The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry 58:10 (2017), pp 1132–1140 doi:10.1111/jcpp.12762

Improving language comprehension in preschool children with language difficulties: a cluster randomized trial

Åste M. Hagen,[1] Monica Melby-Lervåg,[2] and Arne Lervåg[1]
[1]Department of Education, University of Oslo, Oslo; [2]Department of Special Needs Education, University of Oslo, Oslo, Norway

Background: Children with language comprehension difficulties are at risk of educational and social problems, which in turn impede employment prospects in adulthood. However, few randomized trials have examined how such problems can be ameliorated during the preschool years. **Methods:** We conducted a cluster randomized trial in 148 preschool classrooms. Our intervention targeted language comprehension skills and lasted 1 year and 1 month, with five blocks of 6 weeks and intervention three times per week (about 75 min per week). Effects were assessed on a range of measures of language performance. **Results:** Immediately after the intervention, there were moderate effects on both near, intermediate and distal measures of language performance. At delayed follow-up (7 months after the intervention), these reliable effects remained for the distal measures. **Conclusions:** It is possible to intervene in classroom settings to improve the language comprehension skills of children with language difficulties. However, it appears that such interventions need to be intensive and prolonged. **Keywords:** Language difficulties; vocabulary; language comprehension; randomized trial.

Introduction

The ability to comprehend and use language effectively for communication is a fundamental part of child development. Language skills are important in and of themselves, and they create a foundation for later educational success (Foorman, Koon, Petscher, Mitchell & Truckenmiller, 2015; Fuchs, Fuchs, Compton, Hamlett & Wang, 2015) and for the ability to participate in society and work life (Heckman, 2000). Thus, difficulties in language comprehension are likely to affect many areas of learning and social interaction and, career opportunities and employment prospects in adulthood.

Recent studies show that language comprehension depends on a range of skills including vocabulary knowledge and, syntactic (i.e. structure of sentences) and morphological (i.e. form of words) skills, which develop hand in hand and are closely related (Bornstein, Hahn & Putnick, 2016; Bornstein, Hahn, Putnick & Suwalsky, 2014; Klem et al., 2015; Lervåg, Hulme & Melby-Lervåg, 2017). Between the ages of 2 and 6 years, children show both large variation and a rapid growth in language comprehension and its component skills. (Melby-Lervåg et al., 2012). The rank order in children's language comprehension skills appears to be relatively stable, suggesting that children with the poorest skills remain disadvantaged throughout their lives (e.g. Klem et al., 2015; Melby-Lervåg et al., 2012). These findings have implications for interventions for children with poor language skills: to succeed in improving language skills, a broad set of skills need to be targeted at an early age since stability tends to increase over time (Bornstein et al., 2014, 2016).

Interventions targeting language comprehension

Several meta-analyses have summarized the effects of interventions targeting language comprehension and related skills (e.g. Blok, 1999; Elleman, Lindo, Morphy & Compton, 2009; Lonigan, Shanahan & Cunningham, 2008; Marulis & Neuman, 2010; Mol, Bus & de Jong, 2009). Studies in this area take three main approaches to training language comprehension: (a) dialogic book reading (originating from Whitehurst et al.,1988) focuses on actively involving children during reading, (b) direct teaching of language comprehension skills, often with emphasis on training vocabulary, or (c) a broader approach that combines book reading, vocabulary instruction and other exercises for language comprehension-related skills. A meta-analysis examining the effects of dialogic reading showed moderate effects on expressive [d = .62 (95% CI 0.29–0.95)] and receptive language [d = .45 (95% CI 0.22–0.68)] (Mol et al., 2009). For vocabulary instruction, another meta-analysis showed moderate to strong effects for custom vocabulary measures (d = 0.79) and small to moderate effects on standardized vocabulary tests (d = .29) (Elleman et al., 2009).

However, there are two main difficulties in drawing clear conclusions from the many meta-analyses in this area (for details, see Rogde, Hagen, Melby-Lervåg & Lervåg, 2016). First, many of these

Conflict of interest statement: No conflicts declared.

meta-analyses include studies with no control group or are quasi-experiments with poor baseline controls; therefore, they tend to overestimate the effects (e.g. Elleman et al., 2009). Second, when computing a mean effect size several of the meta-analyses combine language measures that are custom-made for that study with global standardized language measures. To get effects on measures that contain the words included in the intervention is obviously easier than to get effects on distal standardized language measures (see Elleman et al., 2009). Therefore, collapsing near and distal measures can also lead to overestimates of the true effects (Mol et al., 2009).

In fact, very few randomized trials of language comprehension interventions for young children with language difficulties exist (Rogde, Hagen et al., 2016). One such study is a cluster randomized trial that examined whether it is possible to enhance vocabulary in Head Start preschoolers by teaching novel words through categorization for 12–15 min a day (Neuman, Newman & Dwyer, 2011). The intervention was organized by topic and built around working with categorizations of related constructs. This study did not find any effect on a standardized vocabulary test, the Woodcock-Johnson Picture Vocabulary, immediately after the intervention ($d = .07$, ns). There were, however, effects on three custom measures of words used in the study [*ds* ranging from .32 (95% CI 0.39–0.71) to 0.55 (95% CI 0.16–0.48)]. Another study examined children who, upon school entry, have poorly developed oral language skills (Bowyer-Crane et al., 2008). The children received either an intervention focusing on a) vocabulary and narrative skills or b) phonology and reading for 20 weeks, in a combination of small group and individual teaching sessions. The language comprehension intervention, in comparison to the phonology with reading intervention, showed improvements in taught vocabulary [$d = 1.00$ (95% CI 0.66–1.34)] and expressive grammar skills [$d = .31$ (95% CI −0.01 to 0.63)].

Finally, a third randomized trial examined the effects of an oral language intervention delivered to children with weak oral language skills in the last term of preschool (aged 4 years) and during the first two terms of reception class (the first year in primary school; pre-Year 1) (Fricke, Bowyer-Crane, Haley, Hulme & Snowling, 2013). The intervention program included vocabulary training, narrative work (e.g. creating stories), listening activities as well as training in the alphabetic principle, letter-sound knowledge and phoneme awareness. Children in the intervention group showed significantly better performance [$d = .8$ (95% CI 0.50–1.10)] on a latent language comprehension factor that included measures of vocabulary, listening comprehension and expressive language and on a latent spoken narrative factor [$d = .38$ (95% CI 0.09–0.68)]. The effects were maintained at a 6-month follow-up test.

Although these are promising findings, the knowledge about how to best help children develop oral language skills is still sparse and the studies are diverse. For example, the study by Fricke et al. (2013) included a combination of phoneme awareness and letter knowledge tasks, in addition to language comprehension tasks, while the comparison group in Bowyer-Crane et al. (2008) received phonological training. Also, the study by Neuman et al. (2011) was with children with low SES and second language learners and not children with language difficulties per se. Thus, there is clearly a need to examine effects from language comprehension interventions in children with language difficulties.

The current study

This study assesses the effects of a language comprehension intervention on children with language difficulties. Based on the evidence reviewed above, we developed an intervention that was delivered by preschool teachers and included comprehensive, systematic and explicit vocabulary instruction in combination with shared reading and other language activities. The effects of the intervention were examined with measures that were near, intermediate and distal from the tasks in the intervention. We examined the degree to which the language intervention improved the language comprehension skills of preschoolers with language difficulties and whether any effects found were durable.

Method

Participants

All children who were born in 2009 and attending preschools in two municipalities in Norway were invited to participate in the study. Notably, in Norway, there is no kindergarten year. Children attend preschool until the year they turn 6, then they start first grade. This resulted in 860 initial participants. Figure 1 shows the details of the recruitment, allocation and flow of the participants throughout the study in accordance with the CONSORT guidelines (Schulz, Altman & Moher, 2010). Ethical approval was obtained from the Norwegian Social Science Data Services, and informed parental consent was obtained for each child in the study. The children were screened with a measure consisting of 29 items from the British Picture Vocabulary Scale II (BPVS-II) (Dunn, Dunn, Whetton & Burley, 1997) and 12 items from the picture-naming subtest of the Wechsler Preschool and Primary Scale of Intelligence-III (Wechsler, 1989). The reliability of the screening measure was .67 (Cronbach's alpha). We identified the 35% of children with the lowest scores on the vocabulary screening measure ($n = 301$, 49.4% girls) for further participation in the study (mean age 57.84 months, $SD = 3.39$).

The children attended 150 classrooms in 77 different preschools. The average number of children per preschool and classroom was 3.7 and 1.9, respectively. We randomized children at the classroom level to avoid contamination between the intervention and the control groups. Figure 1 shows that a

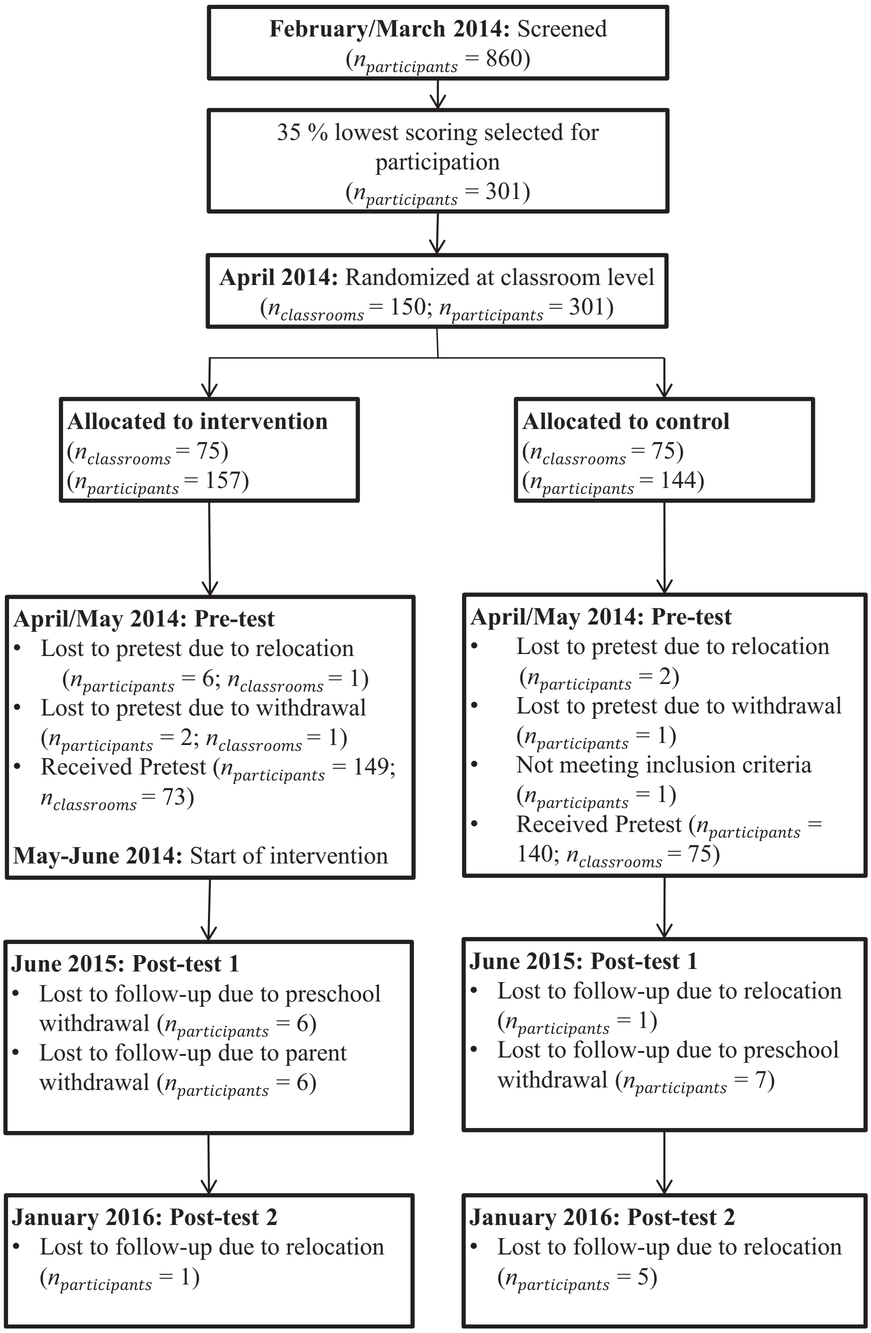

Figure 1 CONSORT diagram showing flow of participants through RCT study

total of 37 children (21 in the intervention group and 16 in the control group) were lost from the trial by the end of the study. However, Little's MCAR test on the pretests showed that the children with missing data were not significantly different from the children without missing data χ^2 (246) = 276.95, p = .085.

Measures

The measures for the study were selected and constructed based on how far they were from the taught words and material. Near measures consisted purely of taught words, intermediate measures were passages or sentences that contained the words children came across during the intervention, and distal measures were in general standardized language measures that did not contain taught words. See Tables 1 and 2 for details about the reliability of the measures (Cronbach's alphas).

Near measures.

Taught vocabulary was assessed using a task that required the children to give a definition for words that were explicitly taught in the intervention. The test included a random sample of 30 words from a list of 90 taught words. The response to each word was scored on a scale from 0 to 3 points. One point was given when the answer was a demonstration or a simple example, two points were given for a good example or explanation, and three points were given for a synonym or a full definition.

Intermediate measures.

Listening comprehension was measured using a test developed for the study and included taught vocabulary, both in stories and in questions. The first test consisted of short stories and questions associated with each one. An adult read the story to the child and then asked questions about the story. To answer the questions required a

Table 1 Means, standard deviations, and effect sizes from individual language measures containing taught vocabulary

	Mean (*SD*)				
	Intervention group	Control group	Cronbach's alpha	Cohen's *d*[a] (*Z*-value)	*p*
Listening comprehension					
Pretest	16.41 (5.37)	15.65 (6.77)	.83		
Post-test 1	29.82 (5.36)	26.84 (5.90)	.82	.468 (3.56)	<.001
Post-test 2	32.34 (5.40)	30.60 (6.07)	.85	.288 (2.63)	.008
Vocabulary definitions					
Pretest	5.46 (4.19)	5.90 (4.28)	.70		
Post-test 1	28.77 (13.94)	18.28 (9.27)	.87	.828 (9.36)	<.001
Post-test 2	32.73 (12.72)	27.82 (11.62)	.86	.426 (3.36)	.001
Morpheme generation					
Pretest	9.10 (4.40)	10.05 (5.16)	.80		
Post-test 1	16.94 (4.39)	15.58 (4.81)	.79	.413 (3.37)	.001
Post-test 2	20.48 (4.31)	20.08 (4.29)	.79	.179 (1.35)	.179

Effect size[a] = the standard deviation difference between the intervention and the business-as-usual control group, controlling for the pretest using robust clustered (Huber-White) standard errors.

Table 2 Means, standard deviations, and effect sizes from the distal individual language measures

	Mean (*SD*)				
	Intervention group	Control group	Cronbach's alpha	Cohen's *d*[a] (*Z*- value)	*p*
Listening comprehension					
Pretest	16.61 (5.89)	17.28 (6.14)	.83		
Post-test 1	26.88 (5.39)	23.65 (5.67)	.90	.614 (6.06)	<.001
Post-test 2	29.99 (4.69)	28.64 (5.57)	.79	.312 (2.81)	<.001
Narrative skills (Bus Story)					
Pretest	12.13 (6.94)	13.77 (7.12)	.77		
Post-test 1	22.47 (7.80)	19.56 (8.37)	.78	.448 (3.84)	<.001
Post-test 2	26.88 (8.51)	23.33 (9.76)	.83	.359 (2.90)	.004
Syntactic Skills (TROG II)					
Pretest	44.63 (17.14)	41.25 (15.53)	.96		
Post-test 1	64.18 (12.48)	59.12 (13.40)	.95	.304 (1.83)	.067
Post-test 2	65.32 (11.65)	63.82 (11.64)	.95	.055 (.500)	.617
Morpheme Generation (ITPA-GC)					
Pretest	11.87 (3.45)	12.15 (4.23)	.75		
Post-test 1	16.62 (4.10)	16.33 (4.10)	.75	.097 (.755)	.145
Post-test 2	19.98 (3.61)	19.50 (4.42)	.79	.166 (1.17)	.242
Vocabulary definition (WPPSI/WISC)					
Pretest	15.56 (7.00)	16.50 (5.98)	.81		
Post-test 1	25.08 (6.22)	23.88 (8.25)	.83	.194 (1.456)	0.145
Post-test 2	26.54 (5.17)	25.18 (6.12)	.77	.279 (2.31)	0.021
Vocabulary Breadth (BPVS II)					
Pretest	55.42 (11.36)	53.86 (12.95)	.90		
Post-test 1	74.61 (11.21)	72.80 (11.44)	.86	.133 (.867)	0.386
Post-test 2	84.06 (9.71)	83.42 (10.06)	.85	.031 (.216)	0.829

Effect size[a] = the standard deviation difference between the taught and the business-as-usual control group, after controlling for pretest using robust clustered (Huber-White) standard errors.

combination of recall and inference. At pretest, the test had 10 stories with three to five questions each (total number of items 36). At the second time point, another story and six more difficult questions were added to the test to avoid ceiling effect. Therefore, the number of items at both the immediate post-test (post-test 1) and follow-up post-test (post-test 2) was 42.

Morpheme generation was measured with a custom-made measure that included only taught vocabulary. This test had 30 items and the task was to complete phrases by saying the missing parts.

Distal measures. *Word definition skills* were also measured with a standardized measure using a selection of words from the vocabulary tests of the Wechsler Preschool and Primary Scale of Intelligence-III (Wechsler, 1989) and the Wechsler Intelligence Scale for Children-III (Wechsler, 2003).

Morpheme generation was assessed using the grammatic closure subtest of the Illinois Test of Psycholinguistic Abilities (ITPA) (Kirk, McCarthy & Kirk, 1968).

Receptive vocabulary was assessed with the first 144 words from a Norwegian translation of the British Picture Vocabulary Scale II (BPVS-II) (Dunn et al., 1997). Because the participants had poor vocabulary skills, each child started at the easiest level (item 1 in set 1) and was stopped after eight wrong items on two consecutive sets.

Verbal comprehension of syntax was assessed using the Test for Reception of Grammar, version 2 (TROG-2) (Bishop, 2003).

Listening comprehension was assessed using a test developed by the researchers following the same design as described for listening comprehension of intermediate measures, with the exception that it did not contain the taught words.

Narrative skills were measured using the Renfrew Bus Story Test (Renfrew, 1997). In this test, the children are told a story while looking at illustrative pictures; then, they are instructed to retell it. The children's retellings were transcribed verbatim and scores were given based on vocabulary/key words and story structure.

Procedure

The children were assessed individually on the language comprehension measures at preintervention, at immediate post-test in the end of their last year of preschool, and at a follow-up 7 months after the intervention ended. All testing was conducted by trained research assistants in the children's preschools.

Intervention programme

Children in the intervention group took part in a 30-week language programme delivered by trained preschool teachers. The 30 weeks were split into five blocks of 6-7 weeks each and were delivered with approximately 2-week breaks between each block (except for summer and Christmas holidays). The mean age at the beginning of the study was 57.84 months, *SD* = 3.39. There were a total of 90 sessions, and each of the 30 intervention weeks consisted of two 30-min group sessions (of 3-5 children) and one 15-min individual session. Teachers received training prior to the intervention and approximately halfway through the intervention. For each language programme block, a detailed scripted manual described the activities and procedures and included materials to minimize teacher preparation time.

The language programme aimed to improve aspects of the children's oral language, such as vocabulary skills, narrative skills and active listening skills. One goal was also to let children practice independent speaking. The programme was adapted from a previous randomized controlled trial with second language learners (for details, see Rogde, Hagen et al., 2016; Rogde, Melby-Lervåg & Lervåg, 2016). The first component—dialogic reading—was based on procedures described by Whitehurst et al. (1988). A number of age-appropriate short stories were designed for use in the intervention. To provide opportunities for the children to participate, the stories included engaging themes, rich language and opportunities to draw inferences. After reading a story, the teacher would ask the children questions and help them to draw inferences about the course of the story, why certain things happened, and the meanings of novel words.

The second component involved more direct instruction to develop the children's vocabulary skills, grammar skills and narrative skills. A set of 90 age-appropriate words was selected for the purpose of the study, that is, three to four words per week. The words were based on the concept of tier-2 words (Beck, McKeown & Kucan, 2002; Biemiller, 2009), which are abstract words that children will not easily learn by themselves in a general preschool setting but that are highly relevant for building more abstract language and for later school performance. Words were selected on the basis of school textbooks and age-appropriate children's books. The words were embedded in short stories, in addition to being taught directly. Activities involved various themes (e.g. travel, food, emotions and animals) and included diverse tasks, such as listening activities, exercises on knowledge of grammar, classification of words and concepts, and story structuring and sequencing. Some of these tasks were created for the intervention, while others were based on material from sources such as Black Sheep Press and Taskmaster.

The control group followed a business-as-usual regime in which the children received their ordinary preschool programme. This programme also involved reading and language activities but in a much less explicit and structured manner than our intervention programme. For a detailed description of reported language activities in control classrooms, see Hagen (2017).

Treatment fidelity

The teachers were asked to maintain logs and all sessions were audio recorded. A random selection of 5% of the sessions across all preschools and at least one session per teacher was checked; these sessions showed 100% consistency between the audio recordings and the events reported in the logs. To preserve treatment fidelity, the research group also held a joint meeting with preschool teachers about halfway through the intervention. The average number of completed sessions per child was 50.56 (*SD* = 30.78) or 56%.

Results

Because we were interested in the effects of the intervention at the individual level, we first calculated the intraclass correlation for measures across preschools. This was done to estimate the design effect in order to decide whether we needed to control for clustering at both the preschool and classroom levels. The design effect incorporates the ICCs and quantifies the effect of violating the statistical independence that is caused by clustering and that can affect standard error estimates. It is an estimate of the multiplier that needs to be applied to standard errors to correct for the negative bias that results from nested data (Peugh, 2010). Simulation studies suggest that design effects above 2 indicate a need for either the correction of standard errors or multilevel modelling (Muthén, 1991, 1994; Muthén & Satorra, 1989).

Because the design effects in the current study varied from 1 to 2.053 at the kindergarten level and from 1 to 1.25 at the classroom level, we decided to only control for dependence at the kindergarten level. Furthermore, because the average number of participants within the clusters in our study was low (3.7 per kindergarten and 1.9 per classrooms), adjusting standard errors was considered a better solution than estimating multilevel models (McNeish, Stapleton & Silverman, 2017). Thus, all subsequent analyses were done with full information maximum likelihood (FIML), using clustered robust (Huber-White) standard errors to control for dependency at the kindergarten level. We used intention to treat (ITT) analyses that included all the 289 children that received the pretest, irrespective of how many sessions the children had actually participated in.

Effects on near and intermediate measures of language

Table 1 shows means, standard deviations, reliabilities and effect sizes, with baseline controls, for the variables that contained taught words. Table 1 shows moderate to strong effects on all variables on both the immediate post-test and the post-test follow-up 7 months later (the one exception was morpheme generation at the follow-up). The results are, as expected, stronger for defining the taught words (near measures) than for tasks that only contained the taught words (intermediate measures).

To examine the effect on a common language construct that consists of measures that were intermediate from the tasks and words that were actually trained, we estimated a structural equation model (*SEM*) where the listening comprehension and morpheme generation tasks were indicators of the same latent variable. Such a measure assesses an underlying language factor that captures the common variance shared by two intermediate language measures.

As Figure 2 shows, there were clear improvements from the intervention on both the immediate and follow-up post-tests. The effect size (standardized on the outcome factor, making it equivalent to Cohen's *d*) was $d = .662$ (95% CI 0.365–0.959) at the immediate post-test and $d = .482$ (95% CI 0.168–0.796) at the 7-month follow-up. A Wald test confirmed that the effect at the immediate post-test was stronger than the effect at the 7-month delayed post-test (χ^2 (1) = 4.62, $p = .032$).

Furthermore, the analyses suggest that the intervention worked equally well for children with different levels of language skills at pretest, because there was no significant interaction between differences in children's starting levels and gains from training (Wald test: χ^2 (2) = 4.13, $p = .127$). This model (Figure 2) showed an excellent fit to the data (χ^2 (10) = 11.03, $p = .356$; RSMEA = .000 (90% CI: 0.000–0.068); CFI = .998; SRMR = .033). In this model, there was metric (invariant factor loadings) but not scalar (noninvariant intercepts) invariance across time suggesting that parts the composition of the language factor varied over time. See Appendix S1 for invariance details.

Effects on distal measures of language

Table 2 shows the means, standard deviations, reliabilities and effect sizes, with baseline controls, for the distal language tests. There was variability across the different tests, in that stronger effects were seen on the broader measures, such as listening comprehension and narrative retelling, then the narrower measures, such as grammar and vocabulary.

Our primary goal was to examine whether the intervention produced improvements on a broad language factor that did not contain taught words. We therefore estimated a *SEM* model where the language factor at each time point loaded on our six distal language measures. Figure 3 shows clear improvements from the intervention on this latent language factor $d = .563$ (95% CI 0.280–0.846) at the immediate post-test and $d = .340$ (95% CI 0.089–0.591) at the 7-month follow-up. The difference in the size of the effect between the immediate post-test and the follow-up post-test was not significant (Wald test: χ^2 (1) = 2.53, $p = .112$).

Further analyses showed no significant interaction between the training effects and baseline language level (Wald test: χ^2 (2) = 5.21, $p = .074$), so the intervention appears to have worked equally well

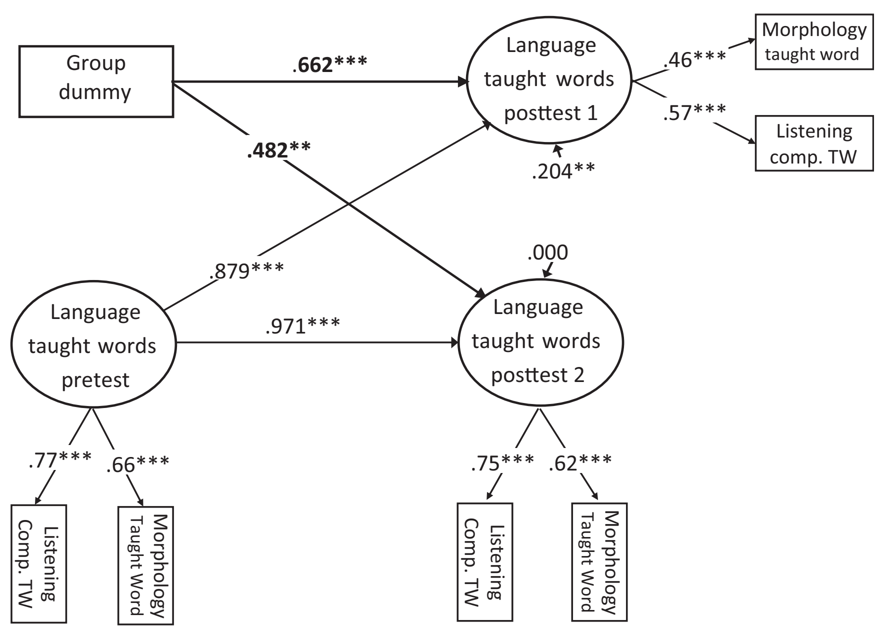

Figure 2 Model showing the effects of the intervention on intermediate language skills containing thought words at immediate post-test and delayed follow-up. Standardized coefficients are shown (except for dummy variables where *y*-standardized values are shown). A number of covariances between the same measures at adjacent time points were significant and included in the model, but not shown in the diagram

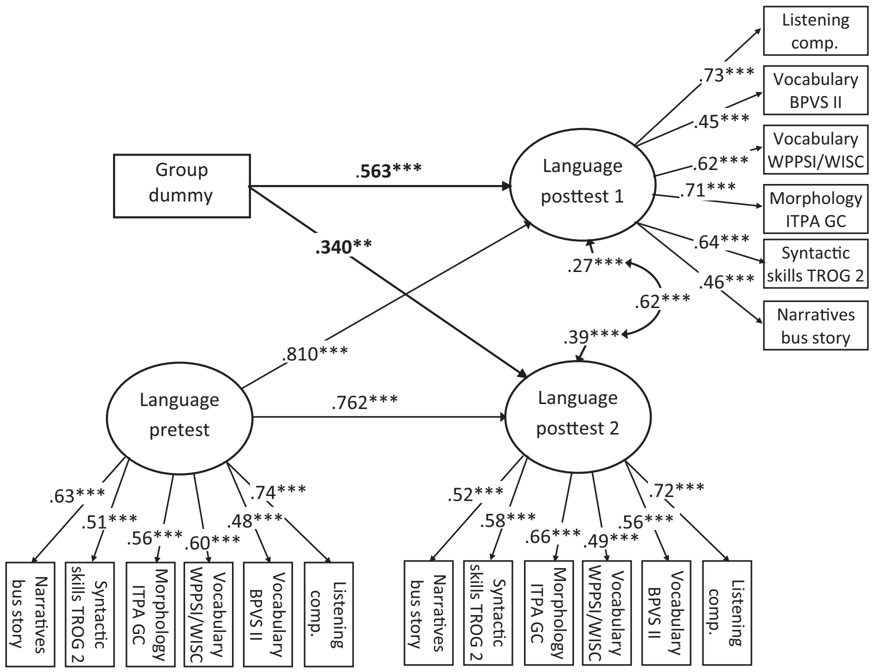

Figure 3 Model showing the effects of the intervention on distal language skills at immediate post-test and delayed follow-up. Standardized coefficients are shown (except for dummy variables where *y*-standardized values are shown). A number of covariances between the same measures at adjacent time points were significant and included in the model, but not shown in the diagram

for children with different levels of initial language skills. This is similar to the findings for the intermediate measures. This model (Figure 3) fitted the data very well: χ^2 (141) = 251.23, $p < .001$; RSMEA = .052 (90% CI: 0.041–0.062); CFI = .926; SRMR = .058. In this model, there was partial metric invariance (morpheme generation and BPVS vocabulary varied across time points) but not scalar invariance across time suggesting that parts the composition of the language factor varied over time. See Appendix S1 for invariance data.

Discussion

This trial evaluated the effects of a 30-week language programme designed to help children with language difficulties. The intervention produced clear benefits on near, intermediate and distal measures of language comprehension. While the effects on the near and intermediate constructs were reduced at the follow-up compared to the immediate post-test, the effects on the more important distal measures were maintained.

Our findings are in line with those from two earlier studies targeting a broad set of language comprehension skills (Bowyer-Crane et al., 2008; Fricke et al., 2013). However, our findings show considerably better effects than the trial with Head Start children that focuses on building taxonomies of constructs (Neuman et al., 2011). One reason for this could be that building construct taxonomies is perhaps too specific and related to too few constructs for the results to be visible on global measures of language comprehension. As for follow-up versus immediate effects, our effects also lasted at follow-up; in fact, for the distal measures, there were no significant differences between immediate and follow-up effects. This is in line with one other trial where the effects remained unchanged from immediate to follow-up testing (Fricke et al., 2013). Both of these trials show powerful effects.

Note that the children in our study were selected because they were below the 35th percentile in language skills. Thus, this is not a clinical sample, but a broader group of children that we hypothesized would benefit from the intervention due to relatively poor language skills. In addition, our screening measure had somewhat lower reliability than desired probably leading to a less precise inclusion of children around the cut-off. However, the effect of the intervention did not vary as a function of the severity of the children's language problems suggesting that the findings should be relevant to clinically referred cases as well.

In field trials in educational settings, we do not expect full compliance from every participant. In this trial, the average number of completed sessions was 51 (of a maximum of 90). Therefore, it might be that the effect sizes found here would have been even higher if we had full compliance. To get an indication of this, we compared the children who participated in 60 sessions or more with the control group on the distal measures. In these post hoc analyses, we found that the effect size increased to .909 (95% CI 0.654–1.164) for immediate effects on the distal measures and to .401 (95% CI 0.081–0.721) at follow-up. This might indicate that the effect sizes would have been stronger with full compliance. However, these post hoc analyses ignore the randomization, so selection bias might be present and the findings should be interpreted with caution.

Even if our latent variable models give an estimate of the size of change in language skills produced by our intervention, the invariance tests suggest differential rates of improvement across measures. Thus, we cannot make strong claims about the intervention had effects on a same unitary underlying language factor across time. This is a result shared with other similar studies (e.g. Fricke et al., 2013).

Still, our study has important theoretical and applied implications. Theoretically, although children's early language skills show a high degree of longitudinal stability (Klem et al., 2015) this should not be interpreted to mean that the language skills of those with language learning difficulties cannot be improved by intervention (cf. studies showing equivalent effects for IQ, e.g. Duyme, Dumaret & Tomkiewicz, 1999). In relation to clinical practice, our results add to a growing body of evidence that language intervention in the preschool/early school years can produce substantial and lasting improvements in children's language skills (see also Bowyer-Crane et al., 2008; Fricke et al., 2013). A major aim for future studies should be to conduct longer term follow-up assessments of such effects and, ideally, to conduct studies in which language interventions are delivered over longer periods of time. Current evidence suggests that such interventions could have considerable educational benefits and be highly cost effective.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Supplemental information about the measurement invariance tests done for Figures 2 and 3.

Table S1. CONSORT 2010 checklist.

Acknowledgements

The research was funded by two grants from the Research Council of Norway's UTD2020/FINNUT programmes: Grant: 203335 and Grant: 237724. The authors thank many research assistants with data collection, and the preschool teachers, preschools, municipalities, children and parents for their participation. The authors have declared that they have no competing or potential conflicts of interest in relation to this article.

Correspondence

Arne Lervåg, Department of Education, University of Oslo, P.O. Box 1092 Blindern, NO-0317 Oslo, Norway; Email: a.o.lervag@iped.uio.no

Key points

- We know that children with language comprehension difficulties risk low performance in school, problems with social interaction, and that this in turn will also affect their ability to find employment and to participate in society.
- Few randomized field trials have examined how such problems can be ameliorated.
- In this teacher-led language intervention for preschool children with language difficulties, we find effects both on immediate and 7-month follow-up measures of language function.
- This intervention shows that it is possible to improve language comprehension in children with language difficulties in a classroom setting.

References

Beck, I.L., McKeown, M.G., & Kucan, L. (2002). *Bringing words to life*. New York: Guilford Press.

Biemiller, A. (2009). *Words worth teaching*. Columbus, OH: SRA/McGraw-Hill.

Bishop, D.V.M. (2003). *Test for reception of grammar. Trog-2 manual (Version 2)*. London, UK: Harcourt Assessment.

Blok, H. (1999). Reading to young children in educational settings. *Language Learning*, *49*, 343–371.

Bornstein, M.H., Hahn, C.S., & Putnick, D.L. (2016). Stability of core language skill across the first decade of life in children at biological and social risk. *Journal of Child Psychology and Psychiatry*, *57*, 1434–1443.

Bornstein, M.H., Hahn, C.S., Putnick, D.L., & Suwalsky, J.T. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development*, *85*, 1346–1356.

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.

Dunn, L.M., Dunn, L.M., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale II*. Windsor, UK: National Foundation for Educational Research-Nelson.

Duyme, M., Dumaret, A.C., & Tomkiewicz, S. (1999). How can we boost IQs of "dull children"?: A late adoption study. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 8790–8794.

Elleman, A.M., Lindo, E.J., Morphy, P., & Compton, D.L. (2009). The impact of vocabulary instruction on passagelevel comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*, 1–44.

Foorman, B.R., Koon, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). Examining general and specific factors in the dimensionality of oral language and reading in 4th–10th grades. *Journal of Educational Psychology*, *107*, 884–897.

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.

Fuchs, L.S., Fuchs, D., Compton, D.L., Hamlett, C.L., & Wang, A.Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading*, *19*, 204–223.

Hagen, Å.M. (2017). Improving the odds: Identifying language activities that support the language development of preschoolers with poorer vocabulary skills. *Scandinavian Journal of Educational Research*. Advanced online publication. https://doi.org/10.1080/00313831.2016.1258727.

Heckman, J.J. (2000). Policies to foster human capital. *Research in Economics*, *54*, 3–56.

Kirk, S.A., McCarthy, J.J., & Kirk, W.D. (1968). *Illinois test of psycholinguistic abilities*. Urbana, IL: University of Illinois Press.

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.

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2017). Unpicking the developmental relationships between oral language skills and reading comprehension: Its simple, but complex. *Child Development*. Advanced online publication. https://doi.org/10.1111/cdev.12861.

Lonigan, C.J., Shanahan, T., & Cunningham, A. (2008). *Impact of shared-reading interventions on young children's early literacy skills*. Developing Early Literacy: Report of the National Early Literacy Panel, 153–171.

Marulis, L.M., & Neuman, S.B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *30*, 300–335.

McNeish, D., Stapleton, L.M., & Silverman, R.D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*, 114–140.

Melby-Lervåg, M., Lervåg, A., Lyster, S., Klem, M., Hagtvedt, B., & Hulme, C. (2012). Nonword repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science*, *23*, 1069–1098.

Mol, S.E., Bus, A.G., & de Jong, M.T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*, 979–1007.

Muthén, B.O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*, 338–354.

Muthén, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376–398.

Muthén, B.O., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R.D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 87–99). San Diego: Academic Press.

Neuman, S.B., Newman, E.H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly*, *46*, 249–272.

Peugh, J.L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*, 85–112.

Renfrew, C.E. (1997). *Bus story test—A test of narrative speech*, 4th ed.. Bicester, UK: Winslow Press.

Rogde, K., Hagen, Å., Melby-Lervåg, M., & Lervåg, A. (2016). *The effect of linguistic comprehension training on language and reading comprehension: A systematic review*. Protocol Campbell Library. Retrieved from https://www.campbellcollaboration.org/media/k2/attachments/Rogde_Linguistic_Comprehension_Training_Protocol.pdf

Rogde, K., Melby-Lervåg, M., & Lervåg, A. (2016). Improving the General Language Skills of Second-Language Learners in Kindergarten: A Randomized Controlled Trial. *Journal of Research on Educational Effectiveness*, *9*, 150–170.

Schulz, K.F., Altman, D.G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, *8*, 1.

Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence (Revised)*. San Antonio: The Psychological Corporation: Harcourt Assessment.

Wechsler, D. (2003). *Administration manual (Norwegian version)*. San Antonio, TX and Assessio, Norway: The Psychological Corporation.

Whitehurst, G.J., Falco, F., Lonigan, C.J., Fischel, J.E., DeBaryshe, B.D., Valdez-Menchaca, M.C., & Caulfield, M. (1988). Accelerating language development through picture-book reading. *Developmental Psychology*, *24*, 552–555.

Accepted for publication: 9 May 2017
First published online: 3 July 2017