes relating the outcome to the covariate: posttest scores for the intervention group were 1.96 points higher for the intervention group than the control group ($z = 4.24$, $p < .001$; $d = .30$).

However, as is apparent from the figures in Table 1, equivalent models confirmed that there were no significant effects of the programme on teacher ratings of behavioural adjustment (whole-class programme difference in marginal means = −0.45, $z = -0.55$, $p = .58$, $d = -0.08$; whole-class + targeted programme difference in marginal means = 0.13, $z = 0.15$, $p = .88$, $d = 0.02$).

Non-preregistered analyses

Differences on two non-preregistered analyses are also of interest. These analyses used mixed effects ANCOVA models as above. Narrative skills were measured by getting the child to retell a story aided by picture prompts. The measure used was the total number of words uttered. The whole-class intervention group made more progress in developing their narrative skills than children in the control group: (difference in marginal means = 10.95, $z = 3.78$, $p = .001$, $d = 0.67$) but this effect was smaller and non-significant for the whole-class + targeted intervention group (difference in marginal means = 4.62, $z = 1.93$, $p = .053$, $d = 0.37$).

Finally, self-regulation, as assessed by the HTKS-R, improved significantly as a result of the intervention, but only for the whole-class strand. The ANCOVA model for this strand showed a significant negative interaction term reflecting the fact that the intervention was more effective for children with lower scores at pretest. For this strand the difference in marginal means at the mean of the covariate (6.62) was significant ($\chi^2 = 5.62$, $p = .0178$, $d = 0.31$). In contrast, the model for the whole-class + targeted strand showed a positive interaction reflecting the fact that the intervention was most effective for children with better pretest scores. However, for this strand the difference in marginal means at the mean of the covariate (3.40) was not significant ($\chi^2 = 2.72$, $p = .0989$, $d = 0.22$).

Intervention dosage

Child-level attendance registers were completed for 368 children receiving NELI Preschool (95.3% of children in the intervention group of the trial). Registers showed that dosage remained high throughout the programme (whole-class children's sessions 88.3 (*SD* = 10.9) out of 100), which equates to 88.3% of the programme. Whole-class + targeted children's dosage: mean whole-class sessions = 84.8 (*SD* = 13.6) out of 100, mean small group sessions = 47.9 (*SD* = 10.2) out of 60, and mean

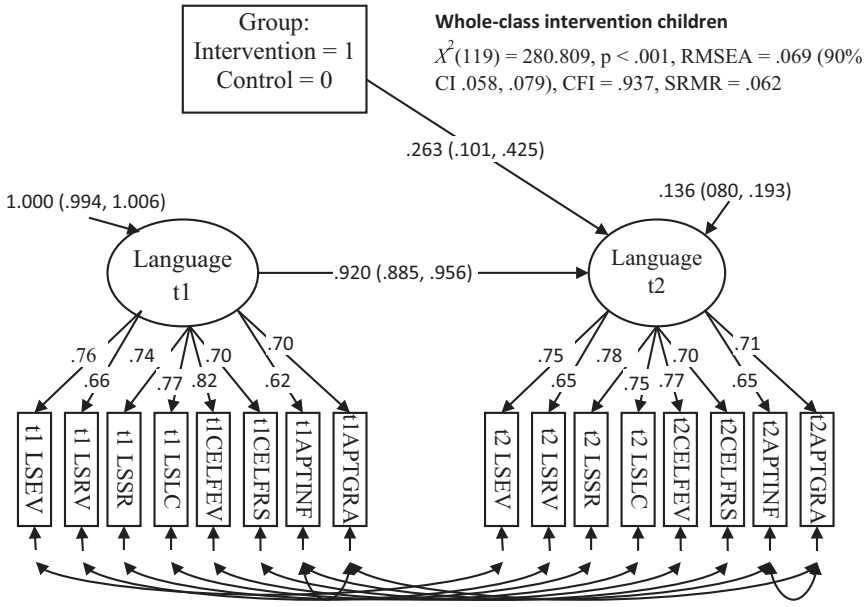

Figure 3 Path diagram for the pre-reregistered primary outcome for the trial showing the effect of the intervention on children receiving the whole-class enrichment element of the programme. The effect of the intervention is shown by the path weight from Group (dummy coded) to language at posttest. The path weight is y-standardised and equivalent to Cohen's *d*, with 95% robust confidence intervals accounting for clustering within schools shown in brackets. APTGRA, Action Picture Test Grammar score; APTINF, Action Picture Test Information score; CELFEV, CELF Expressive Vocabulary; CELFRS, CELF Recalling Sentences; LSEV, LanguageScreen Expressive Vocabulary; LSLC, LanguageScreen Listening Comprehension; LSRV, LanguageScreen Receptive Vocabulary; LSSR, LanguageScreen Sentence Recall

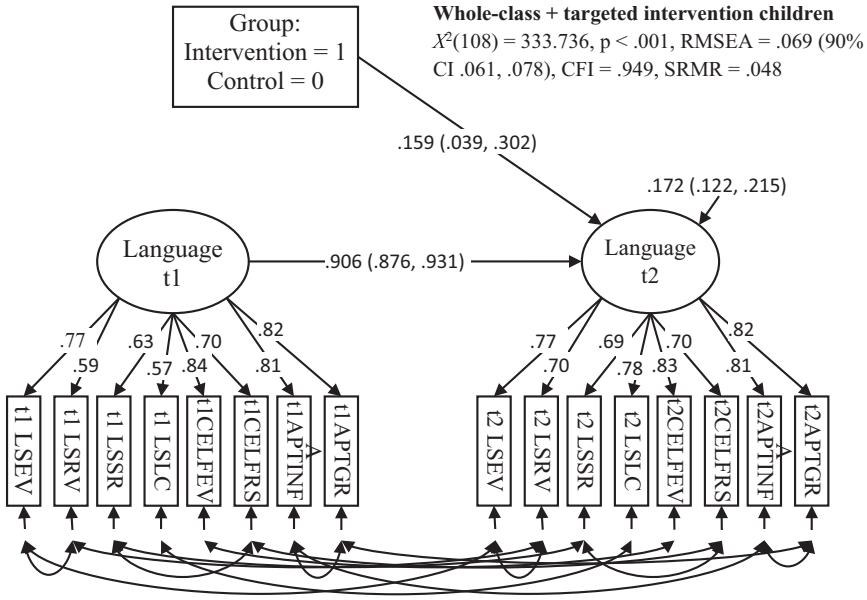

Figure 4 Path diagram for the pre-reregistered primary outcome for the trial showing the effect of the intervention on children with weak language skills of receiving the whole-class + Targeted support. The effect of the intervention is shown by the path weight from Group (dummy coded) to language at posttest. The path weight is y-standardised and equivalent to Cohen's *d*, with 95% robust confidence intervals accounting for clustering within schools shown in brackets. APTGRA = Action Picture Test Grammar score; APTINF, Action Picture Test Information score; CELFEV, CELF Expressive Vocabulary; CELFRS, CELF Recalling Sentences; LSEV, LanguageScreen Expressive Vocabulary; LSLC, LanguageScreen Listening Comprehension; LSRV, LanguageScreen Receptive Vocabulary; LSSR, LanguageScreen Sentence Recall

individual sessions = 15.6 (*SD* = 4.1) out of 20. This equates to 84.8%, 80.0% and 77.9% dosage for whole-class + targeted children receiving whole-class, small group and individual sessions respectively.

Discussion

The NELI Preschool programme evaluated here produces clear improvements in children's oral language skills. The size of effects for the whole class programme (*d* = .26) and the whole class + targeted

programme (d = .16) are substantial and of educational significance (Kraft, 2020; What Works Clearing House, 2014). The programme did not produce any improvements in behavioural adjustment and it is too early to assess any possible effects on measures of later school attainment including literacy and numeracy. The size of improvements seen in the whole-class programme are striking and have important implications for policy.

To our knowledge this is the first RCT to demonstrate substantial improvements in children's oral language skills from a whole-class language enrichment programme. Several other studies have shown positive effects of whole-class vocabulary instruction. For example, Zucker, Cabell, Justice, Pentimonti, and Kaderavek (2013) found improvements in knowledge of words directly taught after just four weeks of a whole-class programme, but there was no generalisation to other measures of language ability. Similarly, Moore, Hammond, and Fetherston (2014) contrasted three different methods of direct vocabulary instruction over 18 weeks, compared to a business-as-usual control. All three methods improved knowledge of the words taught, but no measures of generalisation to other measures of language were included. The current study, however, takes a much broader approach than these earlier studies which focused exclusively on vocabulary teaching. The NELI Preschool programme involves direct teaching of new vocabulary, but in addition involves instruction in narrative (story telling) and active listening skills. Throughout the programme there is an emphasis on children's ability to use language productively, with supportive feedback from teaching staff. Arguably, the programme emphasises a range of meta-cognitive skills (active listening strategies, the need to put new words into context when learning them, the role of temporal structure (beginning, middle, end) when producing a narrative) that bring about generalised improvements in children's language learning ability.

The most directly comparable study to the current one is an RCT by Bleses et al. (2018), in which they examined the effects of a scripted programme that targeted pre-reading (phoneme awareness and print awareness) and language (vocabulary and narrative) skills and was delivered either in small groups or to the whole class in different preschools. Both the whole-class and small group programmes produced moderate improvements in children's pre-reading skills that not did differ in size (d = .20 and .18 respectively). However, for language skills, both programmes had non-significant negative effects (d = −.11 and − .08, respectively). A major limitation of this study is that several of the measures showed ceiling effects at pretest, and given that the same measures were used at posttest, this severely limits the ability to detect improvements. Puzzlingly, the study reported stronger improvements in language skills (d = .30) when teachers were trained in the methods of the intervention and delivered lessons of their own devising to small groups, without being provided with scripted lessons. This might suggest that the scripted lessons were not effective. In summary, this study failed to show any benefits of a combined whole-class pre-literacy and oral language intervention on children's oral language skills.

The NELI Preschool programme is a two-tiered programme (Catts, Petscher, Schatschneider, Sittner Bridges, & Mendoza, 2009), in which high quality teaching is offered to all children (Tier 1) followed by more intensive instruction in small groups or individual sessions (Tier 2) for children who most need help. Such an approach seems ideal as a way of providing a rich language curriculum that will be of benefit to all children in the early years of education. Given the strong association between social deprivation and early language skills this approach may be particularly beneficial to children from disadvantaged backgrounds.

We found smaller benefits of the whole-class + targeted programme (d = .16) compared to the whole-class programme alone (d = .26). We had expected greater benefits from the whole-class + targeted programme where children were receiving more help. However, the differences between these two conditions are difficult to interpret because the children receiving the whole-class + targeted programme had poorer language skills. A further trial will be necessary to compare the relative effects of the whole-class only, versus whole-class + targeted programme for children with weak language skills, in order to assess whether there is any additional benefit from the targeted intervention for children with poor language skills. Such a study would allow us to assess the benefit of providing additional targeted intervention to children with poor language prior to school entry and clarify whether the cost and time-burden involved in delivering it is justified.

Clinically, it was clear that many of the children allocated to the whole-class + targeted programme had very limited language ability. It is possible that such children may not be at a developmental a stage where they can benefit from the highly structured language intervention offered here and might benefit more from language intervention slightly later in development.

Limitations

We note that further evidence is needed to clarify the effects of the small group and individual teaching sessions. Ideally, in future studies, longer term follow-up data will be able to assess the durability of the effects obtained.

Additionally, as the number of children with language weaknesses at baseline differs across settings, it would be helpful to assess the efficacy and utility of a version of the programme that permits settings to include additional children in the targeted intervention where necessary, while taking care not to diminish the small-group nature of this aspect of the programme.

Finally, previous research into targeted language intervention for 4–5-year-old children in Reception found no difference in response to intervention between monolingual children or those with EAL (West et al., 2021). However, given the high proportion of children with EAL in nursery settings in this trial, future research should be powered to assess whether the response of EAL children to intervention in preschool is equivalent to that of their monolingual peers.

Conclusions

This study provides evidence from a well-powered RCT that whole-class language enrichment can be effective in preschool. The materials and online training used make the intervention scalable and cost effective. Our finding that a whole-class language enrichment programme can produce educationally significant improvements in language skills (d = .26) has important implications for educational and social policy. Such programmes clearly have the potential to improve educational outcomes and psycho-social development and might be particularly important for immigrant children and those from socially disadvantaged backgrounds.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Table S1. Details of analyses to establish configural variance.

Table S2. Details of analyses to investigate metric invariance.

Acknowledgements

This trial was funded by the Nuffield Foundation (Grant number: EDU 43387). The authors are grateful to all the preschools, educators and children who took part.

C.H. and G.W. are Directors of OxEd and Assessment Ltd, a University of Oxford spin-out company founded to distribute LanguageScreen as a commercial product. M.D. is a shareholder in OxEd and Assessment Ltd. G.W., C.W., R.G. and C.H. are authors of the NELI Preschool programme, which is licenced to and distributed by OxEd and Assessment Ltd. Additionally, Arne Lervåg's work is funded by The Research Council of Norway, Centres of Excellence, Grant 331640.

Data availability statement

The anonymised data that support the findings of this study are available in the UK Data Service ReShare repository at https://reshare.ukdataservice.ac.uk/cgi/users/home?screen=EPrint::View&eprintid=856802#t.

Correspondence

Gillian West, Department of Language and Cognition, University College London, 2 Wakefield Street, London WC1N 1PF, UK; Email: g.west@ucl.ac.uk

Key points

- Oral language skills are critical for education and psychosocial development.
- Data from a cluster randomised controlled trial show that a 20-week language enrichment programme delivered in preschool to all 3–4 year-olds can produce educationally meaningful improvements in their oral language skills.
- An automated screening app allows school staff to identify children with language difficulties in preschool in order to provide them with additional targeted language support within the programme.

References

Beck, I.L., McKeown, M.G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction.* New York, NY: Guilford Press.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, *98*, 4–62.

Bleses, D., Højen, A., Justice, L.M., Dale, P.S., Dybdal, L., Piasta, S.B., … & Haghish, E.F. (2018). The effectiveness of a large-scale language and preliteracy intervention: The SPELL randomized controlled trial in Denmark. *Child Development*, *89*, e342–e363.

Bowles, D., Radford, J., & Bakopoulou, I. (2018). Scaffolding as a key role for teaching assistants: Perceptions of their pedagogical strategies. *British Journal of Educational Psychology*, *88*, 499–512.

Catts, H.W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, *42*, 163–176.

Chow, J.C., & Ekholm, E. (2019). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*, *46*, 179–186.

Coyne, M.C., McCoach, D.B., Loftus, S., Zipoli, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal*, *110*, 1–18.

Egert, F., Fukkink, R.G., & Eckhardt, A.G. (2018). Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Review of Educational Research*, *88*, 401–433.

Gonzales, C.R., Bowles, R., Geldhof, G.J., Cameron, C.E., Tracy, A., & McClelland, M.M. (2021). The Head-Toes-Knees-Shoulders Revised (HTKS-R): Development and psychometric properties of a revision to reduce floor effects. *Early Childhood Research Quarterly*, *56*, 320–332.

Guo, G., & Harris, K.M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography*, *37*, 431–447.

Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes.

Hjetland, H.N., Brinchmann, E.I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, *30*, 100323.

Hornburg, C.B., Schmitt, S.A., & Purpura, D.J. (2018). Relations between preschoolers' mathematical language understanding and specific numeracy skills. *Journal of Experimental Child Psychology*, *176*, 84–100.

Hughes, C., Daly, I., Foley, S., White, N., & Devine, R.T. (2015). Measuring the foundations of school readiness: Introducing a new questionnaire for teachers–The Brief Early Skills and Support Index (BESSI). *British Journal of Educational Psychology*, *85*, 332–356.

Hulme, C., Nash, H.M., Gooch, D., Lervåg, A., & Snowling, M.J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science*, *26*, 1877–1886.

Kraft, M.A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*, 241–253.

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, *108*, 1–24.

MacDonald, M.C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

Moore, W., Hammond, L., & Fetherston, T. (2014). Strengthening vocabulary for literacy: An analysis of the use of explicit instruction techniques to improve word learning from story book read-alouds. *Australian Journal of Learning Difficulties*, *19*, 153–172.

Muthén, B.O., & Muthén, L. (1998–2016). *Mplus user's guide* (7th edn). Los Angeles: Muthén & Muthén.

Nation, K., Dawson, N.J., & Hsiao, Y. (2022). Book language and its implications for Children's language, literacy, and development. *Current Directions in Psychological Science*, *31*, 375–380.

Nielsen, D.C., Friesen, L.D., & Fink, J. (2012). The effectiveness of a model of language-focused classroom instruction on the vocabulary and narrative development of kindergarten children. *Journal of Education*, *192*, 63–77.

Norbury, C.F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., … & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*, 1247–1257.

Renfrew, C. (2003). *Action picture test*. Milton Keynes, UK: Speechmark.

Roulstone, S., Law, J., Rush, R., Clegg, J., & Peters, T. (2011). *Investigating the role of language in children's early educational outcomes*. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181549/DFE-RR134.pdf

Sampson, R.J., Sharkey, P., & Raudenbush, S.W. (2008). Durable effects of concentrated disadvantage on verbal ability among African–American children. *Proceedings of the National Academy of Sciences*, *105*, 845–852.

Semel, E., Wiig, E., & Secord, W. (2006). *Child evaluation of language fundamentals-preschool UK* (2nd edn). Oxford: Pearson Assessment.

Siraj, I., Melhuish, E., Howard, S.J., Neilsen-Hewett, C., Kingston, D., De Rosnay, M., … & Luu, B. (2023). Improving quality of teaching and child development: A randomised controlled trial of the leadership for learning intervention in preschools. *Frontiers in Psychology*, *13*, 1092284.

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453.

Snow, P.C. (2016). Elizabeth Usher Memorial Lecture: Language is literacy is language-positioning speech-language pathology in education policy, practice, paradigms and polemics. *International Journal of Speech-Language Pathology*, *18*, 216–228.

Stadler, M.A., & Ward, G.C. (2005). Supporting the narrative development of young children. *Early Childhood Education Journal*, *33*, 73–80.

StataCorp. (2019). *Stata statistical software: Release 16*. College Station, TX: StataCorp LLC.

van Agt, H., Verhoeven, L., van den Brink, G., & de Koning, H. (2011). The impact on socio-emotional development and quality of life of language impairment in 8-year-old children. *Developmental Medicine & Child Neurology*, *53*, 81–88.

West, G., Lervåg, A., Snowling, M.J., Buchanan-Worster, E., Duta, M., & Hulme, C. (2022). Early language intervention improves behavioral adjustment in school: Evidence from a cluster randomized trial. *Journal of School Psychology*, *92*, 334–345.

West, G., Snowling, M.J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., … & Hulme, C. (2021). Early language screening and intervention can be delivered successfully at scale: Evidence from a cluster randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *62*, 1425–1434.

What Works Clearing House. (2014). *Procedures and standards handbook*. Available from: https://ies.ed.gov/ncee/wwc/Handbooks#procedures [last accessed 02 August 2023].

Yew, S.G.K., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry*, *54*, 516–524.

Zucker, T.A., Cabell, S.Q., Justice, L.M., Pentimonti, J.M., & Kaderavek, J.N. (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental Psychology*, *49*, 1425–1439.

Accepted for publication: 22 November 2023