*2) Title Page (WITH Author Details)



# How do Verbal Short-Term and Working Memory Relate to the Acquisition of Vocabulary and Grammar? A Comparison between First and Second Language Learners

Josje Verhagen[1] and Paul Leseman[1]

[1]Utrecht University, Department of Special Education: Cognitive and Motor Disabilities

Heidelberglaan 1, 3584 CS Utrecht, the Netherlands

Email addresses: J.Verhagen@uu.nl, P.P.M.Leseman@uu.nl

Corresponding author:

Dr. Josje Verhagen

Department of Special Education: Cognitive and Motor Disabilities, Utrecht University,

Heidelberglaan 1, 3584 CS Utrecht, the Netherlands

Phone number: +31-(0)30-2531511, Email address: J.Verhagen@uu.nl

Word count: 8,183

Abstract

Previous studies show that verbal short-term memory (VSTM) is related to vocabulary learning, while verbal working memory (VWM) is related to grammar learning in children learning a second language (L2) in the classroom. In this study, we investigate if the same relationships apply to children learning an L2 in a naturalistic setting and to monolingual children. We also investigate if relationships with verbal memory differ depending on the type of grammar skill investigated (i.e., morphology vs. syntax). Participants were 63 Turkish children who learned Dutch as an L2 and 45 Dutch monolingual children (mean age: five years). Children completed a series of VSTM and VWM tasks, a Dutch vocabulary task, and a Dutch grammar task. A confirmatory factor analysis showed that VSTM and VWM represented two separate latent factors in both groups. Structural equation modeling showed that VSTM, treated as a latent factor, significantly predicted vocabulary and grammar. VWM, treated as a latent factor, only predicted grammar. Both memory factors were significantly related to the acquisition of morphology and syntax. There were no differences between the two groups. These results show that (i) VSTM and VWM are differentially associated with language learning and (ii) the same memory mechanisms are employed for learning vocabulary and grammar in L1 children and L2 children who learn their L2 naturalistically.

Keywords: verbal short-term memory, working memory, vocabulary, grammar, first language learners, second language learners, structural equation modeling [221 words]

# How do Verbal Short-Term and Working Memory Relate to the Acquisition of Vocabulary and Grammar? A Comparison between First and Second Language Learners

There is increasing evidence that verbal working memory is related to the acquisition of vocabulary and grammar in both first language (L1) and second language (L2) learning (Adams & Gathercole, 1996, 2000; Baddeley, Gathercole, & Papagno, 1998; French & O'Brien, 2008; Masoura & Gathercole, 2005). Significant correlations with language learning have been found for both components of verbal memory, that is, for verbal short-term memory (VSTM), referring to the capacity to store verbal information, and for verbal working memory (VWM), or the ability to process verbal information while it is being stored. However, few studies have simultaneously examined effects of VSTM and VWM on language learning in the same sample. Also, to the best of our knowledge, no previous studies have investigated (i) whether relationships between VSTM or VWM and language learning are the same for L1 and L2 children, and (ii) whether these relations are similar for vocabulary and grammar.

VSTM has been considered important for the development of stable phonological representations in long-term memory that are needed for vocabulary and grammar learning, based on studies with L1 or L2 children (Baddeley et al., 1998; Speidel, 1989). VWM has been considered important for grammar learning through its involvement in noticing (Mackey, Philp, Egi, Fuji, & Tatsumi, 2002) and processing of linguistic structures (Harrington & Sawyer, 2002; Sunderman & Kroll, 2009), but these claims have been almost exclusively based on L2 classroom studies. There is some evidence, however, that VWM is related to grammar learning more strongly in explicit L2 learning conditions than in implicit learning conditions (Tagarelli, Borges-Mota, & Rebuschat, 2011), in line with the idea that explicit learning requires the control of attention, an important function of VWM.

In this study, we aim to obtain a more complete picture of how verbal memory relates to language learning than in previous studies by investigating how the two components of verbal memory (VSTM and VWM) relate to two domains of language (vocabulary and grammar) in two learner groups (L1 children and naturalistic L2 children). In so doing, our goals are to obtain a better understanding of a potentially major source of individual differences in L1 and L2 vocabulary and grammar learning as well as shed more light on whether the same verbal memory processes are involved in L1 and naturalistic L2 learning.

To the best of our knowledge, only two previous studies have simultaneously looked at effects of VSTM and VWM in the same sample, but both looked at L2 children acquiring their L2 in the classroom (Engel de Abreu & Gathercole, 2012; Kormos & Sáfár, 2008). For these children, a division of labor between the two memory components was found: while VSTM was associated with L2 vocabulary, VWM was associated with L2 grammar. In this study, we investigate if the same relationships apply to children acquiring their L2 naturalistically, without formal instruction, and children acquiring their native language.

**Verbal Working Memory and Language Learning**

A common view on the structure of working memory holds that working memory is not a single store, but a system containing separate but interacting components (Baddeley & Hitch, 1974). Besides a domain-general component, termed central executive, there are two domain-specific storage components for verbal and visuospatial information. The verbal storage component, or ‘phonological loop’, allows the storage of verbal information for short periods of time; the visuospatial sketchpad enables the storage of visual and spatial representations. The central executive is a domain-general component responsible for a range of processes, such as controlling and monitoring information, retrieving information from long-term memory and attentional control (Baddeley & Logie, 1999). Studies on young children using confirmatory factor analyses have shown that all working memory components are in place from age four on (Alloway, Gathercole, Willis, & Adams, 2004; Alloway, Gathercole, & Pickering, 2006).

VSTM has typically been measured through simple span tasks that require the storage of verbal units, such as nonwords or digits. VWM has been measured through complex span tasks that require the simultaneous short-term storage and processing of information. For example, in sentence span tasks, participants are asked to recall the last word of each sentence in a series of sentences, while at the same time judging whether each sentence is true or false. So, like VSTM tasks, VWM tasks require phonological storage, but in addition, they tap executive processes that are responsible for maintaining information active and updating the information that is stored.

For monolingual children, previous studies show that both VSTM and VWM are involved in language acquisition. VSTM shows strong associations with word learning (Gathercole, et al., 1992; see Gathercole, 2006 for an overview) as well as with the production of long and grammatically complex sentences (Adams & Gathercole, 1996, 2000; Blake et al., 1994). VWM is related to grammar learning, more specifically, to monolingual children's grammaticality judgements (Gottardo, Stanovich, & Siegel, 1996), receptive syntax (Ellis Weismer, Evans, & Hesketh, 1999), and sentence comprehension (Montgomery, 1995).

Most previous studies have looked at composite measures of grammatical ability, however, rather than at specific grammatical constructions. An exception to this is McDonald (2008), which investigated how differences in VWM were associated with children's judgment accuracy of various grammatical structures. McDonald found that differences in VWM predicted performance on constructions that require children to keep in mind multiple sentence parts, while checking the consistency of information across these parts (e.g., regular past tense, third person agreement), but not on other constructions (e.g., yes/no-questions).

For L2 children, previous work on the role of VSTM has also shown significant relationships with L2 vocabulary (Cheung, 1996; Masoura & Gathercole, 2005) and L2 grammar (French & O'Brien, 2008; Kormos & Sáfár, 2008; Service & Kohonen, 1995; Verhagen, Messer, & Leseman, 2015). However, as for L2 grammar, studies differ as two whether effects of VSTM were independent of L2 vocabulary knowledge: Whereas, in some studies, effects of VSTM on L2 grammar could be explained by L2 vocabulary knowledge (French, 2006; Service, 1992; Service & Kohonen, 1995), other studies reported direct effects of VSTM on L2 grammar, independent of L2 vocabulary (French & O'Brien, 2008; Verhagen et al., 2015). Verhagen et al. (2015), for example, studied the same L2 children as the current study at a younger age (age four) and found that VSTM predicted children's production of L2 grammatical structures such as subject-verb agreement and word order, even after differences in L2 vocabulary were controlled.

Studies on relationships between VWM and L2 grammar have focused on adult learners, showing, for example, that adults rely on VWM when learning a novel language (Martin & Ellis, 2012; Williams & Lovatt, 2003). Previous work on L2 adults also suggests that VWM is related to grammar learning more strongly in situations of explicit learning rather than implicit learning (Tagarelli, Borges-Mota, & Rebuschat, 2011), presumably due to the fact that explicit learning requires attentional control, an important function of VWM. Specifically, Tagarelli et al. (2011) found that VWM predicted adults' learning of a syntactic pattern in a semi-artificial language when participants were asked to discover the word-order rules of this language, but not when they were not told about these rules, but instead, listened to the sentences for meaning. This raises the question if VWM is related to grammar learning in learners who acquire their L2 naturalistically, without explicit L2 instruction.

To the best of our knowledge, only two studies have simultaneously investigated effects of VSTM and VWM on language learning in the same sample. Kormos and Sáfár (2008) investigated effects of VSTM and VWM on a range of L2 language skills, including L2 vocabulary and grammar. Participants were fifteen- and sixteen-year-old Hungarian students who learned English through an intensive instruction program. VSTM was measured through nonword repetition and VWM was assessed through backward digit recall. Children's L2 proficiency was measured through a proficiency test that assessed L2 grammar and vocabulary in an integrated manner. This study showed that VSTM was significantly related to test scores in intermediate learners, but not in beginning learners. VWM was significantly correlated with test scores in all learners. However, as the test assessed both vocabulary and grammar, this study leaves unclear whether the two components of verbal memory have differential effects on L2 skills. Moreover, the participants in this study learned their L2 via explicit instruction, which may have led them to apply metalinguistic rules and strategies, implying stronger involvement of VWM (Linck & Weiss 2011; Tagarelli et al., 2011).

The second study comparing effects of VSTM and VWM on L2 (and L3) vocabulary and grammar was conducted by Engel de Abreu and Gathercole (2012). These authors examined data from trilingual Luxembourgian children who were seven or eight years old and learned their L2 (German) and L3 (French) at school. The study showed that VSTM, assessed through nonword repetition and digit recall and treated as a latent factor, was uniquely related to L1 and L2 vocabulary. VWM, assessed through counting recall and backward digit recall and also treated as a latent factor, was a unique predictor of grammar in children's L1, L2, and L3. So, in none of children's languages, there were effects of VSTM on grammar after vocabulary was controlled or effects of VWM on vocabulary, after grammar was controlled. However, as the children in this study learned their L2 and L3 through explicit, rule-based instruction, this study leaves open how VSTM and VWM are related to L2 language learning in naturalistic situations.

A few questions remain from these previous studies. First, how do VSTM and VWM relate to vocabulary and grammar in L2 children who acquire their L2 in a naturalistic setting rather than in an L2 classroom? One possibility is that naturalistic L2 learners rely more on phonological storage, also in grammar learning, than classroom learners. More specifically, such learners may – in the absence of intensive instruction involving metalinguistic L2 knowledge – mainly acquire the L2 by storing chunked stretches of speech in VSTM that they analyze only later on, with increasing L2 proficiency. Similar ideas have been advanced by Speidel (1989, 1993) and Speidel and Herreshoff (1989) as well as in connectionist accounts of L2 learning (Ellis & Sinclair, 1996; Martin & Ellis, 2012). Speidel (1989; 1993), for example, argued that L2 learners store grammatical constructions in VSTM, in the same way that they store words. In so doing, they build a "store house" of constructions in long-term memory from which they can extract patterns to support their spontaneous L2 speech. Indeed, Verhagen et al. (2015) found that VSTM was a significant predictor of L2 grammar (independent of L2 vocabulary) in naturalistic child L2 learners of Dutch. As for VWM, previous studies have only looked at L2 classroom children and there is some tentative evidence from adults that VWM is related to artificial grammar learning in explicit learning but not in implicit learning (Tagarelli et al., 2011), so it is an open question if VWM is related to L2 grammar in naturalistic L2 learners.

A second question that remains from earlier work is whether relationships between VSTM and VWM on the one hand, and vocabulary and grammar on the other, are the same for

L2 children and their monolingual peers. None of the previous studies included a monolingual comparison group so, it is as yet unclear whether the two components of verbal memory relate to vocabulary and grammar in the same way in L1 and L2 children. An L1-L2 comparison would enable us to see if L1 acquisition is less affected by memory processes than L2 acquisition, as assumed in mentalist approaches holding that L1 learners use innate principles, whereas L2 learners use general problem-solving and processing principles (Bley-Vroman, 1990).

Finally, a question that has hitherto not received much attention is how VSTM and VWM relate to the acquisition of grammatical sub-skills. As outlined above, McDonald (2008) found a relationship between VWM and L1 children's ability to judge grammaticality of constructions for which information across constituents had to be checked for consistency (e.g., verb morphology), but not for other types of constructions. Addressing which domains of L1 and L2 grammar are susceptible to individual differences in verbal memory skill is important, as it may shed more light on why verbal memory is related to grammar learning.

## This Study

In this study, we investigate how VSTM and VWM relate to the acquisition of vocabulary and grammar in five-year-old L1 and L2 children. An advantage of studying children at this relatively young age as compared to earlier studies is that monolingual children could be included as a comparison group. Monolingual five-year-olds are still in the process of acquiring relatively basic vocabulary and grammar and thus do not perform at ceiling on vocabulary and grammar tests.

The current L1 children came from monolingual Dutch families in the Netherlands. The L2 children came from Turkish immigrant families in the Netherlands in which Turkish was the main language of communication. While the L2 children had been exposed to the majority language Dutch from birth through television, contacts outside the home and siblings, systematic exposure to Dutch only started at the age of three or four when children entered preschools or kindergarten. At that age, they were immersed in a Dutch-speaking (pre-)school environment where they did not receive explicit instruction in Dutch, but learned the language naturalistically, through every-day communication with their teachers and peers.

The L1 and L2 children were given a number of VSTM and VWM tasks as well as vocabulary and grammar assessments. The following research questions guided our study:

1. How do VSTM and VWM relate to the acquisition of vocabulary and grammar in naturalistic Turkish child L2 learners of Dutch?
2. Do relationships between VSTM and VWM on the one hand, and vocabulary and grammar on the other, differ between these L2 children and L1 Dutch children?
3. Do relationships between VSTM or VWM and grammar in the two groups vary depending on the grammatical sub-skills investigated (i.e., morphology or syntax)?

As for the first question, we predict that VSTM is related to L2 vocabulary as well as L2 grammar, based on the assumption described above that naturalistic learners store words and grammatical constructions in VSTM, to support language learning (Ellis & Sinclair, 1996; Speidel, 1989). Regarding VWM, we do not expect a relationship with L2 vocabulary, as such a relationship was not found in earlier studies and there is no reason why differences in processing skill would be related to L2 vocabulary. It is an open question if VWM is related to L2 grammar. Previous work finding such a relationship only looked at L2 classroom children (Engel de Abreu & Gathercole, 2012; Kormos & Sáfár, 2008) and there is some evidence from adults that VWM is involved in explicit grammar learning but not implicit learning (Tagarelli et al., 2011). However, as VWM is involved in the acquisition of at least some grammatical constructions in L1 (McDonald, 2008), which involves implicit learning, VWM may also be involved in naturalistic L2 grammar learning.

As for the second question, we expect that the same relationships between VSTM and VWM on the one hand, and vocabulary and grammar on the other, will be found in L1 children. Like naturalistic L2 children, L1 children learn words and grammar through implicit processes, so we expect that the same memory processes will be involved in both learner groups.

Finally, regarding our third question, we predict that VWM is related to the acquisition of grammatical constructions that require linguistic information to be checked across sentence parts, but not or less strongly related to other types of grammatical constructions, based on McDonald (2008). Specifically, in our study, we expect the strongest relations for the acquisition of syntax and either no or weaker relations for plural noun morphology and past participles. No differences are expected between the groups.

To investigate these questions, we applied a latent factor approach (see also Engel de Abreu & Gathercole, 2012). That is, we used multiple tasks for assessing VSTM and VWM, and constructed latent factors for both memory components. The advantage of this approach is that it allowed us to confirm whether these factors, as measured through our tasks, constituted two different constructs in the L1 and L2 children. Another advantage of adopting a latent structure approach rather than looking at true task scores is that it reduces the variance in scores due to task properties. As a result, any relationship found between a latent factor representing, for example, VSTM, and another variable (e.g., vocabulary) is more likely to represent a true relationship between this underlying construct and the variable than a relationship that can be explained by specific task characteristics.

# Method

## Participants

Participants were 63 Turkish L2 learners of Dutch with a mean age of 63 months (*SD* = 2.5, range = 59 – 75, 59% boys) and 45 L1 learners of Dutch with a mean age of 62 months (*SD* = 2, range = 58 – 66, 69% boys). These children constituted a subset of the participants studied in Messer et al. (2010). Children were excluded from the larger sample in Messer et al. if their parents reported the use of other languages than Dutch (L1 group) or if Turkish was not the main language they spoke (i.e., >= 75% of the time in everyday communicative situations at home) to their children (L2 group). These criteria were applied to make sure that children clearly fell within one of the two groups. Detailed information about parents' language use was obtained through questionnaires that were administered in oral interviews with children's primary caregivers (Messer et al., 2010). These interviews were conducted by research assistants who were fluent in both Dutch and Turkish.

Family SES was computed on the basis of parents' highest completed level of education as well as a measure of their job status. More precisely, parents' highest completed level of education was coded as ranging from 1 (no education) to 7 (university degree) and the status of their jobs on the Dutch national job index list was coded on a scale from 1 (unemployed) to 6 (academic job level). Subsequently, SES was computed as the mean of parents' level of education and job. This yielded a mean value of 4.12 for the L1 children (*SD* = 1.19, range = 2 – 6) and a mean value of 2.38 for the L2 children (*SD* = 1.05, range = 0 – 4.50). Since this difference was significant ($F(1,92) = 59.46$, $p <. 001$, $\eta^2_p = .40$), SES was included as a covariate in the analyses.

Data were collected in three waves when children were four, five and six years old (for more details, see Messer 2010). For the current study, however, only data from wave two were taken when children were five years old. Data from the first wave were not included because no grammar assessment was included at this age and one of the VWM tasks appeared to be too difficult. Data from the last wave were not analyzed either, due to a ceiling effect on the vocabulary and one of the grammar tests in the L1 group at this age. Informed consent was obtained from the parents of all participating children.

**Measures**

**Verbal memory.** The verbal memory measures used were taken from the Automated Working Memory Assessment (AWMA; Alloway, 2007) and translated into Dutch (cf. Messer, 2010). The AWMA is a computerized test battery for assessing visuospatial and verbal short-term and working memory in children between ages four and eleven. The verbal memory tasks in this battery measure VSTM with simple span tasks that assess the storage of information, and VWM with complex span tasks that assess the simultaneous storage and processing of information. For the current study, three simple span tasks and two complex span tasks from the AWMA were taken. For all tasks, the original procedures of the AWMA were followed for instruction and scoring. Psychometric quality of the AWMA assessments is satisfactory to good (Alloway, 2007; see also Automated Working Memory Assessment (AWMA) – Reliability and Validity, 2014).

***VSTM.*** In the *Word recall* task, children were presented with sequences of highly frequent words of increasing length and asked to repeat these sequences in the correct order.

Items had been pre-recorded by a female native speaker of Dutch. The task started with a block of one item and built up to a block of seven items in a row. Each block contained six trials that were scored as correct when all items were repeated in the correct order. When the first four trials within a block were recalled correctly, the child automatically received a score of 6 and proceeded to the next block. Testing stopped after three incorrect recalls within one block. The maximum score was 42. In the *Dutch-like nonword recall* task, children repeated monosyllabic Dutch-like nonwords in lists of increasing length, starting with a block of one nonword and building up to a block of five nonwords in a row. The nonwords used in this task were novel words containing highly frequent phoneme combinations in Dutch and thus were similar to Dutch words (see Messer et al., 2010 for more details about the items in this task). The nonwords had been pre-recorded by the same female speaker used for the items in the word recall task. As in the word recall task, each block consisted of six trials. A trial was awarded a score of 1 when none of the nonwords were omitted, when the sequence of nonwords was correct, and when each nonword was recalled correctly. Each phoneme of a nonword needed to be recalled correctly for a positive score, with the exception of consistently substituted phonemes resulting from articulation problems. With a total of six trials per block, the maximum score per block was 6. When the first four trials within a block were recalled correctly, the child automatically received a score of 6 and proceeded to the next block. Testing stopped after three incorrect trials within one block. The maximum score was 30. Finally, the *Dutch-unlike nonword recall* task was exactly the same as the Dutch-like nonword task, except that nonwords were used that contained phoneme combinations that are infrequent in Dutch (for more details, see Messer et al., 2010; Messer et al., 2015).

***VWM.*** In the *Backward digit recall* task, children were presented with sequences of spoken digits and asked to recall these sequences in reverse order. Trials began with two digits, and increased by one digit in each block. Trials were scored as correct if children recalled all digits in backward order. As in the other AWMA tests, testing stopped when children recalled only three correct trials of a block, and proceeded to the next block if they recalled four trials in correctly. Scores were calculated as the number of correct trials for each child. In the *Listening recall* task, children were presented with a series of spoken sentences. Their task was twofold: (1) verify the sentence by saying ‘true’ or ‘false’ and (2) recall the first word of each sentence in the correct sequence.[1] Test trials began with one sentence, and proceeded with additional sentences in each block until the child was unable to recall three correct trials in a block. Trials were scored as correct if children recalled all words in the sentences. Testing stopped when children recalled only three correct trials in a block, and continued to the next block if they recalled four trials correctly, as in the other AWMA tests. Scores were calculated as the number of correct trials for each child.

**Grammar.** The Dutch standardized language test *Taaltoets alle kinderen* (TAK) (Language test for all children, Verhoeven & Vermeer, 2002) was used to assess children’s grammar skills. This test has been designed and normed for L1 and L2 children in the Netherlands and consists of various sub-tests. For the current study, three sub-tests were selected, two at the morphological level and one at the sentence level. Instruction and scoring procedures

[1] Usually, in this type of working memory task, participants are asked to recall the last word of each sentence rather than the first word of each sentence. However, in the current project, the L2 children performed the listening recall task twice, once in Dutch and once in Turkish. In Turkish, verbs are inflected for person, number and tense, and they appear at the sentence end. In order to avoid asking children to repeat the verb and all its inflections in the Turkish version of the task, and have comparable procedures in both languages, children were asked to recall the first word of each sentence in both language versions of the task.

of the original test were applied for all sub-tests. The TAK has excellent reliability and good validity (Verhoeven & Vermeer, 2006).

***Noun plurals.*** The plural noun morphology sub-test of the TAK assessed children's production of plural nouns. In this test, children were presented with pictures of objects and a prompt sentence from the experimenter. To elicit noun plurals, prompt sentences of the following type were used: *Dat is een X, dat zijn twee* … 'This is an X, these are two…' Children's production of three different plural forms was assessed: forms ending in the suffix '–en' without a vowel change in in the stem (*stoel-stoelen* 'chair-chairs'), forms ending in the suffix '–s' (*vlinder-vlinders* 'butterfly-butterflies'), and forms ending in '–en' that undergo a vowel change (*dak-daken* [dAk-da:ke] 'roof-roofs'). There were four items of each type, resulting in a total of twelve items. Children's scores were computed as the total number of correct responses for each child.

***Past participles***. The past participle sub-test of the TAK elicited children's use of past participle forms. Children were presented with a picture and a prompt sentence like *Rosita is een bal aan het gooien. Gisteren heeft zij ook al een bal* ... 'Rosita is throwing a ball. Yesterday she has also … a ball'. The task elicited three types of past participles: regular forms taking the circumfix 'ge–' and '–t' (*koken-gekookt* 'cook-cooked'), irregular forms with a vowel change in the stem (*vliegen-gevlogen* 'fly-fled'), and irregular forms with both a vowel and consonant change (*brengen-gebracht* 'bring-brought'). There were three items of each type, resulting in twelve items in total. Children's scores were computed as the total number of correct responses for each child.

***Sentence production.*** In the sentence production sub-test, the experimenter read out a complex sentence to the child and then asked the child to repeat this sentence as accurately as possible. While such sentence repetition tasks have also been used to assess children's VSTM (see Discussion), the current test is assumed to provide a measure of children's grammatical proficiency. [2] In particular, it assesses children's knowledge of function words (e.g., connectives) and syntactic structures (e.g., subject-verb inversion). Children's scores were computed as the total number of targeted function words and syntactic patterns that were repeated correctly by the child. Changes or omissions other than those involving the targeted structures were not taken into account in computing children's scores. Specifically, for each item, children received zero points if both the function word and syntactic pattern were incorrect, one point if either the function word or the sentence pattern was repeated correctly, and two points if both the function word and the sentence structure were repeated correctly. The task contained 20 items, so the maximum score was 40.

**Vocabulary.** Dutch receptive vocabulary was assessed through the *Diagnostische Toets Tweetaligheid* (Test for Bilingual Development, Verhoeven, Narrain, Extra, Konak, & Zerrouk, 1995) that was specifically developed for research with bilingual immigrant children in the Netherlands. In this test, children choose one out of four picture drawings after an orally presented word. To minimize fatigue and reduce testing time, the test was split in two parts, one part containing the odd items and the other half containing the even items (the correlation between the two parts was $r$ = .71 in a sample of 162 children, cf. Scheele et al., 2010). The test contained 30 items and was supplemented with fifteen items from a vocabulary test of the TAK, in order to avoid ceiling effects, as the TAK vocabulary test is applicable to a broader age range than the *Diagnostische Toets Tweetaligheid*.

---

[2] Based on a detailed investigation of the psychometric quality of the TAK, Verhoeven and Vermeer (2006) report good convergent validity for the sentence production task, as indicated by positive, significant, moderate to strong correlations with tasks assessing grammatical ability, narrative skills and spontaneous speech in both L1 and L2 children (*r*s between .38 and .75).

**Procedure**

Children were assessed individually by trained research assistants in a quiet room at their schools. Test sessions took place on two separate days that were on average one week apart. Both sessions lasted for approximately 75 minutes, including play breaks and tasks that were part of another study. The nonword recall tasks were videotaped for scoring purposes. The tasks were administered in a fixed order to vary task demands across successive tasks and minimize fatigue. The order of the tasks reported in the current study was as follows: Dutch vocabulary, Dutch-like nonword recall, listening recall, word recall (Day 1); TAK noun plurals, TAK past participles, TAK sentence production, backward digit recall, Dutch-unlike nonword recall (Day 2). To keep children motivated, they were given a sticker after each task.

**Analyses**

First of all, children's scores were inspected for outliers, normality, and missing data. For all tasks, there were no outliers greater than three standard deviations below or above the mean and standardized measures for skewness and kurtosis did not exceed the value of .3 in both groups, with the exception of listening recall task in the Turkish group on which many children obtained low scores (skewness: 2.2, kurtosis: 9.3). In the L1 group, two children did not complete the nonword recall tasks and one child did not complete the grammar sub-tests of the TAK.

As a first step, we tested for differences between the groups with a MANOVA in which 'group' was the between-subjects factor and children's task scores were the dependent variables. After an effect of group was found, we ran separate ANOVA's on the various task scores. We also calculated all zero-order (Pearson) correlations among the tasks for the two groups separately.

As a second step, to see whether VSTM and VWM could be considered separate constructs in the current sample of L1 and L2 children, we investigated the factor structure of verbal working memory using confirmatory factor analysis (CFA) in Mplus (Version 7.11, Muthén & Muthén, 1998-2010). Specifically, we tested a two-factor model in which the simple span tasks loaded on one factor (VSTM) and the complex span tasks loaded on a second factor (VWM). Full maximum likelihood estimation was used to deal with missing data (Enders, 2010), so all available information was fully used, including data from participants with missing data. In our CFA analysis, we tested a multi-group model in which all factor loadings and intercepts were constrained to be equal in the L1 and L2 groups, to test for measurement invariance. Full measurement invariance indicates that the same structural model can be applied to both groups and direct comparisons between groups can be made. Model fit was evaluated through various fit indices. As a rule of thumb, model fit is considered good if $\chi^2$ is not significant, RMSEA is below .08, and CFI and TLI are above .90 (Kline, 2005).

Subsequently, we used structural equation modeling (SEM) to investigate the relationships between the two latent memory factors on the one hand, and vocabulary and grammar on the other hand, in the two groups separately. The measurement model of VSTM and VWM obtained through the CFA-analysis above was used in this model to predict children's scores on grammar and vocabulary. For grammar, mean scores were computed on the basis of the three sub-tests of the TAK. SES was included as a covariate in the model, because this factor differed significantly between the two groups and was significantly correlated with some of the memory and language scores. As in the previous analyses, we used full maximum likelihood estimation in Mplus. All factor loadings and intercepts were constrained to be equal across groups, to investigate whether there was measurement invariance between groups. Subsequently, this constrained model was compared through a chi-square difference test against a model in which all regression estimates were freely estimated. As this test was not significant, we opted for the former, more parsimonious model, in which all regression estimates were constrained and direct comparisons between the two groups could be made.

Finally, we fitted a multi-group SEM model to investigate whether the two latent memory factors predicted children's scores on the grammar sub-tests. Specifically, we investigated whether VSTM and VWM differentially predicted children's on the morphology and sentence production sub-tests. For morphology, mean scores on the noun plurals and past participles sub-tests were taken. Dutch vocabulary and SES were entered as covariates in the model, since both factors were significantly related to some of the variables of interest. Again, we compared two nested models. In our first model, all factor loadings and intercepts were constrained to equality in both groups, while in the second model these were freely estimated. Again, the models did not differ on a chi-square difference test, so the former, more constrained model that allowed for direct comparisons between the two groups was chosen.

For all models, we tested whether bootstrapping would yield different results. Bootstrapping allows comparison of parametric values over repeated samples that have been drawn from the original sample (Byrne, 2001), and can be used if data are from a small sample or non-normally distributed. In the present analyses, bootstrapping with 1000 iterations did not yield different results (i.e., same confidence intervals and p-values), so the models without bootstrapping are reported.

## Results

### Descriptive Statistics and Correlations

Table 1 presents mean scores and standard deviations on all tasks for the L1 and L2 children separately.

[Insert Table 1 about here]

A MANOVA with group as the between-subjects factor, SES as a covariate, and all task scores as dependent variables showed a strong effect of group, $F(9,91) = 15.76$, $p < .001$, $\eta^2_p = .61$. SES was a significant covariate in this analysis, $F(9,91) = 3.33$, $p = .001$, $\eta^2_p = .25$. Subsequent ANOVAs with group as the between-subjects factor and SES as a covariate showed no effects of group for the verbal memory measures (all *p*s > .1), but clear effects for the language measures. Specifically, the L1 children scored significantly higher than the L2 children on the two morphology sub-tests, $F(1,105) = 36.37$ $p < .001$, $\eta^2_p = .28$ for noun plurals and $F(1,105) = 41.68$, $p < .001$, $\eta^2_p = .29$ for past participles, as well as on the sentence sub-test, $F(1,105) = 25.57$, $p < .001$, $\eta^2_p = .20$. They also performed significantly better on the Dutch receptive vocabulary task than the L2 children, $F(1,106) = 30.16$, $p < .001$, $\eta^2_p = .23$.

Zero-order Pearson correlations between all measures are presented in Table 2 for the two groups separately. This table also shows correlations with age and SES. The only significant correlation found for age involved listening recall scores in the L1 group. SES correlated moderately and significantly with most of the verbal memory, grammar, and vocabulary measures in the L1 group, but not in the L2 group, where it only correlated significantly with listening recall.

For the variables of interest, there were moderate and significant correlations between most of the memory and the grammar measures for both groups. Vocabulary correlated significantly with almost all variables in the L1 group, but not in the L2 group where significant correlations were only found with SES, word recall, listening recall, and sentence production. Overall, correlations between tasks that are supposed to measure the same construct (i.e., VSTM or VWM) tended to be higher than correlations between tasks that are assumed to measure different constructs.

[Insert Table 2 about here]

**Confirmatory Factor Analysis**

Results of our CFA in which the simple span tasks and complex span tasks were loaded on two factors (representing VSTM and VWM) showed that a multi-group, two-factor model fitted the data well, $\chi^2 = 16.19$, $p > .1$, RMSEA = .05, CFI/TLI = .97/.96. Full measurement invariance was found, indicating that the same measurement model could be used for both groups, with factor loadings and intercepts constrained to equality in the two groups. Figure 1 displays the model. It must be noted that, in this model, only standardized factor loadings are presented. These loadings differ between the groups, presumably due to differences in error variances of the observed variables between the groups. Unstandardized factor loadings that were constrained to be equal between the two groups are not presented.

[Insert Figure 1 about here]

This model shows, first, that all tasks loaded significantly on the two latent factors and second, that there was a significant correlation between the latent factors in both groups. But this correlation was higher in the L1 group than in the L2 group. To test whether this difference was significant, a multi-group model was run which was the same, except that the correlation between the two latent memory factors in the two groups was constrained to equality (rather than freely estimated as in the previous model). This new model fitted the data less well than the previous model, $\chi^2 = 21.90$, $p > .1$, RMSEA = .09, CFI/TLI = .92/.89, as evidenced by a higher AIC-value (2373.4 vs. 2369.7), indicating that the difference between the correlations in the two groups was significant.

**Relationships between Verbal Memory and Vocabulary and Grammar**

A multi-group model predicting children's grammar and vocabulary scores from the latent VSTM and VWM factors with SES as a covariate was fitted to the data. The model, depicted in Figure 2, fitted the data well, $\chi^2 = 40.95$, $p > .1$, RMSEA = .05, CFI/TLI = .98/.96. As above, only standardized factor loadings are presented; unstandardized loadings that were constrained to equality between the two groups are not presented.

[Insert Figure 2 about here]

This model shows that, in both groups, VSTM was a significant predictor of both vocabulary and grammar, whereas VWM was a significant predictor of grammar only. It also shows that SES was significantly related to both memory components in the L1 group. In the L2 group, SES was only related to VWM and even showed a negative relationship with grammar. Finally, the model shows that vocabulary and grammar were not significantly correlated in either of the two groups.

Taken together, these results show very similar predictive relationships between VSTM and VWM, on the one hand, and vocabulary and grammar, on the other hand, in the two groups. While regression coefficients seem to differ across groups with higher coefficients for vocabulary in the L1 group and higher coefficients for grammar in the L2 group, these differences are only apparent (and due to differences in scale variances), because coefficients in the two groups were constrained to be equal in the model.

**Relationships between Verbal Memory and Grammar Sub-Skills**

Finally, a multi-group model was fitted in which we distinguished between morphology and syntactic sub-skills. Specifically, in this model we predicted word-level and sentence-level grammar from the two latent memory factors, after controlling for vocabulary and SES. The model, depicted in Figure 3 below, showed excellent data fit, $\chi^2 = 42.80$, $p > .1$, RMSEA = .02, CFI/TLI = 1.00/.99. As earlier, only standardized factor loadings are given.

[Insert Figure 3 about here]

These results show that relationships with VWM were again very similar for the L1 and L2 children. In both groups, VSTM and VWM were significant predictors of both word-level and sentence-level grammar scores, after controlling for vocabulary and SES. The strengths of these relationships did not differ between groups, just as in the previous analysis, given that all coefficients were constrained to be equal in the two groups.

## Discussion

The current study examined how VSTM and VWM were related to language learning in naturalistic Turkish L2 learners of Dutch and Dutch L1 children. Three questions were addressed: (i) How do VSTM and VWM relate to the acquisition of vocabulary and grammar in naturalistic child L2 learners? (ii) Do relationships with VSTM and VWM differ between L2 and L1 children?, and (iii) Do relationships between the two memory components and grammar differ depending on the type of grammatical sub-skill investigated? Earlier studies on L2 classroom children showed a division of labor such that VSTM was related to L2 vocabulary, while VWM was related to L2 grammar. However, no previous studies have directly compared L1 and L2 children or looked at L2 children who learned their L2 without explicit instruction. In this study, we investigated whether language learning under these more implicit learning conditions would show different involvements of VSTM and VWM than found in earlier studies on L2 children learning their L2 through explicit instruction.

A confirmatory factor analysis first showed that the same latent memory constructs could be observed in both learner groups, with full measurement invariance across groups. This is important as it showed that the memory tasks used measured the same two latent constructs in the L1 and L2 children. The correlation between the two memory components was much stronger in the L1 group than in the L2 group, however. One possible explanation of this difference is that the L2 children's lower linguistic proficiency in Dutch, the language used in the memory tasks, introduced additional variance in these children's task scores that did not reflect mere memory capacity, but also linguistic knowledge (Messer et al., 2015). Future research could explore whether other results are obtained when L2 children are tested in their dominant language or whether correlations between VSTM and VWM in L2 children become stronger with increasing L2 proficiency.

As for our first question, we found that, in the L2 children, VSTM significantly predicted both vocabulary and grammar, whereas VWM significantly predicted grammar, but not vocabulary. The finding that VSTM was related to L2 vocabulary while VWM was important for L2 grammar is in line with results of earlier studies on L2 classroom children (Cheung, 1996; Engel de Abreu & Gathercole, 2012; Kormos & Sáfár, 2008; Masoura & Gathercole, 2005; Speciale, Ellis, & Bywater, 2004). However, our finding that VSTM is important for L2 grammar is more controversial. Although this supports earlier findings for the same sample at an earlier age using less controlled measures of grammar (Verhagen et al., 2015) as well as findings from L2 classroom children (French & O'Brien, 2008), it does not support the findings by Engel de Abreu and Gathercole (2012), who found that VSTM predicted vocabulary but not grammar in multilingual children. There are two possible explanations for these contradictory results across studies. First, the discrepancy might be due to the type of L2 learners studied. As hypothesized in this study, naturalistic L2 learners might rely more strongly on phonological storage of syntactic and morphological patterns than instructed L2 learners who might analyze the incoming input on the basis of L2 metalinguistic knowledge and subsequently store smaller chunks of L2 speech, for which well-developed VSTM skills are less needed. However, this cannot explain why French and O'Brien (2008) found a significant association between VSTM and grammar in L2 classroom learners. A second, alternative explanation, then, is that mixed findings across studies are due to the grammar tests used. While both the current study and French and O'Brien (2008) used relatively 'pure' tests of grammar, other studies used tests assessing a mixture of vocabulary and grammar (French, 2006; Service, 1992; Service & Kohonen, 1995). In the latter case, the lack of direct effects of VSTM on L2 grammar (i.e., independent of L2 vocabulary) may be due to the fact that the grammar tests used also drew on children's vocabulary knowledge. The idea of the current grammar task as a relatively pure measure of grammar receives support from our result that no significant correlation was found between vocabulary and grammar in either group.

As for our second question, we found strikingly similar results for the L1 and L2 children. Specifically, we found that individual differences in vocabulary and grammar were correlated with the same memory processes in the L1 and L2 learners, and that relationships were equally strong in the two groups. The result that individual differences in the L1 group were correlated with differences in VSTM and VWM runs counter the idea that L1 acquisition is mainly been based on innate principles, such as Universal Grammar, and in this sense, differs from L2 acquisition, as has been claimed in mentalist approaches to language acquisition.

The finding that VSTM and VWM affect L1 and L2 learning similarly may also have implications for L2 children with poor verbal memory skills. In such children, a cumulative effect of verbal memory might be found such that children's L2 proficiency is directly affected by poor verbal memory, and indirectly, through reduced L1 knowledge. This reduced L1 knowledge could, itself, also be the result of poor verbal memory skills, and would allow for less positive transfer from the L1 to the L2. Future research is needed to further investigate how L2 children with poor verbal memory skills could be at risk for L2 language delays. Future studies could also address whether the current finding of similar relationships for L1 and L2 children can be generalized to older L2 children, who might apply more analytical, metalinguistic strategies in language learning, and therefore show a stronger reliance on VWM as compared to monolingual children.

With respect to our final question, we found that both VSTM and VWM significantly predicted two different types of grammar skill: knowledge of noun plurals and past participles (word-level grammar) and knowledge of function words and syntactic structures (sentence-level grammar). Again, relationships were very similar for the L1 and L2 children. Contrary to what we had predicted based on work on monolingual children (McDonald, 2008), no differences in the strengths of the relationships were found such that, for example, VWM would be more important for sentence-level grammar than for word-level grammar. Perhaps the fact that noun plurals and past participles also involve information checking – for congruency in number and the dependency between the auxiliary and past participle respectively, might explain why effects were equally strong for both types of structures.

Taken together, the results of this study support earlier studies comparing effects of VSTM and VWM in the same sample, as they show that both components of verbal memory are significantly related to individual differences in vocabulary and grammar learning (Engel de Abreu & Gathercole, 2012; Kormos & Sáfár, 2008). However, in our study, a less strict division of labor was found between both components than in earlier work (Engel de Abreu & Gathercole, 2012), as VSTM was associated with vocabulary and grammar, and VWM was associated with grammar. The finding that VSTM was correlated with learning both words and grammar supports earlier theories assuming that well-developed VSTM skills enable children to store long-term representations of sounds and grammatical templates that help them to acquire new words and grammar (Speidel, 1989; 1993; Adams & Gathercole, 1996; 2000). The finding that VWM was correlated with grammar suggests that children with strong executive processing skills have an advantage in learning new grammatical structures, both at the sentence-level and word-level. Previous work on L2 adults has suggested that effects of VWM may be restricted to grammatical success under explicit learning rather than implicit learning conditions (Tagarelli et al., 2011). However, the current results indicate that VWM is also related to acquiring grammar in situations of implicit learning, at least in L1 and L2 children of a relatively young age.

The study has a few limitations. First and foremost, only concurrent measures were used in a correlational design, so we cannot draw firm conclusions about the causality of the effects. Second, our sentence-level sub-test was a sentence repetition task in which children had to repeat sentences produced by the experimenter. This type of task has also been used as a measure of VSTM (Willis & Gathercole, 2001), casting doubt on its validity as a measure of grammatical proficiency. While we cannot rule out that the observed correlation between this task and the VSTM construct was at least partially due to the sentence repetition test assessing VSTM, we do not think that this can fully explain the relationships found. First, scoring in the task was specifically focused on the use of target structures. Other omissions, additions and changes made to the stimuli were not taken into account, following the standard administration of the task. Second, all sentences were constructed in such a way that they were assumed to be too long to be retained in VSTM, making it unlikely that for some children, storing the whole sentence was possible. Finally, a detailed investigation of the convergent validity of this task has shown that it correlates moderately to strongly with assessments of grammar, narrative skill, and spontaneous speech in four-to-six-year old L1 and L2 children in the Netherlands, suggesting that the task does not assess mere VSTM skill (Verhoeven & Vermeer, 2006, see also footnote 2).

A third limitation of the present study is that sample size was rather low. For CFA and SEM analyses, bigger groups of participants are usually recommended. The fact that there was little missing data and that distributions were not very much skewed allowed us to run multi-group models on the basis of rather small samples and still find good model fit. Importantly, moreover, as there was full measurement equivalence between the two groups in all the models, the final models were based on the whole group of participants rather than the two sub-groups, increasing power in the analyses.

Despite its limitations, we think the current study shows that both components of working memory constitute an important source of individual differences in L1 and L2 acquisition. Taking a latent factor approach, this study is the first to show that the two components of verbal memory are differentially related to individual differences in vocabulary and grammar in the same way in L1 and L2 children. As such, it suggests that L1 and L2 acquisition do not involve fundamentally different processes, at least not when it comes to young L2 children who learn their L2 in a naturalistic setting.

## References

Adams, A. M., & Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology, 49A*, 216-233.

Adams, A. M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language Communication Disorders, 35*, 95-116.

Alloway, T. P. (2007). *Automated Working Memory Assessment*. London: Psychological Corporation.

Alloway, T. P., Gathercole, S. E., Willis, C. S., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology, 87*, 85-170.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuo-spatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698-1716.

Automated Working Memory Assessment (AWMA) – Reliability and Validity (2014). Retrieved June 12, 2014, from http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildMemory/AutomatedWorkingMemoryAssessment(AWMA)/AutomatedWorkingMemoryAssessment(AWMA).aspx

Baddeley, A., & Hitch, G. J. (1974). Working memory. *The Psychology of Learning and Motivation, 8*, 47-89.

Baddeley, A. D., & Logie, R. H. (1999*)* Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.) *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158-173.

Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development, 17*, 91-107.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*, 3-49.

Blom, E., Küntay, A. K., Messer, M. H., Verhagen, J., & Leseman. P. P. M. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology, 128,* 105-119.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867-873.

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology*, *17,* 367-378.

Ellis, N. C., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax. *Quarterly Journal of Experimental Psychology, 49,* 234-250.

Ellis Weismer, S., Evans, J., & Hesketh, L. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language and Hearing Research, 42,* 1249-1260.

Enders, C. (2010). *Applied missing data analysis.* The Guilford Press.

Engel de Abreu, P. M. J., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology, 104,* 974-986.

French, L. M. (2006). *Phonological working memory and second language acquisition: A developmental study of Francophone children learning English in Quebec*. Lewiston, NY: Edwin Mellen Press.

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29,* 463-487.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513-543.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887-898.

Gottardo, A., Stanovich, K., & Siegel, L. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology, 63*, 563-582.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.

Kline, R. B. (2005). *Principles and practices of structural equation modeling*. New York: The Guilford Press.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*, 261-271.

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In Cristina Sanz & Ronald P. Leow (Eds.), *Implicit and explicit language learning : Conditions, processes, and knowledge in SLA and bilingualism* (pp. 101-113). Washington, DC: Georgetown University Press.

Mackey, A., Philp, J., Fuji, A., Egi, T., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181-208). Philadelphia, PA: John Benjamins.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*, 379-413.

Masoura, E. V., & Gathercole, S. E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory*, *13*, 422-429.

McDonald, J. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language*, *35*, 247-268.

Messer, M. H. (2010). Verbal short-term memory and vocabulary development in monolingual Dutch and bilingual Turkish-Dutch preschoolers. Unpublished doctoral dissertation, Utrecht University, the Netherlands.

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, *105*, 306-323.

Messer, M. H., Verhagen, J., Boom, J., Mayo, A. Y., & Leseman, P. P. M. (2015).

Growth of verbal short-term memory of nonwords varying in phonotactic probability: A longitudinal study with monolingual and bilingual children. *Journal of Memory and Language, 84,* 24-36.

Montgomery, J. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research, 38,* 177-189.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén and Muthén.

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of mono- and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31*, 117-140.

Service, E. (1992). Phonology, working memory, and foreign language learning. *Quarterly Journal of Experimental Psychology, 45A,* 21-50.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16,* 155-172.

Speciale, G., Ellis N., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*, 293-320.

Speidel, G. E. (1989). A biological basis for individual differences in learning to speak. In G. E. Speidel & K. E. Nelson (Eds.), *The many faces of imitation in language learning* (pp. 199-229). New York: Springer Verlag.

Speidel, G. E. (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study. *First Language, 13*, 69-91.

Speidel, G. E., & Herreshoff, M. J. (1989). Imitation and the construction of long utterances. In G. E. Speidel & K. E. Nelson, (Eds.), *The many faces of imitation in language learning* (pp. 181-197). New York: Springer-Verlag.

Sunderman, G., & Kroll, J. F. (2009) Sunderman G, Kroll JF. When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, *30*, 79-99.

Verhagen, J., Messer, M. H., & Leseman, P. P. M. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Language Learning*, *65*, 417-448.

Verhoeven, L., & Vermeer, A. (2002). *Taaltoets Alle Kinderen*. [Language Test for All Children]. Arnhem: Cito.

Verhoeven, L., & Vermeer, A. (2006). *Verantwoording Taaltoets Alle Kinderen* [Language Test for All Children]. Arnhem: Cito B.V.

Verhoeven, L., Narain, G., Extra, G., Konak, O. A., & Zerrouk, R. (1995). *Toets Tweetaligheid* [Test for Bilingual Development]. Arnhem, the Netherlands: Cito.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, *53*, 67-121.

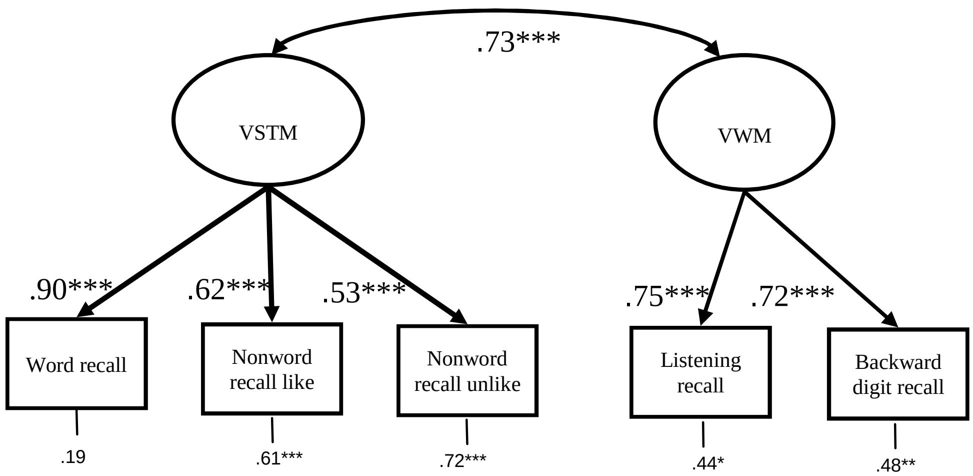
Willis, C. S., & Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, *9*, 349-364.

Highlights

- We investigate relationships between verbal working memory and vocabulary/grammar
- We compare first language (L1) and second language (L2) learning children
- We use structural equation modeling, treating phonological and working memory as latent factors
- Results: phonological memory predicts vocabulary and grammar, while working memory predicts grammar only
- These relationships hold for both groups, and after controlling for SES

**Figure**

L1 children:




L2 children:

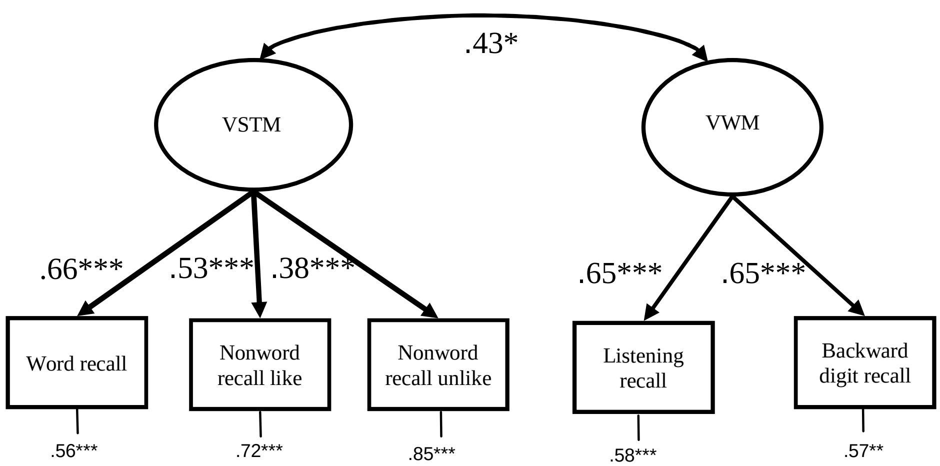
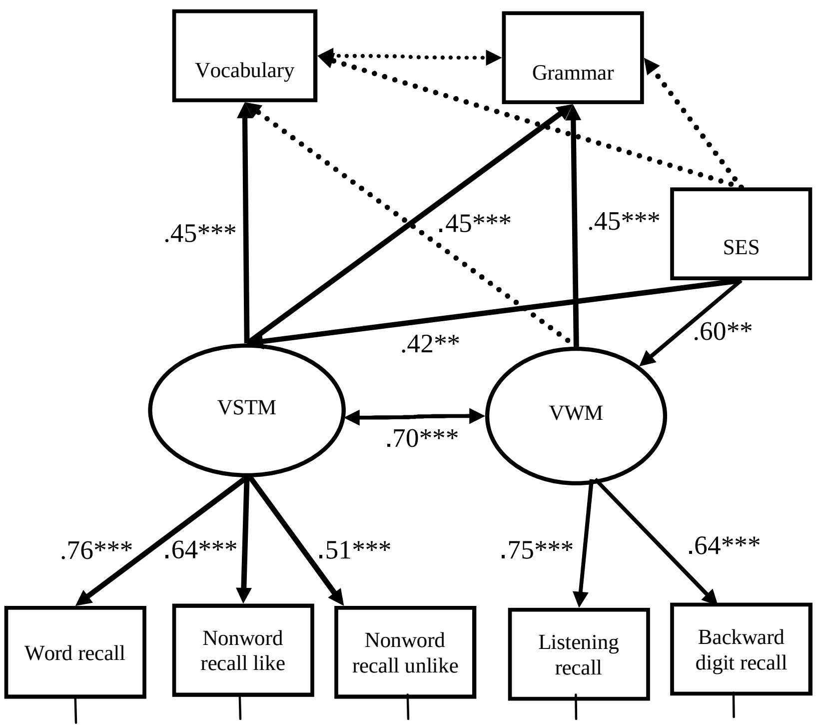



*Figure 1.* Results of CFA: Multi-group two-factor model of simple and complex memory span tasks representing a VSTM and a VWM latent factor.* $p < .05$, ** $p < .01$, *** $p < .001$.

L1 children:




L2 children:

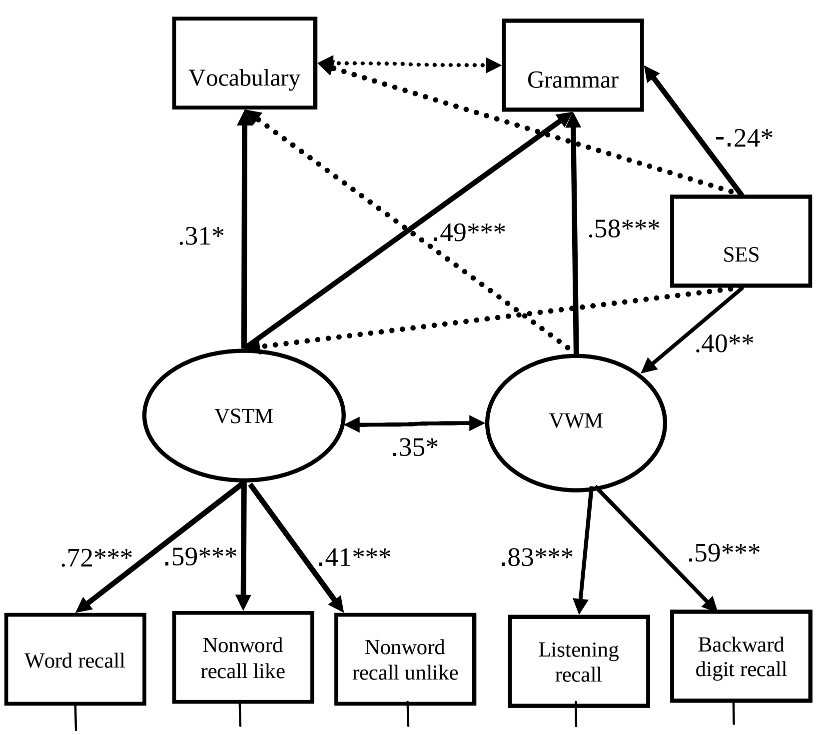
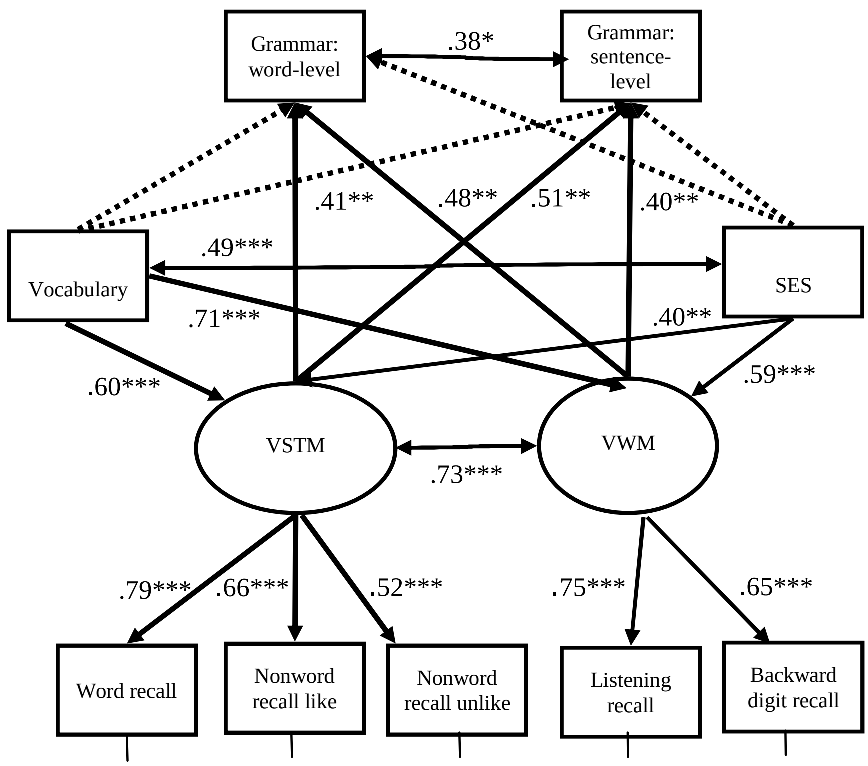


*Figure 2.* Multi-group model predicting vocabulary and grammar from the latent VSTM and VWM factors. Non-significant coefficients (*p* > .05) are presented as dotted lines. * *p* < .05, ** *p* < .01, *** *p* < .001.

L1 children:


Grammar: word-level
.38*
Grammar: sentence-level
.41**
.48**
.51**
.40**
Vocabulary
.49***
SES
.71***
.40**
.60***
.59***
VSTM
VWM
.73***
.79***
.66***
.52***
.75***
.65***
Word recall
Nonword recall like
Nonword recall unlike
Listening recall
Backward digit recall

L2 children:

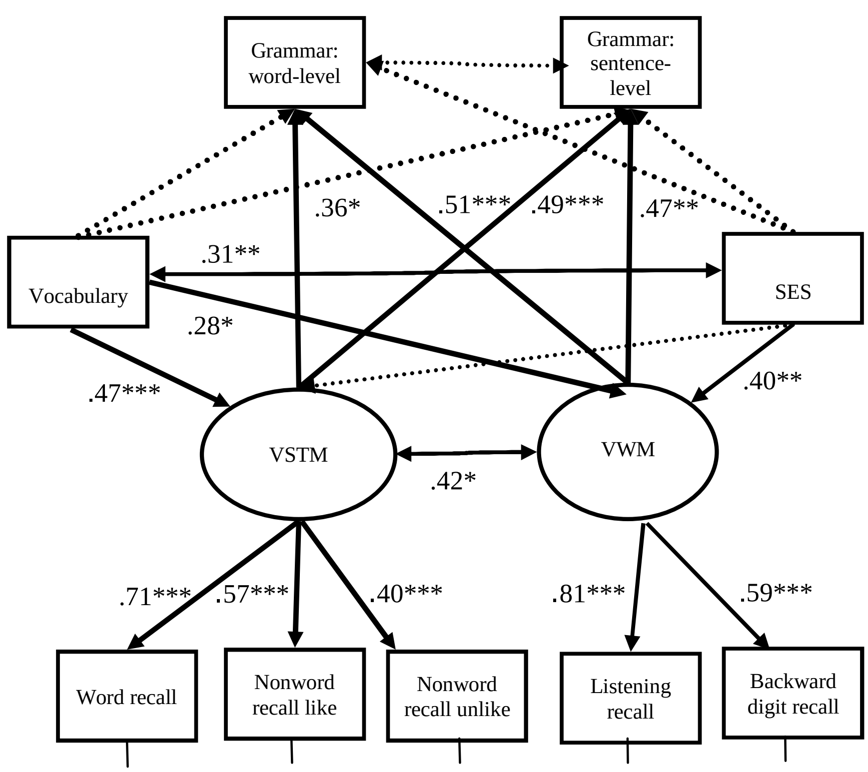

Grammar: word-level
Grammar: sentence-level
.36*
.51***
.49***
.47**
Vocabulary
.31**
SES
.28*
.47***
.40**
VSTM
VWM
.42*
.71***
.57***
.40***
.81***
.59***
Word recall
Nonword recall like
Nonword recall unlike
Listening recall
Backward digit recall

*Figure 2.* Multi-group model predicting word- and sentence-level grammar from the latent memory factors. Non-significant coefficients ($p > .05$) are presented as dotted lines.* $p < .05$, ** $p < .01$, *** $p < .001$.

**Table**

Table 1.

*Descriptive Statistics for All Tasks*

| | L1 Children | | | | L2 Children | | | |
|---|---|---|---|---|---|---|---|---|
| | M | SD | Range | N | M | SD | Range | N |
| *Verbal memory: Simple span* | | | | | | | | |
| Word recall | 16.62 | 4.08 | 8-27 | 45 | 14.79 | 3.37 | 6-23 | 63 |
| Dutch-like nonword recall | 4.69 | 2.02 | 0-9 | 45 | 4.23 | 1.57 | 1-9 | 61 |
| Dutch-unlike nonword recall | 2.16 | 1.33 | 0-6 | 45 | 2.28 | 1.16 | 0-5 | 61 |
| *Verbal memory: Complex span* | | | | | | | | |
| Backward digit recall | 3.64 | 3.01 | 0-10 | 45 | 3.13 | 2.54 | 0-9 | 63 |
| Listening recall | 4.98 | 3.29 | 0-12 | 45 | 4.38 | 2.80 | 0-13 | 63 |
| *Grammar* | | | | | | | | |
| TAK plural | 7.60 | 1.947 | 0-11 | 45 | 3.23 | 2.88 | 0-9 | 62 |
| TAK past participles | 6.67 | 3.261 | 1-12 | 45 | 1.81 | 1.74 | 0-7 | 62 |
| TAK sentence production | 12.00 | 4.452 | 3-19 | 45 | 5.68 | 3.25 | 0-13 | 62 |
| *Vocabulary* | | | | | | | | |
| Dutch vocabulary | 24.70 | 3.75 | 17-30 | 45 | 17.42 | 3.98 | 10-25 | 63 |

Table 2.

*Bivariate Correlations among All Variables for the L1 and L2 Children*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Background variables* | | | | | | | | | | | |
| 1. Age | - | .16 | .12 | -.03 | **.45** | .16 | **.49** | .18 | .06 | .02 | .19 |
| 2. SES | -.11 | - | **.32** | .28 | .19 | **.45** | **.40** | .18 | **.37** | **.41** | **.49** |
| *Verbal memory* | | | | | | | | | | | |
| *Simple span* | | | | | | | | | | | |
| 3. Word recall | -.11 | .17 | - | **.53** | **.52** | **.54** | **.45** | .25 | **.45** | **.49** | **.46** |
| 4. Nonword recall (like) | .12 | .17 | **.35** | - | **.39** | .22 | **.32** | **.36** | **.55** | **.55** | **.35** |
| 5. Nonword recall (unlike) | -.03 | .06 | .20 | **.41** | - | **.31** | .29 | **.45** | **.41** | **.39** | **.46** |
| *Complex span* | | | | | | | | | | | |
| 6. Backward digit recall | .19 | .14 | .10 | .08 | .21 | - | **.54** | **.34** | **.51** | **.44** | **.52** |
| 7. Listening recall | .07 | **.36** | .23 | .22 | .06 | **.42** | - | .26 | **.51** | **.47** | **.50** |
| *Grammar* | | | | | | | | | | | |
| 8. TAK plural | -.07 | .06 | **.36** | .19 | .14 | **.32** | **.44** | - | **.45** | **.46** | **.39** |
| 9. TAK past participles | .03 | .04 | **.41** | **.40** | .09 | .07 | **.37** | **.33** | - | **.67** | **.52** |
| 10. TAK sentence prod. | -.15 | .15 | **.49** | **.30** | .17 | **.32** | **.50** | **.39** | **.42** | - | **.38** |
| *Vocabulary* | | | | | | | | | | | |
| 11. Dutch vocabulary | -.08 | **.32** | **.38** | .18 | .19 | .07 | **.26** | .22 | .18 | **.33** | - |

*Note*. Correlation coefficients for the L1 children are presented in the upper triangle; coefficients for the L2 children are presented in the lower triangle. Correlations in bold-face are significant at $p < .05$.