# The role of existing language knowledge in bilingual and multilingual toddlers' repetition of cross-linguistic and language-specific nonwords

Josje Verhagen and Sible Andringa
University of Amsterdam

Previous studies have shown that bilingual children typically score more poorly on nonword repetition (NWR) tasks than monolingual peers, which has been attributed to bilinguals' lower proficiency in the language that the NWR task is based on. To enable fairer assessments of bilingual children, Cross-Linguistic NWR tasks (CL-NWR tasks) have been developed that are based on the linguistic properties of many languages. The aim of this study is to investigate whether young children's performance on a CL-NWR is less dependent on existing knowledge of a specific language than performance on a Language-Specific (Dutch-based) NWR (LS-NWR). Bilingual and multilingual two- and three-year-olds ($N$ = 216) completed a CL-NWR and LS-NWR, as well as a Dutch receptive vocabulary task. Parents reported the number of languages children spoke other than Dutch. Results of linear mixed-effect regressions showed that Dutch vocabulary scores related to performance on the CL-NWR task less strongly than to performance on the LS-NWR task. The number of non-Dutch languages spoken did not differentially relate to performance on the two tasks. These findings indicate that CL-NWR tasks – at least as used here – allow for more language-neutral NWR assessments within linguistically diverse samples, already at toddler age.

**Keywords:** nonword repetition tasks, language-specificity, bilingualism, toddlers, existing language knowledge

## 1. Introduction

Although the exact skills underlying nonword repetition (NWR) have been subject to debate, there is general consensus that NWR tasks assess, amongst others,

https://doi.org/10.1075/lab.20050.ver | Published online: 20 August 2021
*Linguistic Approaches to Bilingualism* ISSN 1879-9264 | E-ISSN 1879-9272

the ability to store auditory material in verbal-short-term memory (Coady & Evans, 2008; Gathercole, 2006; Rispens & Baker, 2012). Differences in children's NWR abilities are associated with a wide range of language outcomes, including vocabulary knowledge (Gathercole, 2006), grammatical proficiency (Adams & Gathercole, 1996; Verhagen & Leseman, 2016), and novel word learning (Kaushanskaya, 2012). Moreover, previous work has shown that NWR is delayed in children with developmental language disorders as compared to more typically developing children, at least at the group level (Dispaldro, Leonard & Deevy, 2013; Ellis-Weismer et al. 2000; Graf Estes, Evans & Else-Quest, 2007), which has led to the proposal that NWR tasks can be used as clinical tools for diagnosing impaired language development (Archibald, 2008; Conti-Ramsden & Hesketh, 2003; Gray, 2003).

For bilingual children, who grow up with two languages, assessing NWR may be problematic, however. There is now ample evidence that NWR tasks do not only tap verbal short-term memory, but also a number of other processes, such as speech perception and phonological encoding, some of which are dependent on long-term language knowledge (for a review, see Coady & Evans, 2008). Bilingual children often have lower levels of long-term knowledge of the language that the NWR items are based on than monolingual children, that is, they have less knowledge stored in memory about the phonemes, phoneme combinations, syllables, words and phrases in this language. Therefore, they cannot rely on their existing, long-term language knowledge during NWR to the same degree (Duncan & Paradis, 2013; Engel de Abreu, et al. 2013; Kohnert, Windsor & Yim, 2006; Windsor, Kohnert, Lobitz & Pham, 2010). Hence, it is no surprise that bilingual children often perform more poorly on NWR tasks than monolingual peers (Engel de Abreu, et al. 2013; Duncan & Paradis, 2013; Kohnert, et al. 2006). In fact, in a study by Kohnert and colleagues (2006), English monolingual children with developmental language disorders could not be distinguished from English-Spanish bilingual children without such disorders on the basis of their performance on an English-based NWR task. Thus, using NWR tasks for clinical purposes with bilingual children seems problematic (see also Boerma, Chiat, Leseman, Timmermeister, Wijnen & Blom, 2015), and may, in fact, explain why overdiagnosis of language disorders is more common in bilingual children than in monolingual children (Grimm & Schulz, 2014).

To enable fairer assessment of bilingual children's NWR abilities, a Cross-Linguistic NWR (CL-NWR) task has been developed, which draws minimally on knowledge of a specific language (Chiat, 2015). In two previous studies that compared monolingual and bilingual children's performance on this task (Boerma et al., 2015; Chiat & Polišenská, 2016), no effects of bilingualism were found, indicating that the CL-NWR is an appropriate task for bilingual children. However, in these studies, data from a diverse group of bilingual children were collapsed and analyzed at the group level, leaving unaddressed how individual differences in long-term language knowledge *within bilingual groups* relate to performance on the CL-NWR. Also, children between four and six years of age were assessed, leaving open whether the CL-NWR is an appropriate tool for bilingual children at a younger age. Since early diagnosis is important, there is a need for 'language-neutral' NWR tasks that can be used in children's early years. Investigating whether a CL-NWR is less affected by differences in existing language knowledge than a LS-NWR, already at toddler age, constitutes a first step in this direction.

In the current study, we investigate data from a diverse group of bilingual and multilingual toddlers to see how children's existing language knowledge relates to their performance on a CL-NWR task as compared to a Language-Specific (Dutch-based) NWR (LS-NWR). We predict that long-term knowledge of Dutch will be more strongly implicated in children's performance on the LS-NWR than on the CL-NWR. If this prediction is borne out, this would lend further support to the use of the CL-NWR as a better tool with heterogeneous groups of bilinguals and multilinguals. Since younger children are investigated than in earlier work, moreover, our study would extend previous findings to younger populations.

## 1.1 Effects of existing language knowledge on NWR in bilingual children

As pointed out above, NWR tasks are generally assumed to assess the ability to store auditory material in verbal short-term memory, or phonological storage (Coady & Evans, 2008; Gathercole, 2006; Rispens & Baker, 2012). However, NWR tasks also tap into long-term language knowledge. Evidence for this comes from research with monolingual children showing that nonwords that are more word-like, higher in phonotactic probability, and from more dense phonological neighborhoods are repeated more accurately than nonwords that are less word-like, lower in phonotactic probability, and from more sparse neighborhoods, respectively (e.g., Coady & Aslin, 2004; Gathercole, 1995; Metsala & Chisholm, 2010; Munson, Kurtz & Windsor, 2005).

Three different lines of research on bilingual children have provided support for the idea that existing long-term knowledge is implicated in NWR ability. First, in some studies, negative effects of bilingualism on NWR performance disappeared once differences in vocabulary between monolingual and bilingual children were controlled for (Engel de Abreu, 2011; Engel de Abreu et al. 2013; Meir & Armon-Lotem, 2017). This suggests that poorer NWR performance in bilingual as opposed to monolingual children is indeed due to bilinguals' less well-developed long-term knowledge of lexical items, phonemes, and phoneme combinations.

Second, earlier work has shown that bilinguals obtain higher scores on NWR tasks that are based on their stronger language than on NWR tasks based on their weaker language (Masoura & Gathercole, 1999; Messer, Leseman, Boom & Mayo, 2010). Masoura and Gathercole (1999), for example, found that English-Greek school-aged bilinguals who were dominant in Greek were more accurate in a Greek-based NWR task than in an English-based task. Similarly, Messer and colleagues (2010) found that Turkish-Dutch preschool children who were dominant in Turkish obtained higher scores on Turkish-based nonword recall tasks than on Dutch-based tasks. An exception to this pattern comes from Ebert and colleagues (2014), who found that Spanish-English children who were dominant in English performed better on a Spanish-based NWR task than on an English-based NWR task. To explain these findings, the authors proposed that the Spanish-based items had simpler phonological properties, making them easier to repeat. However, it is questionable whether performance on NWR tasks that are based on different languages can be compared directly: other explanations cannot be excluded, for example, that nonwords resembled real words more closely in one of the tasks (see Gibson et al. 2015 for similar claims).

Finally, associations have been reported between bilingual children's vocabulary knowledge in a given language and performance on NWR tasks based on that language (e.g., Brandeker & Thordardottir, 2015; Gutiérrez-Clellen & Simon-Cereijido, 2010; Parra, Hoff & Core, 2011). Parra and colleagues (2011) studied Spanish-English bilingual two-year-olds' performance on an NWR task that included both English-like and Spanish-like nonwords, and found that English vocabulary correlated more strongly with children's performance on the English-like items than Spanish-like items, whereas the reversed pattern emerged for Spanish vocabulary. Messer and colleagues (2010) found that Turkish vocabulary scores were significantly associated with Turkish-Dutch four-year-old children's scores on a Turkish-based NWR task, whereas Dutch vocabulary was associated with children's scores on a Dutch-based NWR task (for similar findings for school-aged children, see Thorn and Gathercole, 1999). An exception to this pattern was reported by Brandeker and Thordardottir (2015), however, who found a significant and moderate correlation between English vocabulary and English-based NWR, but a non-significant and weak correlation between French vocabulary and French-based NWR in a sample of French-English bilingual two- and three-year-olds. The authors suggested that the simple phonological properties of the French-based NWR items enabled even children with low French vocabulary scores to repeat these items, thus weakening the correlation between vocabulary and NWR.

Taken together, earlier findings on NWR in bilingual children indicate that children's vocabulary knowledge of a given language generally supports their repetition of nonwords based on that language. If exceptions to this pattern were found, these were attributed to the nature of the nonwords, for example, by considering these as phonologically simple.

### 1.2 Cross-linguistic NWR

Recently, a Cross-Linguistic NWR task (CL-NWR) has been developed, which is minimally based on participants' knowledge of a specific language (Chiat, 2015). The aim of this task is to enable NWR assessments that are optimally independent of children's existing language knowledge. As such, the task should enable a fairer assessment of bilingual children, and allow for better diagnosis of children at risk of language disorders across linguistically diverse groups (Chiat, 2015). The items in the CL-NWR task are based on linguistic properties that are shared across many languages. Specifically, the task contains CVCV items that are composed of phonemes that occur in many languages and equal stress is placed on each syllable (for more details and the rationale behind the task, see Chiat, 2015).

To date, a handful of studies have reported on CL-NWR tasks (Antonijevic-Elliott et al. 2020; Boerma et al. 2015; Chiat & Polišenská, 2016; White, 2019), three of which were specifically focused on comparing bilingual children's performance on a CL-NWR and a LS-NWR. In the first, Boerma and colleagues (2015) presented the CL-NWR developed by Chiat (2015) - referred to as "Quasi-Universal" NWR in this study - and a Dutch-based LS-NWR to Dutch monolingual five- and six-year-old children and bilingual peers who spoke Dutch and another language. The authors found that the monolingual children did not outperform the bilingual children on the CL-NWR, while they did outperform the bilinguals on the LS-NWR. The authors also found that performance was higher in the CL-NWR than in the LS-NWR, irrespective of group. Finally, comparing children with and without developmental language disorders, Boerma et al. observed that the CL-NWR had higher sensitivity levels than the LS-NWR. That is, the CL-NWR was better suited to identify children with language disorders in the bilingual group than the LS-NWR.

In the second study, by Chiat and Polišenská (2016), NWR performance of monolingual and bilingual four- to seven-year-olds from diverse SES backgrounds was compared across three NWR tasks that varied in the degree to which they were based on English: (i) the English version of the same CL-NWR that was used by Boerma and colleagues and had been developed by Chiat (2015), (ii) a Prosodically Specific NWR, which was an adaptation of the CL-NWR such that it followed English prosody, and (iii) an English-based LS-NWR. While Chiat and Polišenská did find an effect of SES, they did not find differences in performance between monolingual and bilingual children. A possible explanation for this null effect of bilingualism is that subgroups of children in this study (classified according to both SES and bilingualism) were small (i.e., 9 or 12 children per group), which raises the possibility that the numerical differences observed between monolinguals and bilinguals on the LS-NWR (45% vs. 36% correct) and the Prosodically Specific task (53% vs. 62% correct) were non-significant due to low power.

Finally, Antonijevic-Elliott and colleagues (2020) administered an English-based LS-NWR and the same CL-NWR as used by Chiat and Polišenská to English monolingual and multilingual five- to seven-year old children who spoke English and one out of a large set of other languages. The results showed no differences in performance across the two groups on either of the tasks. However, relating age of exposure to English to children's repetition accuracy, the authors found that age of exposure to English predicted children's performance on the LS-NWR, but not on the CL-NWR. The authors concluded that it would be best to use the CL-NWR as a language-neutral task for heterogeneous and linguistically diverse populations.

### 1.3 This study

In the current study, we compared children's performance on a CL-NWR and a Dutch-based LS-NWR task to address an issue that remains open from earlier work on the use of CL-NWR tasks with bilingual children. That is, we investigated how bilingual children's existing language knowledge related to their performance on the CL-NWR as compared to the LS-NWR. Unlike earlier studies using CL-NWR tasks (Boerma et al. 2015; Chiat & Polišenská, 2016), we did not compare groups of monolingual and bilingual children. Rather, we analyzed data from a heterogeneous group of bilingual and multilingual children from diverse language backgrounds who varied greatly in their long-term language knowledge, to allow for an analysis of the role of differences in children's existing language knowledge, and avoid treating bilinguals or multilinguals as a monolithic group. Our research questions were the following:

i. How do differences in Dutch vocabulary knowledge within a linguistically diverse sample of bilingual and multilingual children relate to performance on a CL-NWR task as compared to a (Dutch-based) LS-NWR task?
ii. How does the number of languages other than Dutch spoken by these children relate to their performance on the two NWR tasks?

As for the first question, we hypothesized that Dutch vocabulary knowledge would be associated with children's performance on the LS-NWR, but not, or less so, with performance on the CL-NWR task. Specifically, since the items in the former task, but not the latter, are based on specific properties of Dutch, we expected knowledge of these properties (i.e., through acquired vocabulary items) to aid repetition in the LS-NWR task in particular.

Regarding the second question, we predicted that the number of languages children spoke other than Dutch would be related to children's scores on the CL-NWR more strongly than on the LS-NWR. Specifically, since the CL-NWR contains properties that occur in many languages – rather than all languages, we assumed that knowledge of a larger number of (non-Dutch) languages would especially benefit children's repetition of the CL-NWR items. Thus, we hypothesized that increased familiarity with a broader set of phonemes, phoneme combinations and syllables (through knowing many languages), would facilitate repetition of the CL-NWR nonwords, but not the LS-NWR nonwords, the latter being specific to Dutch.

## 2. Method

### 2.1 Participants

Participants were 216 toddlers who took part in a research project that was primarily aimed at investigating the effects of bilingual daycare on children's language development (i.e, Project MIND (Multilingualism in Daycare), cf. Keydeniers, Aalberse, Andringa & Kuiken (2021)). These children attended daycare centers in the Netherlands that volunteered to participate in this research project and had been selected by the Dutch Ministry of Social Affairs and Employment, which commissioned the project. Children at these daycares were recruited through information letters sent out to their parents, and constituted a subset out of a larger sample of 303 children (71%) enrolled in the MIND project at the first data wave of this project. Children were selected for the current study if (i) their parents had completed a short parental questionnaire in which they reported which language(s) each parent spoke to their child at home, and (ii) if they had completed at least one of the NWR tasks, as well as a Dutch receptive vocabulary test. Mean age was 36.39 months ($SD$ = 6.49 months, min-max = 24–48 months) and there were 104 girls (48%).

All children were exposed to one or more additional language(s) next to the majority language Dutch, the latter of which they learned at home, at daycare, or both. They attended Dutch-only or Dutch-English daycare. Thus, in our study, 'bilingual' was defined as being exposed to two languages and 'multilingual' was defined as being exposed to more than two languages. The language(s) other than Dutch at home were the following: Arabic ($n$ = 2), Bahasa Indonesia ($n$ = 2), Bosnian ($n=1$), Catalan ($n=2$), Chinese ($n=8$), Croatian ($n=2$), Czech ($n=2$) Danish ($n=1$), English ($n=70$), Filipin ($n=1$), Finnish ($n=3$), French ($n=22$), Greek ($n=7$), Hebrew ($n=4$), Hindi ($n=5$), Italian ($n=12$), Japanese ($n=3$), Marathi ($n=1$), Nepalese ($n=1$), Norwegian ($n=2$), Papiamentu ($n=1$), Polish ($n=3$), Portuguese ($n=10$), Romanian ($n=2$), Russian ($n=8$), Serbian ($n=1$), Slovakian ($n=3$), Spanish ($n=14$), Swedish ($n=6$), Tamil ($n=2$), Telugu ($n=1$), Turkish ($n=7$), Urdu ($n=1$), and Vietnamese ($n=1$). Note that these numbers do not add up to the total number of participants ($n=216$), as several families reported to speak more than one of these languages next to or instead of Dutch. Combinations of English and one or two other languages within the home were also frequent ($n=70$). In slightly over half of the families, Dutch was one of the languages spoken by parents ($n=110$); the remaining children ($n=106$) were exposed to other languages at home, and learned Dutch at daycare. Thus, our bilingual group was vastly heterogeneous, and spanned the whole bilingual continuum, with on the one end, children from monolingual Dutch homes who were exposed to English (besides Dutch) at daycare, and on the other end, children who were exposed to two or three languages at home (e.g., French, German) and to two additional languages – Dutch and English – at daycare.

For a subset of the children (143/216, 66%), detailed information about children's language backgrounds was obtained through an extensive parental questionnaire. In this questionnaire, parents provided information, amongst others, about the frequency with which each language was spoken to their child and which language(s) children spoke themselves. Parental education was also assessed, showing that 139 of the children (97%) had at least one parent who had completed higher professional education or university.

## 2.2 Materials

### 2.2.1 *Language-specific nonword repetition task (LS-NWR)*

The LS-NWR task was taken from earlier research on Dutch-speaking children (Verhagen et al. 2019), and adapted in two ways. First, whereas the original task contained different versions depending on children's age (i.e., different versions for two-, three-, four-, and five-year-olds), the current task contained the collapsed set of items for two- and three-year-olds. Second, while the original task contained items that were either low or high in phonotactic probability, for the current study, only the items with high-phonotactic probability were retained. In so doing, we capitalized on the language-specificity of the task, to obtain a maximal contrast with the CL-NWR task.

The LS-NWR consisted of two practice items and twelve test items: four one-syllable nonwords, four two-syllable nonwords, and four three-syllable nonwords (see Table A1 in Appendix A). All items were composed of phonemes that are known to be acquired early by Dutch children, to minimize articulation difficulties (Beers, 1995; Fikkert, 1994). Furthermore, items did not contain consonant clusters and had maximally one diphthong. Stress was always on the penultimate syllable, following the typical stress pattern in Dutch (Daelemans, Gillis & Durieux, 1994). All items had been pre-recorded in a soundproof room by a female speech therapist using a high pitch voice that is typical of child-directed speech.

The task was administered using a procedure designed to keep children engaged in the task: Children watched short video clips on a laptop in which a novel object appeared from a picture of a box. They then heard a prerecorded sentence labeling the object that encouraged them to repeat the nonword (i.e., "Look, a [X]! Say [X]!"). The purpose of playing movie clips and pre-recorded speech was to keep children engaged in the 'game', while also ensuring uniformity of input in terms of rate, pitch, and other phonetic and auditory features that might vary across and within research assistants.

Items were presented in a fixed order, in increasing length, starting with the one-syllable items, followed by the two-syllable items, and, finally, the three-syllable items. This order was chosen in order to minimize data loss, that is, to prevent as much as possible that children would stop repeating the nonwords upon hearing long words. Specifically, our earlier work with toddlers (Verhagen et al., 2019) had shown that if children refuse to repeat an item they often also refuse any subsequent items, so we avoided presenting long items in the first few items of the test, not to discourage the toddlers from repeating the items. If children did not repeat an item, experimenters used the prompt sentence 'What is that?' to elicit a response from the child without repeating the target item. If children still did not repeat the item, the recording was played again, with a maximum of two repetitions. While such repetitions have been used in earlier studies with young children (Hoff, Core & Bridges, 2008; Roy & Chiat, 2004), a more common procedure in NWR tasks is to present children with a single opportunity to repeat the item. Other differences between our task and more typical NWR procedures are that nonwords were presented in a sentence frame rather than in isolation, and were associated with a novel object (but see Hoff et al. 2008 for a similar procedure with two-year-olds).

Children's responses were scored online by trained assistants as either 'correct', 'incorrect', 'no response', or 'not codable'. Responses were scored as correct if they contained all phonemes of the target in the correct order with no additional phonemes (Chiat & Roy, 2007; Gathercole, Willis, Emslie & Baddeley, 1994).

Prosodic changes in children's responses were not taken into account in scoring. An independent rater scored 10% of the data on the basis of video recordings, yielding 81% agreement between the online and offline codes. The intraclass correlation coefficient (absolute) showed substantial agreement (0.76). Internal consistency of the task was sufficient (Cronbach's alpha = .70).

### 2.2.2 *Cross-linguistic nonword repetition task (CL-NWR)*

This task was adapted from the CL-NWR task developed by Chiat (2015) within COST Action IS0804 (Language Impairment in Multilingual Society: Linguistic Patterns and the Road to Assessment), and used with five- and six-year-olds by Boerma and colleagues (2015). The items in the CL-NWR are composed of CVCV sequences and contain a limited set of consonants and vowels that occur in many languages (Chiat, 2015). For each item, four to six variants are available from which a selection can be made. These variants are matched for length, syllable structure, and segmental categories, and enable researchers to select the most appropriate item in case an item is an existing word or one of the phonemes does not occur in the language at stake. The phonemes that the items are composed of carry the phonetic features of the selected language – in this case Dutch. Language-specific prosodic patterns in the items were avoided by placing equal stress on each syllable, resulting in equal length and pitch of each syllable, apart from the final syllable, which characteristically marks the end of an utterance (see Chiat, 2015 for more details).

To make the task appropriate for toddlers, three adaptations were made to the task used by Boerma and colleagues (2015) – see Table A2 in Appendix A for an overview of the items. First, all four- and five-syllable items were left out, as these were too difficult for the current age group, so only the two and three-syllable items were retained. Monosyllabic items were not included in the CL-NWR, because these were expected to be too easy even for the current age group, given that the bilingual five-year-olds in Boerma et al. (2015) performed close to ceiling on the two-syllable items (88% correct) – considerably higher than on the monosyllabic items in the LS-NWR in that study (53% correct). Second, the /z/ phoneme that is known to be difficult to pronounce for young children was replaced with /m/ (i.e., /mibulɑ/ instead of /zibulɑ/). Third, one item (i.e., /nɑki/) was replaced with another item (i.e., /mikɑ/), because it is an existing word in Dutch (meaning nudie). Subsequently, all stimuli were re-recorded to obtain recordings that sounded more natural (i.e., less-alien like), since previous work with four-year-olds had shown that children felt sometimes intimidated by the alienated voice used in the original stimuli (Boerma, personal communication). In these new recordings, intonation was kept as flat as possible, as in the original CL-NWR task. While care was taken to create a task that was optimally language-neutral, it is not inconceivable that particular phonemes in the task do not exist in one or more of the languages of the children in our sample.

Speech rate was very similar to that in the items in the LS-NWR task, and the same carrier sentence was used ("Look a [X]! Say [X]!"). As in the LS-NWR, responses were scored online (i.e., during the task) by trained assistants as 'correct', 'incorrect', 'no response', or 'not codable'. An independent rater coded 10% of the data on the basis of video recordings, yielding 80% agreement between the online and offline codes. The intraclass correlation coefficient (absolute) was substantial (0.76). Internal consistency of the task was sufficient (Cronbach's alpha = .67). All items of the task (i.e., recordings and video clips) can be found at https://osf.io/gdfvu/?view_only=22b16d70d62b44b8a3e0dc9ff3a890e0.

#### 2.2.3 *Dutch receptive vocabulary*

The Dutch Peabody Picture Vocabulary Test (PPVT) was used to assess receptive vocabulary (PPVT-III-NL, Dunn & Dunn, 2005). A receptive task was chosen since young children are often reluctant to speak in tests that require verbal responses as opposed to tests with low response demands such as pointing or looking, which makes it hard to derive reliable estimates of their vocabularies in expressive tasks. In the PPVT, children choose one out of four picture drawings after an orally presented word. Raw scores were computed for each child, since norm data are available for monolingual children only.

#### 2.2.4 *Knowledge of languages other than Dutch*

Parents reported on their child's language background in an electronic questionnaire. For the current study, only parents' answers to a question about childrens's' language productive use were taken (i.e., Which language(s) does your child speak?). Parents were instructed to list all the languages their child spoke irrespective of proficiency level. From this question, a variable 'number of languages other than Dutch spoken' was derived, which was the sum of all languages spoken by the child, excluding Dutch.

### 2.3 Procedure

Children were assessed individually by trained research assistants in a quiet room at their daycare centers, in two sessions. The Dutch PPVT and LS-NWR task were administered in a session in which Dutch language proficiency tests were administered and the language of communication was Dutch; the CL-NWR task in a session in which English language proficiency tests were administered and the language of communication was English. Tasks were administered in a fixed order, in which the Dutch PPVT preceded the LS-NWR task. The two sessions were conducted one to two weeks apart. Children received a sticker after each task and a small gift at the end of each session; parents received a small gift for filling out the questionnaire.

### 2.4 Data screening and analyses

Binary correct/incorrect NWR scores were used in the analyses, following earlier work showing that such scores are equally informative as scores on the phoneme level (Boerma et al. 2015; Roy & Chiat, 2004). An analysis of non-responses showed that non-response rates at the task level were as follows: 47 children (22%) did not attempt the LS-NWR, and 13 children (6%) did not attempt the CL-NWR. The remaining 156 children (72%) performed both tasks. The analysis also showed that non-responders were comparable in age to responders and that boys were somewhat overrepresented (60% in non-response groups vs. 51% in the group who attempted both tasks). Finally, the data showed that children who did not attempt the LS-NWR were less likely to be exposed to Dutch at home (17% heard Dutch at home) than children who did not attempt the CL-NWR (50%) or children that attempted both tasks (34%). For children who attempted the tasks, non-response occurred in 12% and 11% of the items for the LS-NWR and CL-NWR, respectively, which is lower than in previous NWR tasks with similarly-aged children (Brandeker & Thordadottir, 2015; Chiat & Roy 2007). In our analyses, non-responses were not taken into account, and neither were imitations that involved real words (e.g., *bloem* "flower" for *loen*), existing words that children produced (e.g. "bye!") (2% of all responses in the LS-NWR; 2.5% in the CL-NWR), and 'not codable' responses (less than 1% in each task).

To address our research questions, two generalized linear (logistic) regression models with mixed effects were carried out. The dependent variables in these models were children's binary correct/incorrect scores for each item. Both models were fully specified, including by-subject and by-item random intercepts and random slopes for the variables of interest, as well as all interactions, following recommendations to run maximal models (Barr et al. 2013; Crawley, 2012; Jaeger, 2010). Analyses were carried out in *R* (R Core Team, 2019), using the *lme4* package (Bates, Maechler, Bolker & Walker, 2015).

To answer our first question concerning the effect of Dutch vocabulary knowledge on children's scores in the LS-NWR and CL-NWR, we ran a model on children's accuracy scores for each nonword (correct vs. incorrect) with Dutch vocabulary (PPVT raw sum scores), task (LS-NWR vs. CL-NWR), item length (two-syllable vs. three-syllable nonwords), and children's age (in months) as fixed-effect factors. We were specifically interested in the interaction between task and Dutch vocabulary, as we expected that Dutch vocabulary would be more strongly associated with correct responses on the LS-NWR task than on the CL-NWR task. Scores for one-syllable items in the LS-NWR were not taken into account in the model, as such items were only used in the LS-NWR. As random effects, we included by-subject and by-item random intercepts, as well as a by-subject random slope for task*item length and by-item random slopes for Dutch vocabulary and age.

To address our second question on how children's knowledge of languages other than Dutch children related to performance in the two NWR tasks, an additional fixed-effect factor was added to the model: the number of non-Dutch languages spoken by the child. We were particularly interested in the interaction between this variable and type of NWR task, because we predicted that speaking more languages would facilitate children's performance on the CL-NWR more strongly than on the LS-NWR. Since the variable 'number of non-Dutch languages' was not continuous (i.e., ranged from 0 to 4), it was recoded to make it a binary variable with values 0 (child spoke 0 or 1 languages other than Dutch) or 1 (child spoke 2 to 4 languages other than Dutch). Thus, two groups were obtained that were relatively equal in size. The number of children in each category was as follows: 0 ($n=34$, 24%), 1 ($n=55$, 38%), 2 ($n=50$, 35%), 3 ($n=3$, 2%), 4 ($n=1$, 1%). Note that the number of non-Dutch languages spoken could be 0, even though all children in our sample were bilingual or multilingual, because parental reports showed that not all children spoke all the languages they were exposed to actively themselves. Specifically, children were assigned a 0 if they were reported to be exposed to Dutch and another language, but only spoke Dutch actively themselves. As in the previous model, Dutch vocabulary and age were added as fixed-effect factors. As random effects, we included by-subject and by-item random intercepts, as well as a by-subject random slope for task and item length, and by-item random slopes for number of non-Dutch languages, and Dutch vocabulary.

In both models, orthogonal sum-to-zero contrast coding was applied to our binary fixed effects (i.e., task, item length, number of non-Dutch languages) and all continuous variables were centered around zero (Schad et al. 2020). To solve convergence issues, we increased the number of iterations to 100,000 (Powell, 2009). Prior to running our models, we computed the bivariate correlations between NWR scores and vocabulary, as well as among the predictor variables, to see whether there were indications of multicollinearity in our data.

## 3. Results

### 3.1 Descriptive statistics and correlations

Table 1 presents mean proportions correct for both NWR tasks for each item length. These data show that children's scores on the CL-NWR are higher than on the LS-NWR. Standard deviations are generally large, as is common in NWR tasks in young children (Hoff et al. 2008; Parra et al. 2011). Table 2 presents descriptive statistics for the variables representing Dutch vocabulary and the number of languages other than Dutch children spoke. The number of responses for the second variable was lower than for the first, due to missings in the parent questionnaire (see "Participants"). Note also that, in Table 2, the number of non-Dutch languages spoken can be 0, even though all children in our sample were exposed to more than one language, because parents reported that not all children spoke all the languages they were exposed to actively themselves.

**Table 1.** Descriptive statistics (proportions correct) for the LS-NWR and CL-NWR tasks ($n=153$ for LS-NWR, $n=194$ for CL-NWR)

| | LS-NWR | | CL-NWR | |
|---|---|---|---|---|
| | *M* | *(SD)* | *M* | *(SD)* |
| 1-syllable items | 0.66 | (0.48) | – | – |
| 2-syllable items | 0.44 | (0.50) | 0.73 | (0.49) |
| 3-syllable items | 0.40 | (0.49) | 0.51 | (0.50) |

**Table 2.** Descriptive statistics for Dutch vocabulary and the number of languages other than Dutch spoken by children

| | *M* | *SD* | *min-max* | *n* |
|---|---|---|---|---|
| Dutch vocabulary (Dutch PPVT) | 31.28 | 14.54 | 0–85 | 216 |
| Number of languages other than Dutch | 1.17 | 0.84 | 0–4 | 143 |

Table 3 presents bivariate correlations between these variables, as well as with age. These data show that Dutch vocabulary correlated positively and significantly with both NWR tasks, but more strongly with the LS-NWR ($r=.47$) than CL-NWR ($r=.16$). There was a significant and negative correlation between Dutch vocabulary and the number of languages other than Dutch spoken by the child ($r=-.36$). The data show, furthermore, that the number of non-Dutch languages spoken was positively and weakly correlated with children's performance on the NWR tasks ($r=.23$ for LS-NWR; $r=.16$ for CL-NWR). Finally, age correlated significantly and positively with Dutch vocabulary ($r = .46$) and with both NWR tasks ($r = .42$ for LS-NWR; $r = .24$ for CL-NWR), but not with the number of other languages spoken by the child ($r = -.05$). Since all correlations between the predictor variables were weak to moderate, no multicollinearity was assumed.

**Table 3.** Bivariate correlations for all measures

| | 1. | 2. | 3. | 4. | 5. |
|---|---|---|---|---|---|
| 1. Dutch vocabulary | – | $-.36^{***}$ | $.47^{***}$ | $.16^{*}$ | $.46^{***}$ |
| 2. Number of languages other than Dutch | | – | $.23^{*}$ | $.16^{*}$ | −.05 |
| 3. LS-NWR | | | – | $.29^{**}$ | $.42^{**}$ |
| 4. CL-NWR | | | | – | $.24^{**}$ |
| 5. Age (in months) | | | | | – |

*Note.*
* $p < .05$ ** $p < .01$ *** $p < .001$, $n$ ranges between 122 and 216

## 3.2 Effects of long-term language knowledge on NWR performance

To investigate the effect of Dutch vocabulary knowledge on children's performance on the NWR tasks, we constructed a model assessing the likelihood of a correct NWR response given the child's Dutch vocabulary score, the child's age, the type of NWR task, and the length of the nonword items. This model showed a main effect of Dutch vocabulary ($\beta = 0.296$, SE = 0.081, $z = 3.653$, $p < .001$), which indicated that children with higher Dutch vocabulary scores were more likely to provide a correct response on the NWR tasks than children with lower Dutch vocabulary scores. The model also showed main effects of task ($\beta = 1.310$, SE = 0.292, $z = 4.492$, $p < .001$), item length ($\beta = -0.825$, SE = 0.266, $z = -3.103$, $p = .002$), and age ($\beta = 0.078$, SE = 0.018, $z = 4.288$, $p < .001$). These effects indicated that children were more likely to repeat the CL-NWR items accurately than the LS-NWR items, and more likely to repeat shorter nonwords accurately than longer nonwords. Older children were more likely to provide an accurate response than younger children. The model also yielded two interaction effects. First, the interaction between Dutch vocabulary and task approached significance, which indicated that the effect of Dutch vocabulary tended to be larger for the LS-NWR than CL-NWR ($\beta = -0.266$, SE = 0.138, $z = -1.929$, $p = .054$). Second, there was a significant interaction between item length and vocabulary, which indicated that the effect of item length was stronger for children with higher vocabulary scores than for children with lower vocabulary scores ($\beta = 0.242$, SE = 0.104, $z$ = 2.330 $p$ = .020). The remaining interactions were not significant. See Table B1 in Appendix B for the full results.

To address our second question regarding the effect of knowledge of languages other than Dutch, an additional fixed-effect factor was included in the model: the number of languages other than Dutch spoken by the child. The aim of this analysis was to see whether this factor was differentially related to performance on the CL-NWR and LS-NWR. This model, run on only those children whose parents had filled out the electronic questionnaire ($n$ = 143), showed a significant main effect of the number of languages spoken other than Dutch ($\beta$ = 0.595, SE = 0.295, $z$ = 2.018, $p$ = .044), which indicated that children who spoke two or more languages besides Dutch at home were more likely to repeat the NWR items accurately than children who spoke no or only one language besides Dutch, irrespective of task. As in the previous model, there was a main effect of task ($\beta$ = 0.970, SE = 0.323, $z$ = 3.004, $p$ = .003), as well as of item length ($\beta$ = −0.696, SE = 0.285, $z$ = −2.445, $p$ = .015), Dutch vocabulary ($\beta$ = 0.041, SE = 0.012, $z$ = 3.509, $p$ < .001), and age ($\beta$ = 0.062, SE = 0.029, $z$ = 2.145, $p$ = .032). There were no significant interactions in this model (see Table B2 in Appendix B for the full results).[1]

## 4. Discussion

In this study, we investigated how individual differences in long-term language knowledge related to young bilingual and multilingual children's performance on two NWR tasks: a Cross-Linguistic nonword repetition task (CL-NWR) and a Language-Specific (Dutch-based) nonword repetition task (LS-NWR). The overall aim of our study was to examine whether performance on the CL-NWR, which is based on the linguistic properties of many languages, would be related differentially to children's levels of existing language knowledge in a highly diverse sample of bilingual and multilingual toddlers. We considered two factors representing children's long-term language knowledge: (i) Dutch vocabulary and (ii) the number of languages other than Dutch children spoke. We predicted that the first factor (i.e., Dutch vocabulary) would be related to children's scores on the LS-NWR more strongly than on the CL-NWR, whereas the second factor (i.e., number of

1. Since items were presented in order of increasing length, item order and item length partially overlapped in our task, thus potentially creating a confound. To assess whether item order affected our outcomes, we re-ran our models with an additional fixed-effect factor: item order. The results of these models showed that item order did not have an effect on repetition accuracy and that the effects of the other factors were overall highly similar to those of the models in which item order was not included.

languages other than Dutch) would be related to the CL-NWR more strongly than to the LS-NWR.

As for the first factor, our results showed that Dutch vocabulary indeed correlated more strongly with children's scores on the Dutch-based LS-NWR than on the CL-NWR. A mixed-effect regression analysis in which children's age and item length were taken into account showed a marginally significant interaction between Dutch vocabulary and task that confirmed this pattern: the effect of Dutch vocabulary as a positive predictor of children's NWR accuracy tended to be larger for the LS-NWR than for the CL-NWR. Taken together, these correlational and regression results suggest that performance in the CL-NWR is less contingent on children's long-term lexical and sub-lexical knowledge of Dutch than their performance in a Dutch-based NWR task, already at toddler age. As such, the CL-NWR task seems to allow for a more language-neutral assessment than the Dutch-based LS-NWR task, at least in its current form and with the current group of bilingual and multilingual toddlers. Creating such a level playing field, the CL-NWR task may provide a better alternative to assessing NWR in bilinguals' two languages, as has been recommended earlier (Gutiérrez-Clellen & Simon-Cereijido, 2010).

Our results supplement those in earlier work by Boerma and colleagues (2015), who compared a (Dutch-based) LS-NWR and CL-NWR in monolingual and bilingual groups of five to six-year-old children. Comparing Dutch vocabulary between these groups, these authors found that the bilinguals had lower Dutch vocabulary scores than the monolingual group. Group comparisons also indicated that the bilinguals scored more poorly than the monolinguals on the LS-NWR, but not on the CL-NWR. However, unlike in the current study, relationships between vocabulary knowledge and NWR performance were not examined. Hence, the current study adds to this earlier study by providing more direct evidence that children's existing language knowledge (Dutch vocabulary) affects children's repetition of CL-NWR items, and, in fact, indicates that tasks that draw on language features that are present in many languages create a more level playing field across monolingual and bilingual populations, as well as within such populations. In addition, the current results indicate that this differential effect of vocabulary on performance in the two NWR tasks is found in children as young as two or three years of age.

Regarding the second factor related to long-term knowledge investigated in our study – knowledge of languages other than Dutch – our analyses showed that children who spoke two or more languages besides Dutch performed more accurately on the NWR tasks than children who spoke no or one language besides Dutch, irrespective of task. This result is not in line with our prediction that knowledge of other languages than Dutch would support children's performance on the CL-NWR in particular.

Several explanations of this result are possible. First, children who were reported to speak two or more other languages than Dutch may have been the ones who also spoke Dutch rather well, which might explain their relatively high scores on the LS-NWR. This account is unlikely, however, given that we found a negative correlation between the number of non-Dutch languages spoken by children and Dutch vocabulary in our data ($r = -.36$). A second possible explanation is that the direction of the relationship between children's language knowledge and NWR performance is reversed: children with well-developed verbal short-term memory skills would learn to speak several languages more readily than children with less well-developed verbal short-term memory skills. While we cannot exclude this possibility, it does not seem likely either, given that the number of languages children spoke, as reported by their parents, overlapped considerably with the languages spoken in their homes (i.e., 70% of the children were reported to either speak all or all but one of the languages spoken at home). This suggests that whether children spoke a given language was largely a function of receiving input in this language at home (rather than to some third factor, such as verbal-short term memory ability), in line with a wealth of studies showing moderate to strong associations between parental input and proficiency in bilingual children (e.g., Hoff et al. 2012; Unsworth, 2015). A final possibility is that the items in the LS-NWR task overlapped to some extent with the phonological properties of the other languages children spoke (e.g., German or French), such that children's knowledge of these other languages facilitated their repetition of the LS-NWR items. On the basis of the current data, we cannot test this idea. Our 'number of non-Dutch languages' variable was a crude measure, which did not take into account which actual languages were spoken by the child and how well children spoke these languages. Therefore, the exact role that children's knowledge of non-Dutch languages played in their repetition of the CL-NWR and LS-NWR items remains unclear.

In addition to the effects of children's long-term language knowledge on CL-NWR and LS-NWR that were directly relevant to our research questions, our analyses showed a number of other effects. First, children performed significantly better on the CL-NWR than on the LS-NWR, as in earlier studies using slightly different versions of this task with older children (Boerma et al. 2015; Chiat & Polišenská, 2016, but see Antonijevic-Elliott et al. 2020). This is likely due to the fact that phonological properties that are shared across many languages are typically perceptually more readily distinguished and easier to produce (Haspelmath, 2006; Toombs, Singh & Hayden, 1981).

A second effect that we found was that children with higher vocabulary levels performed better on the NWR tasks than children with lower vocabulary levels, irrespective of task. As outlined in the introduction, effects of long-term knowledge on NWR (in particular, vocabulary knowledge) have been observed across studies (Gathercole, 2006; Jones, 2016). To explain such effects, two main ideas have been put forth. First, on the assumption that NWR tasks mainly tap phonological storage (i.e., the ability to store speech sounds in verbal short-term memory), it has been proposed that the more long-term knowledge participants have available, the more readily they can restore decayed material in verbal short-term memory. Specifically, this idea holds that material stored in short-term memory decays rapidly and that long-term lexical representations that best match the short-term memory trace are used to restore the memory trace: hence, the more lexical representations are available, the easier it is to restore decayed material (Thorn, Gathercole & Frankish, 2005). A second proposal is that children develop more detailed phonological representations as their vocabularies expand, since their growing lexicons increasingly require high-quality phoneme representation to store words and distinguish between them. The higher the quality of these phonological representations, the easier it is to process and repeat novel words (Metsala, 1999; Rispens & Baker, 2012). Thus, on both accounts, increases in vocabulary knowledge result in higher accuracy on a task that assesses both verbal short-term memory and phonological processing abilities, such as NWR. Interestingly, in the current study, the effect of Dutch vocabulary was also found for repetition accuracy on the CL-NWR task, which, at least in part, might be due to the fact that phonemes were produced with the phonetic properties of Dutch, besides the involvement of the memory-based processes described above.

Finally, in our study, effects were found for factors that are commonly reported to affect NWR accuracy in young children: age (Coady & Evans, 2008; Gathercole & Baddeley, 1989) and item length (Gathercole, 2006). Older children were more likely to repeat the NWR items accurately than younger children, and shorter items were more likely to be accurately responded to than longer items. In our study, the effect of item length was stronger for children with lower vocabulary scores than for children with higher vocabulary scores. This interaction effect fits well with earlier observations that children of low verbal ability are more sensitive to word length than children of higher verbal ability (Briscoe, Bishop & Norbury, 2001; Gathercole & Baddeley, 1990). Specifically, earlier work (on monolingual children) has shown that children with low verbal ability (e.g., language disorders or low vocabulary scores for their age) have more problems when repeating longer nonwords than children with higher verbal ability, but not necessarily when repeating shorter nonwords. Since effects of item length on NWR performance are generally considered effects of phonological storage, these results have been taken to suggest that poorer word learners have more limited phonological storage capacities than better word learners (Gathercole, 2006). The finding that, in our study, the interaction was found for tasks varying in language-specificity lends further support to the idea that differences in children's word learning ability are, at least in part, due to differences in phonological storage abilities, rather than some property of the longer or shorter nonword items in the task. Note, however, that only two- and three-syllable items were presented (with only four nonwords per category), which limits the conclusion that can be drawn about the effects of item length. Furthermore, since shorter words are more frequent across languages than longer words (Zipf, 1935), effects of item length might also reflect familiarity with two-syllable items, rather than merely phonological storage.

There are a number of limitations to our study. First, the current participants were from families in which parents were highly educated, and thus are not representative of the larger population. Second, in our study, we did not look into effects of language disorders on performance on the two NWR tasks. In fact, it cannot be excluded that some children in our sample had (an increased risk of) language disorders, but, given their very young age, were not yet diagnosed as such. Since CL-NWR tasks hold the most promise for use with multilingual children with language disorders, future work could compare performance across CL-NWR and LS-NWR tasks in toddlers with an increased risk of language disorders or language delays, to see whether CL-NWR tasks have better sensitivity and specificity than LS-NWR tasks. Finally, in our study, a rather global analysis was performed, rather than a detailed comparison of the effects of children's specific language backgrounds and proficiency levels in each language. Future work could assess to what extent children from different language backgrounds and proficiency levels in their respective languages perform differently on CL-NWR tasks. Specifically, future studies could adopt detailed error analytical approaches taking into account the phonological and phonotactic properties of the nonwords, to see whether children from different language backgrounds (i.e., either specific languages, or language families) pattern together with respect to the errors made. In such studies, expressive measures (i.e., real word repetition tasks) could be included, moreover, to see whether children are able to produce the phonemes in the NWR task.

To conclude, the aim of the current study was to examine how differences in long-term language knowledge related to performance on a CL-NWR task as opposed to a LS-NWR task *within* a highly diverse group of bilingual and multilingual children. To investigate this, we used a CL-NWR task based on earlier tasks used with slightly older children (Boerma et al. 2015; Chiat, 2015) and made it appropriate for use with toddlers. Our results indicated that performance on the CL-NWR was affected by children's existing language knowledge less strongly than their performance on the LS-NWR. Future research with bilingual children with and without language disorders is needed to establish whether the results generalize to children from less high-educated families, whether the task can be used for clinical purposes, and to what extent performance on the CL-NWR task might be subject to more fine-grained influences from children's knowledge of specific languages.

## Funding

Research funded by This work was supported by the Dutch Ministry of Social Affairs and Employment under grant number 201700273.510. to Sible Andringa.

## References

Adams, A. M., & Gathercole, S. E. (1996). Phonological memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology*, *49*, 216–233. https://doi.org/10.1080/713755610

Archibald, L. (2008). The promise of nonword repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology*, *32*, 21–28.

Antonijevic-Elliott, S., Lyons, R., Pat O' Malley, M., Meir, N., Haman, E., Banasik, N., Carroll, C., McMenamin, R., Rodden, M., & Fitzmaurice, Y. (2020). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks, *Clinical Linguistics & Phonetics*, *34*, 293–311. https://doi.org/10.1080/02699206.2019.1637458

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effect models using lme4. *Journal of Statistical Software*, *67*, https://doi.org/10.18637/jss.v067.i01

Beers, M. (1995). The phonology of normally developing and language-impaired children. Doctoral dissertation, University of Amsterdam.

Boerma, T., Chiat, S., Leseman, P. P. M., Timmermeister, M., Wijnen, F. N. K., & Blom, W. B. T. (2015). A cross-linguistic nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language and Hearing Research*, *58*, 1747–1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058

Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers: Performance on nonword repetition and lexical tasks. *American Journal of Speech-Language Pathology*, *24*, 126–138. https://doi.org/10.1044/2015_AJSLP-13-0106

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry*, *42*, 329–340. https://doi.org/10.1111/1469-7610.00726

Chiat, S. (2015). Nonword repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 125–150). Multilingual Matters. https://doi.org/10.21832/9781783093137-008

Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language and Hearing Research*, *59*, 1179–1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, *50*, 429–443. https://doi.org/10.1044/1092-4388(2007/030)

Coady, J.A., & Aslin, R.N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, *89*, 183–213. https://doi.org/10.1016/j.jecp.2004.07.004

Coady, J.A., & Evans, J.L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairment (SLI). *International Journal of Communication Disorders*, *43*, 1–40. https://doi.org/10.1080/13682820601116485

Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: A study of young language-learning children. *International Journal of Language and Communication Disorders*, *38*, 251–263. https://doi.org/10.1080/1368282031000092339

Crawley, M.J. (2012). *The R book*. John Wiley & Sons. https://doi.org/10.1002/9781118448908

Daelemans, W., Gillis, S., & Durieux, G. (1994). The acquisition of stress: A data-oriented approach. *Computational Linguistics*, *20*, 421–451.

Dispaldro, M., Leonard, L.B., Deevy, P. (2013). Real word and nonword repetition in Italian-Speaking children with specific language impairment: A study of diagnostic accuracy, *Journal of Speech, Language, and Hearing Research*, *56*, 323–336. https://doi.org/10.1044/1092-4388(2012/11-0304)

Duncan, T.S., & Paradis, J. (2013). English language learners' nonword repetition performance: The influence of age, L2, vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech Language and Hearing Research*, *59*, 39–48. https://doi.org/10.1044/2015_JSLHR-L-14-0020

Dunn, L., & Dunn, L.M. (2005). *Peabody Picture Vocabulary Test-III-NL. Nederlandse versie door Liesbeth Schlichting* [Dutch version by Liesbeth Schlichting]. Harcourt Assessment B.V.

Ebert, K.D., Pham, G., Kohnert, K. (2014). Lexical profiles of bilingual children with primary language impairment. *Bilingualism*, *17*, 766–783. https://doi.org/10.1017/S1366728913000825

Engel de Abreu, P.M.J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, *19*, 529–537. https://doi.org/10.1080/09658211.2011.590504

Engel de Abreu, P.M., Baldassi, M., Puglisi, M.L., & Befi-Lopes, D.M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, *56*, 630–642. https://doi.org/10.1044/1092-4388(2012/12-0079)

Fikkert, P. (1994). On the acquisition of prosodic structure. Doctoral dissertation, University of Leiden.

Gathercole, S.E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*, *23*, 83–94. https://doi.org/10.3758/BF03210559

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*, 200–213. https://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S., & Baddeley, A. (1990). Phonological memory deficits in language disordered s children: Is there a causal connection? *Journal of Memory and Language*, *29*, 336–360. https://doi.org/10.1016/0749-596X(90)90004-J

Gibson, T., Summers, C., Peña, E. D., Bedore, L. M., Gillam, R. B., & Bohman, T. (2015). The role of language structure and experience in bilingual children's nonword repetition performance. *Bilingualism: Language and Cognition*. *18*, 551–560. https://doi.org/10.1017/S1366728914000248

Graf-Estes, K., Evans, J. L., Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *Journal of Speech Language and Hearing Research*, *50*, 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Gray, S. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, *36*, 129–151. https://doi.org/10.1016/S0021-9924(03)00003-0

Grimm, A., & Schulz, P. (2014). Specific language impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indicators Research*, *7*, 821–84. https://doi.org/10.1007/s12187-013-9230-6

Gutiérrez-Clellen, V. F., & Simon-Cerejido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English-speaking children: Does the language of assessment matter? *Learning Disabilities Research & Practice*, *25*, 48–58. https://doi.org/10.1111/j.1540-5826.2009.00300.x

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*, 25–70. https://doi.org/10.1017/S0022226705003683

Hoff, E., Core, C., & Bridges, K. (2008). Nonword repetition assesses phonological memory and is related to vocabulary development in one-year-olds. *Journal of Child Language*, *35*, 903–916. https://doi.org/10.1017/S0305000908008751

Hoff, E., Core, C., Place, S., Rumiche, R., Senor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*, 1–27. https://doi.org/10.1017/S0305000910000759

Jones, G. (2016). The influence of children's language exposure to language from two to six years: The case of nonword repetition. *Cognition*, *153*, 79–88. https://doi.org/10.1016/j.cognition.2016.04.017

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition*, *15*, 470–489. https://doi.org/10.1017/S1366728911000472

Keydeniers, D., Aalberse, S., Andringa, S., & Kuiken, F. (2021). Bilingual daycares in the Netherlands: an analysis of the implementation of bilingual input and underlying ideologies. *Current Issues in Language Planning*. https://doi.org/10.1080/14664208.2021.1939988

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice*, *21*, 19–29. https://doi.org/10.1111/j.1540-5826.2006.00204.x

Masoura, E.V. & Gathercole, S.E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology*, *34*, 383–388. https://doi.org/10.1080/002075999399738

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, *8*: 1442. https://doi.org/10.3389/fpsyg.2017.01442

Messer, M.H., Leseman, P.P.M., Boom, J., & Mayo, A.Y. (2010). Phonotactic probability effect in nonword recall and its relation with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, *105*, 306–323. https://doi.org/10.1016/j.jecp.2009.12.006

Metsala, J.L., & Chisholm, G.M. (2010). The influence of lexical status and neighborhood density on children's nonword repetition. *Applied Psycholinguistics*, *31*, 489–506. https://doi.org/10.1017/S0142716410000081

Munson, B., Kurtz, B.A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *48*, 1033–1047. https://doi.org/10.1044/1092-4388(2005/072)

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology*, *108*, 113–125. https://doi.org/10.1016/j.jecp.2010.07.011

Powell, M.J.D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. Seminar scholar. https://pdfs.semanticscholar.org/0d2e /dc46f81f9a0b0b62937507ad977b46729f64.pdf?_ga=2.268128462.1175451878.1541705909-955068333.1521386318

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rispens, J., & Baker, A. (2012). Nonword repetition: the relative contributions of phonological short-term memory and phonological representations in children with language and hearing impairment. *Journal of Speech, Language, and Hearing Research*, *55*, 683–694. https://doi.org/10.1044/1092-4388(2011/10-0263)

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2-to-4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, *47*, 223–234. https://doi.org/10.1044/1092-4388(2004/019)

Schad, D.J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Thorn, A., & Gathercole, S.E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology Section A*, *52*, 303–324. https://doi.org/10.1080/713755823

Thorn, A., Gathercole, S.E., & Frankish, C.R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology*, *50*, 133–158. https://doi.org/10.1016/j.cogpsych.2004.07.001

Toombs, M., Singh, S., & Hayden, M. (1981). Markedness of features in the articulatory substitutions of children. *Journal of Speech and Hearing Disorders*, *46*, 184–191. https://doi.org/10.1044/jshd.4602.184
Unsworth, S. (2015). Amount of exposure as a proxy for dominance in bilingual language acquisition. In C. Silva-Corvalan & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 156–173). Cambridge University Press.
Verhagen, J. & Leseman, P.P.M. (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
Verhagen, J., & Boom, J., Mulder, H., de Bree, E.H., & Leseman, P.P.M. (2019). Reciprocal relationships between nonword repetition and vocabulary during the preschool years. *Developmental Psychology*, *55*, 1125–1137. https://doi.org/10.1037/dev0000702
White, M.J. (2019). The development of English proficiency and working memory in 5–6 year old ELLs in their first year of formal education, *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2019.1571009
Windsor, J., Kohnert, K., Lobitz, K.F., & Pham, G.T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, *19*, 298–310. https://doi.org/10.1044/1058-0360(2010/09-0064)
Zipf, G.K. (1935). (reprinted 1965). *The psycho-biology of language*. MIT Press.

## Appendix A. Items in the NWR tasks

**Table A1.** Items of the Language-Specific NWR Task

| Item length | Item | IPA |
|---|---|---|
| 1-syllable | loen | /lun/ |
| | jaat | /jɑt/ |
| | seun | /søn/ |
| | peek | /pek/ |
| 2-syllable | hiemup | /'himʏp/ |
| | natep | /'nɑtɛp/ |
| | holin | /'holIn/ |
| | kepon | /'kepOn/ |
| 3-syllable | liepoetaan | /li'put ɑn/ |
| | pelanot | /pe'l ɑnOt/ |
| | sitalon | /si't ɑlOn/ |
| | jakotis | /ja'kotIs/ |

**Table A2.** Items of the Cross-Linguistic NWR Task (adapted from Boerma et al. 2015)

| Item length | Item | IPA |
|---|---|---|
| 2-syllable | lietaa | /lita/ |
| | noelie | /nuli/ |
| | miekaa | /mika/ |
| | sieboe | /sibu/ |
| 3-syllable | naaliedoe | /nalidu/ |
| | baamoedie | /bamudi/ |
| | loemiekaa | /lumika/ |
| | mieboelaa | /mibula/ |

## Appendix B. Overview of Results from Mixed-Effects Models

**Table B1.** Results of a Generalized Mixed-Effects Model on Children's Correct/Incorrect Responses in the NWR Tasks with Task (LS-NWR vs. QU-NWR), Item Length (2 Syllables vs. 3 Syllables), and Dutch Vocabulary (PPVT Sum Scores), and Age as Fixed-Effect Factors

| | $\beta$ | SE | $z$ | $p$ |
|---|---|---|---|---|
| Intercept | 0.122 | 0.157 | 0.775 | .438 |
| Task | 1.310 | 0.292 | 4.492 | <.001 |
| Item length | −0.825 | 0.266 | −3.103 | .002 |
| Dutch vocabulary | 0.296 | 0.081 | 3.653 | <.001 |
| Age | 0.078 | 0.018 | 4.288 | <.001 |
| Task*Item length | −0.944 | 0.525 | −1.798 | .072 |
| Task*Dutch vocabulary | −0.266 | 0.138 | −1.929 | .054 |
| Item length*Dutch vocabulary | 0.242 | 0.104 | 2.330 | .020 |
| Task*Age | 0.006 | 0.031 | 0.184 | .854 |
| Item length*Age | 0.018 | 0.024 | 0.767 | .443 |
| Dutch vocabulary*Age | −0.014 | 0.011 | −1.252 | .211 |
| Task*Item length*Dutch vocabulary | 0.080 | 0.196 | 0.406 | .685 |
| Task*Item length*Age | −0.013 | 0.046 | −0.294 | .769 |
| Task* Dutch vocabulary*Age | −0.003 | 0.019 | −0.144 | .886 |
| Item length*Dutch vocabulary*Age | −0.015 | 0.015 | −0.990 | .322 |
| Task*Item length*Dutch vocabulary*Age | −0.001 | 0.028 | −0.053 | .958 |

*Note*. This model was based on 2366 observations, 210 subjects, and 16 items.

**Table B2.** Results of a Generalized Mixed-Effects Model on Children's Correct/Incorrect Responses in the NWR Tasks with Number of Languages Other than Dutch, Task, Item Length, Dutch Vocabulary and Age as Fixed-Effect Factors

| | *β* | SE | *z* | *p* |
|---|---|---|---|---|
| Intercept | 0.118 | 0.182 | 0.651 | .515 |
| Item length | −0.696 | 0.285 | −2.445 | .015 |
| Task | 0.970 | 0.323 | 3.004 | .003 |
| Number of non-Dutch languages | 0.595 | 0.295 | 2.018 | .044 |
| Age | 0.062 | 0.029 | 2.145 | .032 |
| Dutch vocabulary | 0.041 | 0.012 | 3.509 | <.001 |
| Item length*Task | −0.855 | 0.557 | −1.536 | .125 |
| Item length*Number of non-Dutch languages | 0.503 | 0.387 | 1.300 | .194 |
| Task*Number of non-Dutch languages | −0.545 | 0.489 | −1.114 | .265 |
| Item length*Age | −0.021 | 0.039 | −0.527 | .598 |
| Task*Age | 0.006 | 0.049 | 0.113 | .910 |
| Number of non-Dutch languages*Age | 0.006 | 0.055 | 0.110 | .912 |
| Item length*Dutch Vocabulary | 0.028 | 0.016 | 1.769 | .077 |
| Task*Vocabulary | −0.019 | 0.020 | −0.996 | .319 |
| Number of non-Dutch languages*Vocabulary | −0.026 | 0.023 | −1.152 | .249 |
| Age*Vocabulary | −0.002 | 0.002 | −0.997 | .319 |
| Item length*Task*Number of non-Dutch languages | −1.063 | 0.736 | −1.444 | .149 |
| Item length*Task*Age | −0.021 | 0.075 | −0.280 | .779 |
| Item length*Number of non-Dutch languages*Age | 0.014 | 0.093 | 0.152 | .879 |
| Item length*Task*Dutch vocabulary | −0.031 | 0.030 | −1.029 | .304 |
| Item length*Number of non-Dutch languages*Dutch vocabulary | 0.023 | 0.029 | 0.793 | .428 |
| Task*Number of non-Dutch languages*Dutch vocabulary | 0.011 | 0.038 | 0.297 | .767 |
| Item length*Age*Dutch vocabulary | −0.003 | 0.003 | −1.049 | .294 |
| Task*Age*Dutch vocabulary | −0.001 | 0.003 | −0.359 | .720 |
| Number of non-Dutch languages*Age*Dutch vocabulary | 0.000 | 0.004 | −0.030 | .976 |

| | $\beta$ | SE | $z$ | $p$ |
|---|---|---|---|---|
| Item length*Task*Number of non-Dutch languages*Age | −0.197 | 0.136 | −1.452 | .147 |
| Item length*Task*Number of non-Dutch languages*Dutch vocabulary | 0.096 | 0.055 | 1.723 | .085 |
| Item length*Task*Age*Dutch vocabulary | −0.007 | 0.005 | −1.464 | .143 |
| Item length*Number of non-Dutch languages*Age*Dutch vocabulary | −0.004 | 0.005 | −0.886 | .376 |
| Task*Number of non-Dutch languages*Age*Dutch vocabulary | −0.007 | 0.006 | −1.024 | .306 |
| Item length*Task*Number of non-Dutch languages*Age*Dutch vocabulary | −0.010 | 0.010 | −1.075 | .282 |

*Note*. This model was based on 1615 observations, 143 subjects, and 16 items.

## Address for correspondence

Josje Verhagen
University of Amsterdam
Spuistraat 134
Amsterdam 1012 VB
The Netherlands
J.Verhagen@uva.nl

## Co-author information

Sible Andringa
University of Amsterdam
s.j.andringa@uva.nl

## Publication history

Date received: 18 June 2020
Date accepted: 4 July 2021
Published online: 20 August 2021