

Early language screening and intervention can be delivered successfully at scale: evidence from a cluster randomized controlled trial

Gillian West,¹  Margaret J. Snowling,¹ Arne Lervåg,² Elizabeth Buchanan-Worster,¹ Mihaela Duta,¹  Alexandra Hall,³ Henrietta McLachlan,³ and Charles Hulme¹ 

¹University of Oxford, Oxford, UK; ²University of Oslo, Oslo, Norway; ³Elklan Training Ltd, Cornwall, UK

Background: It is well established that oral language skills provide a critical foundation for formal education. This study evaluated the effectiveness of the Nuffield Early Language Intervention (NELI) programme in ameliorating language difficulties in the first year of school when delivered at scale. **Methods:** We conducted a cluster randomized controlled trial (RCT) in 193 primary schools (containing 238 Reception classrooms). Schools were randomly allocated to either a 20-week oral language intervention or a business-as-usual control group. All classes ($N = 5,879$ children) in participating schools were screened by school staff using an automated App to assess children's oral language skills. Screening identified 1,173 children as eligible for language intervention: schools containing 571 of these children were allocated to the control group and 569 to the intervention group. **Results:** Children receiving the NELI programme made significantly larger gains than the business-as-usual control group on a latent variable reflecting standardized measures of language ability ($d = .26$) and on the school-administered automated assessment of receptive and expressive language skills ($d = .32$). The effects of intervention did not vary as a function of home language background or gender. **Conclusions:** This study provides strong evidence for the effectiveness of a school-based language intervention programme (NELI) delivered at scale. These findings demonstrate that language difficulties can be identified by school-based testing and ameliorated by a TA delivered intervention; this has important implications for educational and social policy. **Keywords:** Language; RCT; education.

Introduction

Language skills are fundamental to many aspects of cognitive and psychosocial development. Language skills also provide a critical foundation for formal education: they are vital for the development of both word reading and reading comprehension skills (Hjetland, Brinchman, Scherer, Hulme, & Melby-Lervåg, 2020; Hulme, Nash, Gooch, Lervåg, & Snowling, 2015), as well as numeracy and mathematical skills (Chow & Ekholm, 2019; Hornburg, Schmitt, & Purpura, 2018). Language is also crucial for social and emotional development, with poor language skills affecting children's friendships and behaviour (van Agt, Verhoeven, van den Brink, & de Koning, 2011; Norbury et al., 2016). Language difficulties, therefore, may place children at risk of educational failure, social difficulties and reduced employment prospects (Clegg, Hollis, Mawhood, & Rutter, 2005).

Epidemiological studies show that some 7 to 10% of children can be considered to have clinically significant language difficulties (Norbury et al., 2016; Tomblin et al., 1997). Such difficulties reflect multiple causal risk factors including both genetic and environmental influences (Hayiou-Thomas, Dale, & Plomin, 2012). Two important environmental influences are socioeconomic status (SES) and home

language background. Language skills show a moderate to strong relationship with SES (Guo & Harris, 2000; Hart & Risley, 1995; Sampson, Sharkey, & Raudenbush, 2008; Sirin, 2005). Furthermore, many countries have children from immigrant backgrounds entering school without adequate proficiency in the language of instruction (Castro, Paez, Dickinson, & Frede, 2011; Melby-Lervåg & Lervåg, 2014). These children often display language weaknesses that can be long lasting (Mancilla-Martinez & Lesaux, 2011) and are associated with persistent difficulties with reading comprehension (Melby-Lervåg & Lervåg, 2014).

Language difficulties show high stability across development and can be identified relatively early (at around the time of entry to formal education). Given evidence for their serious and wide-ranging effects, it becomes critical to assess whether interventions in the first year of school can ameliorate language difficulties. As yet, however, evidence for the effectiveness of language interventions is mixed. Several previous meta-analyses of the effects of language intervention have included studies without an appropriate control group, for example, within-subject designs. However, a recent meta-analysis assessing the effects of interventions on linguistic comprehension and vocabulary knowledge that included only RCTs and QEs with a control group and measures of baseline differences (Rogde, Hagen, Melby-Lervåg, & Lervåg, 2019) reported a small overall effect of interventions on language skills

Conflict of interest statement: See Acknowledgements for full disclosures.

($g = .16$, CI [.12, .30]); however, those studies with high-quality implementation ($d = .24$) and where interventions were delivered to small groups ($d = .25$) showed larger, educationally significant effects.

Given, the mixed evidence to date, it is critical to design studies with high statistical power to examine the extent to which intervention programmes delivered at scale can ameliorate language difficulties. In this paper, we report a study of the effectiveness of the Nuffield Early Language Intervention (NELI) programme. NELI is a 20-week oral language intervention for children who show poor oral language skills in Reception class (first year of formal education in the UK). The programme targets vocabulary knowledge, narrative and active listening skills and combines small group with one-to-one sessions delivered by trained teaching assistants. The programme includes high-quality training and ongoing support for the teaching assistants delivering the intervention.

Several earlier randomized controlled trials (RCTs) have assessed the efficacy of different variants of the NELI programme. In an initial small-scale study, Bowyer Crane et al. (2008) showed that children with language weaknesses who received a 20-week version of the NELI language intervention programme made more progress in vocabulary and grammatical skills than children receiving a phonology and reading skills programme. Fricke, Bowyer-Crane, Haley, Hulme, and Snowling (2013) evaluated a 30-week version of the programme spanning Nursery (pre-school) and Reception, finding large effects of the intervention ($d = .80$ on a latent language variable at immediate post-test; $d = .83$ at a 6-month delayed follow up). Reading comprehension also improved at delayed follow up ($d = .52$). A subsequent, efficacy trial involving 394 children in 34 schools, compared the 30-week NELI language intervention programme starting in Nursery with the 20-week version starting in Reception (Fricke et al., 2017). Both programmes produced significant improvements in oral language skills (30-week programme $d = .30$; 20-week programme $d = .21$) which were maintained at 6-month follow-up. The lower effect sizes in the Fricke et al. (2017) study compared to Fricke et al. (2013) appeared to reflect lower implementation quality in the later study.

Since these earlier trials, the NELI programme has been updated and published in its final form as a 20-week programme for Reception class children (Fricke, Bowyer-Crane, Snowling, & Hulme, 2018). Minor amendments were made to the programme, based on recommendations from an advisory panel of educational experts and findings from previous trials. Additionally, a new training and support model for teachers and teaching assistants (TAs) were developed and are now provided to schools alongside the published programme (<https://global.oup.com/education/content/primary/series/nuffie>

Id-intervention/?region=uk). The current trial seeks robust evidence for the effects of this updated and improved 20-week version of the programme delivered at scale in a large number of schools. It will also conduct exploratory analyses to investigate possible moderators of response, including home language background, gender and initial levels of language difficulty. We believe this study is the largest RCT to have assessed the efficacy of language intervention.

Method

A cluster randomized controlled trial (RCT) was conducted in 193 state primary schools (containing 238 Reception classrooms) from 13 geographical areas in the UK (Bristol, Cornwall, Durham, Essex, Herts, London, Manchester, North Tyneside, Blackpool and North West, Northamptonshire, Surrey, Warwickshire, Wolverhampton). The trial was preregistered on the ISRCTN registry (<https://doi.org/10.1186/ISRCTN12991126>). In our previous trial (Fricke et al., 2017), there was better than 80% power to detect a difference between groups equivalent to $d = .29$ ($p < .05$, two-tailed) with $N = 120$ children per arm. For this trial, the aim was to recruit at least 200 schools, with at least 25 children in the Reception class. Selecting the 5 children in each classroom with the poorest language skills would give a total sample of 1,000 children (500 intervention and 500 control).

Following recruitment, schools were randomly allocated, within geographical area, to either a 20-week oral language intervention group or a business-as-usual control group. Schools in the intervention group delivered the Nuffield Early Language Intervention programme with training and delivery support provided by an independent training consultancy (Elklan; <https://www.elklan.co.uk/>). Schools in the control group delivered their usual school provision and received payment to purchase the programme at the end of the trial if they wished.

Assessments took place before the start of the intervention at screening (t_0) for all children in participating classrooms and at pretest (t_1) for children selected via screening (together defining performance at baseline for participating children) and immediately following the intervention (post-test, t_2). The timeline is presented in Figure 1.

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 Reception classrooms were enrolled on an opt out basis. The majority of schools enrolled a single class ($N = 151$ schools; 3745 children); 39 schools enrolled 2 classes ($N = 1,902$ children); and 3 schools enrolled 3 classes ($N = 232$ children), resulting in a total of 5,879 children.

School staff initially screened all children in participating Reception classrooms (t_0) using a computerized language assessment App (LanguageScreen, <https://www.languagescreen.com/>) with four subtests: expressive vocabulary; receptive vocabulary; sentence repetition and listening comprehension. Scoring of the subtests is automated by the App and uploaded to secure servers for analysis. Only children with severe visual, auditory or attentional problems that would prevent them from accessing LanguageScreen were excluded from the screening. All exclusions were discussed with the research team. The use of LanguageScreen facilitated a rigorous and controlled screening process of a large cohort of children in a short period of time. The five children in each class with the lowest LanguageScreen scores were identified as eligible for NELI. In eight schools with fewer than 10 Reception

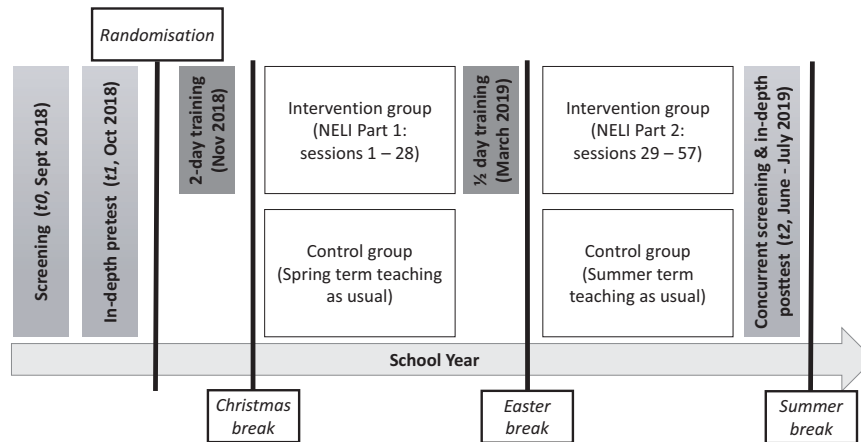


Figure 1 Timeline of trial showing assessment, training and intervention phases

pupils, we selected just 3 children to receive intervention. In one school with Reception classes on different sites, we selected 9 children to receive intervention.

Screening resulted in 1,173 children being identified as eligible for the NELI programme (643 male, 530 female; mean age = 53.31 months, SD = 3.49 months). These children then received individual language testing (*t1*) conducted by speech and language therapists trained by the research team. LanguageScreen was validated against the individually administered language tests. A latent language variable derived from LanguageScreen correlated strongly ($r = .95$) with a latent language variable derived from the four individually administered language tests (see details below) in the sample of 1,156 children who received both sets of tests. These data show that LanguageScreen gives a highly accurate assessment of children's oral language skills which is comparable to that from individually administered standardized language measures.

After completion of *t0* and *t1* testing, schools were randomized to intervention or control group by an independent evaluator. Randomization was stratified by geographical area and the number of classes participating in each school (dichotomized: 1, or more than 1, class). Post-testing (*t2*) using the App was completed by teachers, but individual post-testing by trained testers was done blind to treatment arm. Details of recruitment, selection, randomization and the flow of participants through the study are shown in the CONSORT diagram in Figure 2 (Schulz, Altman, & Moher, 2010).

Assessment measures

As outlined in the trial preregistration, the primary outcome measures were the four standardized tests of language ability administered at *t1* and *t2* to children identified as eligible for the NELI programme. Secondary outcome measures were the school administered LanguageScreen scores and word reading ability (*YARC Early Word Reading* subtest).

Primary outcome measures. Language skills were assessed with the Expressive Vocabulary subtest from the Child Evaluation of Language Fundamentals (CELF) Preschool II^{UK} (Semel, Wiig, & Secord, 2006), *The Renfrew Action Picture Test* (APT; Renfrew, 2003; information and grammar scores) and the *Recalling sentences* subtest from the Child Evaluation of Language Fundamentals (CELF) Preschool II^{UK} (Semel et al., 2006).

Secondary outcome measures. The LanguageScreen assessment comprises 4 subtests: Expressive Vocabulary (naming a series of pictures); Receptive Vocabulary (matching a series of spoken words to one of 4 pictures); Sentence

Repetition (repeating each of a series of sentences verbatim) and Listening Comprehension (listening to 3 spoken stories and answering questions tapping both literal and inferential comprehension).

Word reading ability was measured using the *YARC Early Word Reading* subtest (Hulme et al., 2009).

The Nuffield Early Language Intervention Programme (NELI)

NELI is a 20-week programme for children with poor oral language skills. The programme comprises small group and individual sessions focussed on improving children's vocabulary, developing their narrative skills, encouraging active listening and building confidence in independent speaking. The programme includes a total of 57 small group sessions, each lasting 30 min and 37 individual sessions, each lasting 15 min (total intervention time: small group sessions 28.5 h; individual sessions 9.25 h). It was designed with reference to the Primary Framework for Literacy and Mathematics (DfES, 2006), the Statutory Framework for the Early Years Foundation Stage (DCSF, 2008) and in consultation with teachers and language therapists.

The programme aims to develop children's vocabulary and language skills within a structured framework that follows established principles for teaching listening, vocabulary and narrative skills. Sessions are arranged within topic areas, and vocabulary is taught using a multi-contextual approach within a repetitive framework (Beck & McKeown, 2007; Beck, McKeown, & Kucan, 2013; Carroll, Bowyer-Crane, Duff, Hulme, & Snowling, 2011; Locke, 2006). Narrative work gives children opportunity to practice taught vocabulary in connected speech and introduces them to key story elements and the sequencing of events, while encouraging expressive language and grammatical competence. Listening work targets children's active listening skills and incorporates auditory discrimination, memory and sequencing as well as rhyming activities. In the last 10 weeks, activities promoting phoneme awareness (blending and segmenting) and letter-sound knowledge are introduced to support early literacy instruction.

Training and support

Regional tutors provided training to teaching assistants (TAs) delivering the programme and ongoing support throughout delivery. Regional tutors were experienced speech and language therapists and specialist teachers recruited and trained by Elklan. TAs in intervention schools attended two days of training with classroom teachers attending the first half-day session. The training focussed on deepening teachers' and TAs'

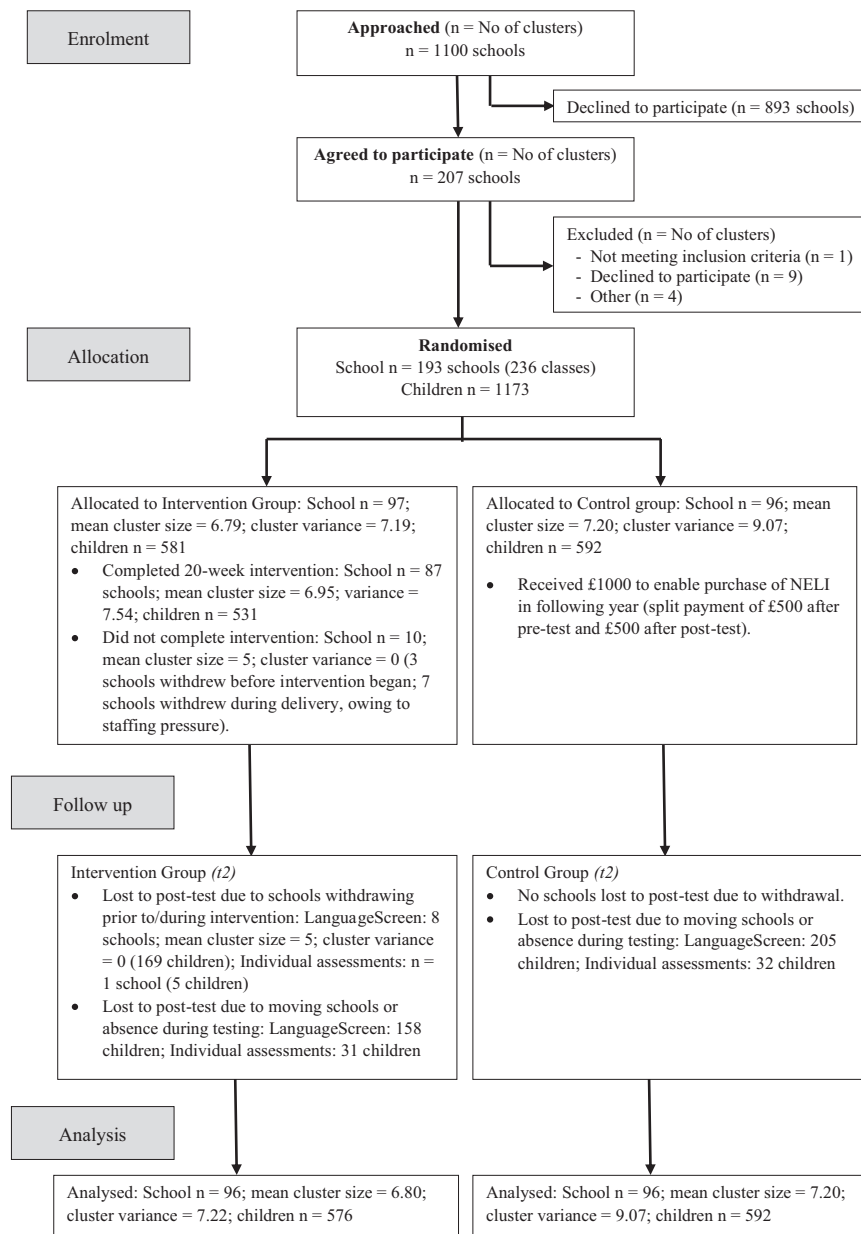


Figure 2 CONSORT Diagram showing flow of participants through the cluster RCT

understanding of oral language and its importance for education and well-being, as well as giving detailed instruction on programme delivery. An additional half day of training midway through the programme trained TAs in delivery of the phonological strand introduced in weeks 11 to 20 of the programme.

TAs were offered ongoing support by their delivery tutors via one-to-one email; a series of three support Webinars and a closed Facebook group to provide a forum for peer-to-peer mentoring.

Results

Analysis plan

The analyses followed the preregistered plan (<https://doi.org/10.1186/ISRCTN12991126>). The primary outcome measure is a latent variable, defined by loadings from individually administered language tests administered at pretest and post-test. The secondary outcome measure is a latent

variable defined by loadings from the subtests of the LanguageScreen App given at screening and post-test. Changes in word reading scores from the EWR test are also assessed. The effects of clustering within schools are accounted for by using robust (Huber-White) cluster standard errors.

All analyses were performed on an intention-to-treat basis. The majority of the analyses were conducted in Stata 16.0 (Stata Corp, College Station, TX, USA). Structural equation models (SEM) were constructed using Mplus 8.4 (Muthen & Muthen, 1998-2019) with Full Information Maximum Likelihood estimators to allow for missing data.

At screening, 592 control and 581 intervention children were identified, based on their LanguageScreen scores, as eligible to receive the NELI language intervention (mean LanguageScreen total score for children allocated to NELI was 27.07

($SD = 9.03$), compared with a sample mean for all 5879 children of 43.49 ($SD = 13.21$). At pretest ($t1$), we obtained data using individually administered language tests from 571 control and 569 intervention children. Subsequently, 30 (5%) control children and 36 (6%) intervention children were lost to follow up (a nonsignificant difference $z = -.78$; $p = .438$). Critically, there were no significant differences at pretest in gender $\chi^2(1) = .18$; $p = .675$), age ($t = -.90$; $p = .368$) or language factor scores derived from standardized language tests ($t = 1.36$; $p = .173$) between children who completed the study and those who dropped out at post-test. In short, there is no evidence that attrition will have biased the estimates of effect sizes reported below.

Descriptive statistics for all measures at baseline (screening and pretest) and post-test for both groups are shown in Table 1. It is clear that the groups are well equated on language skills at baseline. The intervention group shows improvements on all language measures with varying effect sizes ($ds = -.10$ to $.44$). The intervention group also shows a slightly greater improvement on the Early Word Reading test at post-test.

Primary outcome measure

Our primary outcome measure was a language latent variable defined by the standardized measures of language ability (i.e. CELF *expressive vocabulary*, CELF *recalling sentences and APT information and grammar scores*). This variable captures the common variance shared by the different language measures. The model used is shown in Figure 3 and provides an adequate fit to the data ($\chi^2(118, N = 1,173) = 291.96$; $p < .001$; RMSEA (Root Mean Square Error of Approximation) $.035$ [90% CI $.030-.041$]; CFI = $.97$; TLI = $.96$). (The stratification variables used at randomization (area and number of classes within a school) were included as predictors of outcome in this and other models (see, Kahan & Morris, 2011) and allowed to correlate with the pretest latent variable. However, it should be noted that including the stratification variables in these models had no appreciable effect on the estimates of the size of intervention effects obtained or their standard errors.)

It is notable from Figure 3 that the language factor shows considerable longitudinal stability. The most critical result from this analysis is that the intervention group shows a significantly greater increase in language scores than the control group at post-test ($d = .26$ [95% CI $.017, .36$]). A critical assumption for this analysis is that there are equivalent slopes between language pretest and post-test factor scores across groups. An analysis which included the interaction term between the language pretest factor scores and group confirmed that slopes relating pretest to post-test language scores did not differ between groups (standardized slope for

interaction = $-.057$; $p = .284$). In other words, children with the weakest language skills at pretest responded to the programme to the same degree as children with better language skills.

Two other findings are of interest. First, some 34% of the sample came from homes where English is an additional language (EAL). We, therefore, did exploratory analyses to assess whether the intervention was equally effective for EAL and non-EAL children. An initial multigroup model explored whether our language latent variable showed metric invariance between pretest and post-test: unfortunately, this was not the case due to differing degrees of improvement on the different measures (see Table 1). We therefore performed multigroup models comparing the degree of improvement due to intervention on each of the separate measures. In these models, the unstandardized regression of intervention on the post-test measure were freely estimated; in no case did the effect sizes differ between the two groups (CELF-EV (EAL $d = .28$ 95% CI $[.11, .45]$; monolingual $d = .21$ 95% CI $[.08, .33]$; Wald test $\chi^2 = .59$, $d.f. 1$; $p = .44$); CELF-RS (EAL $d = .08$ 95% CI $[-.08, .23]$; monolingual $d = .10$ 95% CI $[-.01, .21]$; Wald test $\chi^2 = .06$, $d.f. 1$; $p = .80$); APT information (EAL $d = .20$ 95% CI $[.02, .38]$; monolingual $d = .21$ 95% CI $[.07, .36]$; Wald test $\chi^2 = .003$, $d.f. 1$; $p = .96$); APT grammar (EAL $d = .31$ 95% CI $[.14, .48]$; monolingual $d = .28$ 95% CI $[.14, .42]$; Wald test $\chi^2 = .15$, $d.f. 1$; $p = .70$).

Secondly, we checked whether the intervention was equally effective for boys and girls. In this case we had metric invariance for the language latent variable at pretest and post-test and the model showed equivalent effect sizes for each gender (boys $d = .26$ 95% CI $[.14, .37]$; girls $d = .28$ 95% CI $[.15, .40]$; Wald test $\chi^2 = .02$, $d.f. 1$; $p = .89$).

Secondary outcome measures

Our secondary language outcome measure is from the LanguageScreen test administered by school staff. As per our preregistered analysis plan, we created a LanguageScreen latent variable defined by the four subtests (*expressive vocabulary, receptive vocabulary, recalling sentences and listening comprehension*). The model used is shown in Figure 4 and provides an adequate fit to the data ($\chi^2(111, N = 1,173) = 142.80$; $p < .001$; RMSEA (Root Mean Square Error of Approximation) $.016$ [90% CI $.006-.023$]; CFI = $.99$; TLI = $.98$).

The pattern of results for the LanguageScreen measure shows striking similarities to the results for the standardized measures reported above. The LanguageScreen latent variable shows very high stability, and most critically the intervention group shows a significantly greater increase in their scores than the control group ($d = .32$ [95% CI $.20, .44$]). This estimate of effect size is for children at the mean of the pretest language scores, since in this model

Table 1 Mean raw scores (*SD*) for intervention and waiting control groups for primary and secondary outcome measures preintervention (t0, t1) and postintervention (t2), with effect sizes for intervention effects

	Reliability	<i>N</i>	Intervention <i>n</i> = 581		Control Group <i>n</i> = 592		Cohen's <i>d</i> [95% CI]	
			<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>		<i>SD</i>
Age (months)								
t0			53.22	3.50	53.40	3.49		
LanguageScreen (t0 & t2)								
Expressive vocabulary	.84 ^a							
t0 – (24)		563	7.25	4.16	585	7.15	3.92	.21 [.10, .32] ^c
t2 – (24)		496	11.84	3.80	543	10.99	4.17	
Receptive vocabulary	.75 ^a							
t0 – RV (31)		581	15.73	4.15	591	15.71	4.16	
t2 – RV (31)		503	21.40	3.88	555	20.36	4.13	.26 [.13, .39] ^c
Sentence repetition	.87 ^a							
t0 – SR (12)		579	2.47	2.45	585	2.44	2.46	
t2 – SR (12)		503	6.80	2.83	550	6.31	3.27	.19 [.05, .33] ^c
Listening comprehension	.77 ^a							
t0 – LC (12)		580	1.90	1.99	585	1.89	1.90	
t2 – LC (12)		503	6.12	2.89	553	5.30	2.87	.30 [.17, .43] ^c
In-depth tests (t1 & t2)								
CELF-EV	.78 ^a							
t1 – (40)		569	10.43	6.27	571	10.48	5.91	
t2 – (40)		545	16.85	7.07	560	15.37	6.66	.22 [.19, .32] ^c
CELF-RS	.87 ^a							
t1 – (22)		569	8.11	6.32	571	7.87	6.23	
t2 – (22)		545	14.25	7.46	560	13.65	7.48	.08 [–.01, .17] ^c
APT information	.86 ^a							
t1 – (40)		569	19.19	7.84	571	20.09	7.36	
t2 – (40)		545	26.24	5.87	560	25.38	6.32	.20 [.09, .32] ^c
APT grammar	.74 ^a							
t1 – (38)		569	11.80	6.86	571	12.11	5.57	
t2 – (38)		545	18.89	6.13	560	17.25	6.48	.30 [.19, .41] ^c
YARC-Early Word Reading	.94 ^b							
t1 – (30)		569	0.60	2.97	571	0.49	2.38	
t2 – (30)		545	8.96	7.60	560	8.06	6.77	

EWR is close to floor at *t1*, so we do not report an effect size for this measure; Maximum subtest scores in parentheses.

^aCronbach's alpha calculated at *t1*.

^bCronbach's alpha calculated at *t2*.

^cEffect size for the intervention based on difference in progress between groups from ANCOVA model divided by pooled *SD* for the measure at *t1* (see Morris, 2008), robust standard errors are used to correct for clustering within schools.

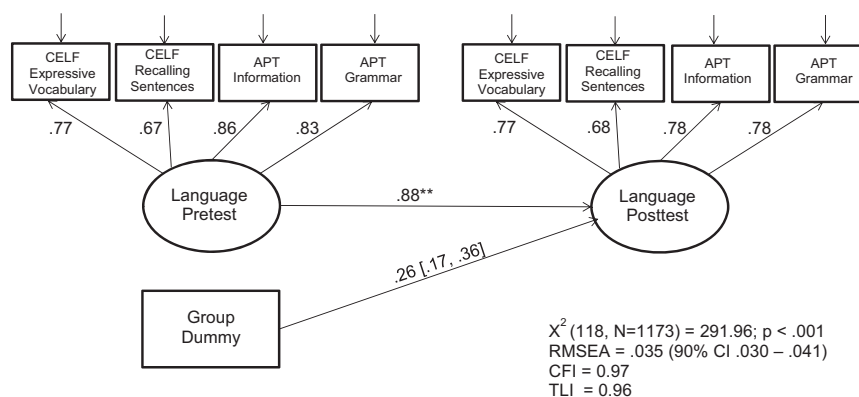


Figure 3 Path diagram showing the effects of Intervention on the individual language measures. All coefficients are standardized except for the path coefficient for Group which is y-standardized (equivalent to Cohen's *d*). The 95% robust (Huber-White) CI is shown for the Group coefficient. Stratification variables (Area, Number of Classes in a school – dummy coded) were included as additional covariates but are not shown in the model. The following correlations between residuals were included in the model: CELFEV1-CELFV2 (.48), CELFRS1-CELFRS2 (.56), APTinfo1-APTinfo2 (.14), APTgram1-APTgram2 (.11), APTinfo1-APTgram1 (.36) and APTinfo1-APTgram2 (.42) but are not shown in the diagram

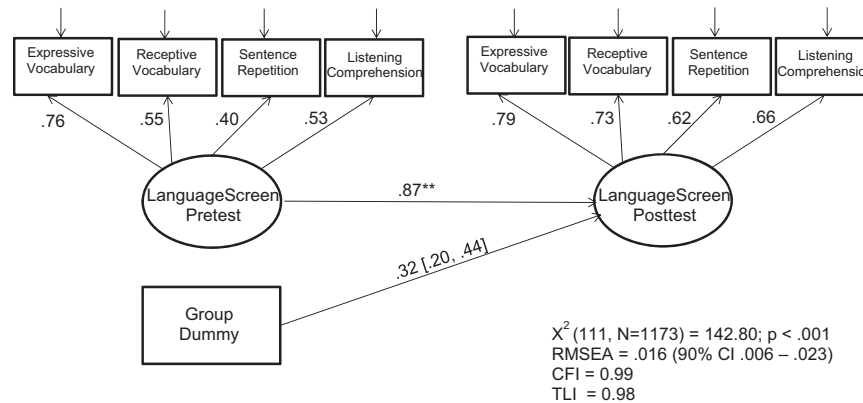


Figure 4 Path diagram showing the effects of intervention on the LanguageScreen measure. All coefficients are standardized except for the path coefficient for Group which is γ -standardized (equivalent to Cohen's d). The 95% robust (Huber-White) CI is shown for the Group coefficient. Stratification variables (Area, Number of Classes in a school – dummy coded) were included as additional covariates but are not shown in the model. The following correlations between residuals were included in the model: EV1-EV2 (.56), RV1-RV2 (.17), SR1-SR2 (.33), LC1-LC2 (.19), SR1-LC1 (.14) and SR2-LC2 (.21) but are not shown in the diagram

there was a significant group by pretest score interaction (standardized slope for interaction = $-.076$; $p = .003$). This interaction reflects the fact that children with the weakest language skills showed the largest improvements from the intervention.

Once again we checked whether the intervention was equally effective for EAL and non-EAL children. Using a multigroup model (EAL versus monolingual) based on that in Figure 4, we obtained full metric and partial scalar invariance when the intercepts (means) for expressive vocabulary were freely estimated in the two groups. In this model, the regression coefficients from intervention to post-test were freely estimated for the two groups and the effect size for the intervention was slightly larger in the EAL ($d = .45$; 95% CI [.27, .63]) than the monolingual group ($d = .31$; 95% CI [.16, .46]) though this difference was not significant (Wald test $\chi^2 = 2.2$, $d.f. = 1$; $p = .12$). A further model comparing boys and girls showed metric invariance, and revealed no significant difference in the size of the intervention effect as a function of gender (boys' $d = .33$ 95% CI [.21, .45]; girls' $d = .28$ 95% CI [.16, .42]; Wald test $\chi^2 = .15$, $d.f. = 1$; $p = .70$).

We also measured single word reading ability (YARC Early Word Reading) at pretest and post-test. At pretest, this measure displayed a floor effect with some 87% of the sample being unable to read a single word on the test. At post-test, scores had improved considerably and only 13% of the sample could not read any word on the test. We ran a mixed effects regression model controlling for stratification variables, with EWR and the Language factor score at pretest as covariates and school as the grouping variable. Given the highly non-normal distributions we used bootstrapped standard errors. This model revealed a significantly larger improvement in word reading for the intervention group (marginal mean difference = $.83$ [95% CI .10, 1.56] more words read

correctly; $z = 2.62$, $p = .009$). Given the floor effect at time 1 on EWR, we do not report a standardized effect size; but the differences in progress on this measure are clearly small (see Figure S1 in Supporting Information). We conclude that the intervention produces, at best, very small improvements in word reading, at least at this stage when children are just at the end of their first year in school.

Fidelity

NELI is a 20-week programme including both group and individual teaching sessions. TAs were asked to record the number of group and individual sessions delivered to each child, but data were missing for 20% of children. The children for whom we had data (whose TAs we presume were more engaged in delivering the programme) received a mean of 46.42 ($SD = 8.05$)/56 group sessions (83%) and 23.49 ($SD = 7.91$)/36 (65%) individual sessions. It is clear therefore that in this trial most children received less than the full dose of the programme. Because of concerns about the accuracy of the reporting of fidelity measures, we did not analyse these data further. However, the effects obtained clearly reflect a situation in which children typically receive only a subset of the NELI programme sessions.

Discussion

This study evaluated the effectiveness of the Nuffield Early Language Intervention (NELI) programme when delivered at scale in educationally realistic circumstances. The programme produced significant improvements in language skills ($d = .26$ on standardized tests; $d = .32$ on a school administered test) and small improvements in children's word reading skills.

These findings largely confirm, but considerably strengthen, the conclusions drawn from two earlier

studies of the NELI programme. Fricke et al. (2013) evaluated a 30-week version of the programme spanning the nursery and Reception year of school. They found a strong effect ($d = .80$) on language skills. In a larger trial with a modified version of the programme, Fricke et al. (2017) found improvements in language from 30-week ($d = .30$) and 20-week ($d = .21$) versions of the programme. The differences in effect sizes between these trials appeared to reflect differences in implementation quality. The current trial extends findings from previous trials and establishes that the published 20-week version of the NELI programme can produce significant improvements in children's oral language skills in a short period of time. The effects obtained here can be considered to be of educational significance since, according to the What Works Clearinghouse (2014), an effect size of 0.25 standard deviation units or larger can be considered 'substantively important'. The size of effects obtained here align well with those from a meta-analysis by Rogde et al. (2019), which indicated that interventions aimed to improve language comprehension produce the best effects if they are of high quality ($d = .24$) and delivered to small groups ($d = .25$). The current trial also shows that children who speak English as an additional language benefit as much as native speakers of English, and hence, programmes such as this may help to address some of the problems encountered by such children in school (Whiteside, Gooch, & Norbury, 2017). In addition, boys and girls responded equally well to the intervention in the current trial.

Language skills are a critical foundation for education, and for aspects of psychosocial development, including friendship formation (Hulme, Snowling, West, Lervåg, & Melby-Lervåg, 2020). Although language is a complex system, involving phonology, semantics and grammar, individual differences in language skills in children appear well described by a unidimensional construct (Klem et al., 2015; Tomblin & Zhang, 2006). Given this prior evidence, we used latent variable models to measure language skills in the current trial at pretest and post-test. Our trial was preregistered and specified the exact analytic approach used and the measures contributing to our latent variables, but not their exact factor loadings. The current study provides further evidence for the unidimensionality of children's language skills (see the high factor loadings for the language latent variables shown in Figures 3 and 4) and confirms that language skills were measured by our latent variables with very high reliability (longitudinal stabilities of $r_s = .88$ and $.87$ respectively).

The current study has also demonstrated that school staff can identify children with language difficulties effectively and that language intervention can be delivered successfully by TAs working in schools if they are given appropriate materials and

training. Since early language difficulties have strong, and likely causal, associations with later educational attainment in word reading (Hulme et al., 2015), reading comprehension (Clarke, Snowling, Truelove, & Hulme, 2010; Fricke et al., 2013; Hjetland et al., 2020) and numeracy (Hornburg et al., 2018), the ability to identify children with such difficulties is vitally important. Moreover, the improvements in oral language skills found here are sizeable and could have meaningful effects on later educational attainment (though future studies are required to confirm this). Language skills are also related to psychosocial development. This evidence, therefore, suggests that interventions to improve language skills at school entry could have important, and quite broad, effects on children's later educational development and wellbeing. More specifically, the acceleration in language skills we report here could make a real difference to how prepared a child is for education in the first years of school (see Hulme et al., 2020).

More generally, the current findings have implications for attempts to reduce social inequalities in educational outcomes. Programmes to ameliorate early language weaknesses clearly have the potential to help reduce the social gradient in language skills at school entry. Similarly, the current trial shows that the NELI programme can improve the language proficiency of minority speakers in the language of instruction.

Limitations

The current paper reports a large cluster randomized trial in which language intervention was delivered in some 100 schools and confirms that implementation at scale is practicable. However, inevitably there were some limitations. First, while clinicians typically refer children for intervention if they fall below certain agreed criteria (such as falling below age-level on 2/5 tests of receptive and/or expressive language), here children were selected according to their ranking relative to peers in the same classroom on an automated language screening measure. Arguably, selecting children on the basis of a cut-off score might be more common clinically, although educationally the procedure used here of identifying the children in each classroom who might benefit from language intervention is perhaps more realistic. In addition, while the intervention included training in oral narrative, this was not measured comprehensively at pre- or post-test owing to resource limitations. However, we did measure narrative skills in an earlier research trial of NELI using a story-retelling task (Fricke et al., 2013). That study revealed a small but significant gain in mean length of utterance in words, and a moderate increase in the number of different words used when retelling the story ($d = .55$).

Conclusion

This study provides strong evidence that a school-based language intervention programme (NELI) can produce educationally meaningful improvements in children's language skills. Further research is needed to assess the possible long-term effects of such language interventions and their implications for educational and social policy.

Supporting information

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

Figure S1. Violin plots showing gains in early word reading scores at t2 for the control and the intervention group.

Acknowledgements

This trial was funded by the Educational Endowment Foundation (EPR00640). An EEF evaluation report on the trial is available under the Open Government

Licence and can be found here (<https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/nuffield-early-language-intervention-1>). We thank the many schools, teachers, teaching assistants, children and parents for their participation.

C.H., M.S., and G.W. are Directors and M.D. is a shareholder of OxEd and Assessment Ltd, a University of Oxford spin-out company founded to distribute LanguageScreen as a commercial product. H. M. and A. H. work for Elklan Training Ltd, an independent training company which provides training for school staff using the Nuffield Early Language Intervention programme. The programme is published by Oxford University Press. Copyright of the programme is held by the Nuffield Foundation and the authors of the programme receive no royalties from sales. The remaining authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Gillian West, Department of Education, University of Oxford, Oxford, OX2 6PY, UK; Email: gillian.west@education.ox.ac.uk

Key Points

- Oral language skills are critical for education and psychosocial development.
- School staff, using an automated App, can effectively identify children with language difficulties who are eligible for intervention.
- Data from a cluster randomized controlled trial show that a 20-week language intervention delivered in Reception classes can produce educationally meaningful improvements in oral language skills.
- Children with English as an Additional Language can benefit as much from language intervention as monolingual speakers, and there are no gender effects in response to intervention.

References

- Beck, I., & McKeown, M.G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, 107, 251–271.
- Beck, I.L., McKeown, M.G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford Press.
- Bowyer-Crane, C., Snowling, M.J., Duff, F.J., Fieldsend, E., Carroll, J.M., Miles, J., ... & Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry*, 49, 422–432.
- Carroll, J.M., Bowyer-Crane, C., Duff, F.J., Hulme, C.J., & Snowling, M.J. (2011). *Developing language and literacy: Effective intervention in the early years*. Oxford: Wiley-Blackwell.
- Castro, D.C., Páez, M.M., Dickinson, D.K., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice, and policy. *Child Development Perspectives*, 5, 15–21.
- Chow, J.C., & Ekholm, E. (2019). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*, 46, 179–186.
- Clarke, P.J., Snowling, M.J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomized controlled trial. *Psychological Science*, 21, 1106–1116.
- Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders—a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry*, 46, 128–149.
- DCSF (2008). *Statutory framework for the early years foundation stage*. Nottingham: Department for Children, Schools and Families.
- DfES. (2006). *Primary framework for literacy and mathematics*. Norwich: Department for Education and Skills.
- Fricke, S., Bowyer-Crane, C., Haley, A.J., Hulme, C., & Snowling, M.J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, 54, 280–290.
- Fricke, S., Bowyer-Crane, C., Snowling, M.J., & Hulme, C. (2018). *The Nuffield early language intervention*. Oxford: Oxford University Press.
- Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., ... & Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: A randomized controlled trial. *Journal of Child Psychology and Psychiatry*, 58, 1141–1151.

- 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 Publishing.
- Hayiou-Thomas, M.E., Dale, P.S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science*, 15, 233–249.
- 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.
- 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.
- Hulme, C., Snowling, M.J., West, G., Lervåg, A., & Melby-Lervåg, M. (2020). Children's language skills can be improved: lessons from psychological science for educational policy. *Current Directions in Psychological Science*, 29, 372–377.
- Hulme, C., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Snowling, M.J. (2009). *York assessment of reading for comprehension: Early reading*. London: GL Assessment.
- Kahan, B.C., & Morris, T.P. (2011). Stratified randomisation: a hidden form of clustering? *Trials*, 12, A22.
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.A.H., Gustafsson, J.E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, 18, 146–154.
- Locke, A. (2006). *One step at a time*. London: Network Continuum Education.
- Mancilla-Martinez, J., & Lesaux, N.K. (2011). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study. *Child Development*, 82, 1544–1560.
- Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 140, 409.
- Morris, S.B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386.
- Muthén, B.O., & Muthén, L. (1998–2016). *Mplus user's guide*, 7th ed. Los Angeles: Muthén & Muthén.
- 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 Publishing.
- Rogde, K., Hagen, Å.M., Melby-Lervåg, M., & Lervåg, A. (2019). The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review. *Campbell Systematic Reviews*, 15(4), e1059.
- 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.
- Schulz, K.F., Altman, D.G., & Moher, D. & Consort Group (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Trials*, 11, 32.
- Semel, E., Wiig, E., & Secord, W. (2006). *Child Evaluation of language fundamentals-preschool UK*, 2nd ed. Oxford: Pearson Assessment.
- Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75, 417–453.
- Tomblin, J.B., Records, N.L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 40, 1245–1260.
- Tomblin, J.B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, 49, 1193–1208.
- 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.
- What Works Clearing House. (2014). Procedures and standards handbook. Available from: <https://ies.ed.gov/ncee/wwc/Handbooks#procedures>. [last accessed 11 March 2021].
- Whiteside, K.E., Gooch, D., & Norbury, C.F. (2017). English language proficiency and early school attainment among children learning English as an additional language. *Child Development*, 88, 812–827.

Accepted for publication: 29 January 2021