Early language intervention improves behavioral adjustment in school: Evidence from a cluster randomized trial

Gillian West[1], Arne Lervag[2], Margaret J. Snowling[1], Elizabeth Buchanan-Worster[1], Mihaela Duta[1], and Charles Hulme[1,2]

[1]University of Oxford, UK

[2]University of Oslo

Author Note

This trial was funded by a grant from the Education Endowment Foundation (EPR00640).
Correspondence concerning this article should be addressed to Gillian West, Department of Education, University of Oxford, OX2 6PY; Telephone: 08615 274024**;** Email: gillian.west@education.ox.ac.uk.

The Nuffield Early Language Intervention program is published by Oxford University Press, copyright of the program is held by the Nuffield Foundation and the authors of the program receive no royalties from sales. CH, MJS and GW are Directors of OxEd and Assessment, a University of Oxford spin-out company founded to distribute LanguageScreen commercially, as a measure for schools to use to assess children's language skills.

Abstract

Oral language skills are critical for psychosocial development and children with language difficulties are more likely than peers to experience behavioral problems. This study investigated the effects of an oral language intervention on behavioral adjustment. We collected teacher ratings of behavioral adjustment for 1173 children taking part in a cluster randomized trial of the Nuffield Early Language Intervention (NELI) program in 193 primary schools. Ratings were collected before and immediately after the 20-week intervention. Children receiving the language program showed significantly greater improvements than the untreated control group on a latent variable reflecting behavioral adjustment ($d = .23$). However, the improvements in behavioral adjustment for children receiving language intervention were not mediated by improvements in language. We suggest that the improvements in behavioral adjustment are a consequence of the small group and individual teaching sessions in the language intervention program, which emphasizes the need to pay attention and regulate behavior. This emphasis appears to produce generalized improvements in children's behavior regulation outside of the targeted language teaching sessions.

Keywords: language, behavior, behavioral adjustment, RCT, education

Language intervention can improve behavioral adjustment in the early school years: Evidence from a cluster randomized trial

Oral language skills are critical predictors of educational success. Language skills provide the foundation for literacy and numeracy development (Chow & Ekholm, 2019; Hulme et al., 2015) and appear important for psychosocial development (Norbury et al., 2016; van Agt et al., 2011). More generally, language and behavioral skills co-develop, and it is likely that there are reciprocal relationships between them (Chow et al., 2020).

Evidence for an association between language difficulties and socio-emotional and behavioral problems in children comes from numerous studies (e.g., Chow & Wehby, 2019; Hollo et al., 2014; Morgan et al., 2015; St Clair et al., 2011; Spilt et al., 2015). A meta-analysis of 27 longitudinal studies (Chow et al., 2018) reported a small but significant negative correlation between language skills and later behavior (r = -.14, 95% CI [-.16, -.11]), such that poor early language skills were associated with higher rates of later behavior problems. Similarly, a meta-analysis by Yew and O'Kearney (2013) estimated that children with language impairments were between 1.84–2.26 times more likely than peers to exhibit behavior problems in later childhood and adolescence.

Language difficulties may also lead to problems in forming and maintaining social relationships between children and their peers (Mok et al., 2014; Fujiki et al., 2004) and teachers (Justice et al., 2008; Spilt et al., 2015; White, 2013). For example, longitudinal research in children with developmental language disorder (Mok et al., 2014) has suggested a causal process from poor language skills to less skilled social interactions with peers, which in turn results in fewer opportunities to use and improve language skills. Similar reciprocal effects of teacher-child relationships on children's language skills have also been suggested (Pianta & Stuhlman, 2004). Furthermore, there is evidence that positive emotional support provided in the classroom by teachers moderates the negative effects of poor language on behavior (Qi et al., 2020). More generally, good child-teacher relationships may act as a protective factor for children with language difficulties (Rhoad-Drogalis et al., 2018).

Understanding the relationships between language and behavior has important implications in relation to the identification and assessment of language disorders and the implementation of interventions. Despite the established relationship between language and behavior regulation (Robson et al., 2020), many children with language difficulties are not identified until they are referred for behavioral problems (Bishop et al., 2016). Furthermore, a recent meta-analysis showed a positive association between vocabulary skills in children aged 4–7 years and later self-regulation, defined as the ability to control thoughts, feelings, and behavior. However, few studies have investigated the possible mechanisms involved. According to Gross (1998), an individual can modify their state of emotion several times in the course of an emotional experience, and the ability to do so is fostered through parent-child interactions; typically, such interactions involve language. It follows that the effects of early language skills on children's emotional and behavioral development could be mediated by the process of self-regulation (Salmon et al., 2016).

Using this framework, Curtis and colleagues (2019) investigated the effect of a parent-delivered language intervention (i.e., Enhanced Milieu Training; EMT) on the language skills and behavioral outcomes of preschool children who had received the intervention as compared with controls. EMT involves training parents to use strategies to facilitate their child's language development by noticing and responding to child-led communications, modelling language, expanding child utterances, and using language prompts. At the end of the intervention, the use of strategies for caregivers had increased significantly and children who had received the intervention had better receptive language (but not expressive) language skills than controls (Roberts & Kaiser, 2015). At follow-up 12 months later, a secondary analysis revealed that the intervention had generalized to improvements in parental ratings of behavior. In turn, these changes in behavior were partially mediated by improved language, as measured by the number of utterances spoken by the child during two play sessions. These findings are important because they provide support for the hypothesis that language difficulties are causally related to behavioral problems. However, the study was relatively small in scale (N = 97) and used a parent delivered language intervention with very young children (mean age 30 months at the beginning of the study). Within the present study, we report the findings from a study using a larger sample of older children who had received a school-based language intervention program; we also used a wider range of language outcome measures that included screening of expressive and receptive language skills, as well as individually administered standardized language assessments.

The Present Study

The present study reports data from a cluster randomized control trial that assessed the effects of a 20-week language intervention program on teachers' ratings of children's behavioral adjustment. The study investigated whether any improvements in behavior could be attributed to changes in language skills and assessed two main hypotheses: (1) early language difficulties would be causally related to problems of behavioral adjustment in children in the first year of formal schooling; if this was the case, then language intervention should lead to improvements in behavioral adjustment; and (2) improvements in behavior as the result of language intervention would be mediated by gains in language skills (i.e., Language Intervention -> Improved Language Skills -> Improved Behavioral Adjustment).

Method

Trial Design

We report analyses of data from a cluster randomized controlled trial of the Nuffield Early Language Intervention program (NELI; West et al., 2021). NELI is a 20-week program for children with poor oral language skills that is delivered by trained teaching assistants to children in the first year of formal schooling (Fricke et al., 2018). The program follows a pull-out model in which children who are identified as having language weaknesses participate in three small group sessions and two individual sessions per week. Teaching assistants (TAs) are trained to implement the program and use strategies that scaffold and facilitate children's language and communication using modelling and prompts, and also encourage the children to take turns in communicating and to listen attentively. Sessions focus on improving children's vocabulary, developing their narrative skills, encouraging active listening, and building confidence in independent speaking. The program is divided into topic areas. *Vocabulary* is taught using a multi-contextual approach within a repetitive framework (Beck & McKeown, 2007; Beck et al., 2013; Carroll et al., 2011; Locke, 2006). *Narrative work* introduces children to the sequencing of events while encouraging expressive language competence, grammatical competence, and enabling children to practice taught vocabulary in connected speech. *Listening work* targets children's active listening skills and includes auditory discrimination, memory, sequencing, and rhyming activities. To support early literacy instruction, *phoneme awareness* (blending and segmenting) and *letter-sound knowledge* are introduced in the last 10 weeks of the program.

The trial was conducted in 193 primary schools in 13 geographical clusters across England (*N* = 238 classes). Schools were randomly allocated by an independent evaluator to a 20-week language intervention or a business-as-usual control group. Randomization was stratified by the number of participating classes in each school (dichotomized as one class or more than one class) and geographical area. Language assessments took place at screening (*t0*), at pre-test (*t1*), and immediately following the intervention (*t2*). Teacher ratings of participants' behavioral adjustment to school were gathered at *t1* and *t2*.

A preregistration of this trial (https://doi.org/10.1186/ISRCTN12991126) did not include reference to the measures of adjustment that are used in this study.

< Figure 1 approximately here >

Participants

An unselected sample of 5879 children (boys *N* = 2955 (50%); English as an additional language *N* = 1001 (17%)) took part in the initial language screening at *t0*. Following screening, 1173 of these children were identified as eligible for intervention (boys *N* = 643 (55%); English as an additional language *N* = 399 (34%)). Of 581 children subsequently allocated to the intervention arm of the trial and 592 children allocated to the control arm, behavior ratings at both *t1* and *t2* were available from 486 children in the intervention arm and 522 children in the control arm. Details of the flow of participants through the study are shown in the CONSORT diagram in Figure 2 (Schulz et al., 2010).

< *Figure 2 approx here* >

Measures

Language Screening

Screening was conducted by school staff using the LanguageScreen app, a 10-min screening assessment consisting of four subtests (i.e., Expressive Vocabulary (naming 24 pictures), Receptive Vocabulary (matching each of 31 spoken words to one of four pictures), Sentence Repetition (repeating each of a 12 sentences verbatim), and Listening Comprehension (answering 12 questions on spoken stories that tap literal and inferential comprehension). Scoring was automated and results uploaded to a secure website (LanguageScreen.com). LanguageScreen reliability was high (pretest screening Cronbach's α = .84) and concurrent validity was also good, with latent variables derived from with the individually administered standardized assessments of language correlating strongly with a latent variable derived from the four LanguageScreen subtests (r = .95).

Individual Language Assessment

Language skills were assessed with the Expressive Vocabulary and Recalling Sentences subtests from the Child Evaluation of Language Fundamentals (CELF) Preschool II^{UK} (Semel et al., 2006) and the Renfrew Action Picture Test (APT; Renfrew, 2003; information and grammar scores) The CELF Preschool II^{UK} is used to assess the language ability of children between the ages of 3 – 6 years. The Expressive Vocabulary subtest comprises of 20 items measuring children's referential naming ability. The Recalling Sentences subtest includes 13 items measuring children's ability to listen to and repeat spoken sentences without changing their semantics, morphology or syntax. The standardization by the test developers demonstrates good reliability (CELF Expressive Vocabulary and Recalling Sentences subtests Cronbach's α = .82 and .88 respectively) and concurrent validity with PLS-4 and CELF-4 (r's = .85 - .93). The APT is a measure of children's spoken language (aged 3 – 8 years old), including words used to convey information, use of tenses, and sentence construction. Children are shown 10 picture cards and asked a question on each one. Answers are scored for information content and grammatical complexity. Cronbach's alphas for Information and Grammar scores at pretest were r = .86 and r = .80 respectively.

Behavioral Adjustment Ratings

Children's class teachers completed the 12-item Behavioral Adjustment sub-scale of the Brief Early Skills and Support Index (BESSI; Hughes et al., 2015) at pretest (*t1*) and posttest (*t2*). The scale consists of 12 items rated on a 4-point scale ranging from *strongly agree* (1) to *strongly disagree* (4) with various items being reverse scored. BESSI items were grouped into 3 subsets, including (a) Behavioral Regulation ("Good at waiting patiently when this is required", "Has temper tantrums", "Is good at calming down when asked to do so"; "Responds poorly to reprimands [e.g., backchat, anger]", "Gets easily frustrated if a task is too difficult", "Grabs other children's belongings", "Is respectful towards adults"), (b) Attention/Hyperactivity ("Has trouble sitting still when required", "Often interrupts conversations inappropriately", "Is easily distracted"), and (c) Sociability ("Can play with lots of different children of his or her own age", "Is usually happy to share with peers"). This 12-item scale has good reliability (Cronbach's α = .90). The scores for each of the BESSI subsets were used as indicators of a behavioral adjustment factor.

Procedures

Ethical permission for the study was granted by the Research Ethics Committee of the University of Oxford. Head teachers provided consent to take part in the trial. Children in the first year of formal schooling in participating classrooms were enrolled to take part in the initial language screening at *t0* on an opt-out basis. School staff screened all children using a language assessment app (LanguageScreen: https://www.languagescreen.com/) running on an Android tablet. The five children with the lowest LanguageScreen scores in each class were identified as eligible for the NELI language intervention. In eight schools with a class size of under 10 children, three children per class were selected and in one school with two classes on different sites, 9 children were allocated across both classes. At pretest (*t1*), these children's language skills were individually assessed by speech and language therapists trained by the research team and ratings of behavioral adjustment for the children were completed by their teachers. The same individual language assessments and behavioral adjustment ratings were collected at posttest (*t2*) immediately after the intervention.

Analyses

All analyses were performed on an intention-to-treat basis. Analyses using observed variables were conducted in Stata 16.1 (StataCorp, 2019). Structural equation models (SEM) using latent variables were constructed using Mplus 8.4 (Muthen & Muthen, 1998-2017) with Full Information Maximum Likelihood estimators to allow for missing data.

Data consisted of longitudinal assessments of language skills and behavioral regulation (i.e., BESSI ratings) at pretest and posttest. Before assessing possible changes in these constructs associated with the intervention, we first conducted longitudinal Confirmatory Factor Analysis (CFA) models for each construct.

To assess whether the language intervention was associated with improvements in behavioral adjustments ratings, we estimated a latent variable analysis of covariance (ANCOVA) model (see Figure 3) with group (Intervention vs Control) dummy coded. In this model, the behavioral adjustment construct was reflected by measures using the Behavioral Regulation, Attention/Hyperactivity, and Sociability subscales of the BESSI at both pre and posttest.

To test whether any potential improvements in behavioral adjustment caused by the language intervention were mediated by children's improvements in language skills, we used a latent variable version of Valante and MacKinnon's (2017) ANCOVA model for estimating the mediated (indirect) effects of an intervention. In this model, the indirect effect of language intervention on behavioral adjustment at posttest (mediated via language skills at the posttest) is estimated after controlling for both behavioral adjustment and language skills at pretest. Conversely, to address our hypothesis about the effect of the intervention on language skills being mediated through the effect on behavioral adjustment, we estimated an equivalent latent variable ANCOVA model where we estimated the indirect effect of the language intervention on behavioral adjustment at posttest through language skills, after controlling for both behavioral adjustment and language skills at pretest. In both models, behavioral adjustment was measured as described above and the language latent variable by the two CELF tests and the two APT tests at both pre and posttest.

To control for the clustering of children within schools, we used robust (Huber-White) standard errors in all the latent variable models.

Results

Individual language assessments (*t1*) were obtained from 571 control and 569 intervention children identified during screening. BESSI ratings were completed for 522 control and 486 intervention children. Of children with completed *t1* BESSI ratings, 68 control (13%) and 97 (20%) intervention children did not receive ratings at follow up. No statistically significant group differences were found on pretest BESSI ratings (t = -.74; p = -.463; d = -.12, 95% CI [-0.43, 0.19]) or language skills (language composite Z score t

= -.08; *p* = 0.353; *d* = -.15, 95% CI [-0.16, 0.46]) between children lost to follow up on the BESSI ratings and those retained in the sample.

Descriptive statistics and reliabilities for all measures at pretest and posttest for both groups are shown in Table 1. Higher behavioral ratings indicate poorer behavior. It was clear that the groups were well equated on ratings of language skills and behavioral regulation at pretest. In addition, it appears that the language intervention group showed larger improvements on measures of language skills and ratings of behavioral adjustment (decreases in ratings) at posttest compared to controls.

< *Table 1 approx here* >

Prior to the main analyses, longitudinal CFA models examining behavioral adjustment (BESSI ratings) and language skills were conducted. In these models, as is typical in longitudinal CFA models (Brown, 2006), correlated errors between the same indicators at both time points were included to account for shared method variance between these indicators. For the BESSI ratings, the longitudinal correlations between the residuals of *t1* ADHD and *t2* ADHD and *t1* Sociability and *t2* Sociability were significant and thus were retained in the model (see Figure 3). However, the longitudinal correlation between the residuals of *t1* Behavior Control and *t2* Behavior Control was small and not significant, and thus was dropped from the model. This model showed metric invariance because there was no significant difference between a freely estimated model and a model where the unstandardized factor loadings were fixed to be equal across time, $\Delta\chi 2(2) = 2.892$, $p = .2849$, ΔCFI = -.001, ΔRMSEA =- .006, ΔSRMR = .001. The model also showed partial scalar invariance (i.e., invariant loadings and intercepts across time for both ADHD and Behavior Control, but not for Sociability) because a model where the intercepts for these two indicators were constrained to be equal did not differ appreciably from the previous model with unconstrained intercepts, ΔCFI = .004, ΔRMSEA = .008, ΔSRMR = .003; $\Delta\chi 2(1) = 11.350$, $p = .0008$. Note that in this case, we relied on the small changes in CFI and RMSEA to judge changes in model fit, because with a large sample that results in high power, the test of the change in $\chi 2$ may be too stringent (see Chen, 2007).

Our first major question was to assess whether changes in the behavioral adjustment latent variable were brought about by the intervention. We used the longitudinal CFA model of the BESSI scores as the basis for an ANCOVA model to assess the size of change attributable to the intervention (see Figure 3).

< insert Figure 3 approx here >

As indicated in Figure 3, the children in the intervention group had significantly lower (i.e., better) scores ($d = -.23$, 95% CI [-.10, -.37]) on behavioral adjustment at posttest as compared to the control group. This effect was negative because higher ratings indicate poorer behavioral adjustment. An interaction between the behavioral adjustment pretest factor scores and group confirmed that slopes relating to pretest-to-posttest behavioral adjustment scores did not differ between groups. In other words, children with the poorest behavior regulation showed equivalent improvements at posttest to children with better behavior regulation. This model fitted the data well, $\chi 2(14) = 50.338$, RMSEA = .049, 90% CI [.035, .064], CFI = .985, SRMR = .037. It should be noted that the size of changes in ratings on the different BESSI items varied, but no systematic differences were observed (see Appendix A for details).

For language skills, we also conducted a longitudinal CFA model prior to the mediation models that we wanted to conduct. In this CFA model, correlated errors between the same indicators at both time points were initially included, as well as correlated errors between the APT information and APT grammar measures (given that they were derived from the same test) to account for shared method variance between these indicators (see Figure 4). All correlations between the residuals were significant and were retained in the model. We assessed whether we had metric and scalar invariance by first constraining the unstandardized factor loadings for each indicator to be equal across time and subsequently constraining the intercepts for each indicator to be equal across time. These analyses revealed that a model with partial metric and scalar invariance provided an adequate fit to the data. In this model all factor loadings and intercepts were constrained to be equal across time except for the APT information indicator. Partial metric invariance was demonstrated by the fact that constraining the other three factor loadings to be equal across time, whereas freely estimating the factor loading for APT information resulted in no significant loss of fit (ΔCFI = -.006, ΔRMSEA = .014, ΔSRMR = .025 but not for: $\Delta\ \chi 2(2) = 37.170(2)$, $p < .0001$). Similarly, partial metric and scalar invariance were demonstrated by subsequently constraining the three intercepts to be equal across time, while freely estimating the intercept for APT information, which again resulted in no appreciable loss of fit ($\Delta\chi 2(2) = 5.249$, $p < .0725$, ΔCFI = -.000, ΔRMSEA = -.004, SRMR = .001).

We used this longitudinal CFA model as the basis for an ANCOVA model assessing the size of change in language scores attributable to the intervention.

< insert Figure 4 approx here >

As seen from the model in Figure 4, children in the intervention group had significantly better language scores (d = .26, 95% CI [-.16, -.36]) at posttest as compared to the control group. Given that the language factor only displays partial metric invariance, we cannot conclude that an underlying language factor had been changed by the intervention; this lack of metric invariance across time appears to reflect the fact that larger improvements were made on some of the language measures (i.e., CELF Expressive Vocabulary and APT Grammar) as compared to others (i.e., CELF Recalling Sentences and APT Information; see Table 1 for effect sizes). This model fitted the data well: $\chi 2(24) = 104.755$, RMSEA = .053, 90% CI [.043, .064], CFI = .984, SRMR = .040. We found no interaction between the pretest and the treatment.

It should be noted that the model in Figure 4 with correlated errors between the residuals for the APT Information and APT Grammar Scores is equivalent to a 2-factor model in which we separate the two pairs of indicators (APT Information and APT Grammar vs CELF Expressive Vocabulary and CELF Recalling Sentences) to define two separate language factors: APT Language and CELF Language. We fitted this model to our data but do not report it here because there is strong evidence from previous research that diverse language measures correlate strongly with each other and define a meaningful common language factor (Hulme et al., 2020; Tomblin & Zhang, 2006). Furthermore, although perhaps statistically separable in our data, these two language factors correlated very highly together ($r = .94$ at *t1*) and showed equivalent amounts of improvement from the intervention (APT Language $d = .27$, 95% CI [.15, .40]; CELF Language $d = .23$, 95% CI [.13, .33]).

Finally, we considered whether changes in behavior regulation were mediated by improvements in language (or vice versa). In other words, did improvements in language from the intervention indirectly lead to improvements in behavior regulation (Figure 5), perhaps as a result of changes in emotion regulation associated with the intervention (e.g., more use of inner speech to regulate behavior)? Figure 5 shows the path diagram for this model. The indirect effect was negative as improvements in language skills were associated with reductions (i.e., improvements) in behavioral adjustment scores, but these changes were not significant (d = -0.044, 95% bootstrapped CI [-0.098, 0.010], p = 0.146). An equivalent model assessed whether the improvements in language skills produced by the intervention were mediated by the improvements in behavioral adjustment (see Figure 6); once again this indirect effect was not significant (d = 0.015, 95% bootstrapped CI [-0.001, 0.031], p = 0.137).

< *insert Figure 5 and Figure 6 approx here* >

The conclusion from these models is that there are moderately sized direct effects of the intervention on both language skills (Intervention -> Language) and behavior regulation (Intervention -> Behavioral Adjustment); however, these effects were largely independent of each other (i.e., the improvements in behavioral adjustment produced by the intervention were not mediated by improvements in language skills and vice versa). These models, therefore, did not provide any support for causal effects such that improvements in language led to improvements in behavioral adjustment, or vice versa. However, the design here is not capable of identifying the direction of any such hypothetical causal effects if they were found; for that purpose, we would require an additional follow-up measurement in which both language and behavioral adjustment were assessed.

Discussion

This study investigated the hypothesis that a language intervention delivered to young children with language difficulties would improve their behavioral adjustment. The intervention produced moderate and significant effects both on language skills (d = .26) and on teacher ratings of behavioral adjustment (d = .23) in comparison to the control group. The implemented intervention directly targeted language skills so that improvements in language were to be expected. Improvements in behavioral adjustment at first seemed more difficult to explain; however, the intervention involved individual and small group work with a teaching assistant, which involves teaching children to keep still and to attend and to speak at appropriate times in the sessions. An important finding from this study is that these individual and small group teaching activities were associated with generalized improvements in behavioral adjustment in school. These improvements in behavior, however, were not accounted for by improvements in language skills.

The Association Between Language Difficulties and Behavioral Adjustment

The findings from this study are consistent with a large body of evidence showing an association between language difficulties and children's behavioral adjustment (Chow et al., 2018; Chow & Wehby, 2019; Morgan et al., 2015; Yew & O'Kearney, 2013). A plausible explanation for this well-established association is that language skills have a causal effect on behavior. A prominent theory (Gross, 1998) is that the mechanisms involved include a range of strategies to regulate emotional experiences as they unfold. Further, language (perhaps 'inner speech') may provide the structure through which these processes operate (Salmon et al., 2016; Vygotsky, 1962; Zadeh et al., 2007).

A previous study by Curtis et al. (2019) that followed pre-school children 1 year after a parent-delivered language intervention had finished found that improvements in language in the intervention group partially mediated gains in behavior. Our present study found positive effects of a language intervention on both language and behavioral adjustment, but we found no evidence for a mediated or for a partially mediated effect (Language intervention -> Language skills -> Behavioral adjustment), similar to the effect Curtis et al. had found. It is unclear why Curtis et al. found evidence for a partially mediated effect whereas the current study did not. The two studies differ in several ways. The Curtis et al. study involved pre-school children and an intervention delivered by parents at home, whereas the current study involved older children with a school-based intervention delivered by school staff. Furthermore, language was measured in the Curtis et al. study by counting the number of utterances produced by a child in a free play session with a parent, whereas in the current study we used standardized assessments of the quality of children's language. Finally, our assessment of behavior was made at the end of the intervention rather than at a later follow-up point as in the Curtis et al. study; it is plausible that the effect of better language skills may take time to influence behavior regulation, and this was not explored in our study.

In summary, although we failed to find evidence of a partially mediated effect (i.e., Language intervention -> Language skills -> Behavioral adjustment) as found by Curtis et al. (2019), it remains plausible that such effects exist. However, clearly other factors are likely to be involved. In school-aged children, for example, teachers and other school staff play an important role in instilling good behavior through strategies and dialogue, such as 'interpreting' the causes of a child's tantrum or by ensuring they understand class instructions. Following an investigation of children from low-income backgrounds and the environments in their Head Start settings, Qi et al. (2020) reported that the association of language and behavior problems varied as a function of emotional support provided by the teacher to the class. Here, emotional support was assessed through observations of classroom climate, teacher sensitivity and regard for student perspectives.

We speculate that the mechanisms responsible for the improvements in behavioral regulation observed in the current study reflect aspects of the way the NELI program is delivered. The program is a pull-out program in which children with language difficulties are given additional small group and individual language teaching by a trained teaching assistant. These individual and small group teaching sessions involve intensive and structured interactions between children and the teaching assistant. Alongside language-focused vocabulary and narrative work, the program continually reinforces the active listening and attentional skills required for learning. The program also gives guidance to teaching assistants in how to foster good listening behavior in children and how to reward children for displaying it. In this way, the regular NELI sessions may promote pro-social and classroom-appropriate behavior in children, independently from language learning outcomes. Additionally, NELI enables a child to work closely with a teaching assistant over an extended period, thereby establishing the foundation for the development of a positive adult-child relationship that the child may then be able to use as a model for other relationships going forward.

Implications for School Psychology

Our findings have important implications for understanding and managing the effects of language difficulties in school children and for education policy. Language difficulties are common among children in the early school years and can be ameliorated by high quality interventions (for a review, see Hulme et al., 2020). Our results confirm that language intervention, delivered in the early school years, can have positive effects on children’s language skills. In addition, we have shown that consistent with Curtis et al. (2019), language interventions may have broader effects in that the requirements of the NELI program for children to sit still and listen, coupled with rewarding children for following these requirements, appears to produce generalized improvements in children’s behavior in school.

Finally, the current study highlights the vital importance of early identification of children's language problems. Interventions, such as NELI, show promise not just in ameliorating language weaknesses in children, but also in mitigating the adverse educational and social consequences of poor behavioral adjustment in school. Our study used a language screening app (i.e., LanguageScreen) administered by school staff. The development of easy-to-use, reliable screening tools brings benefits in allowing school staff to identify children with language weaknesses. Such screening measures, more broadly, are also potentially important in raising awareness amongst school staff of how common language weaknesses are in children of school age.

Future Directions

We believe that future research examining the effects of language interventions should include a wider range of outcome measures (i.e., measures of language and literacy skills, as well as measures of child behavioral outcomes, socio-emotional development, and wellbeing). In addition, it is important for research in this area to include longer term follow-ups of children. The current study demonstrated educationally meaningful improvements in children's language and behavior regulation, but we need to know whether such effects are *durable*. In addition, developing a more complete understanding of the mechanisms operating in the NELI intervention that result in improvements in both language and behavior is an important issue for future research.

Limitations

This study evaluated behavioral adjustment concurrently with language outcomes immediately following an intervention. Ideally, a later follow-up measure should have been included. In addition, the study used a single measure of teacher-rated behavioral adjustment. Chow and Wehby (2019) showed that the relationship between language and behavior varies according to the type of behavioral measure used as they found a relationship between language and behavior only when using a direct observational measure and not when relying on teacher ratings. Although the current study found no evidence of a mediating effect of language on behavioral adjustment, the study only followed children for a short period of time, and our design with only two times of measurement cannot provide any direct test of causation. It remains for future studies to clarify whether, and under what circumstances, language may exert a causal influence on children's behavior regulation, and on broader measures of psycho-social adjustment and wellbeing. It is quite possible that any such effects may be reciprocal. Notwithstanding these limitations, it is important to note that the current findings come from a large randomized controlled trial and they demonstrate clear evidence that a school-delivered language intervention program can have positive effects on children's language and behavioral adjustment.

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*(3), 251–271. DOI:10.1086/511706

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

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE Consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLOS One*, *11*(7), e0158753.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.

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*. Wiley-Blackwell.

Chen F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. doi: 10.1080/10705510701301834.

Chow, J. C., Cunningham, J., & Wallace, E. (2020). Interaction-centered model for language and behavioral development. In T. Farmer, B. Farmer, K. Sutherland, & M. Conroy (Eds.), *Handbook of research on emotional & behavioral disabilities: interdisciplinary developmental perspectives on children and youth* (pp. 1-28). Routledge. https://doi.org/10.4324/9780429453106.

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

Chow, J. C., Ekholm, E., & Coleman, H. (2018). Does oral language underpin the development of later behavior problems? A longitudinal meta-analysis. *School Psychology Quarterly*, *33*(3), 337.

Chow, J. C., & Wehby, J. H. (2019). Profiles of problem behavior in children with varying language ability. *Journal of Emotional and Behavioral Disorders*, *27*(2), 110–118.

Curtis, P. R., Kaiser, A. P., Estabrook, R., & Roberts, M. Y. (2019). The longitudinal effects of early language intervention on children's problem behaviors. *Child Development*, *90*(2), 576–592. doi:10.1111/cdev.12942

Fricke, S., Bowyer-Crane, C., Snowling, M., J., & Hulme, C. (2018). *The Nuffield early language intervention*. Oxford University Press.

Fujiki, M., Spackman, M. P., Brinton, B., & Hall, A. (2004). The relationship of language and emotion regulation skills to reticence in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47, 637–646.

Gross, J.J. (1998) The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Hollo, A., Wehby, J. H., & Oliver, R. M. (2014). Unidentified language deficits in children with emotional and behavioral disorders: A meta-analysis. *Exceptional Children*, *80*(2), 169–186.

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*(12), 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*(4), 372–377.

Justice, L. M., Cottone, E. A., Mashburn, A., & Rimm-Kaufman, S. E. (2008). Relationships between teachers and preschoolers who are at risk: Contribution of children's language skills, temperamentally based attributes, and gender. *Early Education & Development*, *19*, 600–621. doi: 10.1080/10409280802231021

Locke, A. (2006). *One step at a time*. Network Continuum Education.

Mok, P. L., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2014). Longitudinal trajectories of peer relations in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, *55*(5), 516–527.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, *86*(5), 1351-1370.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364-386.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide* (8th ed.). Los Angeles, CA: 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*(11), 1247–1257.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review*, *33*(3), 444–458.

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.

Qi, C. H., Zieher, A., Lee Van Horn, M., Bulotsky-Shearer, R., & Carta, J. (2020). Language skills, behavior problems, and classroom emotional support among preschool children from low-income families. *Early Child Development and Care*, *190*(14), 2278–2290, doi: 10.1080/03004430.2019.1570504.

Renfrew, C. (2003). *Action picture test.* Speechmark Publishing.

Rhoad-Drogalis, A., Justice, L. M., Sawyer, B. E., & O'Connell, A. A. (2018). Teacher–child relationships and classroom-learning behaviors of children with developmental language disorders. *International Journal of Language & Communication Disorders*, *53*(2), 324–338.

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, *146*(4), 324.

Roberts, M. Y., & Kaiser, A. P. (2015). Early intervention for toddlers with language delays: A randomized controlled trial. *Pediatrics*, *135*(4), 686–693. doi: 10.1542/peds.2014-2134

Salmon, K., O'Kearney, R., Reese, E., & Fortune, C. A. (2016). The role of language skill in child psychopathology: Implications for intervention in the early years. *Clinical Child and Family Psychology Review*, *19*(4), 352–367.

Schulz, K. F., Altman, D. G., Moher, D., & Consort Group. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials*, *11*(1), 32.

Semel, E., Wiig, E., & Secord, W. (2006). *Child Evaluation of Language Fundamentals-Preschool UK* (2nd ed.). Pearson Assessment.

Spilt, J. L., Koomen, H. M., & Harrison, L. J. (2015). Language development in the early school years: The importance of close relationships with teachers. *Developmental Psychology*, *51*(2), 185–196.

St Clair, M. C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders*, *44*(2), 186–199.

StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.

Valente, M. J., & MacKinnon, D. P. (2017). Comparing models of change to estimate the mediated effect in the pretest–posttest control group design. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(3), 428–450.

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*(1), 81–88.

Vygotsky, L. S. (1962). *Thought and language*. MIT Press.

West, G., Snowling. M., J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., McLachlan, H., & Hulme, C. (2021). Early language screening and intervention can be delivered successfully at scale: Evidence from a cluster randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *62*(12), 1425-1434.

White, K. M. (2013). Associations between teacher–child relationships and children's writing in kindergarten and first grade. *Early Childhood Research Quarterly*, *28*(1), 166-176.

Yew, S. G. K., & O'Kearney, R. (2013). Emotional and behavioral 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*(5), 516–524.

Zadeh, Z. Y., Im-Bolter, N., & Cohen, N. J. (2007). Social cognition and externalizing psychopathology: An investigation of the mediating role of language. *Journal of Abnormal Child Psychology*, *35*(2), 141–152.

Table 1

Mean Raw Scores (SD) for Intervention and Waiting Control Groups for Primary and Secondary Outcome Measures Pre-Intervention (t0, t1) and Post-Intervention (t2), with Effect Sizes for Intervention Effects

	Reliability	Intervention n=581			Control Group n=592			Cohen's *d*
		N	*M*	SD	N	*M*	SD	
Age (months) at *t0*			53.22	3.50		53.40	3.49	
LanguageScreen (t0)	.83[a]	581	29.37	11.93	592	29.18	11.55	
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	0.25[b]
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	0.10[b]
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	0.16[b]
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	0.28[b]
BESSI – Behavioral adjustment	.89[a]							
• t1 (12)		486	14.20	6.73	522	13.79	7.06	
○ Regulation (7)		493	6.91	4.03	523	6.86	4.26	
○ ADHD (3)		496	4.71	2.40	527	4.74	2.40	
○ Sociability (2)		500	2.51	1.34	528	2.21	1.32	
• t2 (12)		430	10.45	6.82	490	11.77	7.26	-0.25[b]
○ Regulation (7)		441	5.06	3.94	501	5.89	4.33	-0.20[b]
○ ADHD (3)		439	3.69	2.51	501	4.17	2.43	-0.22[b]
○ Sociability (2)		441	1.74	1.20	500	1.74	1.31	-0.11[b]

Note. [a] Cronbach's alpha calculated at Time 1; [b] Effect 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); Maximum scores in parentheses.

Figure 1

Timeline of Trial Showing Assessment, Training, and Intervention Phases

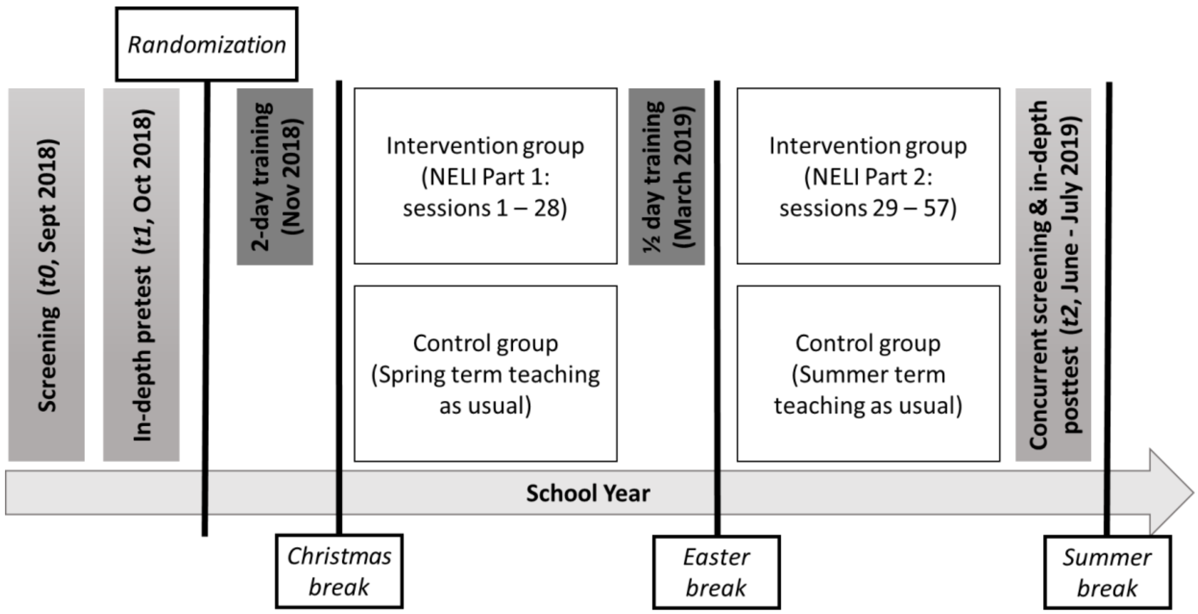


Figure 2

CONSORT Diagram Showing Flow of Participants Through RCT

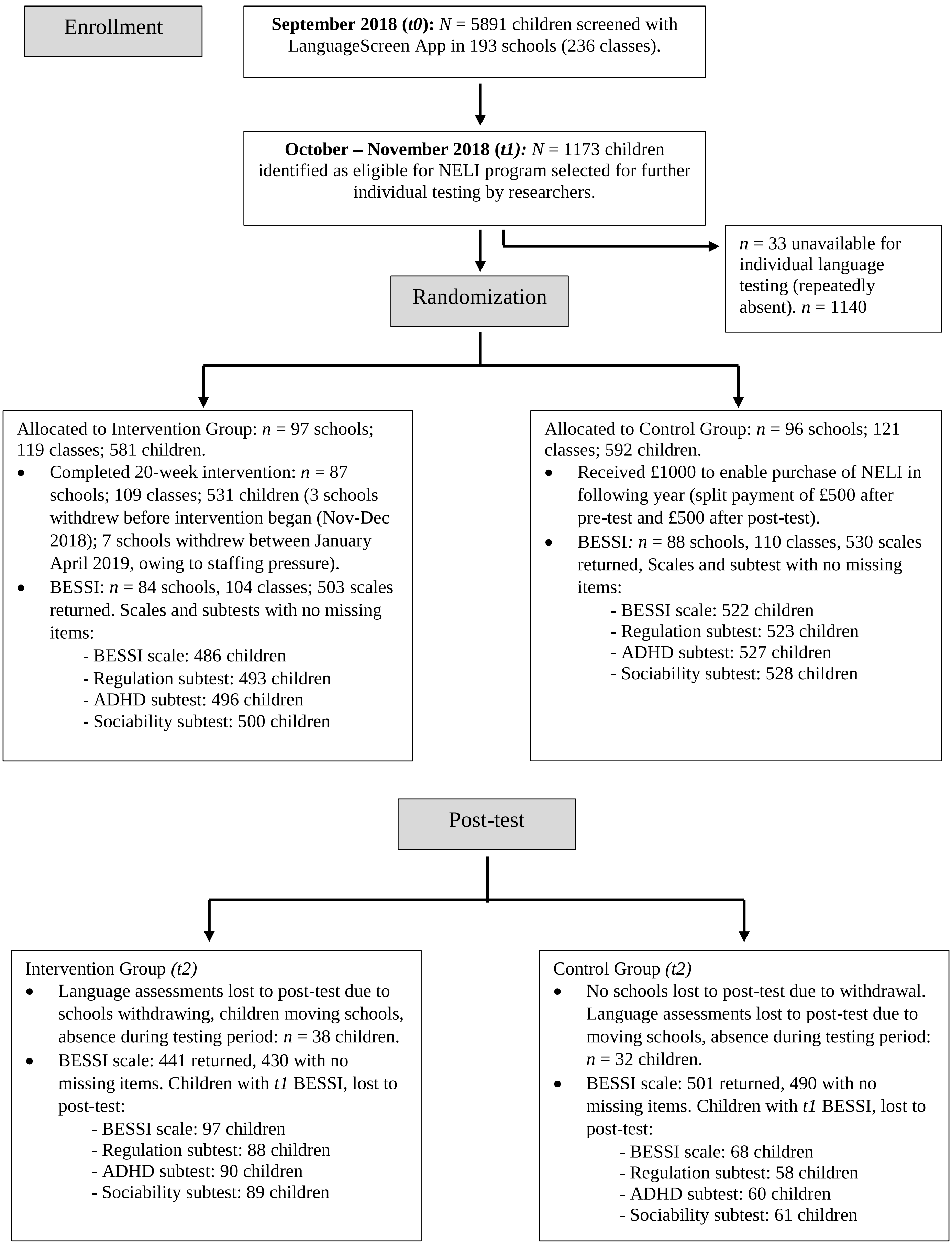
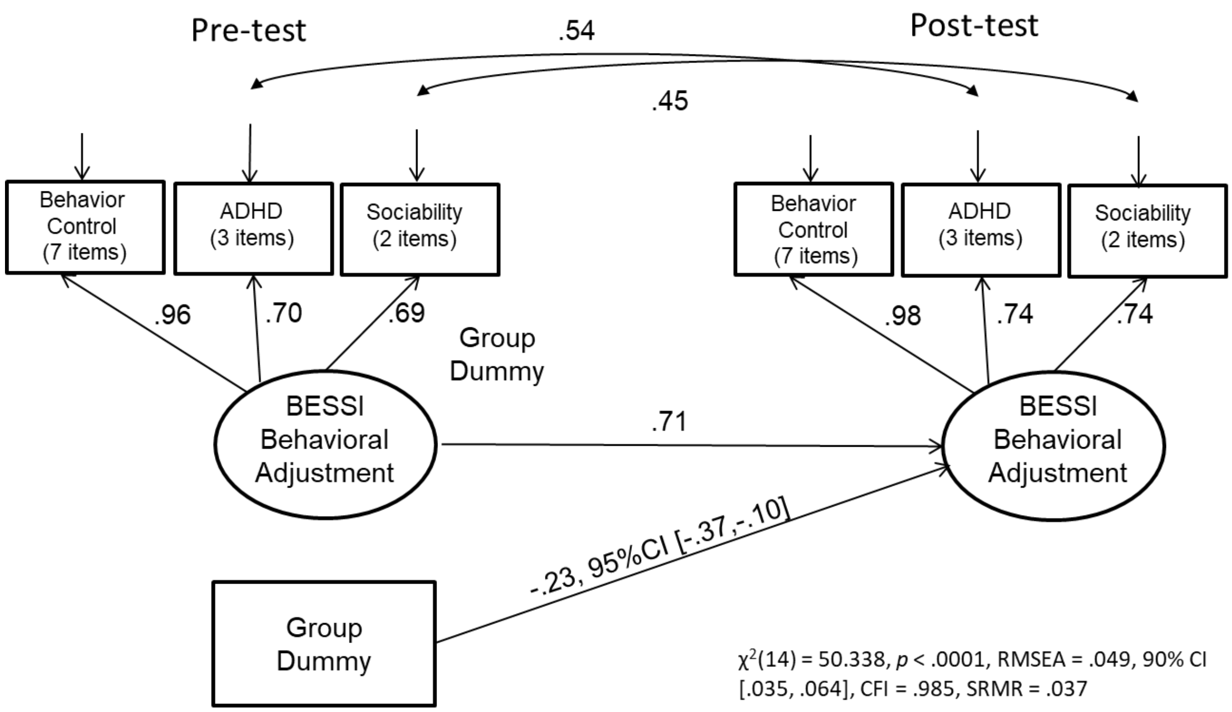

Enrollment
September 2018 (t0): N = 5891 children screened with LanguageScreen App in 193 schools (236 classes).
October – November 2018 (t1): N = 1173 children identified as eligible for NELI program selected for further individual testing by researchers.
n = 33 unavailable for individual language testing (repeatedly absent). n = 1140
Randomization
Allocated to Intervention Group: n = 97 schools; 119 classes; 581 children.
• Completed 20-week intervention: n = 87 schools; 109 classes; 531 children (3 schools withdrew before intervention began (Nov-Dec 2018); 7 schools withdrew between January–April 2019, owing to staffing pressure).
• BESSI: n = 84 schools, 104 classes; 503 scales returned. Scales and subtests with no missing items:
- BESSI scale: 486 children
- Regulation subtest: 493 children
- ADHD subtest: 496 children
- Sociability subtest: 500 children
Allocated to Control Group: n = 96 schools; 121 classes; 592 children.
• Received £1000 to enable purchase of NELI in following year (split payment of £500 after pre-test and £500 after post-test).
• BESSI: n = 88 schools, 110 classes, 530 scales returned, Scales and subtest with no missing items:
- BESSI scale: 522 children
- Regulation subtest: 523 children
- ADHD subtest: 527 children
- Sociability subtest: 528 children
Post-test
Intervention Group (t2)
• Language assessments lost to post-test due to schools withdrawing, children moving schools, absence during testing period: n = 38 children.
• BESSI scale: 441 returned, 430 with no missing items. Children with t1 BESSI, lost to post-test:
- BESSI scale: 97 children
- Regulation subtest: 88 children
- ADHD subtest: 90 children
- Sociability subtest: 89 children
Control Group (t2)
• No schools lost to post-test due to withdrawal. Language assessments lost to post-test due to moving schools, absence during testing period: n = 32 children.
• BESSI scale: 501 returned, 490 with no missing items. Children with t1 BESSI, lost to post-test:
- BESSI scale: 68 children
- Regulation subtest: 58 children
- ADHD subtest: 60 children
- Sociability subtest: 61 children

Figure 3

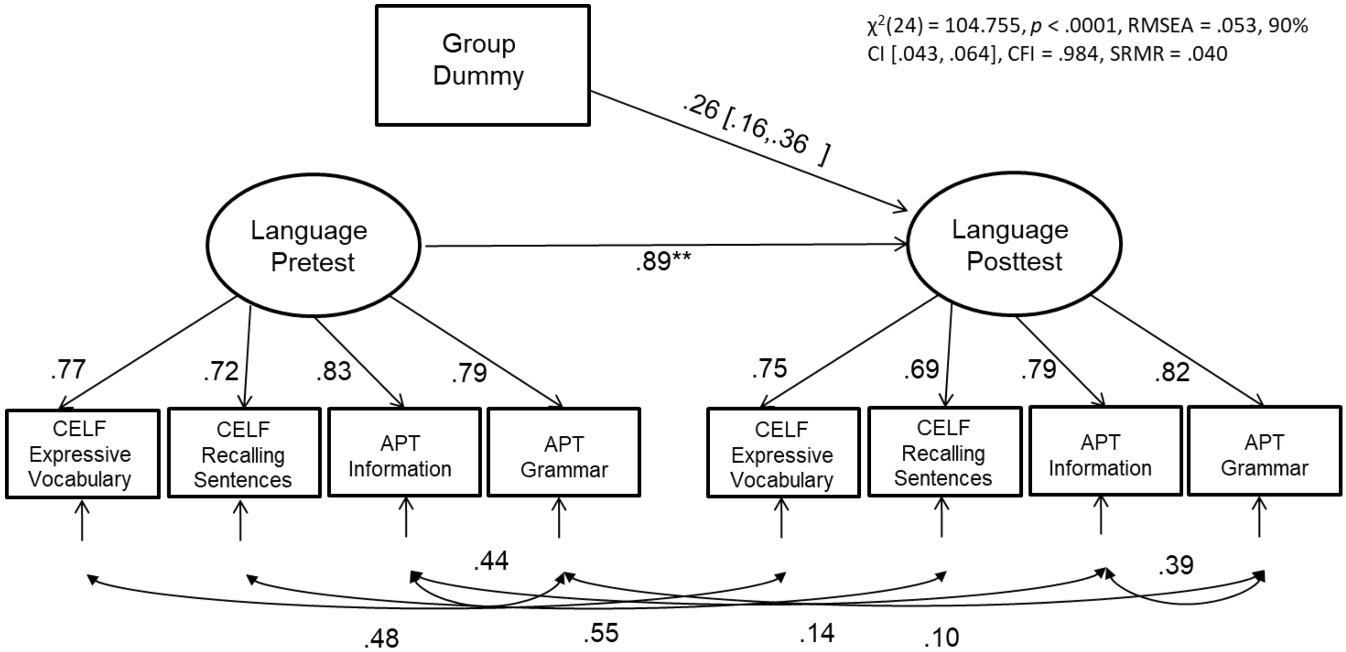
Path Diagram Showing the Effects of Intervention on Ratings of Behavioral Adjustment (BESSI)



Note. All coefficients are standardized except for the path coefficient for Group, which is y-standardized (equivalent to Cohen's *d*). The standard error for the treatment effect uses a robust (Huber-White) cluster estimator.

Figure 4

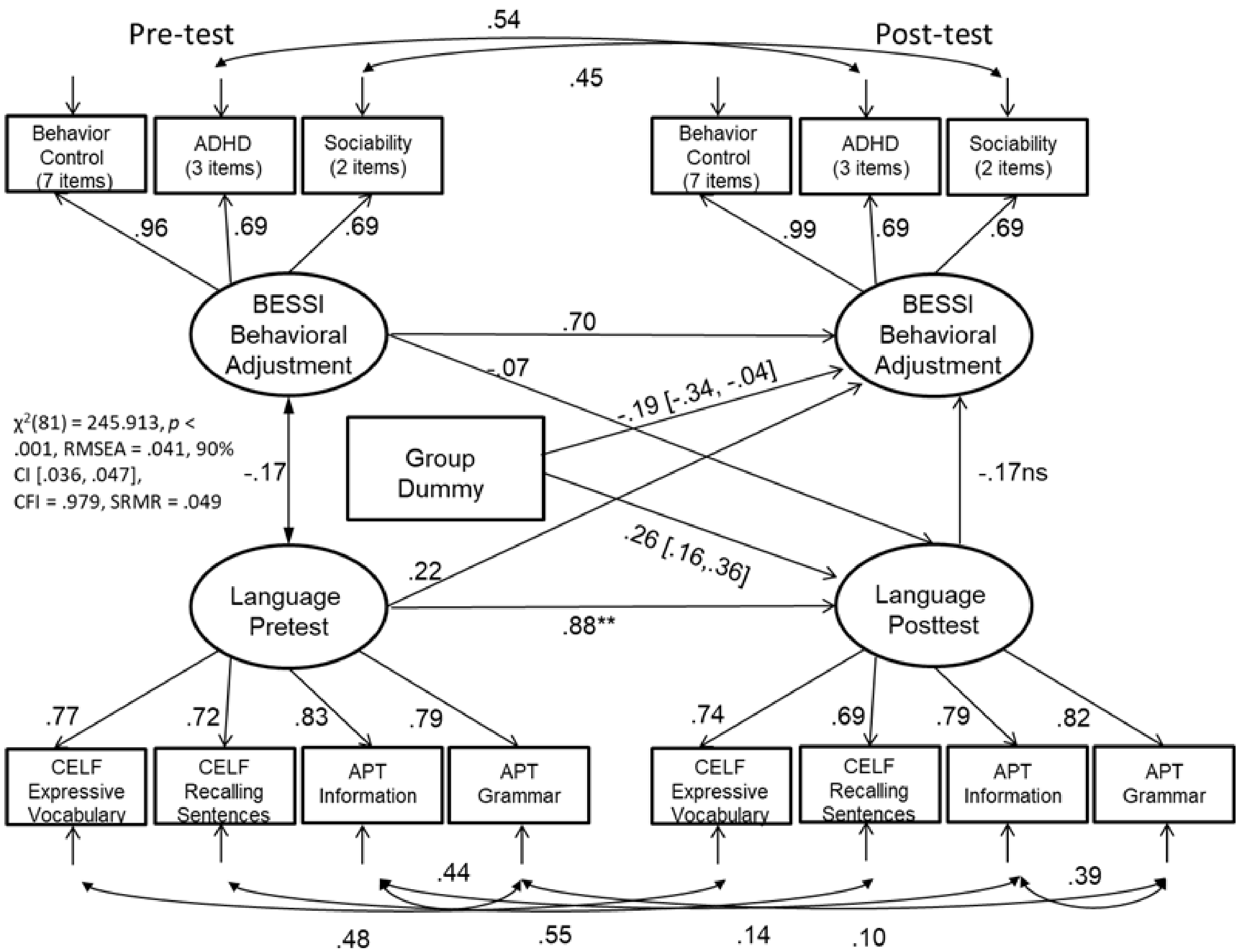
Path Diagram Showing the Effects of Intervention on Language Skills



Note. All coefficients are standardized except for the path coefficient for Group, which is y-standardized (equivalent to Cohen's d). The standard error for the treatment effect uses a robust (Huber-White) cluster estimator.

Figure 5

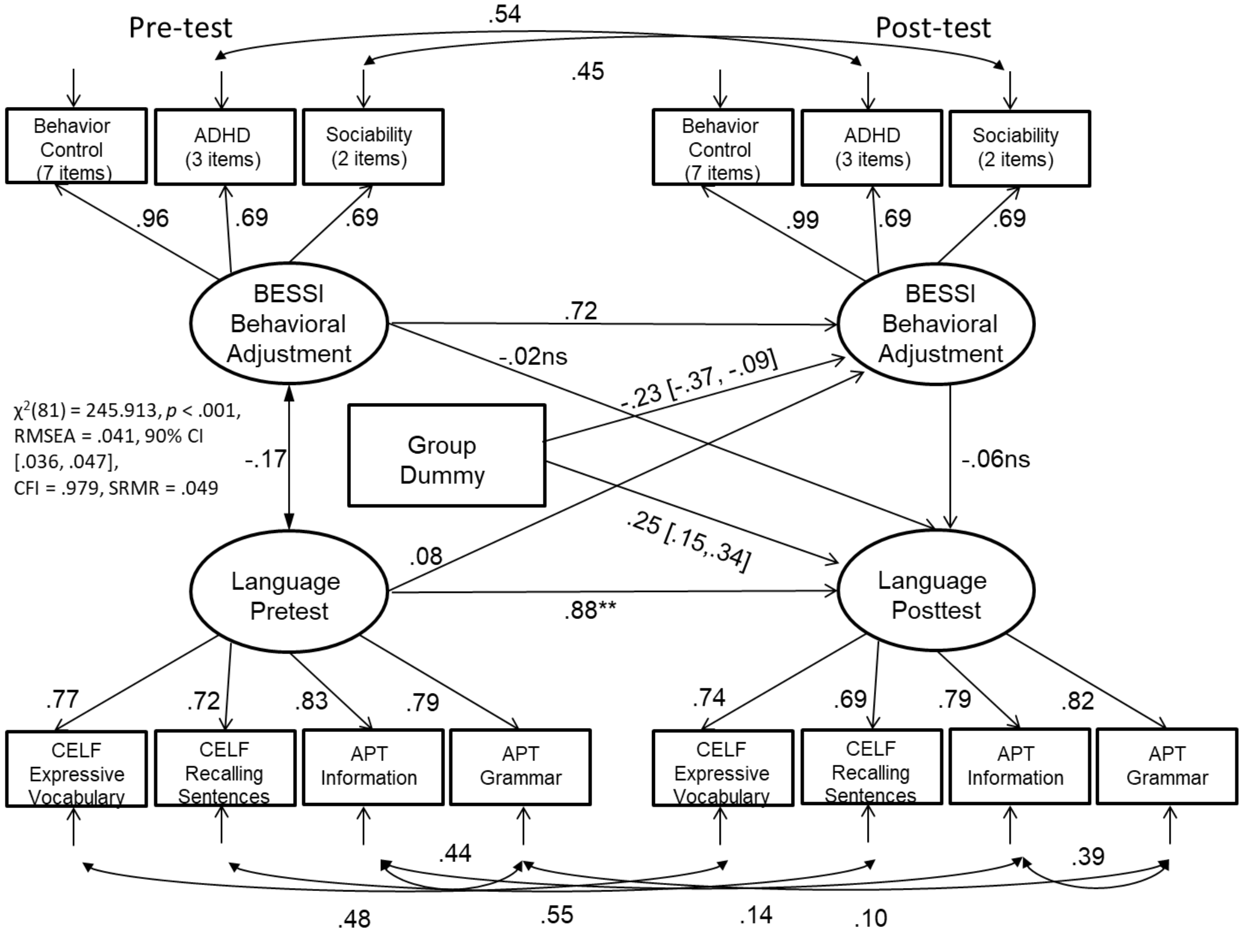
Mediation Model Examining the Indirect Effects from Language Intervention to Improvements in Ratings of Behavioral Adjustment (BESSI)



Note. All coefficients are standardized except for the path coefficient for Group, which is y-standardized (equivalent to Cohen's d). The standard errors in the model are based on bootstrapped estimates.

Figure 6

Mediation Model Examining the Indirect Effects from Rated Behavioral Adjustment to Language Outcome



Note. All coefficients are standardized except for the path coefficient for Group, which is y-standardized (equivalent to Cohen's d). The standard errors in the model are based on bootstrapped estimates.