Disentangling the Far Transfer of Language Comprehension Gains Using Latent

Mediation Models

Authors

Corresponding author:

Monica Melby-Lervåg

Department of Special needs education

University of Oslo

Pb 1140 Blindern

0318 OSLO

Email: monica.melby-lervag@isp.uio.no

Phone 0047 22855050

Åste Mjelve Hagen

Department of Special needs education

University of Oslo

Email: ammhage@isp.uio.no

Arne Lervåg

Department of education

University of Oslo

Email: a.o.lervag@iped.uio.no

Research Highlights

- This study reveals interesting knowledge about potential drivers behind gains in standardized general language measures from language comprehension interventions.

- Both the effects and transfer effects are mainly generated through expressive rather than receptive language measures.

- The effects on far transfer measures are mediated through gains on the specific words that are trained in the intervention.

- In order to optimize intervention effects, future studies should focus on expressive language.

Abstract

While we know that interventions targeting oral language can be effective, little is known about what drives these effects. In this study, we examine whether gains in transfer measures are mediated through the specific words that are trained in a language intervention. Based on a large-scale randomized controlled trial of language intervention in four- to five-year-old children, latent mediation models were used to disentangle oral language gains in transfer measures. The results first showed that the effects of the language intervention and the transfer effects are generated through expressive rather than receptive measures of language. Second, we found that the effects of the intervention on intermediate transfer measures of language were mediated through the ability to define the trained words. Third, and critically, for far transfer measures that did not contain any of the trained words, the effects were mediated through the trained words. The findings relate to theories of transfer and support the idea that far transfer is possible, at least within the same domain. In addition, it seems that effects on receptive language skills are difficult to obtain and that what is improved is instead the children's ability to express themselves and use procedures to explain words. Thus, in order to optimize intervention effects, future studies should focus on expressive language.

Keywords: Language, intervention, transfer, comprehension, vocabulary

Disentangling the Far Transfer of Oral Language Gains Using Latent Mediation Models

Introduction

In today's digital society, language and reading abilities are critical life skills. Due to the foundational nature of language comprehension for human development, it is the main component in crystallized intelligence and is commonly measured with tests related to vocabulary, such as picture vocabulary, word definitions, or verbal fluency (Wechsler, 2003). Well-developed language are also critical for reading comprehension, and in older children, language comprehension and reading comprehension are close to fully overlapping constructs (Foorman, Koon, Petscher, Mitchell, & Truckenmiller, 2015). Despite the importance of language comprehension, cognitive training studies have merely focused on training working memory and executive functions, and transfer to fluid intelligence (e.g., see Karbach & Verhaegen, 2014; Melby-Lervåg, Redick & Hulme, 2016). Here, we focus on training oral language and language comprehension, as well as the relationship between the effect of training on measures related to the specific words that are trained and the general measures of language comprehension. No studies have examined the mechanisms behind these effects using latent mediation models to determine whether gains in specific trained words mediate gains in more general language measures and, if so, what can explain this.

Theoretical Accounts to Explain Cognitive Transfer

The extent to which we can generalize knowledge to a broader range of skills is a topic that has held the interest of psychologists for decades. A main reason for this is that the ability to generalize knowledge and improve cognitive skills in children at risk can have important individual repercussions and facilitate better school and life outcomes, in turn even fostering the economy and employment rates (Doyle, Harmon, Heckman, & Tremblay, 2009; Heckman, 2000).

The generalizability of knowledge is closely linked to the transfer of learning, which occurs when one set of skills can transfer to another set of skills, either within the same domain or even between domains (Barnett & Ceci, 2002). It is common to view transfer as a continuum, from near transfer, via intermediate transfer, to far transfer (Barnett & Ceci, 2007; Lee, 1989). *Near transfer* is transfer between skills that are highly related. One example of this is that working memory training seems to result in improvement on working memory tests that closely resemble the tasks that are trained (Melby-Lervåg, et al., 2016). *Far transfer* is transfer between domains that are rather different—for example, from playing chess to problem solving in general. Naturally, *intermediate transfer* falls somewhere in between the two.

Throughout the years, there have been shifting views concerning the conditions under which far or intermediate transfer is possible, and what can explain and promote it. In a seminal article, Thorndike and Woodworth (1901) stated a rather pessimistic view about far transfer. In their common elements theory, they concluded that transfer typically happens within one domain between knowledge that share common elements but that far transfer is rare. A more contemporary adaptation of this theory suggests that a production rule is the critical element when attempting to produce transfer (Anderson, 2007; Anderson, Reder, & Simon, 1997; Singley & Anderson, 1989). *Production rules* are representations of knowledge that specify how to solve tasks that demands several actions or steps. These production rules are thought to coordinate the exchange of information between specialized cognitive systems. Typically, production rules are specific to a particular task; it is not often that you can use task-specific knowledge developed in one task to solve another. For instance, even if you know the production rules that represent the word through rote learning, this strategy will not help you much if you are going to construct sentences. Therefore, Singley and Anderson (1989), in line with Thorndike and Woodworth (1901), claim

that far transfer is quite rare. The assertion that far transfer is difficult to obtain has support across several different domains in which a lack of, or limited, transfer from one domain to another has been demonstrated, such as chess (Sala & Gobet, 2016), working memory training (Melby-Lervåg, et al., 2016), music (Sala, & Gobet, 2017), and video games (Sala, Tatlidil, & Gobet, 2017).

In contrast, the more recently developed primitive elements theory provides more optimistic prospects for transfer, including far transfer (Taatgen, 2013; Taatgen, 2016). According to this theory, if "training on a particular task evolves the available set of operators toward that task, and if those operators and their partial predecessors are also useful for a new task, there will be transfer" (Taatgen, 2013, p. 2). The difference between the primitive elements theory and Singley and Anderson's (1989) production rule theory is not entirely clear. However, Taatgen (2016) argues that the primitive elements theory is concerned with smaller elements of skills than those represented in a production rule. While a production rule typically performs several functions simultaneously, the primitive elements theory divides these into even more basic information processing units (Taatgen, 2016). Taatgen (2013) exemplifies this with memory: in a working memory task, copying the item to be remembered visually is considered the first of many primitive information processing units (PRIMS); in a production rule, the task is not broken down into so many units. Thus, according to the primitive elements theory, transfer is possible between tasks that are quite different in content but that share the same basic underlying structure. This model has performed well in different studies using computational modelling to simulate a variety of tasks, and it   accounts for far transfer (Taatgen, 2013; Taatgen 2016).

Sternberg and Frensch (1993) take a third route to explain the mechanisms behind transfer, building on theories about memory recall and learning. According to their view, the degree of transfer is dependent on four mechanisms. The first is *encoding specificity*: the transfer of learning

is more likely to happen if something is taught in a way that makes it encoded into memory not only in one specific way. Second, whether transfer and retrieval will occur depends on how information is *organized* in memory. An example of this is that it is easier to remember words organized in a fixed list than if they are presented in a random order in each exposure. Third, *discrimination*—meaning whether the techniques or strategies used in one situation is applicable to another—is also important for transfer; for example, problem-solving strategies used to comprehend expository texts might not be effective for comprehending narrative texts. Finally, *setting,* or the way in which one sees how a task or situation might carry over to other situations, also matters. This is supported by a study in which participants were either primed to view two situations as one interconnected problem-solving task or told that the two tasks were unrelated. Participants who perceived the tasks as interconnected showed excellent transfer. In contrast, participants who saw the tasks as unrelated showed no sign of transfer (Sternberg & Frensch, 1993). Thus, Sternberg and Frensch's theory is based on more general principles from classic learning and memory theories than are the theories of Taatgen (2013) or Singley and Anderson (1989).

As for language learning more specifically, one theory that has attempted to explain the transfer of language skills is vector semantics (Jurafsky & Martin, 2014). According to this theory, the chances of learning a new word increases if the meaning or grammatical features of the word are related to those of other words. In vector semantics, the probability of learning a new word can be computed from the distribution of words with a similar meaning or surrounding syntax. Thus, words are represented as a vector or array of numbers related in some way to counts (Jurafsky & Martin, 2014). Therefore, we can calculate the probability that a new word will be learned based on the relationship between the new word and the already known words. Transfer can occur when

a new word is related to similar words that are already known. In addition, learning new words improves the learner's understanding of the words they already know because the meaning or syntax of the words are related

Finally, it has been argued that transfer needs to be considered in a broader way than any of the theories discussed above (i.e., the common elements/production rule theory [Thorndyke & Woodworth, 1901; Andersson 2007], the primitive elements theory [Taatgen, 2013,2016], the theory of transfer mechanisms [Sternberg and Frensch's, 1993] or vector semantics [Jurafsky & Martin, 2016]). Critics argue that transfer has traditionally been studied on a micro level in expert performers, which blinds us from actually identifying transfer (Bransford & Schwartz, 1999; Detterman, 1993). In line with this argument, studying transfer by training one skill and testing whether this skill is directly applicable to another skill is too restrictive (Bransford & Schwartz, 1999; Detterman, 1993; Lee, 1989). Thus, it is suggested that transfer occurs on a more general level, related to meta-cognition, verbal reasoning, and critical thinking (Bransford & Schwartz, 1999; Halpern, 1988). By studying transfer in a broader way, it is argued that we can detect transfer that is hidden from us when using the traditional paradigm to determine whether procedures in one task are directly applicable to another. In support of this, it is worth mentioning that a number of studies have failed to find far transfer from different specific tasks to IQ tests, even though it is clear from a number of studies that schooling increases intelligence (Ritchie & Tucker-Drob, 2017).

Cognitive Transfer in Oral language

A recent review of the effects of language comprehension interventions shows two important findings (Rogde et al., in press). First, although overall there are indications that effects from language interventions generalize to expressive but not receptive language, the results in the

primary studies are highly variable on this point. Some studies show effects only on expressive language (Rogde, Melby-Lervåg & Lervåg, 2016; Farver, Lonigan & Eppe, 2009), some only on receptive (Gonzalez et al. 2010) and some on both (Valdez-Menecha & Whitehurst 1992). However, several of these studies are not randomized, and this could potentially create bias. Thus, it is first important to clarify to what extent generalize across different measures of language and this is the focus in the first research question.

Secondly, the review by Rogde et al., (in press) found that studies typically show large gains in the specific words that are trained in the intervention (mean d = 1.17, Rogde, Hagen, Melby-Lervåg & Lervåg, in press). In contrast, the gains in measures that do not contain the trained words are smaller but still reliable (mean d = 0.16). The studies in the review also showed large heterogeneity in the size of the effects on standardized measures. However, none of the included language intervention studies examined whether there was a relationship between gains in the trained words and gains in more general language tests that do not contain the trained words. There were also few studies examining the theoretical underpinnings of why there may be a relationship between these gains and if so what can cause it. Thus, whether gains on intermediate and far transfer measures mediated by gains in the trained words is the focus in the second research question.

The theories discussed previously (i.e. the common elements theory/production rules [Thorndyke & Woodworth, 1901; Anderson, 2007], the primitive elements theory [Taatgen, 2013], the theory of transfer mechanisms [Sternberg and Frensch's, 1993], vector semantics [Jurafsky & Martin, 2016], and theory of broader transfer [Bransford & Schwartz, 1999; Detterman, 1993]) were mainly developed to explain cognitive skills, such as memory, nonverbal reasoning, and problem solving, rather than language skills. If applying these theories to language skills, there are

several reasons for why we can expect training generalize to different types of language measures, but perhaps mainly expressive language (research question 1) and from trained words to general language measures (research question 2).

First, in line with the theory of common elements (Thorndike & Woodsworth, 1901) transfer and generalization may be enhanced because it happens within one domain (i.e., language). Also in line with the production rules of Anderson (2007), if the tasks are within the same domain this increases the likelihood that it can be established production rules that specify how to solve tasks that demands several actions or steps. Further, based on the primitive elements theory (Taatgen, 2013), transfer can be predicted because the intervention may improve not only children's ability to explain the specific words that are explicitly trained but also their ability to explain words in general. This prediction can be deduced based on the primitive elements theory; transfer occurs when the set of procedures learned with the trained words can also be used for untrained words. If transfer occurs because children get better at explaining words, this may be because the set of procedures that they learn with the trained words also can be used for untrained words (Taatgen, 2013).

That transfer will occur because children get better at explaining words in general can also be related to Sternberg and Frensch's (1993) account of transfer. The intervention aimed to promote a more general encoding of information so that information can be stored in different ways. This, in itself, will promote transfer and the generalizability of knowledge, according to Sternberg and Frensch (1993). For example, to achieve this in the intervention of the present study, auditory information was supported with pictures and picture book stories. The language intervention also focused on *discrimination*, meaning teaching children that techniques or strategies used in one situation are applicable to other tasks or settings. This also enhances transfer

according to Sternberg and Frensch (1993). For example, not only were the definitions of words presented but also how the words could be used in different settings and texts. In line with Sternberg and Frensch (1993), this focus on general encoding and discrimination may have enhanced transfer.

Importantly, in line with the primitive elements theory (Taatgen, 2013), it can also be deduced that if transfer is related to enhancing the procedural knowledge of explaining the meaning of a word we might expect that the effects and transfer effects will be strongest in the expressive measures. This could be because the receptive measures do not require procedures in the same way as the expressive measures do, since receptive measures often concerns pointing at pictures. In this case, effects on expressive, rather than receptive, language will give support to the primitive elements theory (Taatgen, 2013), because transfer happens when explanatory procedures learned with trained words can also be applied to untrained words. As such, the transfer then can be due to that the explanatory procedures learned with the trained words can also be used for untrained words. In contrast, if the effects and transfer effects are stronger in receptive measures that do not require a verbal response, this would give support to the vector semantics theory. This could be because the training of specific words has given an increased understanding of other words, as the words are related in a network, resembling each other in meaning and/or grammar.

The second research question concerns whether gains in intermediate and far transfer measures are mediated by gains in the trained words, i.e., if is it likely that transfer happens at all between the trained words and the untrained words. One reason for why transfer between trained and untrained words can occur be generated based on vector semantics (Jurafsky & Martin, 2016). In line with this theory, transfer will occur because learning new words gives children an improved understanding of the words they already know. In line with predictions related to vector semantics

(Jurafsky & Martin, 2014), if a new word is similar either in syntax or semantics to a word that a child already knows, this increases the probability that the child will learn the new word. For instance, if the child knows the word "damp," it will be easier to learn the word "moist" because the two have a related meaning. Similarly, learning the word "moist" might also give the child a more nuanced understanding of "damp."

Importantly, it can also be deduced from the theory of broader transfer (Bransford & Schwartz, 1999; Detterman, 1993) that there is no relationship between effects on trained words and on general language tests. In line with theories by Bransford and Schwartz (1999) and Detterman (1993), it is not possible to detect transfer by training one skill and narrowly testing whether it is directly applicable to another skill. This way of evaluating transfer is too restrictive because transfer occurs on a more general level and affects broader skills such as critical thinking and meta-cognition (Bransford & Schwartz, 1999; Detterman, 1993; Lee, 1989).

The Current Study

In the current study, we examine the two research questions outlined above based on data from a cluster randomized controlled trial of four- to five-year-old children with weaker vocabulary skills who received an oral language intervention (for more details about the study, see the "Method" section, below). A prior publication based on the dataset used in the current study showed that, immediately after the intervention, there were moderate effects on near ($d = 0.82$), intermediate ($d = 0.66$), and distal ($d = 0.56$) language measures, with latent variables of language performance (Hagen, Melby-Lervåg & Lervåg, 2017). However, these analyses of overall effects do not explain the mechanisms that drive the effects and the extent to which effects on near, intermediate, and far transfer are related. In the current study, we therefore analyze these questions in detail. The current study examines the following research questions:

1) Do the effects of training generalize to different types of language measures? Notably, in this study, the language measures broadly belong to two different categories: a) expressive language measures that require a verbal response (for instance, requiring that the children explain a word, retell a story, or verbally construct a sentence), or b) receptive language measures that require little or no verbal response (for instance, requiring that the children point at the correct picture when presented with a word or sentence). This is summarized in Table 3.

2) Are gains in intermediate and far transfer measures mediated by gains in the trained words, (i.e., is it likely that transfer happens at all between the trained words and the untrained words)? The measures for which this mediation was significant are shown in the models in Figures 2a, 2b, and 2c.

Method

Participants

The participants ($n = 301$; 49.4% females) in this study were a sub-sample of a cohort of children from preschools in two municipalities in Norway. For the flow of participants through the study, see Figure 1. The participants were at or below the 35th percentile on a vocabulary screening tool. The screening tool consisted 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 (WPPSI-III; Wechsler, 1989). The reliability of the screening tool was 0.67.

The children attended 77 different preschools in a total of 150 classrooms. The average number of children per preschool and classroom were 3.7 and 1.9, respectively. The children's

average age at the beginning of the study was 57.84 months ($SD = 3.39$). To avoid contamination between the intervention and control groups, children were randomized at the classroom level. A 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. We performed Little's missing completely at random (MCAR) test on the pretests, which indicated that the children with missing data were not significantly different from the children without missing data ($\chi^2[246] = 276.95$; $p = 0.085$).

Measures

The measures for the study were categorized according to how close the tasks in the measure were to the taught words and material. One measure consisted purely of taught words (i.e., the near-transfer measure). This measure simply reflects how able the children were to learn definitions of the words that were explicitly trained in the intervention. The two intermediate measures included passages or sentences that contained taught words but did not test the understanding of the words directly. The distal measures included six general language measures, mostly standardized tests, which did not include taught vocabulary.

Taught vocabulary (near-transfer measure).

Taught vocabulary was measured using a word definition task in which the children were to give definitions for taught words. The test included 30 words randomly selected from the 90 taught words. The response to each word was scored on a scale from 0 to 3 points. Zero points were given for wrong or missing answers; 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 language measures.

Morpheme generation was measured using a task created for this study that included only taught vocabulary. This test had 30 items. The children were instructed to complete phrases by saying the missing parts of the phrases.

Listening comprehension was measured using a custom-made test that included taught vocabulary. The adult read a story to the child and then asked questions about the story. Questions were a combination of recall and inference and required the child to retell key information (recall) or produce new sentences (infer) to answer. The test was therefore considered to measure expressive language skills. At the pretest, the test had 10 stories, with three to five questions each (i.e., a total of 36 items). At a second time point (i.e., after the pretest), another story and six more difficult questions were added to the test to avoid the ceiling effect. Therefore, the total number of items included in the post-test was 42.

Distal language measures.

Word definition skills were tested with a standardized measure using a selection of words from the vocabulary tests of the WPPSI-III (Wechsler, 1989) and the Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 2003).

Receptive vocabulary was assessed using the first 144 words from a Norwegian translation of the BPVS-II (Dunn et al., 1997). Because the children were expected to have poor vocabulary skills, all children began at the easiest level (item 1 in set 1) and were stopped after answering eight items incorrectly on two consecutive sets.

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

Morpheme generation was assessed using the grammatic closure subtest of the Illinois Test of Psycholinguistic Abilities (ITPA; Kirk, McCarthy, & Kirk, 1968). The ITPA has been adapted to Norwegian conditions and was translated into Norwegian by Gjessing and Nygaard (1975).

Verbal comprehension of syntax was assessed using the Test for Reception of Grammar, version 2 (TROG-II; Bishop, 2003), in which children listen to a sentence and must then point out the picture, amongst four pictures, that best represents the sentence.

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

Procedure

The children were assessed individually on the language measures at pre-intervention and at immediate posttest, at the end of their last year of preschool. Trained research assistants conducted all testing in the children's preschools. The testing lasted a maximum of 45 minutes per session, and the children had breaks when necessary.

Intervention program.

Children in the intervention group took part in a language program delivered by trained preschool teachers. The program was spread over one year and one month. It began at the end of the student's second-last year of preschool and lasted throughout their last year of preschool. The program lasted for 30 weeks, split into five blocks of six to seven weeks each. The blocks were delivered with approximately two-week breaks between (except for summer and Christmas holidays, when it was longer). There was a total of 90 sessions, and each of the 30 intervention weeks consisted of two 30 min group sessions (of three to five children) and one 15 min individual

session. Teachers received training prior to the intervention and approximately halfway through the intervention. For each language program block, teachers were given materials and a detailed scripted manual describing the activities and procedures to minimize the preparation time.

The language program aimed to improve aspects of the children's oral language, such as vocabulary skills, narrative skills, and active listening skills. The program was an adaption and extension of a previous randomized controlled trial with second-language learners (for details, see Rogde et al., 2016). The first component—dialogic reading—was based on procedures described by Whitehurst et al. (1988). A number of short stories were designed for use in the intervention. To encourage children to participate actively, the stories included a variety of themes, rich language, and opportunities to discuss and draw inferences. The teacher asked questions to help the children to draw inferences about the course of the story, determine why certain things happened, predict the plot line, and infer the meanings of novel words.

The second component involved more direct instruction to develop children's vocabulary skills, grammar skills, and narrative skills. A set of 90 words was selected for the purpose of the study. The words were based on the concept of tier-two words (Beck & McKeown, 2004; Biemiller, 2009), which are abstract words that children will not easily learn by themselves in a general preschool setting but which are important 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 the short stories, in addition to being taught directly. Activities involved a variety of age-appropriate themes (e.g., travel, food, emotions, and animals) and were diverse, including listening activities, grammar exercises, the classification of words, story structuring, and plot sequencing.

The control group children followed their ordinary preschool program (i.e., business as usual). This program also involved reading and language activities but in a substantially less explicit and structured manner than did our intervention program. There was also a large variation in reading and language activities between the control classrooms. A detailed description of the reported language activities in control classrooms is presented in Hagen (2018).

Treatment fidelity.

The teachers maintained logs and audio-recorded each language session. The logs and the audio files were collected between each of the teaching blocks, at the same time as the materials and instructions for the next block were given to the teachers. This allowed the research team to keep in touch with each teacher during the intervention and provided the teachers with extra opportunities to ask questions and discuss challenges. A random selection of 5% of the sessions across all preschools were listened to as a fidelity check. These showed 100% consistency between the audio recordings and the events reported in the logs. To preserve treatment fidelity, the research team also held a joint meeting with preschool teachers to answer questions and to repeat information about implementation about halfway through the intervention. The average number of completed sessions per child was 50.56 ($SD = 30.78$).

Results

All analyses were done in Mplus (Muthén & Muthén, (1998–2017) using full information maximum likelihood (FIML), with 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.

Descriptive Results

Table 1 presents 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 at the immediate posttest. Not surprisingly, the effect sizes are higher for defining the taught words (i.e., near measures) than they are for tasks that only contained the taught words (i.e., intermediate measures). Table 2 shows the means, standard deviations, reliabilities, and effect sizes, with baseline controls, for the distal language tests. Here, the effects clearly vary between the different measures, with stronger effects on some measures—such as listening comprehension and narrative retelling—than on others—such as grammar and vocabulary. Further, as shown in tables 1 and 2, the reliabilities (using Cronbach's alpha) were satisfactory for all measures.

Effects of the intervention on expressive, minimal expressive, and receptive language outcomes.

To test our first hypothesis about transfer being more likely to be found among expressive, in contrast to receptive, measures, we analyzed the effect of the intervention on near (defining trained words), intermediate (language tasks containing trained words), and distal (language tasks not containing trained words) language outcomes at three different levels of expressiveness: expressive, minimal expressive, and receptive (see Table 3). *Minimal expressiveness* is a morphology test that requires some degree of expressiveness but to a lesser degree than the pure expressive measure. Distal expressive language was measured by a latent variable using vocabulary (WPPSI/WISC), listening comprehension, and narrative retelling (Renfrew Bus Story Test) as indicators, and distal receptive language was measured using the BPVS-II and TROG-II as indicators.

Confirmatory factor analyses (CFAs) estimating both pretest and posttest scores showed scalar invariance for both the distal expressive (χ^2[4] = 6.809; p = 0.146) and the distal receptive (χ^2[2] = 2.895; p = 0.235) factors across time (using Wald tests). In these models, the residuals of the same observed variables were correlated across time, and the model fit was very good in both the distal expressive version, (χ^2[9] = 14.057; p = 0.120; root mean square error of approximation [RMSEA] = 0.044; 90% confidence interval [CI] 0.00–0.086; comparative fit index [CFI] = 0.987; SRMS = 0.039) and the distal receptive version (χ^2[2] = 4.339; p = 0.114; RMSEA = 0.064; 90% CI 0.000–0.147; CFI = 0.978; SRMS = 0.046).

As all the other outcomes were estimated by single variables, we fixed the residuals of these variables to reflect their alpha score in order to avoid misleading consequences due to uncorrected measurement errors (see Cole & Preacher, 2014). In all these analyses, we also controlled for pretest scores. As can be seen from Table 3, the intervention had effects on the near language and on both the intermediate language outcomes. However, among the distal measures that did not contain trained words, the intervention only affected the expressive language; no effects were observed on the minimal expressive or receptive measures that did not contain trained words. Notably, the listening comprehension task is considered a distal expressive language measure since it concerns words not included in the training and also requires expressive abilities.

Mediation models: Are effects on transfer measures mediated through near measures?

To test our second hypothesis, we estimated latent mediation models to determine whether the effects of the intervention on intermediate and distal language were mediated through the ability to define the trained words. To do this, we used the analysis of covariance (ANCOVA) approach, which was recommended as the most efficient in the simulation study by Valente and

MacKinnon (2017). Here, we estimated the indirect effect of the intervention while controlling the mediator (i.e., defining trained words at the posttest) and the language outcomes (i.e., intermediate expressive, intermediate minimal expressive, and distal expressive) for both defining trained words and the corresponding language construct at the pretest. The three models can be seen in Figures 2a, 2b, and 2c, which show intermediate expressive, intermediate minimal expressive, and distal expressive outcomes, respectively. All parameters are fully standardized, with the exception of the paths from the intervention dummy variable to the language outcomes at posttest, which were standardized only on the Y variable.

As the intervention variable is a dummy variable indicating participation in either the control (0) or intervention (1) group, the path from the intervention variable can be interpreted as the effect of the intervention in Cohen's d. As can be seen from these models, the intervention had an effect on the trained words at the posttest, and trained words at the posttest predicted the language outcomes at the posttest in all three models. To estimate the mediation—which is the product of the path from the intervention to the mediator (i.e., defining the trained words at the posttest) and the path from the mediator to the distal outcomes at the posttest—we bootstrapped the confidence intervals, as recommended by the simulation study by Biesanz, Falk, and Savalei (2010).

The results of these estimation showed significant indirect effects of the intervention through trained words on intermediate expressive language ($\beta = 0.334$; 95% CI 0.095, 0.247]), intermediate minimal expressive language ($\beta = 0.172$; 95% CI 0.000, 0.160), and distal expressive language ($\beta = 0.502$; 95% CI 0.303, 0.742). In the mediation models with intermediate and distal expressive language, the direct path from the intervention to expressive language at the posttest became nonsignificant, showing that the impact of the intervention on these two expressive

constructs was fully mediated through the trained words. For minimal expressive language, the effect of the intervention was only partially mediated through the trained words, as there was also a direct effect of the intervention on minimal expressive language at the posttest. The mediation model with distal expressive language had a very good fit to the data ($\chi^2[21] = 37.612$; $p = 0.014$; RMSEA = 0.052; 90% CI 0.023–0.079; CFI = 0.970; SRMS = 0.038), and the two other mediation models were saturated.

Discussion

This study has revealed several interesting findings regarding the effect of language interventions and how knowledge transfers from words that are trained to measures that do not contain these words. First, we found that the effects from the language intervention and the transfer effects were generated through impact on expressive rather than receptive measures of language. Second, we found that the effects of the intervention on the intermediate transfer measures of language were mediated through the ability to define the trained words. This was true for both intermediate expressive measures that contained the trained words in the context of oral text (i.e., listening comprehension). It was also the case, albeit weaker, for the less expressive test using morphology tasks. Third, and critically, for far transfer measures that did not contain any of the trained words, the effects were also mediated through the trained words. As for the effect sizes of the mediation, the impact of the intervention on intermediate and distal expressive language was fully mediated through the trained words. For the morphology test, which was only partly expressive, the mediation was also only partly.

Theoretical Consequences

These findings have several potential theoretical repercussions. The finding that the effects of training and the transfer effects are solely related to expressive measures could indicate that the

primitive elements theory explains this transfer (Taatgen, 2013). Thus, one aspect that seems to improve and transfer to the untrained words is children's ability to develop procedures to give better explanations of words. The primitive elements theory predicts that transfer is possible between tasks that share the same basic underlying structure and similar operators or procedures. Thus, Taatgen (2013) claims that if an operator that is used in one task can be useful in another task as well, transfer will occur. It could therefore be the case that the procedures children use to explain trained words can also be used to explain untrained words. Notably, the results could also potentially fit Singley and Anderson's (1989) theory about production rules. This theory states that transfer will occur if one can use general rules to solve a task and apply those rules to another task. Because explaining the meaning of trained words very much resembles explaining the meaning of untrained words and sentences, it could be that similar procedures can be used to improve performance on the measures with untrained words. Hence, in this case, it can be difficult to distinguish between the production rules theory and the primitive elements theory, as both seem to fit our findings.

However, the finding that transfer is restricted to the expressive measures weakens the explanatory value of vector semantics (Jurafsky & Martin, 2014). If transfer occurs because children learn new words and relate them to other familiar words, we could perhaps expect to see a similar pattern in both receptive and expressive measures. Thus, the fact that we only see improvement in expressive measures may indicate that there is something related to the specific skills or procedures used in these tasks that improve, rather than the more general improvement of the underlying structure or network of word meanings.

The language intervention also targeted receptive language and contained many exercises in relation to this; however, the effects from the intervention were only evident in terms of

expressive measures. Results showing effects on expressive, but not on receptive, language from language interventions are in line with findings from other well-controlled studies (Lonigan & Whitehurst, 1998; Rogde et al. in press). A meta-analysis also reported an overall larger effect on listening comprehension and narrative retelling skills ($d = 0.43$) than on receptive language ($d = 0.17$) (Rogde et al., in press).

One could speculate that this difference between receptive and expressive measures is due to the receptive understanding of language being less malleable through interventions than is expressive language, in which one can improve specific procedures to explain words. This might relate to studies of IQ and schooling, which show that schooling improves scores on IQ tests by as much as one to five IQ points per year (Richie & Tucker-Drob, 2018). This improvement is, among other things, due to an improvement on verbal scales in IQ tests, which often consist of expressive measures (cf. word comprehension in the Wechsler tests that were used in this study). However, when looking at the nature of the improvement of IQ scores due to schooling, it becomes clear that schooling does not improve the underlying general intelligence (g) factor or IQ, per se, but rather the specific skills, strategies, and procedures needed to solve tasks commonly used in IQ tests (Ritchie, Bates, & Deary, 2015). Thus, it is possible that receptive measures draw upon this underlying g factor more than expressive measures do as these tasks are less dependent on specific procedures or skills.

Another important issue is that receptive language seems difficult to target, yet to have well-functioning expressive language, one is dependent on a foundation of well-developed receptive skills. Children very often have a much richer receptive language and language understanding than they are able to use expressively. It could therefore be that our intervention assisted children in exploiting more of their receptive language for expressive purposes. In future

studies, perhaps with even longer and stronger interventions, we may be able to see the effect of interventions on receptive skills as a result of improving expressive skills. Experimental training studies have demonstrated that this is possible (Hopman & MacDonald, 2018).

The findings can also cast light on more general theories of transfer. Thus, in line with the common elements theory (Thorndyke & Woodworth, 1901) and the primitive elements theory (Taatgen, 2013), we see that transfer is possible, at least within one domain. Sternberg and Frensch (1993) suggested that transfer is more likely to happen if a task meets four specific criteria or characteristics: encoding specificity, memory organization, discrimination, and similarity in settings. The language intervention used in this study included *encoding specificity* in that the training of the specific words emphasized how these words could be used in different contexts. Second, whether transfer and retrieval will occur depends on how information is *organized* in the memory, and it is likely that information about both trained words and untrained words is stored in a similar way in the memory. The training of the words also includes what Sternberg and Frensch (1993) labelled *discrimination* because the techniques or strategies used in expressing the trained words are applicable to untrained words. Finally, the children were primed to know that the intervention aimed to improve their language skills, and this might have made them more attentive not only to the training but also to new words in general, helping them use language expressively. This is in line with what Sternberg and Frensch (1993) described as the way in which a task in one *setting* might carry over to other settings. Thus, Sternberg and Frensch (1993) suggested that mechanisms for optimizing transfer seem to be present in language intervention and can perhaps provide an explanation for why transfer occurs.

However, it is important to note that the transfer here from the trained words only partly explains the effects on the intermediate transfer measure, which was only partly expressive (i.e.,

in morphology). This suggests that direct transfer—as measured by improvement from one specific task to another—is difficult to obtain, even within the same domain. Perhaps this is due to the fact that, as suggested by several studies (Bransford & Schwartz, 1999; Halpern, 1988), this way of measuring transfer is somewhat limited because the transfer of learning is subtler and occurs on a more general level, related to meta-cognition, verbal reasoning, and critical thinking.

Implications for Interventions

This study also has consequences for language intervention and instruction. Our findings indicate that most of the effects of language interventions are generated through expressive measures and that effects on receptive measures are not easily obtained. Thus, when constructing language interventions, it might be reasonable to focus on maximizing effects on expressive measures rather than on receptive skills. This can be done by focusing on expressive skills in the intervention, such as retelling stories, constructing sentences, getting the children to create stories orally, and having them finish stories that are already started. This can also potentially have implications for classroom instruction. Because expressive language skills seem effective in promoting children's language skills in general, teachers should aim at planning instruction that allows the children to have rich experiences with such expressive measures.

A meta-analysis of language interventions also indicated that small group interventions, as we also had in our study, are the most effective (Rogde et al., in press). This may be because children have more opportunities to use their language expressively in smaller group settings than they do in whole-classroom settings. Thus, in order to allow children with language challenges to catch up to their peers, instruction in small groups is likely to be most beneficial.

Moreover, in line with theories of transfer (e.g., Sternberg & Frensch, 1993; Taatgen, 2013), it is important that the language interventions are constructed in such a way that transfer

can be enhanced. This means that words should be focused on in context, and the way in which different words can be used in different contexts should be emphasized. It could also be beneficial to focus on the procedures that are used when solving tasks so that children become aware of the procedures they use and, perhaps, become more likely to use them in other tasks. Finally, it could also be useful to prime the children quite specifically about what the intervention aims to do. Awareness that the intervention is targeting their language skills can make them more conscious of how they use language and the language that surrounds them, which may in turn make it more likely that the effects will transfer beyond the three to four hours per week they spend in the language intervention group. Also, to study transfer effects, it can be useful to set up the effect measures in an incremental way, as was done in this study, including near, intermediate, and far transfer measures.

Limitations and Future Directions

First of all, it should be noted that the theories used as a backdrop for the current study were mainly developed based on and to explain cognitive skills, such as memory, nonverbal reasoning, and problem solving, rather than language skills. In the future it is a need for developing this theory and conduct theoretical experiments also on language skills.

Although mediation models can be a useful tool to examine transfer, it is also important to note that even though this is a randomized controlled trial, the effects in a mediation model are not causal; the children are not randomized based on the mediator, and therefore, the models can shed light on longitudinal relationships but not causal effects (Kline, 2015). Thus, there could be third variables underlying both improvements on the trained words and on the transfer measures. One such third variable is likely to be children's general capacity for language learning or the rate at

which children are able to learn new words, and this could be useful to examine in further detail in future studies.

Notably, in the current study, vocabulary was used as a near transfer measure, as the ability to learn new words can be seen as fundamental not only for vocabulary but also for tasks that require sentence construction, grammar, and narratives. However, in future studies, it could also be interesting to review near transfer measures not only for vocabulary but also for narrative skills (i.e., stories that the children explicitly work with in the intervention) and grammar (i.e., sentences and morphemes that are specifically trained in the intervention). By doing this, it may be possible to get even stronger indirect effects.

Moreover, we here examined transfer effects from trained words to more general language skills. Because general language skills are a central component in reading comprehension, it would also be important for future intervention studies to examine whether it is possible to get transfer effects from general language skills to reading comprehension and to set up language interventions so that the possibility for this can be taken full advantage of. Only a few studies to date have examined transfer effects to reading comprehension, and the results are mixed (Rogde et al., in press). However, studies that have used small-group interventions show transfer effects to reading comprehension that give reasons for optimism and that should be further explored in future studies (Brinchman, Hjetland, & Lyster, 2015; Clarke et al., 2010). It should also be mentioned that the language problems in this sample were not particularly severe, and there is a need for more intervention studies of clinically referred children with more severe language problems.

Finally, it is important to note that effects from an intervention that improves specific skills and procedures rather than underlying verbal IQ, per se, do not make the improvement less valuable (Ritchie et al., 2015). Regardless of whether g is affected, improving language on

standardized tests by a quarter of an SD or more can lead to important advantages and potentially make a difference in the lives of children.

Overall, language intervention and instruction for vulnerable groups is a promising and valuable tool to increase educational output, break unfortunate cycles, and promote social equity. However, in order to maximize the effects of these interventions, future studies should also focus on the potential mechanisms that underlie these effects. By doing so, we can construct more targeted interventions. Thus, instead of delivering multicomponent interventions in which some components might be less effective or lack effect entirely, more time could be spent on the aspects of the interventions that are actually responsible for these positive effects.

Data Availability Statement: Data available in the open science framework, https://osf.io/cua4f/

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1997). Rejoinder: Situative versus cognitive perspectives: Form versus substance. *Educational Researcher, 26*(1), 18–21.

Barnett, S. M., & Ceci, S. J. 2002. When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637.

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

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

Biesanz, J. C., Falk, C. F., & Savalei, V. Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research, 45,* 661–701.

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

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*(1), 61–100.

Brinchmann, E. I., Hjetland, H. N., & Lyster, S. A. H. (2016). Lexical quality matters: Effects of word knowledge instruction on the language and literacy skills of third-and fourth-grade poor readers. *Reading Research Quarterly*, *51*(2), 165–180.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, *21*(8), 1106–1116.

Cole. D. A., & Preacher, J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods 19,* 300–315.

Detterman, D. L. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction*,(pp. 1-24). Norwood, NJ: Ablex.

Doyle, O., Harmon, C. P., Heckman, J. J., & Tremblay, R. E. (2009). Investing in early human development: timing and economic efficiency. *Economics & Human Biology*, *7*(1), 1–6.

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

Farver, J. A. M., Lonigan, C. J., & Eppe, S. (2009). Effective early literacy skill development for young Spanish-speaking English language learners: An experimental study of two methods. *Child Development*, *80*(3), 703-719. doi: 10.1111/j.1467-8624.2009.01292.x

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*(3), 884.

Gjessing, H., & Nygaard, H. (1975). *ITPA håndbok (Norsk utgave) [Norwegian version].* Oslo, Norway: Universitetsforlaget.

Gonzalez, J. E., Pollard-Durodola, S., Simmons, D. C., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2010). Developing low-income preschoolers' social studies and science vocabulary knowledge through content-focused shared book reading. *Journal of Research on Educational Effectiveness*, *4*(1), 25-52. doi: 10.1080/19345747.2010.487927

Hagen, _A.M. (2018). Improving the odds: Identifying language activities that support the language development of preschoolers with poorer vocabulary skills. *Scandinavian Journal of Educational Research, 62*(5), 649-663.. https://doi.org/10.1080/00313831.2016.1258727.

Hagen, Å. M., Melby-Lervåg, M., & Lervåg, A. (2017). Improving language comprehension in preschool children with language difficulties: a cluster randomized trial. *Journal of   Child   Psychology and Psychiatry*, *58*(10), 1132-1140.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449.

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

Hopman, E. W., & MacDonald, M. C. (2018). Production practice during language learning improves comprehension. *Psychological Science*, *29*(6), 961–971.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3). London, UK: Pearson.

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychological science*, *25*(11), 2027-2037.

Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology*, *37*(4), 202–213.

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

Lee, A. Y. (1998). Transfer as a measure of intellectual functioning. In S. Soraci & W. J. McIlvane (Eds.), *Perspectives on fundamental processes in intellectual functioning: A survey of research approaches* (Vol. 1) (pp. 351–366). Stamford, CT: Ablex.

Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, *13*(2), 263–290.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512-534

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide version 8* (6th ed.). Los Angeles, CA: Muthén & Muthén.

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

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology, 51*(5), 573–582. http://dx.doi.org/10.1037/a0038981

Ritchie, S. J., & Tucker-Drob, E. M. (2017). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*(8), 135 –1369. doi:10.1177/0956797618774253.

Rogde, K., Hagen, Å. M., Melby-Lervåg, M., Lervåg, A., Hagen, Ã. M., & Melby-Lervåg, M. (in press). The Effect of Language Comprehension Training on Standardized Tests: A   Systematic Review. *Campbell Systematic Reviews*.

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.

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review*, *18*, 46–57.

Sala, G., & Gobet, F. (2017). When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review*, *20*, 55–67.

Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive

ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*,

144(2), 111.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill* (No. 9). Cambridge,

MA: Harvard University Press.

Sternberg, R. J., & Frensch, P. A. (1993). Mechanisms of transfer. In D. K. Detterman & R. J.

Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 25–

38). Westport, CT: Ablex Publishing.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*,

120(3), 439.

Taatgen, N. A. (2016). Theoretical models of training and transfer effects. In T. Strobach & J.

Karback (Eds.), *Cognitive training* (pp. 19–29). Cham, Switzerland: Springer.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental

function upon the efficiency of other functions. II. The estimation of magnitudes.

Psychological Review, *8*(4), 384.

Valdez-Menchaca, M. C., & Whitehurst, G. J. (1992). Accelerating language development

through picture book reading: A systematic extension to Mexican day care.

Developmental Psychology, *28*(6), 1106. doi: 10.1037/0012-1649.28.6.1106

Valente, M. J., and MacKinnon, D. P. (2017). Comparing models of change to estimate the

mediated effect in the pretest–posttest control group design. *Structural Equation

Modeling, 24*, 428–450. doi:10.1080/10705511.2016.1274657

Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence* (Revised). San

Antonio, CA: 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 picturebook reading. *Developmental Psychology, 24*, 552–

555.

Tables and Figures

Table 1

Means, Standard Deviations, and Effect Sizes from the Training on the Individual Language Measures Containing Taught Vocabulary

	Mean (SD)		Cronbach's alpha	Cohen's d (Z-value)	*p*	*Maximum Possible Score**
	Intervention Group	Control Group				
Vocabulary Definitions (near)						
Pretest	5.46 (4.19)	5.90 (4.28)	0.70			90
Posttest	28.77 (13.94)	18.28 (9.27)	0.87	0.828 (9.36)	<0.001	90
Listening Comprehension (intermediate)						
Pretest	16.41 (5.37)	15.65 (6.77)	0.83			36
Posttest	29.82 (5.36)	26.84 (5.90)	0.82	0.468 (3.56)	<0.001	42
Morpheme Generation (far)						
Pretest	9.10 (4.40)	10.05 (5.16)	0.80			30
Posttest	16.94 (4,39)	15.58 (4.81)	0.79	0.413 (3.37)	0.001	30

Note. Effect sizes = SD difference between the trained and the business-as-usual control group, controlling for pretest using robust (Huber–White) standard errors.

*The minimum possible score for all measures was 0.

Table 2

Means, Standard Deviations and Effect Sizes from the Training on the Individual Far transfer Language Measures

	Mean (SD)		Cronbach's alpha	Cohen's d (Z-value)	*p*	*Maximum Possible Score**
	Intervention Group	Control Group				
Listening Comprehension						
Pretest	16.61 (5.89)	17.28 (6.14)	0.83			36
Posttest	26.88 (5.39)	23.65 (5.67)	0.90	0.614 (6.06)	<0.001	42
Narrative Skills (Bus Story)						
Pretest	12.13 (6.94)	13.77 (7.12)	0.77			52
Posttest	22.47 (7.80)	19.56 (8.37)	0.78	0.448 (3.84)	<0.001	52
Syntactic Skills (TROG-II)						
Pretest	44.63 (17.14)	41.25 (15.53)	0.96			80
Posttest	64.18 (12.48)	59.12 (13.40)	0.95	0.304 (1.83)	0.067	80
Morpheme Generation (ITPA GC)						
Pretest	11.87 (3.45)	12.15 (4.23)	0.75			33
Posttest	16.62 (4.10)	16.33 (4.10)	0.75	0.097 (0.755)	0.145	33
Vocabulary Definition (WPPSI/WISC)						
Pretest	15.56 (7.00)	16.50 (5.98)	0.81			66
Posttest	25.08 (6.22)	23.88 (8.25)	0.83	0.194 (1.456)	0.145	66
Vocabulary Breadth (BPVS-II)						
Pretest	55.42 (11.36)	53.86 (12.95)	0.90			144
Posttest	74.61 (11.21)	72.80 (11.44)	0.86	0.133 (0.867)	0.386	144

Note. Effect sizes = SD difference between the trained and the business-as-usual control group, controlling for pretest using robust (Huber–White) standard errors.

*The minimum possible score for all measures was 0.

Table 3

The Effects of the Intervention on All the Language Constructs Where Measurement Errors Have Been Controlled Using Latent Variables.

	Cohen's d[a]	SE	p
Near Language			
Word Definitions (Trained Words)	0.910	0.102	<0.001
Intermediate Language (Containing Trained Words)			
Expressive Measures[b]	0.500	0.148	0.001
Minimal Expressive Measures[c]	0.505	0.139	<0.001
Distal Language (Not Containing Trained Words)			
Expressive Measures[d]	0.839	0.114	<0.001
Minimal Expressive Measures[e]	0.126	0.153	0.410
Receptive Measure[f]	0.206	0.276	0.456

Note. [a] = the Y-standardized effect size of the intervention; [b] = the listening comprehension containing trained words; [c] = grammatical closure containing trained words; [d] = the listening comprehension, vocabulary (WISC), and narrative retelling (Bus Story); [e] = grammatical closure; [f] = BPVS-II and TROG-II.
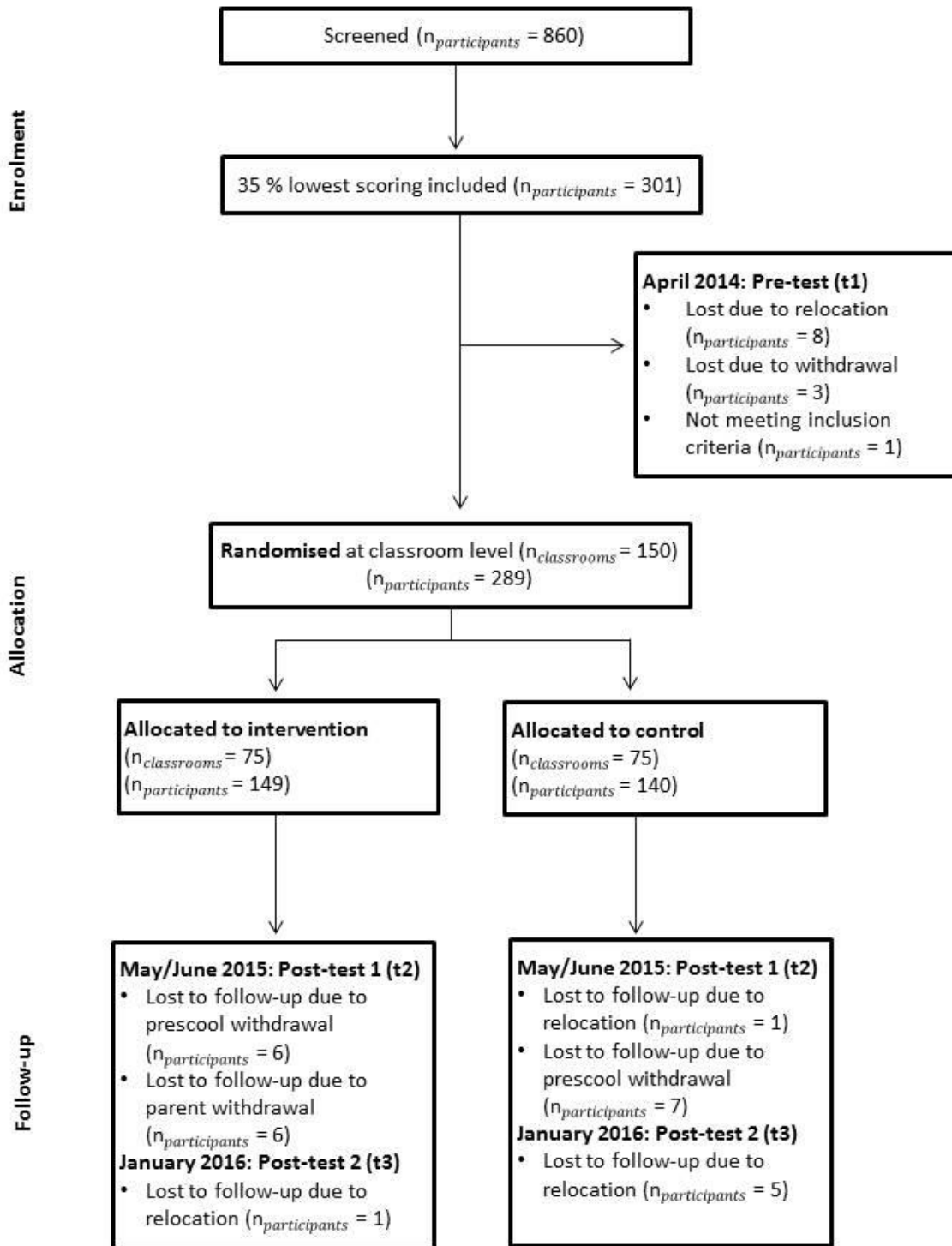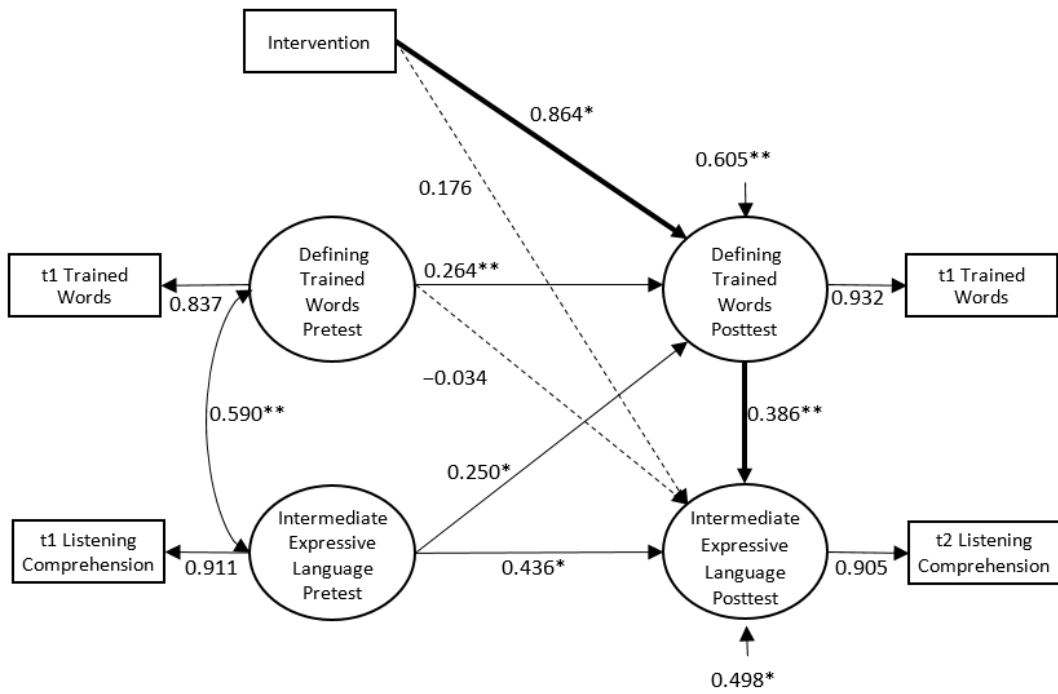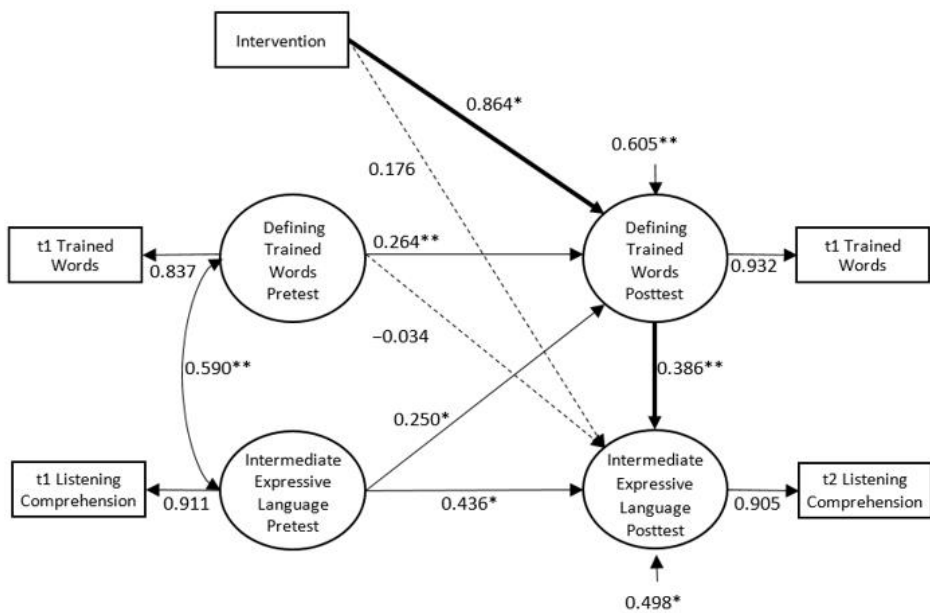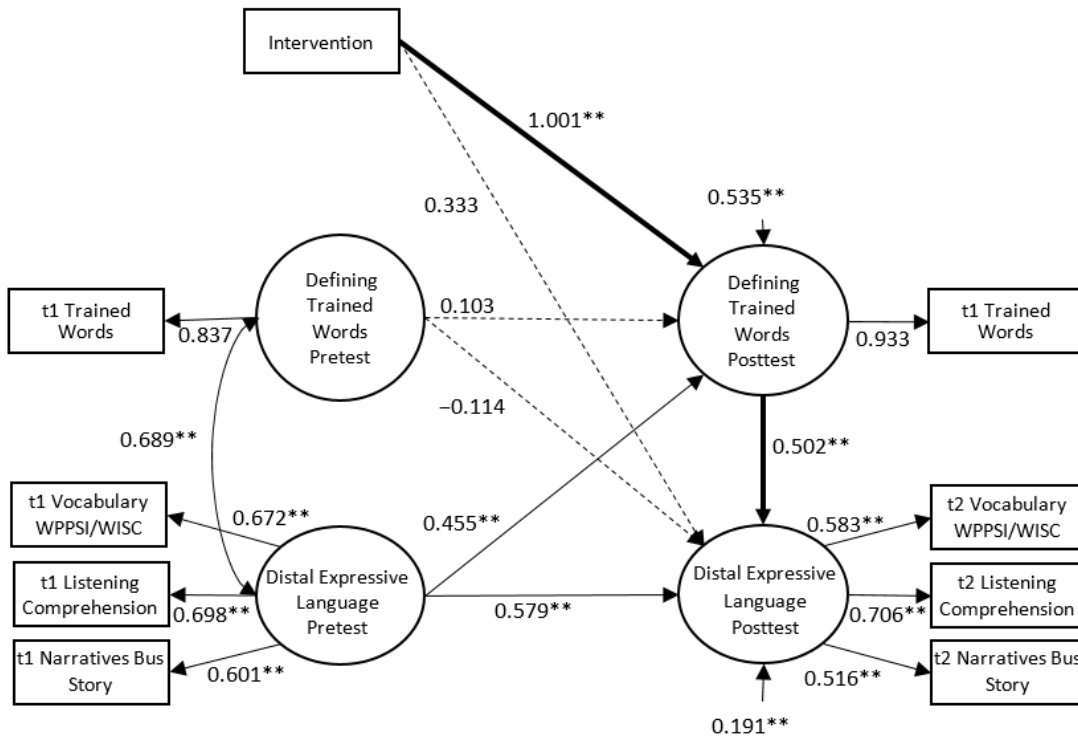
Figure 1

Figure 2a

Figure 2b

Figure 2c

Figure Legends

Figure 1. Structure of the study and flow of participants.

Figure 2. Latent mediation models for the three language constructs (a = intermediate expressive, b = intermediate minimal expressive, and c = distal expressive) that were significantly impacted by the intervention. In these models, rectangles reflect observed variables and ellipses reflect latent variables. The one-arrow paths from the latent variables to the observed variables reflect the factor loadings, and the one-headed arrows from a number to the latent variables reflect the residuals of these variables. The one-headed and two-headed arrows between the latent variables reflect the true-score regressions and correlations between the latent variables, respectively. The one-headed arrow from the intervention variable to other latent variables reflects the impact of the intervention on that particular latent variable and can be interpreted as Cohen's *d*. All other coefficients are fully standardized. The indirect paths are drawn in bold. Dashed paths are estimated but not significant.

*p < 0.05

**p < 0.01