The role of memory in the acquisition of vocabulary and grammar in the first language and in English as a foreign language

Van Koert, M.J.H.; Leona, N.L.; Rispens, J.E.; Tijms, J.; Van der Molen, M.W.; van Daal, V.H.P.; Snellings, P.

DOI
10.1515/9783110743043-003

Publication date
2022

Document Version
Final published version

Published in
Second Language Learning Before Adulthood

Citation for published version (APA):
Van Koert, M. J. H., Leona, N. L., Rispens, J. E., Tijms, J., Van der Molen, M. W., van Daal, V. H. P., & Snellings, P. (2022). The role of memory in the acquisition of vocabulary and grammar in the first language and in English as a foreign language. In V. De Wilde, & C. Goriot (Eds.), *Second Language Learning Before Adulthood: Individual Differences in Children and Adolescents* (pp. 31-64). (Studies on Language Acquisition ; Vol. 65). De Gruyter Mouton. Advance online publication. https://doi.org/10.1515/9783110743043-003

M. J. H. Van Koert, N. L. Leona, J. E. Rispens, J. Tijms,
M. W. Van der Molen, V. H. P. van Daal and P. Snellings

The role of memory in the acquisition of vocabulary and grammar in the first language and in English as a foreign language

Abstract: Previous studies showed that phonological short-term and working memory spans are related to vocabulary and grammar learning in children learning a second language. Typically, short-term storage, as measured by simple span tasks such as non-word repetition, are connected to vocabulary learning. Grammar learning is generally linked to the working memory system. This system is often tested by complex span tasks that require participants to process and store information simultaneously. Yet, few studies have investigated the role of both memory mechanisms in native and foreign vocabulary and grammar learning longitudinally. The current study determines whether phonological short-term and verbal working memory spans contribute differentially to the acquisition of vocabulary and grammar in Dutch as a first language (L1) and in English as a foreign language (EFL). The participants for this study are monolingual Dutch children (N = 138), in grades 4 and 5 (aged 9;0–11;0), learning EFL in the classroom. An L1 and an EFL non-word repetition task were used to measure phonological short-term memory and verbal working memory was measured with a backward digit span task. Receptive vocabulary and production of grammatical knowledge was measured in Dutch and in English, as well as receptive grammar in English. The data indicated that when the same children are longitudinally followed in both L1 and EFL, only past performance is important for L1 vocabulary learning. Phonological short-term memory does not contribute to L1 vocabulary. As expected, working memory span significantly predicts L1 grammar learning. For EFL vocabulary learning, past performance is most important; L1 vocabulary has a smaller but independent role and in line with previous research, phonological short-term memory also has an independent role. For both receptive and productive EFL grammar learning, contrary to expectations, working memory span did not play a role. In contrast, EFL phonological short-term memory had a small but independent role for receptive grammar learning. EFL vocabulary had a similar role for receptive and productive grammar, which was stronger than EFL phonological short-term memory for receptive grammar. However, past performance had the largest role

https://doi.org/10.1515/9783110743043-003

for both types of grammar learning. In sum, when looking at the initial stages of EFL in a formal setting, we only found a role for working memory span in L1 grammar but not in EFL. Phonological short-term memory only had a role in EFL vocabulary and receptive grammar learning, and was language specific. In all, the current data show that in addition to phonological short-term memory, past performance on vocabulary and grammar contributed significantly to vocabulary and grammar learning and that the role of vocabulary and phonological short-term memory and working memory span in grammar learning is language specific.

Keywords: EFL, early second language acquisition, phonological processing, vocabulary, grammar, working memory

1 Introduction

Verbal working memory can be divided into a domain specific short term storage component, also known as the phonological loop, and a domain general component that facilitates both processing and storing verbal information (Baddeley and Hitch 1974; Adams and Gathercole 2000). With regard to language learning most research has focused on the phonological loop, and many studies have also focused on the structure of working memory. The phonological loop represents phonological short-term memory for verbal information and the current study will determine its role in the development of vocabulary and grammar in children. A number of studies have focused on the role of phonological short-term memory in the L1 as a measure of the domain specific short term storage component of verbal working memory. Either studying its effect on vocabulary acquisition (Gathercole et al. 1992), on grammar (Boersma et al. 2018), or on both L1 vocabulary and grammar (Adams and Gathercole 2000). Other studies have focused on phonological short term memory in FL learning, either the effects on vocabulary learning (Cheung 1996; Masoura and Gathercole 2005; Messer et al. 2010), on FL grammar (O'Brien et al. 2006), or on both (Ellis and Sinclair 1996; Paradis 2011). Only few studies (Engel de Abreu and Gathercole 2012) focused on both phonological short-term memory and working memory span, the latter as a measure of the domain general component of verbal working memory. Engel de Abreu and Gathercole (2012) studied only effects on vocabulary, in both L1 and FL. Kormos and Sáfár (2008) studied effects on both vocabulary and grammar, but only in the FL.

In general, working memory span is associated with grammatical proficiency and phonological short-term memory is related to vocabulary knowledge.

However, some studies have also found vocabulary knowledge to be related to working memory span and grammatical proficiency to phonological short-term memory (Archibald and Gathercole 2006; Ellis and Sinclair 1996; Kormos and Sáfár 2008; Martin and Ellis 2012; O'Brien et al. 2006; Paradis 2011; Service and Kohonen 1995). The present study aims to determine whether individual differences in short-term memory and working memory span predict individual differences one year later in both receptive vocabulary knowledge and grammatical proficiency in the first language (L1) and in the initial stages of English as a foreign language (EFL).

1.1 Memory and vocabulary

There are many different ways to measure the span of the phonological loop storage, as it can be quantified by means of serial recall of (non-)words, letters or digits or by non-word repetition (NWR) (Archibald and Gathercole 2006; Crannell and Parrish 1957). Although many studies lump together NWR with other verbal short-term memory tasks (Archibald and Gathercole 2006; Engel de Abreu and Gathercole 2012; Kormos and Sáfár 2008; Paradis 2011; Verhagen and Leseman 2016), there are also studies that argue that they are not identical (O'Brien et al. 2006; Rispens and Baker 2012; Rispens and de Bree 2017). Verbal short-term memory represents phonological processing in the phonological loop. The longer the verbal material is that needs to be retained, the more difficult it is, because the memory traces fade so quickly in the phonological store. Besides, the articulatory rehearsal process is challenged by long verbal material (Baddeley 2003). For example, adults can typically recall a list of seven digits but remembering nine digits is already a more difficult task and retaining fifteen is highly unusual (Crannell and Parrish 1957). Hence, digit span reflects the nature of verbal short-term memory, namely temporary storage. NWR is apart from length of the non-word also susceptible to other factors, such as phonotactic frequency, syllable frequency and wordlikeness, i.e. how much does the non-word look like an existing word (Messer et al. 2010; O'Brien et al. 2006; Rispens and Baker 2012). In fact, Rispens and Baker (2012) found that phonological short-term memory and phonological representations contribute to NWR ability. Thus, there are long-term memory effects that trickle through in NWR (but note that simple span tasks like repetition of existing words are also influenced by long term memory representations, arguably because of chunking, see Jones (2012)). Both digit recall and NWR have been found to be related to vocabulary knowledge (Archibald and Gathercole 2006; Paradis 2011) but their separate contribution to vocabulary knowledge and grammatical performance has – to our knowledge – not been assessed in a sample of young FL learners in an instructed foreign language learning setting. NWR thus measures phonological short-term memory, but as it also depends on phonological processes, it is better described as a task that measures phonological processing, which involves short term memory, but also more long term knowledge.

NWR can be administered in the L1 and/or in the FL. Studies on children learning an FL have shown that they are better at repeating non-words in the language they speak best, underlining the relationship between NWR and language knowledge (Masoura and Gathercole 2005; Messer et al. 2010; Rispens and de Bree 2017). Previous studies that administered NWR in L1 found significant correlations between NWR and FL vocabulary knowledge (Engel de Abreu and Gathercole 2012; Kormos and Sáfár 2008). Prior research that included an NWR in the FL also found significant correlations between NWR and FL vocabulary knowledge (Cheung 1996; Service and Kohonen 1995). Since NWR reflects the storage of long-term phonological representations, it is likely that FL NWR is more tightly related to FL vocabulary than L1 NWR. To our knowledge, only one study, Masoura and Gathercole (2005), investigated the contributions of L1 NWR and FL NWR to FL vocabulary. In that study Greek children aged between 8;6 and 13;7 participated; they were learning English as an FL. The English vocabulary tests consisted of translation assignments, whereby English words had to be translated to Greek and vice versa. Although the focus of the study was on the relative contribution of phonological memory for participants with a large vocabulary and those with a small vocabulary, the correlations between English NWR and the English vocabulary tests are somewhat stronger than those between Greek NWR and English vocabulary tests. The current study aims to compare the relative contributions of L1 and FL NWR to investigate which of the two is a better predictor of FL vocabulary.

Several studies have found that the strength of the relationship between NWR and vocabulary diminishes as the size of the lexicon increases (Cheung 1996; Gathercole et al. 1992; Rispens and Baker 2012). These findings imply that phonological processing is central in the initial stages of vocabulary learning, but when the lexicon reaches a certain threshold, new words seem to be learned on the basis of existing long-term lexical knowledge rather than on the basis of short-term phonological storage. For example, the correlations between L1 NWR and L1 vocabulary knowledge were stronger at age 4 than at age 8 in Gathercole et al.'s (1992) study. Second language acquisition studies found that FL NWR correlated more strongly in learners with low foreign language proficiency than in those with high foreign language proficiency (Cheung 1996; Engel de Abreu and Gathercole 2012; Kormos and Sáfár 2008). The goal of the present study is to investigate the relative contributions of L1 and FL NWR in L1 vocabulary and FL vocabulary. The group of interest are Dutch ten-year-old children learning English as a foreign language. Based on the literature, we expect that FL NWR will contribute to FL vocabulary. However, the literature is less clear on whether L1 NWR also contributes to L1 vocabulary.

1.2 Memory and grammatical performance

Although NWR is typically related to vocabulary acquisition, some studies have also found an influence of NWR on grammatical performance (Archibald and Gathercole 2006; Ellis and Sinclair 1996; Kormos and Sáfár 2008; O'Brien et al. 2006; Paradis 2011; Service and Kohonen 1995). For instance, Service and Kohonen (1995) tested nine-year-old Finnish children by means of an English (as an FL) NWR. Three years later the same children had to complete several tests measuring their knowledge of English, including listening, reading, essay writing and vocabulary. The results showed that NWR was strongly correlated with all of the English tests. Later studies found differences between beginning and more advanced learners. For instance, Kormos and Sáfár (2008) tested Hungarian 16-year-olds on their L1 NWR and working memory span skills and also examined their English (as an FL) grammatical performance. They found that there was a difference between the beginning and the more advanced English language learners: whereas the beginning learners showed no effect of NWR, the more advanced learners did show a correlation between NWR and FL grammar scores. The interpretation of this finding is somewhat complicated, as the grammar scores came from a test which assessed lexical and grammatical knowledge in an integrated manner (Kormos and Sáfár 2008: 268). Kormos and Sáfár assumed that only the lexical retrieval skills differed between the more advanced students and that their grammatical processing skills were similar; this could then explain why an effect of NWR was found in their grammatical performance, as it actually reflected their lexical skills. Another example is the study by O'Brien et al. (2006) who argued that in the initial stages FL learners tend to focus on content words whilst skipping function words. They stated that meaning is more important for beginning FL learners and focusing on meaning consumes most of the working memory space. As FL learners become increasingly more proficient, lexical retrieval becomes easier, making it possible for learners to process words that do not carry meaning, such as function words and grammatical markers. O'Brien et al. (2006) found a relationship between L1 NWR and function words, i.e. grammatical performance, for the more advanced learners, for whom lexical access did not pose any difficulties. Although the present study does not differentiate between beginning and more advanced learners, the contribution of NWR to grammatical performance is assessed.

Grammatical performance is most often linked to the domain general component of working memory (Archibald and Gathercole 2006; Ellis and Sinclair 1996; Martin and Ellis 2012; Verhagen and Leseman 2016). It has been found that working memory-impaired children struggle with parsing and analysing linguistic constructions in their L1 (Marton and Schwartz 2003). Since working memory is linked to noticing and processing new information, it has been argued that working memory is a mechanism to detect new structures in an FL and extract rules from those structures. Grammar is exactly where new structures appear and various studies have found that working memory span influences grammatical performance. For example, Verhagen and Leseman (2016) compared monolingual Dutch children to Dutch-Turkish bilingual children, who had learned Dutch in a naturalistic setting, on their domain general working memory skills in relation to their grammar knowledge in Dutch. Working memory was measured with a backward digit recall task. The grammar task included production of morphologically complex items and sentence repetition. They found no differences between these two groups of 5-year-old children regarding the relationship between working memory and grammar knowledge. For both the monolinguals and the bilinguals, working memory predicted grammar, whereas short-term memory predicted vocabulary and grammar. Verhagen and Leseman (2016) concluded that the same memory mechanisms are employed in L1 and L2 children learning their L2 naturalistically. Another example is a study by Martin and Ellis (2012), who investigated university students implicitly learning an artificial language with intricate morphological rules. They found a robust relationship between vocabulary and grammar scores, suggesting a strong interdependence. They also found that there was a stronger relationship between working memory span and grammar than between short term memory and grammar, implying that the memory mechanisms employed for memorizing items, i.e. vocabulary, are different from abstracting patterns, i.e. grammar. The current study aims to investigate the contribution of working memory span to grammatical performance in both the L1 and in the FL.

Whereas verbal short-term memory and NWR are usually connected with vocabulary learning, working memory span is typically associated with grammar learning. However, the findings of previous studies are inconclusive: both vocabulary and grammar learning can be affected by working memory capacity and NWR skills. The participant's language proficiency level seems to play a role as well. The present study includes verbal short-term memory, NWR and working memory span to determine the effects on L1 vocabulary and grammar and FL vocabulary grammar. Since our study was part of a longitudinal project, all of the predictors were measured a year prior to the outcome measures. To control for past performance, the past value of the outcome variable of interest was also entered into the analyses. Thus we could determine whether the memory measures had any effect on top of past vocabulary and grammatical performance in the L1 and in the FL.

2 The current study

The main research question of the present study was whether phonological short term memory as measured with NWR tasks and domain general working memory as measured with a working memory span task contribute to vocabulary and grammar performance in Dutch (L1) and English as a foreign language (EFL) a year later. We asked two specific research questions:

1. Is the relative contribution of NWR measured in grade 4 taking into account vocabulary performance a year prior significant for vocabulary knowledge measured in grade 5?
2. Is the relative contribution of working memory span measured in grade 4 taking into account grammatical performance a year prior significant for the grammatical performance displayed in grade 5?

We hypothesised that NWR is correlated with and predicts receptive vocabulary knowledge, while working memory span is correlated with and predicts grammatical proficiency. NWR measures the capacity to store verbal information and gives an indication of the strength of long-term phonological representations. In addition, previous research indicated strong connections between phonological processing and FL vocabulary learning. Working memory span measures the ability to process verbal information while it is being stored. It is involved when learners have to notice and process linguistic structures; hence, its importance for grammar learning. Grammar learning is a complex process whereby individual items need to be memorized and the relationships between those items need to be recognized. The first step in the current study was to check whether both predictions were verified for Dutch. The next step was to determine whether these different memory components were differentially employed for English vocabulary learning versus English grammar learning, too.

3 Method

3.1 Participants

Two hundred and ninety-eight grade 4 pupils from seven primary schools in the Randstad area of the Netherlands took part in the first wave of data collection. Two hundred and eight of those were assessed a year later in the 5^{th} grade. Parental permission was obtained for each participant. The study was approved by the Ethical Review Board of the university. The participating schools had different pupil populations regarding social-economic status based on their postal codes, and regarding the number of children with Dutch as an additional language based on the answers the participants gave in the language background questionnaires. These seven schools formed a representative sample of the primary school population in the Netherlands. Four of the seven schools offered English language lessons before grade 4, but no differences on the crucial measures were found between the participants who had received English language lessons and those who had not.[1] In the present study, the data from dyslexic children (N = 16) were excluded, as were the data from children who had a mother tongue other than Dutch (N = 123). Some pupils did not complete all tasks in this study and those data were excluded too (N = 7). Finally, some pupils changed to another school between grade 4 and grade 5 or their permission was withdrawn; therefore, they did not participate in the second round of data collection. Those data were also excluded (N = 6). Hence, the total number of participants was 138 pupils. There were 69 male and 69 female participants. The mean age of the participants was 9 years and 10 months in grade 4 and 10 years and 10 months in grade 5.

3.2 Instruments

All tasks were administered on a tablet, rather than the usual paper and pencil set-up (see Procedure).

1 English as a foreign language is a compulsory subject from grade 5 (ages 10–11 years) onwards in the Netherlands (Toorenburg and van Oostdam 2002). However, primary schools may decide to offer English language lessons from an earlier grade onwards. Of the current 138 participants, 99 pupils received English lessons at school and 39 pupils did not in grade 4. All of the participants received English lessons at school in grade 5. No significant differences were found on the tasks in this study between the pupils who had received English lessons at schools and those who had not.

Non-verbal intelligence. The measure of the non-verbal intelligence was the RAVEN Standard Progressive Matrices (RAVEN-NL SPM, Dutch adaptation by Harcourt Assessment 2006). The RAVEN consists of five blocks, each containing 12 items ($N_{\mathrm{RAVEN}} = 60$). Each item consists of a geometric pattern with one piece missing. Participants have to discover the principles, relationships and rules between the patterns and decide which of six given pieces is the missing one. A test-retest reliability of .88 was reported for the test (Raven 2003). This task was used to measure participants' logical reasoning skills. Participants had 30 minutes to complete the test (Hamel and Schmittman 2006), instead of an unlimited test situation.

Verbal short-term memory. The measure of verbal short-term memory was the digit span forwards from the Wechsler Intelligence Scale for Children (WISC-III-NL, Dutch adaptation by Kort et al. 2005). In the digit span forward, lists of increasing length are presented. Two of each length (2–8). Each digit occurs only once in a list. The lists were pre-recorded by a female native speaker of Dutch. Participants were asked to recall the lists in correct serial order. The test started with a list of two digits and had a maximum of eight digits. At each length two lists were presented separately. Testing was finished when a participant failed to recall both lists of the same length correctly. Each correctly recalled list was awarded with one point.

NWR. There were two measures of non-word repetition: L1 non-word repetition and L2 non-word repetition. The Dutch (L1) non-word repetition (Rispens and Baker 2012) consisted of three trials and 22 items. The non-words were based on Dutch phonotactics. Item length varied between three and five syllables. All syllables had to be repeated correctly for a point to be awarded. The non-words were pre-recorded by a female native speaker of Dutch. The English (FL) non-word repetition was from the Comprehensive Test of Phonological Processing – Second Edition (CTOPP-2, Wagner et al. 2013). It consisted of three trials and 18 items. The non-words were based on English phonotactics. Item length varied between one and six syllables. The items increased in difficulty. All syllables had to be repeated correctly for a point to be awarded. The non-words were pre-recorded by a female native speaker of American English.[2]

Working memory span. The measure of working memory span was the digit span backwards from the Wechsler Intelligence Scale for Children (WISC-III-NL, Dutch adaptation by Kort et al. 2005). It is identical to the digit span

2 The non-word repetition tasks were rated by an independent rater after the tests were administered.

forwards, except that the participant is required to recall the sequence of spoken digits in reverse order and that there is a maximum of seven digits. A trial item is given in order to ensure the participant understands the concept of reverse recall.

L1 vocabulary. The measure of L1 vocabulary was the Peabody Picture Vocabulary Test, Third edition (PPVT-III-NL, Dunn and Dunn 1997; Dutch translation by Schlichting 2005). The test provides an estimate of receptive vocabulary for standard Dutch. Participants saw four yellow and blue coloured pictures and heard a sound file containing the target word that was played automatically. The sound file was recorded by a female native speaker of Dutch. Participants had to select the correct picture, out of four pictures, upon hearing the target word. If they did not respond within five seconds, they received a warning message saying *select a picture*. This test was administered in class, all of the participants completed sets 7 up and including 11 in grade 4 ($N_{\text{PPVT-NLgr4}} = 59$) and sets 8 up and including 12 in grade 5 ($N_{\text{PPVT-NLgr5}} = 59$), regardless of how many mistakes they made in a set.[3] These sets were selected, as they were deemed appropriate for these age groups, whereas normally the starting set is adjusted for each individual.

L1 sentence assembly. L1 sentence production was examined by using the task Sentence Assembly from the Clinical Evaluation of Language Fundamentals (CELF-4-NL, Wiig, Semel and Secord 2003; Dutch adaptation by Kort, Schittekatte and Compaan 2010). There were two trial items and 13 exercises ($N_{\text{CELF-ZS}} = 13$). This test was administered to assess the participants' ability to formulate grammatically acceptable and semantically meaningful sentences by manipulating and transforming given words and word groups. This task was used to assess L1 grammar knowledge in previous studies (Justice et al. 2010; Rescorla 2002; Weber-Fox and Neville 1996). Participants saw word groups on the screen of their tablet and had to type in two grammatically correct sentences for each item. If both sentences were correct, the participant received one point for that item. Secondly, since the CELF-manual allowed very few sentences to count as correct, we listed all the grammatically correct and semantically meaningful versions of the sentences and manually judged those to be correct. Typos, punctuation and spelling errors were ignored.

3 Originally there were 60 items in sets 7–11 of the PPVT-III-NL and all of these items were administered. However, since one item (*groente* 'vegetable') was also used in the English version of the PPVT, it was excluded from the analysis. That item did not appear in sets 8–12; yet, another item (*globe* 'globe') was also used in the English version of the PPVT. Therefore, it was excluded from the analysis.

FL vocabulary. The measure of English (FL) vocabulary was the Peabody Picture Vocabulary Test, Fourth edition (PPVT-4, Dunn and Dunn 2007). The test measures receptive vocabulary knowledge. Participants were presented with four coloured pictures and heard a sound file containing the target word that was played automatically. The sound file was recorded by a female native speaker of British English. Participants had to select the correct picture, out of four pictures, upon hearing the target word. If they did not respond within five seconds, they received a warning message saying *select a picture*. They could play the sound file up to three times. A difference from standard administration was that all of the participants completed sets 1 up and including 7 in grade 4 ($N_{\text{PPVT-ENgr4}} = 61$) and sets 2 up and including 8 in grade 5 ($N_{\text{PPVT-ENgr5}} = 58$), regardless of how many mistakes they made in a set, because the test was administered in class.[4] These sets were selected, as they were deemed appropriate for these age groups.

FL oral comprehension. English (FL) oral grammar comprehension was examined by using the task Sentence Structure from the Clinical Evaluation of Language Fundamentals (CELF-4, Wiig, Semel and Secord 2003). There were three trial items and 26 test sentences ($N_{\text{CELF-SS}} = 26$). The test was administered to assess the participants' ability to interpret spoken sentences and to select the pictures that illustrate referential meaning of the sentences. This test was used to assess English L1 grammar knowledge in previous studies (Justice et al. 2010; Rescorla 2002; Weber-Fox and Neville 1996). Participants were presented with four coloured pictures and heard a sound file containing the target sentence that was played automatically. The sound file was recorded by a female native speaker of British English. Participants had to select the correct picture, out of four pictures, upon hearing the target sentence.

FL morpho-syntax. Word Structure from the Clinical Evaluation of Language Fundamentals (CELF-4, Wiig, Semel and Secord 2003) was used to examine (English) FL grammar production. It was administered to measure the participants' ability to apply morphological rules to mark inflection and comparison

4 Originally there were 84 items in sets 1–7 and sets 2–8 of the PPVT 4th edition and all of these items were administered. However, since two items (*vegetable* and *globe*) were also used in the Dutch version of the PPVT and since other items were deemed Dutch cognates (*ball, foot, banana, cup, bus, cookie, dancing, lamp, penguin, net, tunnel, envelope, calendar, panda, vest, cactus, uniform, gigantic, group, chef flamingo, hyena, river, timer, vase, harp, bloom, heart*), they were excluded from the analysis. The authors independently rated the cognates and then came to an agreement. Baird, Palacios and Kibler (2016: 448) indicate that "cognates [are] words that are semantically and phonologically or orthographically similar in two languages". In the present study, we considered English and Dutch words to be cognates, when a concept is very close in its pronunciation in the two languages (for example *ball* – *bal*).

and their ability to select and use appropriate pronouns to refer to people. This test was used to examine English L1 grammar knowledge in previous studies (Justice et al. 2010; Rescorla 2002; Weber-Fox and Neville 1996). There were 32 items in the original test, but we chose to simplify it, on the basis of a pilot study, by reducing the number of items to 12 ($N_{\text{CELF-WS}}$ = 12). There were two trial items. Participants saw the picture on a laptop screen. Then a sound file containing the target sentence was played. The sound file was recorded by a female native speaker of British English. Participants were asked to complete the target sentence in English. All of the instructions were in Dutch.

3.3 Procedure

All of the tests were part of a larger test battery in a three-year longitudinal study into the predictors of English language learning by Dutch primary school pupils. The test battery consisted of class sessions, in which participants worked by themselves in their classrooms, and individual sessions, in which a test administrator tested a participant individually in a separate room. In the larger test battery, there were five class sessions and six individual sessions in grade 4. Three of the class sessions contained Dutch (L1) language tests and the other two included English (FL) language tests. There were four class sessions and six individual sessions in grade 5. Two of the class sessions contained Dutch language tests and the other two included English language tests. Half of the schools did the Dutch language tests first and the other half did the English tests first. The order of the languages was counterbalanced, meaning that the schools that had started with the Dutch language tests in grade 4 began with the English tests in grade 5 and vice versa. During the class sessions, participants carried out the tests by themselves on a tablet; they wore headphones and a short video served as an instruction for each test. Each participant had a tablet (T550 Galaxy Tab A 9.7) to work on. The non-verbal intelligence test, the Dutch and English vocabulary tests, the Dutch sentence assembly and the English oral comprehension test were part of the classical sessions. All the class tests were converted to an online survey (www.qualtrics.com), so that the participants' answers were recorded and stored automatically and digitally. Each class session took about 45–60 minutes to complete. The language background test, the short-term memory tests, the working memory tests and the English (FL) morpho-syntax test were administered individually. Individual sessions took about 30 minutes to complete. The individual sessions were also counterbalanced, meaning that half of the schools started with English and the other half with Dutch and the order was changed in grade 5.

3.4 Data analysis

First, the participants' scores were inspected for outliers and normality. For all the tests, standardized measures for skewness and kurtosis did not exceed the value of 1.3, with the exception of the L1 vocabulary task in grade 4 (skewness = −0.9, kurtosis = 2.1) and the L1 sentence assembly task in grade 4 (skewness = −1.1, kurtosis = 2.6), on which a few participants obtained low scores. There were eight instances where the outliers were greater than three standard deviations below the mean. Because of these outliers, bootstrapping was used as a method to overcome the problems of normal distribution violations in some of the above measures.

The first step was to carry out Pearson *r* correlation analyses on all the tasks. After significant correlations were found, regression analyses were run. The hierarchical multiple regression analyses were built up step-wise in an identical manner, following theoretical considerations. In this hierarchical regression analysis predictors are entered in separate steps to determine which additional predictors improve the model once predictors in previous steps have been controlled for. The order of the predetermined steps was based on theories of FL learning. An automatic search for the best predictors does not allow for separate steps and therefore does not do justice to theoretical considerations. For example, by entering a past value of the dependent variable first we could control for previous competence in that skill. Also, Dutch primary school children learn their L1 before the FL and previous studies have shown the effect of the L1 on FL aptitude. By entering L1 predictors first we could control for L1 competence and determine the relative contribution of FL predictors. The predictors were only entered into the model when there was a significant ($p < .05$) correlation between the dependent variable and the predictor. First, the autoregressor was entered as a predictor into the model, meaning that a past value, i.e. the value in grade 4, of the variable of interest was used. As a second step, non-verbal intelligence was entered as a predictor into the model, so as to control for intelligence, but only if it correlated significantly with the dependent variable. Then, verbal short-term memory was added (if $p < .05$), as it plays a crucial role in vocabulary learning. The next model always included the L1 NWR but only if it correlated significantly with the dependent variable. To predict L1 grammar, L1 vocabulary was added in the fifth model. The L1 measures were added before the FL measures, because L1 competence is thought to have an effect on FL aptitude. For the FL dependent variables, FL NWR was added in the subsequent model, as this represented the ability to form long-term phonological representations, phonological processing and sublexical skills. Then, FL receptive vocabulary knowledge was added for the FL grammar dependent variables, as it indicated FL lexical knowledge. Finally, in all the analyses, working memory was added in the last model, as it was a measure of the ability to integrate linguistic information. To obtain robust results, all the regression analyses were bootstrapped and the 95% bias corrected and accelerated confidence intervals were reported.

4 Results

The mean scores of the 138 participants on the non-verbal intelligence test, on the memory tasks and on the Dutch (L1) and English (FL) measures are reported in Table 1. The non-verbal intelligence scores fell within the normal range. The verbal short-term memory score and the working memory score were normal for these age groups, as were the scores for NWR in L1 and FL. The mean scores on the L1 and FL vocabulary tests were quite similar in both grades; however, the L1 items were suitable for Dutch native speakers between the ages of 6;6 and 15;11, whereas the FL items were suitable for English native speakers between the ages of 2;6 and 9;0. Hence, the L1 vocabulary items were more

Table 1: The mean, standard deviation, reliability, range and number of items per task.

	Year	*M*	*SD*	α	Range	No. of items
Non-verbal intelligence	grade 4	38.5	6.9	.88[a]	22–53	60
Verbal short-term memory	grade 4	7.7	1.4	.85[a]	4–11	16
Working memory	grade 4	4.7	1.4	.85[a]	2–9	14
L1 NWR	grade 4	9.6	3.2	.64[b]	2–17	22
FL NWR	grade 4	5.2	2.2	.49[b]	1–11	18
L1 vocabulary	grade 4	40.0	5.9	.94[a]	13–51	59
L1 sentence assembly	grade 4	8.8	2.5	.70–.94[a]	0–13	13
FL vocabulary	grade 4	37.3	6.8	.80[b]	18–56	61
FL oral comprehension	grade 4	16.2	4.3	.75[b]	5–26	26
FL morpho-syntax	grade 4	3.7	2.8	.78[b]	0–11	12
L1 vocabulary	grade 5	38.1	6.9	.94[a]	14–54	59
L1 sentence assembly	grade 5	10.0	2.1	.70–.94[a]	3–13	13
FL vocabulary	grade 5	37.8	6.8	.82[b]	22–57	58
FL oral comprehension	grade 5	20.0	4.0	.79[b]	8–26	26
FL morpho-syntax	grade 5	6.9	2.8	.76[b]	0–12	12

Note. $N = 138$.
[a]Reliability values were taken from the manuals of the standardized tests.
[b]Reliability values were calculated on the current group of participants, as the tests were either not standardized or the reliability in the manual was based on native speaker participants.

difficult than the FL items. Furthermore, the vocabulary items that were tested in grade 5 were more difficult than those in grade 4. These two reasons explain why there was little differences between the mean scores on the L1 and FL vocabulary tests in grades 4 and 5. The mean scores for the L1 sentence assembly task in grades 4 and 5 show a difference: on average the participants were able to construct more correct, meaningful sentences in grade 5 than in grade 4. Finally, the participants scored on average around 57–77% correct on all the language measures.

A Pearson *r* correlation analysis was carried out to determine whether there was a relationship between the language measures and the memory measures. Table 2 reports the Pearson *r* correlation values for all the cognitive, L1 and FL measures.

Table 2 shows that there was no significant correlation between verbal short-term memory and L1 vocabulary. However, there was a moderate positive correlation between L1 NWR and L1 vocabulary in grades 4 and 5. Table 2 further shows that there was a moderate positive correlation between working memory and L1 sentence assembly meaning that participants who were relatively good at recalling digits in reverse serial order in grade 4, performed relatively well on the Dutch grammar task in grade 5.

With regard to the FL results, Table 2 shows that there was no correlation between verbal short-term memory and FL vocabulary, or working memory and the FL grammar tasks, nevertheless a moderate positive correlation was observed between FL NWR and FL vocabulary in grades 4 and 5. Both of these results indicate that participants who were better at repeating non-words in English in grade 4, also had a larger vocabulary in English in grade 5.

4.1 Memory and L1

Research question 1 asked whether NWR measured in grade 4 predicted L1 receptive vocabulary knowledge measured in grade 5. A hierarchical multiple regression analysis was run on the data including L1 vocabulary-grade 4, non-verbal intelligence and L1 NWR (predictors) and L1 vocabulary-grade 5 (dependent variable). Table 3 shows the output of the model that had the best fit to the data. Table A in the Appendix presents all of the regression models.

As can be seen in Table 2, L1 receptive vocabulary knowledge measured in grade 5 correlated significantly with L1 vocabulary measured in grade 4, non-verbal intelligence and L1 NWR. Hence, these were the predictors entered in the regression models. Model 1 in Table 3 shows that L1 receptive vocabulary

Table 2: Correlation matrix of the cognitive and L1 and FL language measures.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Non-verbal intelligence	–														
2. Verbal short-term memory	.13	–													
3. Working memory	.17*	.22**	–												
4. L1 NWR	.06	.27**	.16	–											
5. FL NWR	.22**	.26**	.08	.31**	–										
6. L1 vocabulary (grade 4)	.23**	.15	.09	.28**	.19*	–									
7. L1 vocabulary (grade 5)	24**	.16	.10	.29**	.10	.60**	–								
8. L1 sentence assembly (grade 4)	.17*	.31**	.22**	.15	.09	.38**	.13	–							
9. L1 sentence assembly (grade 5)	.30**	.13	.26**	.01	.05	.18*	.09	.28**	–						
10. FL vocabulary (grade 4)	.16	.09	.16	.18*	.18*	.17*	.22**	.12	.17*	–					
11. FL vocabulary (grade 5)	.15	.03	.15	.23**	.28**	.30**	.26**	.18*	.19*	.78**	–				
12. FL oral comprehension (grade 4)	.17*	−.03	.11	.18*	.16	.18*	.20*	.11	.20*	.62**	.58**	–			
13. FL oral comprehension (grade 5)	.19*	.03	.15	.26**	.28**	.29**	.30**	.16	.36**	.64**	.69**	.67**	–		
14. FL morpho-syntax (grade 4)	.13	.05	.09	.32**	.30**	.31**	31**	.17	.23**	.65**	.68**	.68**	.71**	–	
15. FL morpho-syntax (grade 5)	.23**	.09	.08	.28**	.22**	.28**	.37**	.12	.31**	.68**	.67**	.63**	.73**	.75**	–

Note. $N = 138$.
* $p < .05$, ** $p < .01$.

Table 3: Regression Model 1 of the L1 receptive vocabulary knowledge in grade 5.

	B	SE B	β	*p*
Model 1				
Constant	10.57 (−1.64, 20.73)	5.85		.103
L1 vocabulary (grade 4)	.69 (0.43, 0.99)	.15	.60	.001

Note. $N = 138$. $R^2 = .36$ for Model 1; $\Delta R^2 = .01$ for Model 2 ($p = .126$); $\Delta R^2 < .02$ for Model 3 ($p = .051$).

measured in grade 4 was a significant predictor of L1 receptive vocabulary knowledge measured in grade 5. In the next steps, non-verbal intelligence and L1 NWR were entered separately. None of these models (see Table A in the Appendix) proved to be a better fit; thus, Model 1 is the model we base our results on. Table 3 presents the output of the model that had the best fit to the data. About 36% of the variance of the performance on the L1 receptive vocabulary task in grade 5 was explained by the performance on the L1 vocabulary task in grade 4.

Research question 2 asked whether working memory span measured in grade 4 predicted L1 grammatical performance measured in grade 5. A hierarchical multiple regression analysis was run on the data including L1 sentence assembly – grade 4, non-verbal intelligence, verbal short-term memory, NWR, L1 vocabulary and working memory span (predictors) and L1 sentence assembly (dependent variable). Table B in the Appendix presents the different models of the predictors of L1 grammatical performance. Table 4 presents the output of the model that had the best fit to the data.

As can be seen in Table 2, L1 sentence assembly measured in grade 5 correlated significantly with L1 sentence assembly measured in grade 4, non-verbal intelligence, L1 vocabulary and working memory span. Hence, these were the predictors entered in the regression models. Working memory span was a significant predictor in Model 4, which can be seen in Table 4. Intelligence also remained a significant predictor although its influence became slightly smaller after including working memory span. When working memory span was added, L1 sentence assembly in grade 4 was no longer a significant predictor. About 17% of the variance of the performance on the L1 sentence assembly task in grade 5 was explained by the scores on L1 sentence assembly in grade 4, intelligence, L1 vocabulary and working memory span.

Table 4: Regression model of predictors measured in grade 4 of L1 sentence assembly measured in grade 5.

	B	SE B	β	P
Model 4				
Constant	4.16 (1.75, 6.86)	1.28		.003
L1 sentence assembly (grade 4)	0.15 (−0.01, 0.35)	0.09	.18	.094
Intelligence	0.07 (0.01, 0.12)	0.03	.23	.016
L1 vocabulary	0.02 (−0.05, 0.08)	0.03	.05	.624
Working memory span	0.26 (0.06, 0.46)	0.10	.18	.011

Note. $N = 138$. $R^2 = .08$ for Model 1; $\Delta R^2 = .07$ for Model 2 ($p = .002$); $\Delta R^2 < .01$ for Model 3 ($p = .636$); $\Delta R^2 = .03$ for Model 4 ($p = .031$).

4.2 Memory and FL

Second, we wanted to determine whether NWR and working memory span were differentially employed for FL (English) vocabulary learning versus English grammar learning. A hierarchical multiple regression analysis was run on the predictors that correlated significantly with the dependent variable. These included FL vocabulary – grade 4, L1 NWR, L1 vocabulary and FL NWR (predictors) and FL vocabulary – grade 5 (dependent variable). As explained in *Data Analysis*, the first model contained the autoregressor and the other predictors were added one by one in each subsequent model. Table C in the Appendix presents the different models of the predictors of FL receptive vocabulary knowledge. Table 5 presents the output of the model that had the best fit to the data.

About 65% of the variance of the performance on the FL receptive vocabulary task was explained by the scores on FL vocabulary in grade 4, L1 NWR, L1 vocabulary and FL NWR, with FL and L1 vocabulary in grade 4 and FL NWR significantly contributing to FL receptive vocabulary.

Research question 2 asked whether working memory measured in grade 4 predicted FL receptive grammatical performance measured in grade 5, taking into account grammatical performance in grade 4. A hierarchical multiple regression analysis was run on the predictors that correlated significantly with the dependent variable. These included FL oral comprehension – grade 4, non-verbal intelligence, L1 NWR, L1 vocabulary, FL NWR and FL vocabulary (predictors) and FL oral

Table 5: Regression model of predictors measured in grade 4 of FL receptive vocabulary knowledge measured in grade 5.

	B	SE B	Β	*p*
Model 4				
Constant	1.42 (−4.24, 7.73)	2.97		.614
FL vocabulary (grade 4)	0.74 (0.64, 0.85)	0.06	.73	.001
L1 NWR	0.07 (−0.13, 0.28)	0.10	.03	.516
L1 vocabulary	0.16 (−0.02, 0.29)	0.08	.14	.039
FL NWR	0.35 (0.04, 0.68)	0.16	.11	.033

Note. $N = 138$. $R^2 = .61$ for Model 1; $\Delta R^2 < .01$ for Model 2 ($p = .071$); $\Delta R^2 = .02$ for Model 3 ($p = .006$); $\Delta R^2 = .01$ for Model 4 ($p = .042$).

comprehension – grade 5 (dependent variable). Table D in the Appendix presents the different models of the predictors of FL receptive grammatical performance. Table 6 presents the output of the model that had the best fit to the data.

About 56% of the variance of the performance on the FL oral comprehension in grade 5 was explained by the scores on FL oral comprehension – grade 4, intelligence, L1 NWR, L1 vocabulary, FL NWR and FL vocabulary, with FL oral comprehension (grade 4), FL NWR and FL vocabulary significantly contributing to FL oral comprehension.

A hierarchical multiple regression analysis was run in order to answer the question whether working memory span measured in grade 4 predicted FL productive grammatical performance measured in grade 5. Predictors that correlated significantly with the dependent variable were included in the model: FL morpho-syntax – grade 4, non-verbal intelligence, L1 NWR, L1 vocabulary, FL NWR and FL vocabulary (predictors) and FL morpho-syntax – grade 5 (dependent variable). Table E in the Appendix presents the different models of the predictors of FL morpho-syntax. Table 7 presents the output of the model that had the best fit to the data.

About 64% of the variance of the performance on the FL morpho-syntax was explained by the scores on FL morpho-syntax – grade 4, intelligence, L1 NWR, L1 vocabulary, FL NWR and FL vocabulary, with FL morpho-syntax in grade 4 and FL vocabulary being significant predictors.

Table 6: Regression model of predictors measured in grade 4 of FL oral comprehension measured in grade 5.

	B	SE B	Β	P
Model 6				
Constant	1.32 (−1.85, 4.48)	1.66		.443
FL oral comprehension (grade 4)	0.39 (0.24, 0.55)	0.08	.41	.001
Intelligence	0.01 (−0.06, 0.09)	0.04	.02	.777
L1 NWR	0.08 (−0.09, 0.25)	0.08	.06	.318
L1 vocabulary	0.08 (−0.01, 0.16)	0.04	.12	.057
FL NWR	0.19 (0.00, 0.37)	0.09	.10	.038
FL vocabulary	0.19 (0.09, 0.29)	0.05	.33	.003

Note. $N = 138$. $R^2 = .45$ for Model 1; $\Delta R^2 = .01$ for Model 2 ($p = .192$); $\Delta R^2 = .02$ for Model 3 ($p = .028$); $\Delta R^2 = .02$ for Model 4 ($p = .046$); $\Delta R^2 = .01$ for Model 5 ($p = .070$); $\Delta R^2 = .06$ for Model 6 ($p < .001$).

5 Discussion

The aim of the present study was to determine whether phonological processing and working memory span in Dutch speaking pupils measured one year prior to the language measures predict receptive vocabulary knowledge and grammatical proficiency in Dutch (L1) and English (FL) in the initial stages of formal foreign language learning.

5.1 Memory and L1

It was hypothesised that Dutch receptive vocabulary would be predicted by NWR. Our results showed that Dutch receptive vocabulary knowledge in grade 5 correlated significantly with NWR and intelligence. This is in line with Gathercole et al. (1992), who also reported significant correlations between vocabulary, NWR and intelligence. We found that when Dutch vocabulary knowledge in grade 4 was entered as an autoregressor in the regression model, it proved to be the best predictor of receptive vocabulary knowledge in grade 5. Subsequent

Table 7: Regression model of predictors measured in grade 4 of FL morpho-syntax measured in grade 5.

	B	SE B	Β	*P*
Model 6				
Constant	−2.50 (−5.58, 0.99)	1.68		.135
FL morpho-syntax (grade 4)	0.49 (0.33, 0.68)	0.09	.50	.001
Intelligence	0.04 (−0.01, 0.09)	0.03	.11	.093
L1 NWR	0.05 (−0.04, 0.13)	0.04	.05	.305
L1 vocabulary	0.02 (−0.03, 0.08)	0.03	.04	.502
FL NWR	−0.04 (−0.19, 0.09)	0.07	−.03	.570
FL vocabulary	0.13 (0.06, 0.20)	0.03	.33	.001

Note. $N = 138$. $R^2 = .55$ for Model 1; $\Delta R^2 = .02$ for Model 2 ($p = .020$); $\Delta R^2 < .01$ for Model 3 ($p = .479$); $\Delta R^2 < .01$ for Model 4 ($p = .679$); $\Delta R^2 < .01$ for Model 5 ($p = .504$); $\Delta R^2 = .06$ for Model 6 ($p < .001$).

models did not yield a better fit; hence, we found that vocabulary knowledge was the best predictor of future vocabulary knowledge. NWR did not contribute to the present lexical skills on top of the past lexical skills. This seems to be in line with previous findings as L1 phonological processing effects are typically less apparent when the lexicon has reached a certain size (Cheung 1996; Engel de Abreu and Gathercole 2012; Gathercole et al. 1992). The second expectation stated that Dutch grammatical proficiency (measured with sentence assembly) would be predicted by working memory span. Working memory span and intelligence correlated significantly with Dutch grammatical proficiency. The regression models showed that Dutch grammatical performance was predicted by working memory span and intelligence in the current study. Previous studies found a similar relationship between grammatical performance and working memory span (Archibald and Gathercole 2006; Ellis and Sinclair 1996; Martin and Ellis 2012; Verhagen and Leseman 2016). Thus, our first hypothesis was not confirmed, even though significant correlations were observed between NWR and vocabulary, and the second hypothesis was wholly confirmed, showing that in the L1 these memory components contribute differentially to word learning and grammar over a period of one year. These results show that sufficient working memory capacity is important for processing and creating linguistic structures, which is reflected in the performance of our grammatical task. Our longitudinal design shows that the associations between memory and language are long-term as the memory tasks were administered one year prior to the language measures.

5.2 Memory and FL

For English, it was also hypothesised that NWR would predict receptive vocabulary knowledge. Our results showed that English receptive vocabulary knowledge in grade 5 correlated significantly with NWR and with Dutch vocabulary. When English vocabulary knowledge in grade 4 was entered as an autoregressor in the regression model, it predicted English vocabulary knowledge in grade 5 along with English NWR and Dutch vocabulary as predictors. That NWR predicted English vocabulary knowledge is in line with previous research (Cheung 1996; Engel de Abreu and Gathercole 2012; Kormos and Sáfár 2008; Service and Kohonen 1995; Verhagen and Leseman 2016). Verbal short-term memory (measured with the forward digit span in Dutch) did not predict English vocabulary knowledge to the same degree as NWR, meaning that the two measures likely represent different stages during phonological processing. Whereas verbal short-term memory tasks reflect the capacity of the phonological loop (Baddeley 2003), NWR reflects the capacity of the phonological loop as well as the ability to retrieve and manipulate phonological representations stored in the mental lexicon (Rispens and Baker 2012). These phonological representations are the scaffolds of words. It holds that the more often those representations are encountered, the more firmly they are stored in the lexicon and the easier it is to recognise them. Hence, they can be stored more easily in the phonological loop. It also works the other way around: the more vocabulary knowledge there is, the easier it is to extract the representations from that knowledge and to expand the storage of long-term phonological representations. This also explains why verbal short-term memory is a better predictor of vocabulary than NWR in very young children, whose stored vocabulary knowledge containing phonological representations is limited, and why NWR is a better predictor of vocabulary in somewhat older children (Gathercole et al. 1992; Rispens and Baker 2012).

There are differences between the current study and the previous ones. First, our study contained an autoregressor, meaning that we controlled for past English vocabulary knowledge. This was possible since our study was part of a longitudinal project. We found that the pupils' prior English vocabulary knowledge was the only significant predictor and it accounted for a large part of the variance. Second, the current study included both an L1 and a FL NWR, so that the contribution of both NWRs could be compared. When Dutch NWR was added to the model, but English NWR was not yet added, the former significantly predicted English vocabulary knowledge. This is in line with Engel de Abreu and Gathercole (2012), who found that L1 NWR predicted L1 (Luxembourgish) and L2 (German) vocabulary. Interestingly, L1 NWR did not predict L3/FL (French) vocabulary in their study. They suggested that L1 NWR can only predict L2/FL vocabulary when the languages are phonotactically similar, such as German and Luxembourgish, thereby evoking Baddeley (2003) who stated that immediate recall of non-words is better when those non-words are phonotactically similar to the L1. Our results are in line with both Engel de Abreu and Gathercole (2012) and Baddeley (2003), as L1 NWR was related to FL vocabulary yet when FL NWR was taken into account, our results demonstrate that a language specific NWR is a better predictor of FL vocabulary than L1 NWR, because it taps the phonotactics of the target language. This is the case even though FL NWR in our sample was a less reliable measure. This means that beginning learners who have a greater capacity of the phonological loop and who have stored more long-term phonological representations in English, will be better at learning English vocabulary than their peers who have a smaller capacity and who have stored fewer representations. Third, our study included L1 vocabulary knowledge as a predictor, so that the contribution of more general lexical knowledge was taken into account. Previous research established that L1 vocabulary is an indicator of FL aptitude (e.g. Ganschow and Sparks 2001; Skehan and Ducroquet 1988; Sparks, Patton and Ganschow 2012). Our study thus corroborates that finding, as Dutch vocabulary knowledge indeed contributed to the performance of English vocabulary.

Secondly, it was hypothesised that working memory span would predict English grammatical proficiency, for both oral comprehension and morpho-syntax production. The results showed that working memory span did not significantly correlate with English grammatical proficiency. We thus do not have evidence to conclude that working memory span predicts grammatical proficiency in children learning English. Our null results are different from what previous studies obtained (Archibald and Gathercole 2006; Ellis and Sinclair 1996; Martin and Ellis 2012; Verhagen and Leseman 2016). Differences in tasks and methodology may be responsible for the differences in results. For example, the participants in Martin and Ellis' (2012) study were adults learning an artificial language and the participants in the study by Verhagen and Leseman (2016) were five-year-old Dutch monolingual and Dutch-Turkish bilingual children learning Dutch in a naturalistic setting. Our participants' English vocabulary knowledge predicted current English grammatical performance. Hence, it is likely that our participants used different resources than working memory span while they were performing the grammatical tasks. They might very well count on their English vocabulary and their phonological processing skills to retrieve the meaning of the words in the sentences in the oral comprehension task. Indeed, this is what is indicated by the outcome of the regression analysis: English receptive vocabulary and NWR significantly predicted performance on the receptive grammatical task. This could indicate that our participants used a more lexical approach to this task. Of course, it may also be the case that the relation between working memory span and grammar exists in this group, but appears only when working memory span is measured with a different task than the backward digit span.

An influence of NWR on grammatical proficiency has been found before (Archibald and Gathercole 2006; Ellis and Sinclair 1996; Kormos and Sáfár 2008; O'Brien et al. 2006; Paradis 2011; Service and Kohonen 1995). Both Kormos and Sáfár (2008) and O'Brien et al. (2006) suggest that the more advanced learners are also the more proficient ones, meaning that lexical retrieval is easier for them than for the less advanced learners. When less energy can be spent on lexical retrieval, more working memory space is available to process, for example, function words. When function words are processed, Kormos and Sáfár (2008) and O'Brien et al. (2006) found that NWR and grammatical performance were related in more advanced learners. The present study only included beginning learners, but the contribution of NWR to the oral comprehension task could reflect an advantage in grammatical processing. The question is then why we only see a contribution of NWR in the comprehension task and not in the production task. Our participants scored relatively lower on the production task than on the comprehension task (mean percentage correct = 57% vs. mean percentage correct = 77%). It could be that our participants had to pay more attention to lexical retrieval when producing sentences – and, hence, had no working memory space left to focus on grammatical markers – memory space they did have when comprehending sentences. Perhaps the production task was too challenging to see a direct effect of grammatical proficiency. This could be why we only see an effect of past performance on that task and of English vocabulary knowledge when predicting grammatical production.

Besides English receptive vocabulary knowledge for both grammatical tasks and NWR for the oral comprehension task, another significant predictor of our participants' grammatical proficiency was past performance on the tasks. Since past performance contributed significantly, it indicates that besides a lexical approach, these children use grammatical processing to understand these sentences.

6 Conclusion

Past performance on a task is the best predictor for present performance. In addition, we found that L1 phonological processing as measured by a non-word repetition task is correlated with vocabulary in the first language (L1) and vocabulary in a foreign language (FL). However, phonological processing only predicts vocabulary in a foreign language, and only when a language specific FL measure is used. It seems that phonological processing is more important for lexical learning in the initial stages of FL learning than it is for the L1 at a later stage. In the L1 working memory (WM) predicts grammatical proficiency, meaning that children use their WM to hold the information from the grammatical task available so that they can parse the linguistic structures and create meaningful sentences. However, WM does not predict FL grammatical proficiency a year later. Grammatical performance is predicted by past performance on the same task and by FL vocabulary. Only receptive grammatical performance is also predicted by phonological processing, as measured by a language specific non-word repetition task. Whilst beginning learners of an FL seem to activate their grammatical knowledge when faced with a grammatical task, as indicated by the contribution of past grammatical performance, they also seem to use a more lexical approach to these tasks than in their L1.

Appendix

Table A: Linear models of predictors in grade 4 of L1 receptive vocabulary knowledge in grade 5.

	b	SE B	β	*p*
Model 1				
Constant	10.57 (−1.64, 20.73)	5.85		.103
L1 vocabulary (grade 4)	0.69 (0.43, 0.99)	0.15	.60	.001
Model 2				
Constant	7.53 (−3.03, 16.58)	5.02		.155
L1 vocabulary (grade 4)	0.66 (0.42, 0.98)	0.15	.57	.001
Intelligence	0.11 (−0.08, 0.27)	0.09	.11	.236

Table A (continued)

	b	SE B	β	*p*
Model 3				
Constant	6.46 (−3.74, 15.08)	4.78		.201
L1 vocabulary (grade 4)	0.62 (0.35, 0.95)	0.16	.53	.001
Intelligence	0.11 (−0.08, 0.28)	0.09	.11	.237
L1 NWR	0.30 (0.01, 0.59)	0.15	.14	.052

Note. $N = 138$. $R^2 = .36$ for Model 1; $\Delta R^2 = .01$ for Model 2 ($p = .126$); $\Delta R^2 = .02$ for Model 3 ($p = .051$).

Table B: Linear models of predictors in grade 4 of L1 sentence assembly in grade 5.

	b	SE B	β	*p*
Model 1				
Constant	7.97 (6.08, 9.49)	0.86		.001
L1 sentence assembly (grade 4)	0.23 (0.08, 0.43)	0.09	.28	.013
Model 2				
Constant	5.25 (2.88, 7.58)	1.18		.001
L1 sentence assembly (grade 4)	0.20 (0.05, 0.38)	0.08	.23	.016
Intelligence	0.08 (0.02, 0.13)	0.03	.26	.009
Model 3				
Constant	4.86 (2.37, 7.40)	1.29		.001
L1 sentence assembly (grade 4)	0.19 (0.02, 0.39)	0.09	.22	.048
Intelligence	0.08 (0.02, 0.13)	0.03	.25	.011
L1 vocabulary	0.02 (−0.05, 0.07)	0.03	.04	.646

Table B (continued)

	b	SE B	β	*p*
Model 4				
Constant	4.16 (1.75, 6.86)	1.28		.003
L1 sentence assembly (grade 4)	0.15 (−0.01, 0.35)	0.09	.18	.094
Intelligence	0.07 (0.01, 0.12)	0.03	.23	.016
L1 vocabulary	0.02 (−0.05, 0.08)	0.03	.05	.624
Working memory span	0.26 (0.06, 0.46)	0.10	.18	.011

Note. $N = 138$. $R^2 = .08$ for Model 1; $\Delta R^2 = .07$ for Model 2 ($p = .002$); $\Delta R^2 < .01$ for Model 3 ($p = .636$); $\Delta R^2 = .03$ for Model 4 ($p = .031$).

Table C: Linear models of predictors measured in grade 4 of FL receptive vocabulary knowledge measured in grade 5.

	B	SE B	β	*p*
Model 1				
Constant	8.46 (4.19, 12.42)	2.10		.001
FL vocabulary (grade 4)	0.79 (0.68, 0.90)	0.05	.78	.001
Model 2				
Constant	7.12 (2.77, 11.10)	2.11		.002
FL vocabulary (grade 4)	0.77 (0.67, 0.88)	0.05	.77	.001
L1 NWR	0.21 (0.01, 0.41)	0.10	.10	.047
Model 3				
Constant	1.67 (−3.93, 7.89)	2.99		.559
FL vocabulary (grade 4)	0.75 (0.65, 0.86)	0.05	.75	.001
L1 NWR	0.13 (−0.08, 0.33)	0.10	.06	.196

Table C (continued)

	B	SE B	β	*p*
L1 vocabulary	0.17 (−0.00, 0.30)	0.08	.15	.026
Model 4				
Constant	1.42 (−4.24, 7.73)	2.97		.614
FL vocabulary (grade 4)	0.74 (0.64, 0.85)	0.06	.73	.001
L1 NWR	0.07 (−0.13, 0.28)	0.10	.03	.516
L1 vocabulary	0.16 (−0.02, 0.29)	0.08	.14	.039
FL NWR	0.35 (0.04, 0.68)	0.16	.11	.033

Note. $N = 138$. $R^2 = .61$ for Model 1; $\Delta R^2 < .01$ for Model 2 ($p = .071$); $\Delta R^2 = .02$ for Model 3 ($p = .006$); $\Delta R^2 = .01$ for Model 4 ($p = .042$).

Table D: Linear models of predictors measured in grade 4 of FL oral comprehension measured in grade 5.

	b	SE B	β	*p*
Model 1				
Constant	9.93 (7.95, 11.88)	1.02		.001
FL oral comprehension (grade 4)	0.63 (0.5, 0.74)	0.06	.67	.001
Model 2				
Constant	8.24 (5.29, 11.23)	1.51		.001
FL oral comprehension (grade 4)	0.61 (0.50, 0.73)	0.06	.65	.001
Intelligence	0.05 (−0.02, 0.12)	0.04	.09	.160
Model 3				
Constant	7.00 (3.77, 9.89)	1.51		.001
FL oral comprehension (grade 4)	0.59 (0.48, 0.71)	0.06	.63	.001

Table D (continued)

	b	SE B	β	*p*
Intelligence	0.05 (−0.02, 0.12)	0.04	.08	.190
L1 NWR	0.18 (−0.01, 0.36)	0.09	.14	.057
Model 4				
Constant	4.57 (1.24, 7.71)	1.67		.008
FL oral comprehension (grade 4)	0.58 (0.46, 0.70)	0.06	.62	.001
Intelligence	0.03 (−0.04, 0.11)	0.04	.05	.405
L1 NWR	0.14 (−0.05, 0.31)	0.09	.11	.144
L1 vocabulary	0.09 (−0.00, 0.18)	0.05	.13	.044
Model 5				
Constant	4.61 (1.51, 7.60)	1.59		.006
FL oral comprehension (grade 4)	0.57 (0.46, 0.69)	0.06	.61	.001
Intelligence	0.02 (−0.05, 0.10)	0.04	.03	.657
L1 NWR	0.10 (−0.08, 0.27)	0.09	.08	.285
L1 vocabulary	0.09 (−0.01, 0.17)	0.04	.13	.053
FL NWR	0.23 (0.03, 0.42)	0.10	.12	.027
Model 6				
Constant	1.32 (−1.85, 4.48)	1.66		.443
FL oral comprehension (grade 4)	0.39 (0.24, 0.55)	0.08	.41	.001
Intelligence	0.01 (−0.06, 0.09)	0.04	.02	.777
L1 NWR	0.08 (−0.09, 0.25)	0.08	.06	.318
L1 vocabulary	0.08 (−0.01, 0.16)	0.04	.12	.057

Table D (continued)

	b	SE B	β	*p*
FL NWR	0.19 (0.00, 0.37)	0.09	.10	.038
FL vocabulary	0.19 (0.09, 0.29)	0.05	.33	.003

Note. $N = 138$. $R^2 = .45$ for Model 1; $\Delta R^2 = .01$ for Model 2 ($p = .192$); $\Delta R^2 = .02$ for Model 3 ($p = .028$); $\Delta R^2 = .02$ for Model 4 ($p = .046$); $\Delta R^2 = .01$ for Model 5 ($p = .070$); $\Delta R^2 = .06$ for Model 6 ($p < .001$).

Table E: Linear models of predictors measured in grade 4 of FL morpho-syntax measured in grade 5.

	b	SE B	β	*p*
Model 1				
Constant	4.15 (3.55, 4.75)	0.31		.001
FL morpho-syntax (grade 4)	0.73 (0.61, 0.85)	0.06	.75	.001
Model 2				
Constant	2.15 (0.29, 4.10)	0.98		.034
FL morpho-syntax (grade 4)	0.72 (0.59, 0.84)	0.06	.73	.001
Intelligence	0.05 (0.00, 0.10)	0.02	.13	.042
Model 3				
Constant	1.86 (−0.10, 3.91)	1.05		.075
FL morpho-syntax (grade 4)	0.70 (0.57, 0.84)	0.07	.71	.001
Intelligence	0.05 (0.00, 0.10)	0.02	.13	.041
L1 NWR	0.04 (−0.07, 0.12)	0.05	.04	.446
Model 4				
Constant	1.52 (−1.27, 4.28)	1.43		.288

Table E (continued)

	b	SE B	β	*p*
FL morpho-syntax (grade 4)	0.70 (0.56, 0.83)	0.07	.71	.001
Intelligence	0.05 (−0.00, 0.10)	0.03	.13	.049
L1 NWR	0.03 (−0.06, 0.12)	0.05	.04	.488
L1 vocabulary	0.01 (−0.04, 0.08)	0.03	.03	.682
Model 5				
Constant	1.55 (−1.22, 4.36)	1.43		.287
FL morpho-syntax (grade 4)	0.71 (0.57, 0.84)	0.07	.72	.001
Intelligence	0.05 (−0.00, 0.10)	0.03	.14	.042
L1 NWR	0.04 (−0.06, 0.14)	0.05	.05	.433
L1 vocabulary	0.01 (−0.05, 0.08)	0.03	.03	.667
FL NWR	−0.05 (−0.22, 0.11)	0.08	−.04	.507
Model 6				
Constant	−2.50 (−5.58, 0.99)	1.68		.135
FL morpho-syntax (grade 4)	0.49 (0.33, 0.68)	0.09	.50	.001
Intelligence	0.04 (−0.01, 0.09)	0.03	.11	.093
L1 NWR	0.05 (−0.04, 0.13)	0.04	.05	.305
L1 vocabulary	0.02 (−0.03, 0.08)	0.03	.04	.502
FL NWR	−0.04 (−0.19, 0.09)	0.07	−.03	.570
FL vocabulary	0.13 (0.06, 0.20)	0.03	.33	.001

Note. $N = 138$. $R^2 = .55$ for Model 1; $\Delta R^2 = .02$ for Model 2 ($p = .020$); $\Delta R^2 < .01$ for Model 3 ($p = .479$); $\Delta R^2 < .01$ for Model 4 ($p = .679$); $\Delta R^2 < .01$ for Model 5 ($p = .504$); $\Delta R^2 = .06$ for Model 6 ($p < .001$).

References

Adams, A.-M., & Gathercole, S.E. 2000. Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders* **35**(1). 95–116. https://doi.org/10.1080/136828200247278

Archibald, L.M.D., & Gathercole, S.E. 2006. Research report. Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders* **41**(6). 675–693. https://doi.org/10.1080/13682820500442602

Baddeley, A.D. 2003. Working memory: Looking back and looking forward. *Nature* **4**(10).829–839. https://doi.org/10.1038/nrn1201

Baddeley, A.D., & Hitch, G. 1974. Working memory. In G. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory*, 47–89. New York: Academic Press.

Baird, A.S., Palacios, N., & Kibler, A. 2016. The cognate and false cognate knowledge of young emergent bilinguals. *Language Learning* **66**(2). 448–470. https://doi.org/10.1111/lang.12160

Boersma, T., Baker, A., Rispens, J., & Weerman, W. 2018. The effects of phonological skills and vocabulary on morphophonological processing. *First Language* **38**(2). 147–174. https://doi.org/10.1177/0142723717725430

Cheung, H. 1996. Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology* **32**(5). 867–873.

Crannell, C.W., & Parrish, J.M. 1957. A comparison of immediate memory span for digits, letters, and words. *The Journal of Psychology* **44**. 319–327.

Dunn, L.M., & Dunn, D.M. 2007. *Peabody picture vocabulary test*, 4th edn. New York: Pearson.

Ellis, N.C., & Sinclair, S.G. 1996. Working memory in acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology Section A* **49**(1). 234–250. https://doi.org/10.1080/713755604

Engel de Abreu, P.M.J., & Gathercole, S.E. 2012. Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology* **104**(4). 974–986. https://doi.org/10.1037/a0028390

Ganschow, L., & Sparks, R. 2001. Learning difficulties and foreign language learning: A review of research and instruction. *Language Teaching* **34**(2). 79–98. https://doi.org/10.1017/S0261444800015895

Gathercole, S.E., Willis, C.S., Emslie, H., & Baddeley, A.D. 1992. Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology* **28**(5). 887–898.

Hamel, R., & Schmittmann, V.D. 2006. The 20-minute version as a predictor of the Raven advanced progressive matrices test. *Educational and Psychological Measurement* **66**(6). 1039–1046. https://doi.org/10.1177/0013164406288169

Jones, G. 2012. Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed. *Frontiers in Psychology* **3**. 167.

Justice, L.M., McGinty, A.S., Cabell, S.Q., Kilday, C.R., Knighton, K., & Huffman, G. 2010. Language and literacy curriculum supplement for preschoolers who are academically at risk: a feasibility study. *Language, Speech, and Hearing Services in Schools* 141, 161–178. https://doi.org/10.1044/0161-1461(2009/08-0058)

Kormos, J., & Sáfár, A. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* **11**(2). 261–271. https://doi.org/10.1017/S1366728908003416
Kort, W., Schittekatte, M., Bosmans, M., Compaan, E.L., Dekker, P.H., Vermeir G., & Verhaeghe, P. 2005. Wechsler Intelligence Scale for Children III (WISC-III-NL). Amsterdam: Harcourt Test Publisher.
Kort, W., Schittekatte, M., & Compaan, E. 2010. CELF-4-NL | Test voor diagnose en evaluatie van taalproblemen (Nederlandse bewerking). Amsterdam: Pearson.
Martin, I.K., & 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**(3). 379–413. https://doi.org/10.1017/S0272263112000125
Marton, K., & Schwartz, R.G. 2003. Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* **46**(5). 1138–1153. https://doi.org/10.1044/1092-4388
Masoura, E.V., & Gathercole, S.E. 2005. Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory* **13**(3–4). 422–429. https://doi.org/10.1080/09658210344000323
Messer, M.H., Leseman, P.P.M., Boom, J., & Maya, 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**(4). 306–323. https://doi.org/10.1016/j.jecp.2009.12.006
O’Brien, I., Segalowitz, N., Collentine, J., & Freed, B. 2006. Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics* **27**(3). 377–402. https://doi.org/10.1017/S0142716406060322
Oostdam, R., & van Toorenburg, H. 2002. ‘Leuk is not enough’: Het vraagstuk van de positionering van Engels in het basisonderwijs en de aansluiting met het voortgezet onderwijs. *Levende Talen Tijdschrift* **3**(4). 3–18. https://levendetalen.nl/publicaties/tijdschrift/
Paradis, J. 2011. Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism* **1**(3). 213–237. https://doi.org/10.1075/lab.1.3.01par
Raven, J., Raven, J.C., & Court, J.H. 2003. *Manual for Raven’s progressive matrices and vocabulary scales*. San Antonio, TX: Harcourt Assessment.
Rescorla, L. 2002. Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research* 45, 360–371. https://doi.org/10.1044/1092-4388(2002/028)
Rispens, J., & Baker, A. 2012. Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research* **55**(3). 683–694. https://doi.org/10.1044/1092-4388
Rispens, J., & de Bree, E. 2017. Past tense production in children with SLI and bilingual children: The influence of vocabulary and non-word repetition. In E. Blom, L. Cornips, & J. Schaeffer (eds.), *Cross-linguistic influence in bilingualism: In honor of Aafke Hulk*, 259–277. Amsterdam: John Benjamins Publishing Company.
Schlichting, L. 2005. *Peabody Picture Vocabulary Test-III NL*. Amsterdam: Harcourt Test Publisher.

Service, E., & Kohonen, V. 1995. Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics* **16**(2). 155–172. https://doi.org/10.1017/S0142716400007062

Skehan, P., & Ducroquet, L. 1988. A comparison of first and foreign language ability. ESOL Department, Institute of Education, London University: Working Documents. No. 8.

Sparks, R., Patton, J., & Ganschow, L. 2012. Profiles of more and less successful L2 learners: A cluster analysis study. *Learning and Individual Differences* **22**(4). 463–472. https://doi.org/10.1016/j.lindif.2012.03.009

Verhagen, J., & Leseman, P. 2016. How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology* **141**. 65–82. https://doi.org/10.1016/j.jecp.2015.06.015

Wagner, R., Torgesen, J., Rashotte, C., & Pearson, N.A. 2013. *Comprehensive Test of Phonological Processing*, 2nd edn. New York: Pearson.

Weber-Fox, C.M., & Neville, H.J. 1996. Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience* **8**(3), 231–256. https://doi.org/10.1162/jocn.1996.8.3.231

Wiig, E.H., Semel, E., & Secord, W.A. 2003. *Clinical Evaluation of Language Fundamentals*, 4th edn. San Antonio, TX: Pearson/PsychCorp.