Journal of Memory and Language 84 (2015) 24–36

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Growth of verbal short-term memory of nonwords varying in phonotactic probability: A longitudinal study with monolingual and bilingual children

Marielle H. Messer, Josje Verhagen *, Jan Boom, Aziza Y. Mayo, Paul P.M. Leseman

Utrecht University, Department of Special Education: Cognitive and Motor Disabilities, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 16 August 2014
revision received 30 April 2015
Available online 27 May 2015

Keywords:
Verbal short-term memory
Phonotactics
Latent growth modeling
Monolingual children
Bilingual children

ABSTRACT

This study investigates the hypothesis that verbal short-term memory growth in young children can be explained by increases in long-term linguistic knowledge. To this aim, we compare children's recall of nonwords varying in phonotactic probability. If our assumption holds, there should be growth in recall of high-probability nonwords, but no or less growth in recall of low-probability nonwords. Monolingual and bilingual children are compared to see if bilingual children who have less phonotactic knowledge of the target language (Dutch) show different growth patterns than their monolingual peers. Participants were 72 monolingual Dutch children and 69 bilingual Turkish-Dutch children with Dutch as their non-dominant language. Children were assessed at four, five and six years of age. At all ages, they completed serial nonword recall tasks containing Dutch-based high- and low-probability nonwords. They also performed a series of control measures, including a Dutch receptive vocabulary task. Latent Growth Modeling was used to model the data. A model with clear improvement in children's recall of high-probability nonwords, but no improvement in recall of low-probability nonwords in both groups, and equal gains of recall of high-probability nonwords in the two groups, gave good fit to the data. These results indicate that (i) verbal short-term memory growth can be explained by increases in long-term phonotactic knowledge and (ii) bilingual children with lower levels of phonotactic knowledge in the target language benefit from such knowledge to the same degree as monolingual children.

Introduction

Verbal-short term memory is not stable, but develops during childhood such that children are able to remember increasingly longer lists of nonwords or digits (Alloway, Gathercole, & Pickering, 2006; Gathercole, 1998; Gathercole, Pickering, Ambridge, & Wearing, 2004). To explain this growth, a number of factors have been proposed, including increased articulation rate enabling faster rehearsal of verbal material (Hulme, Thomson, Muir, & Lawrence, 1984), slower decay of memory traces (Cowan, Nugent, Elliott, & Saults, 2000; Gomes et al., 1999), increased memory capacity (Cowan & Alloway, 2009), or better developed executive functions (Rypma, Prabhakaran, Desmond, Glover, & Gabrieli, 1999). Another factor that has been proposed relates to the impact of long-term linguistic knowledge on verbal short-term memory performance, in particular of phonotactic knowledge, which refers to the statistical distribution of phonemes and phoneme clusters in a language (Roodenrys, Hulme, & Brown, 1993). In a series of studies, verbal short-term memory performance on nonword

---

* Corresponding author.

E-mail addresses: J.Verhagen@uu.nl (J. Verhagen), J.Boom@uu.nl (J. Boom), P.P.M.Leseman@uu.nl (P.P.M. Leseman).

http://dx.doi.org/10.1016/j.jml.2015.05.001

repetition and nonword recall tasks was enhanced for nonwords containing frequent phoneme combinations as compared to nonwords containing less frequent phoneme combinations, indicating that long-term knowledge about phoneme distributions (or phonotactic knowledge) supports short-term storage (Kovacs & Racsmany, 2008; Majerus, Van der Linden, Mulder, Meulemans, & Peters, 2004; Messer, Leseman, Boom, & Mayo, 2010; Thorn & Frankish, 2005).

Previous studies on the development of verbal short-term memory are typically cross-sectional rather than longitudinal (Alloway et al., 2006; Gathercole et al., 2004) and restricted to monolingual children. In this study, our aim is twofold. First, we examine growth of verbal short-term memory longitudinally in children aged four to six years. We hypothesize that growth in verbal short-term memory during this period can be explained by growing long-term language knowledge. If this is indeed the case, we expect to see growth in children's recall of high-probability nonwords, but no growth or much smaller growth in their recall of low-probability nonwords, as for the latter type of nonwords, not much support from long-term linguistic knowledge about phonotactics is available. Our second aim is to compare monolingual and (sequential) bilingual children to see if bilingual children show the same growth patterns as monolingual children. We hypothesize that sequential bilingual children have had less exposure to the target language, Dutch, and therefore have less (and less well entrenched) Dutch phonotactic knowledge. Previous work on the same sample as studied here has shown that phonotactic knowledge affects verbal short-term memory in bilingual children's non-dominant (or second) language, albeit not as strongly as in monolingual peers (Messer et al., 2010). However, it remains to be investigated how verbal short-term memory develops in these children as a function of growing phonotactic knowledge of their second language.

The current study looks at bilingual children's first years in a rich second language (preschool) environment, when their knowledge of the second language develops rapidly. As for recall of low-probability nonwords, we predict a similar growth rate in the monolingual and bilingual children, because, for both groups, there will be no or only very little support from long-term memory. For recall of high-probability nonwords, there is no clear hypothesis, so we consider three possible outcomes: (i) the bilingual children show a similar growth rate as the monolingual children, since phonotactic knowledge and the support thereof may develop at a similar pace in both groups, (ii) the bilingual children show a slower growth rate than the monolingual children due their comparatively smaller amount of exposure to the second language, and hence, slower development of phonotactic knowledge in this language, or (iii) they show faster growth than the monolinguals due to their immersion in a rich second language environment, causing a catch-up effect of their phonotactic knowledge of the second language, similar to well-observed catch-up effects in Dutch vocabulary in this group during the early school years (Extra, Aarts, van der Avoird, Broeder, & Yagmur, 2001).

Verbal short-term memory development

Cross-sectional studies have shown that verbal short-term memory capacity grows significantly during childhood (Alloway et al., 2006; Gathercole et al., 2004). Children are able to remember increasingly longer sequences of spoken digits and other words, from two to three items at the age of four to about six items when they are twelve years old (Gathercole, 1998). Studies on bilingual children and children learning a second language in a classroom setting have found rather similar growth patterns in verbal short-term memory capacity, both in children's first language (Chincotta & Underwood, 1997; Hu, 2003; Swanson, Saez, & Gerber, 2006) and in their second language (Chiappe, Siegel, & Wade-Woolley, 2002; Chincotta & Underwood, 1997; French & O'Brien, 2008; Service, 1992; Service & Kohonen, 1995; Swanson et al., 2006). Different factors have been proposed to explain this growth, these factors being derived from the different processes that are assumed to be involved in the recognition, encoding and storage of verbal material (for reviews, see Cowan & Alloway, 2009; Gathercole, 1998, 1999).

According to the commonly used working memory model of Baddeley and Hitch (1974), individuals store incoming verbal information temporarily in a phonological form in the storage component of the phonological loop, where the memory trace decays in about two seconds if not rehearsed. A sub-vocal rehearsal system refreshes information in the phonological loop, by rehearsing verbal information out loud or silently, as a strategy to prevent decay of memory traces. Traditionally, it was thought that developmental increases in memory capacity during childhood could be entirely explained by growth in articulation rate supporting the efficiency of (sub-)vocal rehearsal as a strategy to prevent decay. Specifically, since rehearsal takes place in real time, increases in articulation speed would result in a higher number of memory traces that can be refreshed in the phonological store (Hulme et al., 1984). However, there are indications that children use rehearsal strategies only from seven years of age onwards (Baddeley, Gathercole, & Papagno, 1998), making it unlikely that increases in articulation rate explain the substantial growth in memory performance observed during the first years of life.

A number of alternative mechanisms have been proposed to account for increases in short-term memory capacity during childhood. First, it has been assumed that the rate of decay of memory traces decreases as children grow older (Cowan et al., 2000; Gomes et al., 1999). Although Baddeley and Hitch's memory model assumes that the rate of decay is about two seconds in all individuals and does not change with age, behavioral and electrophysiological studies have shown a developmental change in the retention of verbal material in short-term memory (Cowan et al., 2000; Gomes et al., 1999). However, in these studies, children between six and ten years behaved similarly to each other (but differently from adults) in most respects, suggesting that developmental changes in decay rate are not a plausible explanation of verbal short-term memory growth, at least not in school-aged children. A second suggested explanation for growth in short term memory, which remains a topic of debate, is neurological maturation of memory capacity, defined as the number of units that can be held active in short-term memory (Cowan & Alloway, 2009). Cowan (2010) assumes that the number of unrelated information units that can be kept in active memory increases from 1 to 2 in very young children to 3–5 later in life, but it is not immediately clear how this relatively small increase alone would explain the rate of verbal short-term memory growth found in preschool children whose ages range only a few years. Third, research on adults has shown that even in simple verbal short-term memory tasks such as remembering a sequence of letters, parts of the prefrontal cortex are activated when the sequence exceeds an individual's memory capacity (Rypma et al., 1999), which may point to increasing involvement of executive functions, such as focusing of attention, inhibition of irrelevant information, task monitoring, and updating (Smith & Jonides, 1999). A final mechanism that has been proposed is an increase in long-term knowledge that supports short-term storage (e.g., Roodenrys et al., 1993). Roodenrys et al. (1993) studied the development of memory span for words and nonwords, and found that older children showed a larger lexicality effect than younger children, suggesting that long-term knowledge of words aided short-term memory. More recently, such effects have been explained by assuming that long-term language knowledge does not only aid phonological encoding, but is also applied during phonological storage to reconstruct blurred or incomplete memory traces, a process that has been termed "redintegration" or pattern completion (Brown & Hulme, 1995; Thorn, Gathercole, & Frankish, 2005). The exact way in which redintegration works may also be dependent on the nature of the memory task. More specifically, redintegration on the basis of long-term lexical knowledge in word recall tasks would support recall at the retrieval stage, whereas redintegration on the basis of long-term phonotactic knowledge would operate at the storage stage (Thorn, Frankish, & Gathercole, 2009). Consequently, increases in memory capacity for nonwords can be interpreted as being related to effects of phonotactic knowledge at the storage stage.

Two previous studies indicate that developmental increases in verbal short-term memory can be explained by growing long-term language knowledge support. First, Ottem, Lian, and Karlsen (2007) found in a cross-sectional study on children aged three to sixteen that the model explaining most variance in verbal short-term memory performance was one with two separate factors: a capacity factor and a language factor. The capacity factor was not related to age, but the language factor was, indicating that the measure of verbal short-term memory that was free of language knowledge did not increase over time, at least not after age three, whereas the other factor did. Second, Jones, Gobet, and Pine (2007) developed a computational model to explain the relation between verbal short-term memory and long-term knowledge. In their model, a verbal short-term memory component was included that was subject to decay of information after two seconds and could store only a limited amount of chunks. Language production data of mothers interacting with their two- and three-year-olds and additional words from a dictionary were used as input, from which the model built a hierarchy of phonemes and phoneme sequences. At different stages of input, the model performed a nonword repetition task. The modeled nonword repetition data resembled experimental nonword repetition data collected from children very accurately. Specifically, the model, just as the experimental data, indicated that improved short-term memory performance reflected the amount of information in long-term memory rather than increased working memory capacity. Together, these two studies suggest that developmental increases in performance on verbal short-term memory tasks can be explained by a fixed but inter-individually differing capacity on the one hand, and a growing body of long-term knowledge, on the other.

Further, more indirect, evidence for this idea comes from longitudinal studies investigating the impact of growing language knowledge on verbal short-term memory. Gathercole (1995), for example, found greater improvement for wordlike nonwords than for less wordlike nonwords over time in four- and five-year olds' nonword repetition skills, and thus an increasing wordlikeness effect with age. In a study on second language learners of English, moreover, French and O'Brien (2008) found improved performance over time for nonwords based on English, but not for nonwords based on Arabic, a language unknown to the children. Both studies present initial evidence that growing long-term language knowledge, which depends on exposure to linguistic input, is the driving force behind verbal short-term memory development.

Phonotactic knowledge development

A series of studies have shown that performance on tasks of verbal short-term memory such as nonword repetition and nonword recall is better for nonwords of high-phonotactic probability than nonwords of low-phonotactic probability. This holds for children as young as three years of age onwards who learn a first language (Coady & Aslin, 2004; Edwards, Beckman, & Munson, 2004) or a second language (Messer et al., 2010). To the best of our knowledge, however, the *development* of phonotactic knowledge support for verbal short-term memory has to date only been investigated in cross-sectional designs. Yet, investigating the degree to which phonotactic knowledge supports verbal short-term memory over the years is important, as it can provide insight into whether increases in verbal-short-term memory with age are indeed due to growing phonotactic support, as was suggested above. Specifically, assuming that phonotactic knowledge support increases with age (due to more language exposure over time), a stronger growth in short-term memory is expected for high-probability nonwords than for low-probability nonwords.

However, earlier, cross-sectional studies on the effects of phonotactics on verbal short-term memory tasks in different age groups have shown mixed results. In a study on 2.5- and 3.5-year-old children, Coady and Aslin (2004) found that 2.5-year-olds showed an effect of phonotactic probability when phoneme frequencies were manipulated in all syllable positions, but not when they were manipulated in a single syllable. 3.3-year-olds, in contrast, were sensitive to both manipulations of phonotactic probability. This suggests that phonotactic knowledge support increases with age. However, Majerus and Van der Linden (2003) studied children from various age groups between six and 22 years, and found, contrary to Coady and Aslin, that phonotactic probability had an effect in all age groups and that recall of high- and low-probability nonwords developed over the years at a similar pace. Similarly, in a cross-sectional study with children aged 3–4, 5–6, 7–8 years and adults, Edwards et al. (2004) (see also Munson, Edwards, & Beckman, 2005) found that the ability to repeat nonwords containing highly frequent phoneme clusters and nonwords containing infrequent phoneme clusters increased over the years at a similar pace, and in adults the effect of phonotactic probability was even smaller. Taken together, these three studies would suggest that the support of long-term phonotactic knowledge in monolingual children grows until around the age of three (Coady & Aslin, 2004), after which it does not increase anymore (Majerus & Van der Linden, 2003), or even decreases (Edwards et al., 2004; Munson et al., 2005).

The current study

In the current study, we investigate the development of verbal short-term memory in a three-wave longitudinal design, assessing monolingual and bilingual children between four and six years of age. All children were first assessed when they had just entered Dutch kindergarten, at four years of age, and then again at five and at six years. By comparing the effects of high- and low-phonotactic probability on the development of recall of nonwords, our aim is to examine if phonotactic knowledge supports verbal short-term memory over the preschool years in mono- and bilingual children separately.

As described above, different mechanisms have been suggested to explain growth in verbal short-term memory performance over the childhood years. If developmental improvement in verbal short-term memory can indeed be largely explained by growing language knowledge, as suggested by some recent studies, we would expect developmental improvement in the recall of high-probability nonwords, but no (or much smaller) improvement in the recall of low-probability nonwords, because for the latter type, virtually no long-term knowledge support can be available. As for the comparison between monolingual and bilingual children, we expect that bilingual children who lag behind in the target language (Dutch) as compared to their monolingual peers show lower performance on high-probability nonwords due to their poorer knowledge of Dutch phonotactics, but not on low-probability nonwords for which there is no or little support from long-term memory knowledge of Dutch. So, for low-probability nonwords, no disadvantage is expected for the bilingual children. Like the monolingual children, moreover, they are expected to show development for the high-probability but not low-probability nonwords as only the former type of nonwords are supported by long-term linguistic knowledge that they develop over time due to increased language exposure. It is an open question, however, if they develop at a similar pace as the monolingual children or whether they are delayed not only in initial level but also in growth of verbal short-term memory of high-probability nonwords, or actually show a faster development in recall of such nonwords, due to their immersion in a rich Dutch (school) environment.

The current bilingual children present an interesting case as they were cultural minority children from Turkish immigrant families in the Netherlands who had been predominantly exposed to Turkish at home prior to the start of the study, but had enrolled in Dutch kindergarten shortly before the first wave of assessment. Previous studies on this population (Appel & Vermeer, 1998; Droop & Verhoeven, 2003) as well as studies on the same sample (Blom, Kuntay, Messer, Verhagen, & Leseman, 2014; Messer et al., 2010) have shown that these bilinguals lag behind their monolingual Dutch peers in both vocabulary and grammar in the early school years. However, they typically catch up after several years of rich Dutch exposure at school after which Dutch becomes their dominant language (Extra et al., 2001).

In a previous study (Messer et al., 2010), we established that the Dutch monolingual children showed an effect of Dutch phonotactic probability on nonword recall at age four. The Turkish children also showed such an effect, albeit smaller, which was attributed to their more limited phonotactic knowledge of Dutch. In the current study, we build on these findings and ask how the effect of phonotactic probability develops in the two groups over the course of three years, as a function of growing language knowledge. Specifically, we investigate if effects of phonotactic knowledge found at four years become more pronounced over time, when children's long-term linguistic knowledge of Dutch increases. We predict that this is the case as long-term knowledge support becomes increasingly available and only supports recall of high-probability nonwords.

In studying these questions, a number of factors that might be related to verbal short-term memory will be controlled for, such as; nonverbal intelligence, visuospatial short-term memory, and verbal short-term memory of well-entrenched digit knowledge, to rule out that any group differences we may find can be attributed to differences in general cognitive abilities. In addition, we compare vocabulary scores at four years between the two groups to see if the bilingual children indeed have lower vocabulary levels than the monolingual children. To answer our research questions, Latent Growth Modeling (LGM) is used, allowing not only an investigation of possible group differences, but also individual differences in children's developmental trajectories.

Method

Participants

The sample consisted of 69 bilingual Turkish-Dutch children (40 boys and 29 girls) and 72 Dutch children (47 boys and 25 girls). Children were tested in three waves with 12 months between the first two waves and 10 months between the last two waves. At the first wave, when children had just entered kindergarten, their mean age was 53 months ($SD = 3$, range = 49–66) in the Turkish-Dutch group and 52 months ($SD = 3$, range = 48–62) in the Dutch group. Children with missing data at one or two waves due to absence from school during one of the testing periods (one child did not participate at waves 1 and 3, another child participated at wave 1 only) or drop out due to moving house (one child dropped out at wave 2; three more children dropped out at wave 3; total attrition rate 2.9%) were included in the LGM analyses (see under 'Analyses').

The children attended kindergarten classrooms of 36 inner-city Dutch primary schools with a moderate to high percentage of ethnic minority children from various ethnic backgrounds (25–100%) and with Dutch as the language of schooling. A short screening questionnaire was administered to children's primary caregivers at recruitment. Children were eligible if at least 75% of the language interactions with the target child in the family context were in Turkish (Turkish-Dutch group) or in Dutch (Dutch group). Children with known serious developmental delays or medical speech or hearing problems were not included. Parents of eligible children were asked for informed consent. The positive response rate was 69% for the Turkish-Dutch group and 80% for the Dutch group.

Extensive questionnaires, administered to children's primary caregivers at wave 1, revealed that the Turkish-Dutch children had been exposed to some Dutch language prior to kindergarten entry: Most children had attended some form of early childhood care and education providing a Dutch immersion context (88% of the Turkish-Dutch group, 90% of the Dutch group, on average 4 half days a week, no statistically significant difference between the groups) and had older siblings who sometimes or always communicated in Dutch with the child (65% of the Turkish-Dutch children, mean number of siblings = 1.6, range = 1–4; 31% of the Dutch group, mean number of siblings = 1.8, range = 1–4).

Procedure

At each wave, trained research assistants who were fluent in both Dutch and Turkish assessed each child individually in a quiet room at children's schools. Testing took place on two days that were approximately one week apart. Each testing session lasted for about 75 min, including play breaks and tests which were part of another study. The tests were intermixed with these other tests and administered in a fixed order: Dutch vocabulary, high-probability nonword recall, Raven, dot matrix (Day 1), Digit recall, low-probability nonword recall (Day 2). To keep children motivated, they were rewarded with a sticker after each test. The participating families received a children's book at each wave and a copy of all video materials at the end of the study to thank them for participation.

Measures

The verbal-short term memory tests were taken from the Automated Working Memory Assessment battery (AWMA) (Alloway, 2007) and translated into Dutch (see Messer et al., 2010). Instruction and scoring procedures of the AWMA were applied. In all tests, children had to recall sequences of stimuli (i.e., nonwords or digits) starting with a block of one item and building up to a block of seven items in a row. Each block consisted of six trials, that were scored as incorrect when one of the items was omitted, was recalled wrongly, or when the sequence of items was incorrect. When the first four trials within a block were recalled correctly, the child automatically received a score of 6 and proceeded to the next block. Note that a score of 6 could therefore represent either six correct one-item trials and no correct two-item trials or four correct one-item trials and two correct two-item trials. Scores could range from 0 (block 1) to 42 (block 7), but testing stopped after three incorrect recalls within one block.

Nonword recall

In the nonword recall tests, children were asked to repeat voice-recorded monosyllabic nonwords in lists of increasing length. Each phoneme of a nonword had to be recalled correctly in order to obtain a positive score. To make sure that short-term memory and not production failure was measured, phonemes that were consistently substituted by a child due to articulation problems or foreign accent were not considered incorrect.

Two tests were used: one test containing nonwords composed of highly frequent phoneme combinations (high phonotactic probability) and another test containing nonwords composed of infrequent phoneme combinations (low phonotactic probability) (see Messer et al., 2010, for details). All nonwords had been voice-recorded by a female native speaker of Dutch. A total of 36 high-probability nonwords and 36 low-probability nonwords were used (in fact, a higher number of nonwords was created, but, since none of the children passed block 3, these were not administered). These nonwords had been constructed on the basis of cross-translated Dutch and Turkish corpora of children's books (for more details, see Messer et al., 2010), and differed significantly in phonotactic probability, $F(1, 70) = 113.74$, $p < .001$, $\eta^2_p = .62$. Also, fourteen Dutch native speakers had rated the wordlikeness of each voice-recorded nonword on a scale of 1 (*Does not sound like a real Dutch word at all*) to 5 (*Sounds a lot like a real Dutch word*). These ratings showed that the high-probability nonwords sounded more like real words to the native speakers than the low-probability nonwords, $F(1, 70) = 37.45$, $p < .001$, $\eta^2_p = .35$. As a further check, the Dutch high- and low-probability nonwords were examined for phonotactic probability in a Turkish corpus (for more details on this corpus, see Messer et al. (2010)). Approximate cross-linguistic frequencies were derived by changing Dutch graphemes in the phonetically best corresponding Turkish graphemes in the following way: aa = ağ; ai = ay; c = k; dj = c; tj = ç; ee = eğ/iğ; i = i (end of word); j = y; zj = j; oo = oğ; sj = ş; u = ı; uu = ü; v = v (end of syllable); w = v (beginning of syllable); eu = ö; oe = u; ie = i (except when end of word). One-way ANOVAs revealed that the two sets of nonwords showed no significant differences in Turkish phonotactic frequency ($M = 98.3$ and $M = 147.7$ for the high- and low-probability items, respectively, $F(1, 70) = 1.0$, $p > .1$).

Because a few of the low-probability nonwords received rather high ratings of wordlikeness by the native speakers, we decided to make minor adjustments to the tests used at waves 2 and 3 by selecting as high-probability items only those nonwords with both high phonotactic probability and high wordlikeness ratings and as low-probability items only those nonwords with both low phonotactic probability and low wordlikeness ratings. This resulted in blocks consisting of four instead of six trials at the last two waves. For these reduced tests, there was again a significant difference between the two nonword sets in both phonotactic probability, $F(1, 78) = 188.43$, $p < .001$, $\eta^2_p = .71$, and wordlikeness, $F(1, 78) = 239.47$, $p < .001$, $\eta^2_p = .75$. To have comparable tasks at all three waves, only the scores of the first four trials per block were used for wave 1. Scores could range from 0 (block 1) to 24 (block 4), but most of the children did not reach the fourth block. Testing stopped after three incorrect recalls within one block. See Appendix A for a list of all items and Table 1 for an overview of main nonword characteristics.

It should be noted that the low-probability nonwords consisted of phoneme combinations which were infrequent and occasionally very infrequent, so one may wonder whether these nonwords were actually low-probable or non-existent (and perhaps even unpronounceable). As it is important to rule out that they were illegal or too difficult to pronounce, we checked the status of the low-probability nonwords in a number of ways. First, we made sure that all nonwords could be articulated. Second, we performed a check of their biphone transitional probabilities against frequency counts of the Corps of Spoken Dutch (CGN, Goddijn & Binnenpoorte, 2003) with the help of the software PhontacTools (Adriaans, 2006) to find out if they actually occurred in spoken Dutch. This check confirmed that the low-probability nonwords had signficantly lower mean biphone frequencies than the high-probability nonwords ($M = .06$ and $M = .10$, $F(1, 70) = 29.91$, $p < .001$, $\eta^2_p = .30$ for wave 1; $M = .06$ $M = .10$, $F(1, 78) = 36.01$, $p < .001$, $\eta^2_p = .32$ for waves 2 and 3). It also showed that, with one exception (i.e., /fx/ in 'fgip'), all biphones had transitional probability values higher than zero, and thus should be considered legal sound combinations. Some consonant clusters such as 'dj', 'pj' and 'mw' typically occur in loan words in Dutch, also in word-initial positon (e.g., 'dj' in "gin", 'pj' in "pie d-à-terre", 'fj' in "fjord", 'mw' in "moyenne"), but all of these can be easily articulated. This was evidenced by a post-hoc analysis of children's articulation rate. Specifically, children's articulation rate of the nonwords was checked across the two stimuli sets. That is, after data collection at wave 1, the video recordings of 20 randomly selected children (10 Dutch, 10 Turkish-Dutch) were used to measure the pronunciation time of each nonword. The differences in children's articulation rate between both stimuli sets were very small and not significant ($p > .1$), suggesting that the low-probability nonwords were not more difficult to pronounce than the high-probability nonwords for the children. Finally, we analyzed the types of errors made in a subsample of 12 randomly selected children (6 Dutch, 6 Turkish-Dutch) and found very similar distributions of phoneme additions, deletions and substitutions across the high-probability and low-probability nonword recall tasks. This suggests that pronunciation problems were rare and did not play an important role in children's repetition of in particular the low-probability nonwords, since, in such a case, different error patterns would have been expected across the two stimuli sets.

To further examine the validity of the nonword tasks (and control for possible confounding factors), length of the nonwords was checked across the two stimuli sets. At wave 1, the high-probability nonwords were slightly longer than the low-probability nonwords (mean number of phonemes 4.3 and 3.9, $F(1, 70) = 5.70$, $p = .020$, $\eta^2_p = .08$). However, since longer nonwords should be more difficult to remember than shorter ones, this effect would work against the hypothesis and thus was not expected to bias our conclusions. At waves 2 and 3, the two sets of nonwords did not significantly differ in length ($p > .1$).

All administrations of the nonword recall tests were videotaped. Video recordings were then used to check and occasionally correct the assistants' scores provided during testing. More specifically, native speakers scored children's responses from the videos, and their scores were checked by the primary investigator for a random sample of 10% at each wave. Interrater reliabilities assessed through bivariate correlations between the coders were high (wave 1 low-probability nonwords $r = .89$, high-probability nonwords $r = .87$; wave 2 low-probability nonwords $r = .88$, high-probability

Table 1
Characteristics of the high- and low-probability nonwords at wave 1 and waves 2/3.

	Low-probability			High-probability		
	M	*SD*	Range	*M*	*SD*	Range
Sum bigram frequency (relative freq. per 10.000)						
Wave 1	26.6	20.2	0.2–66.7	449.3	223.3	119.2–970.8
Waves 2/3	36.0	34.8	0.2–209.8	466.4	195.2	231.2–970.8
Likeness rating (1–5)						
Wave 1	2.6	0.8	1.4–4.0	3.6	0.6	2.2–4.8
Waves 2/3	2.2	0.4	1.4–3.0	3.7	0.5	2.6–4.6

Note. Because no child exceeded block three, only the items of the first three blocks were used to calculate the means at wave 1 ($n = 36$ for each type). At waves 2/3, all nonwords were used ($n = 40$ for each type).

nonwords $r = .90$; wave 3 low-probability nonwords $r = .87$, high-probability nonwords $r = .86$; all $ps < .05$).

Other measures

Verbal short-term memory

The Digit Recall subtest of the AWMA battery (Alloway, 2007) was used to assess verbal short-term recall at all three waves, in addition to the nonword recall tests. In this test, a random sequence of digits ranging from 0 to 9 was presented. Testing followed the format and scoring rules of the AWMA that were described above. The reason for including digit recall as a control measure was that digits from 0 to 9 can be considered highly frequent words that are ubiquitous in (preschool) language, and therefore well-known in both groups. So, we did not expect Dutch and Turkish-Dutch children to differ in their performance on digit recall.

Visuospatial short-term memory

The Dot Matrix of the AWMA (Alloway, 2007) was used to assess visuospatial short-term memory (Alloway, 2007). This task presented children with a 4 × 4 matrix on a laptop screen. Sequences of red dots shortly appeared in the cells of the matrix and children had to recall the sequence in which the dots had appeared in the matrix.

Nonverbal IQ

Raven's Colored Progressive Matrices (Raven, Raven, & Court, 1998) was administered at wave 1 to measure nonverbal fluid intelligence. The task was presented on a laptop computer using the software package MINDS (Brand, 1999). Children had to decide which one of six pieces on the computer screen would best complete the visual pattern from which a piece was missing. Each correct answer was rewarded with a score of 1, yielding to a total score between 0 and 36.

Vocabulary

A Dutch receptive vocabulary test was administered at wave 1 that was part of the Test for Bilingualism (*Toets Tweetaligheid*, Verhoeven, Narain, Extra, Konak, & Zerrouk, 1995), a language test kit specifically designed for research into bilingual development and examined for cultural bias on item level. In the vocabulary test, four line drawings were presented on the computer screen and children were asked to point to the picture they thought corresponded best to the word produced by the research assistant. To avoid fatigue, only half of the test was used (i.e., 30 even items). Each correct answer was rewarded with a score of 1. Testing was stopped when children failed on five consecutive items, after which the remaining items were rewarded with the chance-score of 0.25. The scores could range from 0 to 30. Cronbach's alpha for the receptive vocabulary test was .84 for the Turkish-Dutch group and .73 for the Dutch group.

Analyses

The data of each test were first explored to check for normality, outliers and missing data. For nonword recall, scores were normally distributed in both groups, with standardized skewness and kurtosis measures not exceeding the value of 3. No cases were excluded, since there were no outliers greater than three standard deviations below or above the mean. Missing values (5.6% of all data) were not imputed in advance, but dealt with through full information maximum likelihood estimation, as recommended by Kline (2005). For the four control tasks, 4.9% of the data was missing and inspection of each variable separately revealed no extreme outliers of more than three standard deviations above or below the mean. Missing data were not imputed.

Latent Growth Modeling (LGM; Duncan, Duncan, & Strycker, 2006) was used to model developmental trajectories in recall of nonwords over the three measurement waves. Estimation was done with Mplus version 7.11 (Muthén & Muthén, 1998–2010). To evaluate model fit for maximum likelihood estimation, we used the chi-square (χ^2), the root mean square error of approximation (RMSEA), the comparative fit index (CFI) and the Tucker-Lewis index (TLI). As a rule of thumb, a non-significant χ^2 indicates good model fit, RMSEA below .05 indicates good fit and below .08 reasonable fit, and CFI and TLI greater than .90 indicate acceptable fit (Little, 2013). To compare performance on the two nonword recall tasks between the Dutch and Turkish-Dutch groups, we used a multivariate multi-group model. We evaluated general model fit of the multi-group model and also compared different models to each other (see 'Results'). When the release of a parameter constraint, resulting in the loss of one degree of freedom, did not significantly improve model fit as indicated by a chi-square difference test, the parameter was again constrained in the next steps.

Results

Control tasks

Table 2 presents the means and standard deviations for the control measures.

One-way ANOVA's showed that Dutch receptive vocabulary was significantly lower in the Turkish-Dutch group than in the Dutch group, $F(1,137) = 103.6$, $p < .01$, $\eta^2_p = .42$. Nonverbal IQ and visuospatial short-term

Table 2
Descriptive statistics for control measures for the monolingual and the bilingual children.

Variable	Dutch group ($n = 72$)		Turkish-Dutch group ($n = 69$)	
	M	*SD*	*M*	*SD*
Control measures				
Dutch vocabulary	20.6	3.5	14.3	3.9
Nonverbal IQ	12.1	3.2	12.1	2.7
Visuospatial recall	9.5	3.5	9.6	3.3
Digit recall wave 1	14.7	4.6	14.5	3.9
Digit recall wave 2	18.0	4.0	16.9	3.1
Digit recall wave 3	19.5	3.1	19.0	3.6

Note. Not all children performed the digit recall tasks ($n = 66$ at wave 1, $n = 71$ at waves 2/3 for the Dutch group; $n = 52$ at wave 1, $n = 69$ at wave 2, and $n = 65$ at wave 3 for the Turkish-Dutch group).

memory did not differ significantly between the groups ($ps > .1$). As for digit recall, a repeated-measures ANOVA showed that there was a main effect of time, showing increased performance over time, $F(2, 111) = 97.88$, $p < .001$, $\eta^2_p = .64$, but no effect of group or interaction effect between time and group.

Nonword recall and phonotactic probability: latent growth modeling

Table 3 presents means and standard deviations for the two nonword recall tasks at all three waves for the monolingual and bilingual children. It should be noted that scores are rather low (a point we will come back to in the Discussion). Correlations between all variables are given in Table 4 for both groups separately.

To analyse growth of both recall measures, we integrated both types of nonwords in one model, to allow us to make comparisons between the two variables. Moreover, if we could constrain the two types of recall to share the same intercept and find good model fit, we could validate the idea that the same basic, language-free, capacity is used in both types of recall. We therefore constrained both types of recall to share the same intercept. The measurement occasion specific intercepts of the low-probability nonwords were fixed at 0, as usual, so we could estimate the shared intercept and examine whether the high-probability nonwords had measurement specific intercepts over and above this shared intercept. The conceptual model is depicted in Fig. 1. In this model, both types of recall have their own slopes, with the regression weights fixed at 0, 1 and 1.8 to represent the exact time intervals in years between each of the three measurement occasions. Each variable was allowed to contain time-specific measurement error.

To examine group differences, we successively released the constraints *between groups* of the high-probability nonwords (the measurement specific intercepts, slope mean, slope variance and the covariance of intercept and slope), the constraints between groups of the shared intercept (mean and variance), and the constraints between groups of the low-probability nonwords (measurement specific intercepts, slope mean, slope variance and the covariance of intercept and slope) as recommended by Little (2013), while *within groups* the measurement specific intercepts of the nonword measures were still constrained to be equal over the three measurement waves. Only the release of the measurement specific intercepts of the high-probability nonwords significantly improved model fit, $\Delta\chi^2(1) = 10.16$, $p < .001$. This indicates, first, that the groups did not differ in mean performance on the low-probability nonwords, while they did differ in mean performance on the high-probability nonwords (with the Dutch children outperforming the Turkish-Dutch children), and second, that the mean growth rates on both measures were equal in both groups. Moreover, releasing the equality constraints on the variances of the measures did not result in improved model fit.

Table 3
Descriptive statistics for the nonword recall tasks at all waves for the monolingual and the bilingual children.

Variable	Dutch group ($n = 72$)		Turkish-Dutch group ($n = 69$)	
	M	*SD*	*M*	*SD*
Nonword recall				
Low-probability: Wave 1	2.8	1.8	2.1	1.3
Low-probability: Wave 2	2.2	1.4	2.2	1.1
Low-probability: Wave 3	2.2	1.0	2.1	1.1
High-probability: Wave 1	4.0	1.7	2.8	1.6
High-probability: Wave 2	4.6	2.1	4.2	1.7
High-probability: Wave 3	5.2	2.1	4.5	1.8

Table 4
Correlations among all variables included in the model.

Variable	1	2	3	4	5	6
Low-probability						
1. Wave 1	–	.49**	.26*	.37**	.36**	.28*
2. Wave 2	.24	–	.40**	.43**	.42**	.44**
3. Wave 3	.25	.19	–	.26*	.38**	.38**
High-probability						
4. Wave 1	.49**	.11	.38**	–	.40**	.33**
5. Wave 2	.23*	.44*	.09	.44**	–	.40**
6. Wave 3	.39**	.53**	.28*	.48**	.49**	–

Note. Correlation coefficients for the monolingual group are shown in the upper diagonal; correlations coefficients for the bilingual group are in the lower diagonal.

* $p < .05$.

** $p < .01$.

The estimated slope mean of the high-probability nonwords was positive and significant, indicating that both groups improved in performance on this measure over the years. Because the estimated slope mean of the low-probability nonwords was almost zero and non-significant, we fixed the mean of this slope to zero. This gain of one degree of freedom did not result in a significantly lower χ^2 as assessed through a chi-square difference test, $\Delta\chi^2(1) = 1.64$, $p = .20$. This indicates that the more parsimonious model in which there was no statistically significant growth for the low-probability nonwords in either of the groups fitted the data as well as the less constrained model. Yet, this model still fitted the data not very well, $\chi^2(33) = 51.21$, $p < .05$; RMSEA = .06; CFI = .88; NNFI = .85. Inspection of the raw data showed a decrease in recall of the low-probability nonwords between waves 1 and 2, but only in the Dutch group. Because a decrease in memory capacity is highly unlikely at this age, and because the low-probability nonwords were found more language-like at wave 1 than at waves 2 and 3 due to the adjustments made in the later test versions to tackle this problem (as described in the Method section) and therefore easier for the Dutch group, we relaxed the measurement specific intercept for this observed variable at wave 1 for the Dutch group. This significantly improved model fit, $\Delta\chi^2(1) = 7.39$, $p < .01$, and resulted in a model with adequate fit, $\chi^2(32) = 43.82$, $p > .05$; RMSEA = .05; CFI = .92; NNFI = .90. This final model is depicted in Fig. 1, and the estimated growth curves are depicted in Fig. 2. Indeed, both variables could be constrained to share the same intercept, with a mean value of 2.2 ($p < .001$) for both groups. The observed means of

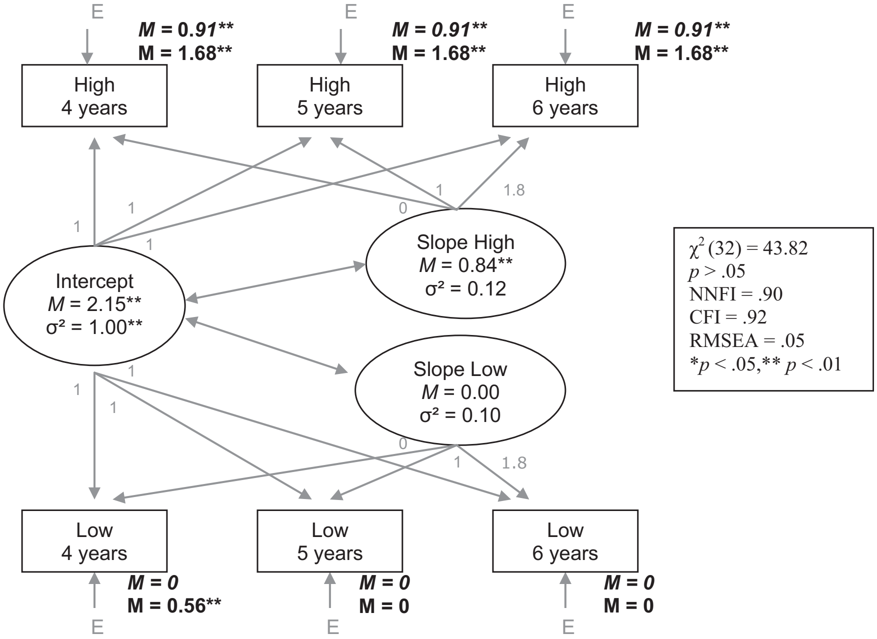

Fig. 1. Model estimated to assess the development of recall of high-probability nonwords (High) and low-probability nonwords (Low) from four to six years. *Bold* = Dutch group, ***Bold-italic*** = Turkish-Dutch group, Non-bold = both groups. E = measurement error.

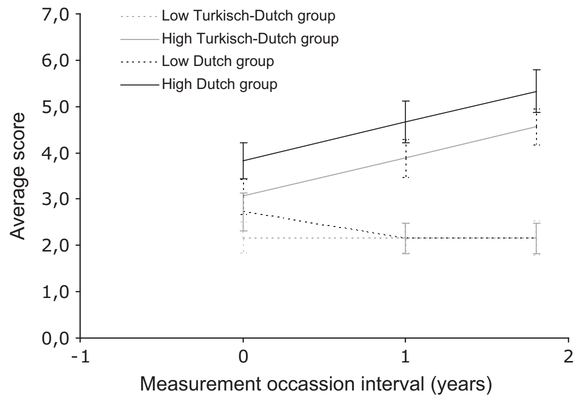

Fig. 2. Estimated growth curves for the recall of high-probability nonwords (High) and low-probability nonwords (Low) in Turkish-Dutch and Dutch children at ages four, five, and six. Error bars represent the 95% confidence intervals around the estimated means.

the high-probability nonwords were significantly higher than the shared intercept of both groups, and were 1.7 ($p < .001$) and .91 ($p < .001$) over and above the shared intercept, for the Dutch group and the Turkish-Dutch group respectively. The slope mean of the high-probability nonwords was the same (.84, $p < .001$) for both groups. The variation around the intercept (1.00, $p < .001$) for both groups showed that the children differed in their initial level of recall of low-probability nonwords. The individual variation around both the slope means was not significant in both groups (high-probability nonwords .12, $p > .05$; low-probability nonwords .10, $p > .05$). However, because the estimated variation for the slopes in the model had relatively large standard errors, the conclusion that the children did not differ in their rate of development over time is too strong; the results just indicate that inter-individual differences in the rate of development over time were small, and could not very well be distinguished from other sources of variance such as measurement error per occasion.

Discussion

In this study, we investigated growth of verbal short-term memory in monolingual and bilingual children from four to six years. We were interested in investigating developmental increases in verbal short-term memory in relation to a specific source of long-term knowledge support: phonotactic knowledge. By comparing recall of nonwords with high phonotactic probability and recall of nonwords with low phonotactic probability in a longitudinal design, we investigated whether long-term phonotactic knowledge is the driving force behind growth in verbal short-term memory over the preschool years.

Latent growth modeling showed that the high-probability nonwords were recalled more accurately than the low-probability nonwords in both groups and at all three waves, with the Dutch group outperforming the Turkish-Dutch group in recall of high-probability nonwords but showing equal performance on low-probability nonwords. In both groups, there was no growth in recall of low-probability nonwords, while there was substantial growth in recall of high-probability nonwords. Our analyses also showed that while there were clear differences between the mono- and bilingual children in their initial level of high-probability nonword recall, between-group differences in the rate of development of high-probability recall were difficult to ascertain. These results suggest that knowledge of the statistical distribution of phoneme clusters in a language becomes increasingly entrenched during the preschool years, and increasingly facilitates verbal short-term memory in both monolingual and bilingual children. The similar growth rates in both groups show that the initial differences between the two groups at the start of kindergarten were not reduced after two years of exposure to the rich linguistic environment provided at school.

The monolingual and bilingual groups did not differ in general cognitive abilities. At the first wave, when children were four years old, measures of nonverbal intelligence and visuospatial short-term memory did not reveal any group differences. Likewise, performance on digit recall, a task that draws on automated digit knowledge, did not differ significantly between the two groups. This is in line with our expectation that, due to the frequent use and ubiquitous presence of Dutch digits at preschool, Dutch and Turkish-Dutch children alike would know these words and would be able to retrieve them from long-term memory at a similar rate. Our findings also showed that performance on digit recall increased over the years in both groups. Digits are highly frequent, well-known words that have a particular meaning and allow for association and chunking. As such, they differ from low-probability nonwords made up of highly infrequent sound combinations. In our study, we found that monolingual Dutch and bilingual Turkish-Dutch children did not differ in performance on low-probability nonwords over the years. When encountering highly infrequent novel phonotactic information, both groups of children thus were similarly skilled to store this information. As expected due to their bilingual input situation, the Turkish-Dutch children had lower vocabulary scores in their second language Dutch compared to the monolingual children. As the two groups did not differ in general cognitive abilities, the differences in verbal short-term memory development are likely to be attributed to differences in knowledge of, and thus exposure to, the Dutch language (see Messer et al., 2010 for relationships between Dutch vocabulary and nonword recall in the same sample).

The finding that the monolingual children showed increased performance on high-probability nonwords but not on low-probability nonwords suggests that language exposure plays a substantial role in the development of phonotactic knowledge support. Our results are in line with those of Coady and Aslin (2004) who found that 3.5-year-olds and 2.5-year-olds did not use phonotactic knowledge to the same degree to support verbal short-term memory in nonword repetition. They also fit well with findings by Gathercole, Frankish, Pickering, and Peaker (1999) who found a more pronounced wordlikeness effect in five-year-olds than in four-year-olds. They do not align with studies by Edwards et al. (2004) and Munson et al. (2005) who found a weaker phonotactic probability effect in older children (and in adults) than in younger children, or studies showing similar effects across ages (Majerus & Van der Linden, 2003).

One possible explanation of these conflicting findings is that not only recall of high-probability nonwords benefits from growing long-term knowledge support, but also recall of low-probability nonwords. In fact, Edwards et al. (2004) set out to study the impact of vocabulary size on phonotactic knowledge, and interpreted their results in exactly this manner. They showed that when dividing the children into groups with larger and smaller vocabularies, the children with smaller vocabularies performed worse on both high- and low-probability nonwords (see also Gathercole et al., 1999), and especially poorly on nonwords containing unattested (zero frequency) phoneme combinations instead of phoneme combinations with low frequency. They concluded that children with larger vocabularies have more robust and flexible representations of phoneme sequences, supporting even the repetition of low-probability nonwords. Indeed, in the current study, we found similar results for the monolingual children at the first wave (Messer et al., 2010), after which the set of nonwords was refined. A possible explanation of the difference between our results and those of previous studies thus may lie in the way the nonwords were constructed across studies. In Majerus and Van der Linden (2003), French monosyllabic, consonant-vowel-consonant (CVC) nonwords were used, with biphones being manipulated as either both high or both low in phonotactic frequency. In Edwards et al. (2004), only one biphone in English-based nonwords of either two or three syllables was manipulated, and both low frequency and zero frequency biphones were used. In the present study, monosyllabic nonwords with different structures were used (CVC, CCVC, CVCC, CCVCC), and the low-probability nonwords contained many very low frequency combinations. The low-probability nonwords in our study thus were more difficult, more infrequent, and therefore probably less bound to long-term memory influences than the nonwords used in Edwards et al. (2004). It is conceivable, however, that perhaps at a later age, in our study performance on low-probability nonwords would also show an increase due to stronger phonotactic knowledge and more fine-grained phoneme representations in children as a function of growing vocabulary knowledge.

An alternative explanation of the lack of growth in low-probability nonwords in our study is that these nonwords were simply too difficult to pronounce for children to show improvement over the years. While we cannot exclude this possibility, we strongly believe that we were measuring memory and not production failure for a number of reasons. First, articulation rate, as measured by pronunciation time during task performance, did not differ between the low- and high-probability nonwords. Second, consistently substituted phonemes, resulting from articulation problems and foreign accents were not scored as incorrect in our coding of children's answers, and thus are not a likely explanation of children's poor performance on the low-probability items. Finally, the fact that recall of both nonword types could be constrained to share the same intercept suggests that in both types of recall the same basic memory mechanisms were employed, and that in the case of the high-probability nonwords, long-term phonotactic knowledge was used as support on top of that.

Another possible explanation for the lack of growth in low-probability nonwords might be that they were simply too difficult to pronounce due to their highly infrequent phoneme combinations, such that the measures suffered from a strong floor effect. Put differently, the low-probability nonword measures might not have been sensitive enough to capture growth in this age range. However, the data gave no indication of floor effects. Mean scores at all waves were significantly different from 0, as became also evident from the shared intercept in the growth model. In addition, the standard deviations on the low-probability nonword measures indicated substantial inter-individual variance, and correlations with concurrent mesures were moderate, showing no restriction of variance. As for the use of (very) infrequent phoneme combinations to construct the low-probability nonwords, we do not think that this constituted a problem. First, virtually all of these phoneme combinations were found in a corpus of spoken Dutch. Second, all low-probability nonwords could be articulated and, in fact, did not differ from the high-probability nonwords in articulation rate in children's productions.

Our finding that children's performance on low-probability nonwords did not increase is in line with accounts that assume a fixed memory capacity from an early age onwards (Cowan, 2005). It also suggests that all the other suggested mechanisms of growth do not play a substantial role in four- to six-year-olds, or are themselves a result of growing language knowledge. It has been shown, for example, that articulation rate increases with increasing language fluency (Standing, Bond, Smith, & Isely, 1980) and the same might be true of memory capacity or decay rate. The crucial difference between these other suggested mechanisms behind short-term memory growth and growing support of long-term language knowledge is that the ability to represent highly infrequent phonotactic information in the correct serial order might be considered as a (possibly highly heritable) neurobiological factor, while the development of linguistic knowledge is highly dependent on input, and thus the linguistic environment of the child. Because one predominant view to date is that verbal short-term memory is highly heritable and free of environmental influences (Gathercole, 2006), an interesting area for future studies could be intervention studies in which specific properties of the linguistic input are trained to see if they result in stronger verbal short-term memory growth.

There are different views on how long-term knowledge may influence short-term recall. According to the probabilistic model of "redintegration" (Brown & Hulme, 1995; Hulme, Maughan, & Brown, 1991; Hulme et al., 1997), encoded memory traces in the phonological store that degrade during recall are repaired with automatically retrieved lexical or phonological representations from long-term memory. The information in the trace is compared to information in long-term memory and when the patterns match, the degraded memory trace is reconstructed. Phonological representations of highly frequent words are more accessible in long-term memory than infrequent words or nonwords, making reconstruction more successful. A slight modification of this account proposes that there are multiple mechanisms operating at different stages (Thorn et al., 2005, 2009). On this account, lexical knowledge (and word frequency) indeed influences short-term memory via redintegration, a process taking place after storage during the retrieval stage, but phonotactic knowledge has its influence at an earlier stage already during encoding and/or storage, as it determines the strength of the initial memory trace. This memory trace is conceived of as a pattern of activations across a network with nodes representing phonological units. In both views, a dichotomy is made between short-term memory and long-term memory, in line with the working memory model of Baddeley and Hitch (1974). In the latter view, however, it is not clear how to interpret phonotactic knowledge support within this dichotomy. This account seems to suggest that phonotactic knowledge forms an integral part of the phonological loop, making it difficult to conceptualize a "pure" phonological loop that is free from long-term memory influences.

The influences of long-term knowledge on short-term memory are easier to interpret within theories postulating that there is no dichotomy and that short-term memory is the activated portion of long-term memory, such as the embedded-processes model of Cowan (1999, 2005). In this model, long-term knowledge necessarily influences short-term recall, with every type of knowledge exerting its influence during the encoding/storage phase. In a similar vein, Martin (2009) proposed that verbal short-term memory is not a separate system, but an emergent property of the temporary activation of phonological, lexical, and semantic representations in the language network.

Interestingly, in the present study, growth curves of recall of high- and low-probability nonwords showed the same patterns for monolingual and bilingual children who started out with less support from long-term phonotactic knowledge in their second language Dutch. As the bilingual children had just been immersed in a Dutch (school) environment and the monolingual children had far more exposure to Dutch and therefore more entrenched knowledge of Dutch phonotactics, these results were not surprising. The differences between both groups at the first wave, however, did not decrease during the kindergarten period in which both groups were exposed to a rich Dutch language environment. That is, the developmental rates of both groups of children were actually equivalent. It should be noted, however, that the monolingual children received more Dutch input during these years as they were exposed to Dutch both at home and at school, unlike the bilingual children whose Dutch input was largely restricted to school. The similar rates of development in the two groups might therefore be taken to suggest that school does compensate for the disadvantages of the Turkish-Dutch children, but not enough for them to reach the level of their monolingual Dutch peers.

Appendix A

See Table A1.

Table A1
Items of the nonword recall tasks.

	Low-probability items				High-probability items			
	Wave 1				Wave 1			
Block 1	Jimf				Zwag			
	Dwup				Grops			
	Pjoef				Zils			
	Fosk				Brof			
	Pifp				Traa			
	Faup				Gleg			
Block 2	Pjosr	Fnup			Grigt	Zwop		
	Fuup	Pjif			Spraam	Kwig		
	Vub	Puif			Zifs	Bropt		
	Fjaip	Dzub			Greel	Knit		
	Fip	Posf			Knog	Glin		
	Pgup	Dwuuf			Ziks	Glof		
Block 3	Mwup	Fjif	Njos		Brop	Sning	Knilk	
	Ims	Fwup	Pjai		Zilg	Brong	Tris	
	Bnup	Osf	Fjeum		Snins	Glirg	Ceng	
	Fwut	Gjuip	Fimk		Fling	Brops	Zwis	
	Djai	Pwut	Fibs		Vlop	Snilg	Kwin	
	Zup	Kjif	Fjui		Zwit	Snint	Dromp	
	Waves 2/3				Waves 2/3			
Block 1	Fnup				Brop			
	Djai				Flit			
	Josf				Gleg			
	Pjif				Zils			
Block 2	Pwut	Kjif			Snint	Brof		
	Pifp	Mwuut			Vlis	Zwag		
	Jimf	Pjai			Bring	Knog		
	Fgip	Njos			Glit	Dromp		
Block 3	Gjif	Djut	Foip		Brong	Knit	Lifs	
	Fwutf	Weum	Sjup		Grilk	Vlin	Snog	
	Jibs	Fosf	Pjoem		Bligs	Zwop	Keng	
	Pwuf	Gjim	Fjaip		Glin	Blof	Zwis	
Block 4	Fwup	Pjosr	Ims	Dwuuf	Blopt	Knig	Ziks	Graar
	Djups	Pimf	Wuip	Fai	Brig	Drof	Grops	Fling
	Kjosf	Bnup	Fwum	Pjoef	Fliks	Glof	Zwit	Vlop
	Fjif	Leels	Posf	Swup	Broft	Kwin	Snig	Vrog

Note. Because at wave 1 no child exceeded block three, only the items of the first three blocks are listed. For waves 2 and 3, all items are listed.

References

Adriaans, F. (2006). *PhonotacTools (Test version) [Computer program]*. Utrecht, the Netherlands: Utrecht Institute of Linguistics OTS.

Alloway, T. P. (2007). *Automated working memory assessment*. London: Psychological Corporation.

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

Appel, R., & Vermeer, A. (1998). Speeding up second language vocabulary acquisition of minority children. *Language and Education, 12*, 159–173.

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

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

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

Brand, A. N (1999). MINDS: Tool for research in health psychology and neuropsychology. In B. P. L. M. d. Brink, P. J. Beek, A. N. Brand, F. J. Maarse, & L. J. M. Mulder (Eds.), *Cognitive ergonomics, clinical assessment and computer-assisted learning* (pp. 155–168). Lisse, the Netherlands: Swets & Zeitlinger.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34*, 594–621.

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6*, 369–400.

Chincotta, D., & Underwood, G. (1997). Speech rate estimates, language of schooling and bilingual digit span. *European Journal of Cognitive Psychology, 9*, 325–348.

Coady, J. A., & Aslin, R. N. (2004). Young childrens sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology, 89*, 183–213.

Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychology Press.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*, 51–57.

Cowan, N., & Alloway, T. P. (2009). The development of working memory in childhood. In M. L. Courage & N. Cowan (Eds.), *The development of memory in infancy and childhood* (pp. 303–342). London: Psychology Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). UK: Cambridge University Press.

Cowan, N., Nugent, L. D., Elliott, E. M., & Saults, J. S. (2000). Persistence of memory for ignored lists of digits: Areas of developmental constancy and change. *Journal of Experimental Child Psychology, 76*, 151–172.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first and second-language learners. *Reading Research Quarterly, 38*, 78–103.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc..

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language and Hearing Research, 47*, 421–436.

Extra, G., Aarts, R., van der Avoird, T., Broeder, P., & Yagmur, K. (2001). Meertaligheid in Den Haag: De status van allochtone talen thuis en op school. *[Multilingualism in The Hague: The status of minority languages at home and at school]*. Amsterdam: European Cultural Foundation.

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

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83–94.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39*, 3–27.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*, 410–419.

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

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 84–95.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 17–190.

Goddijn, S. & Binnenpoorte, D. (2003). Assessing manually corrected broad phonetic transcriptions in the Spoken Dutch Corpus. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 1361–1364). Barcelona.

Gomes, H., Sussman, E., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G. (1999). Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Developmental Psychology, 35*, 294–302.

Hu, C. F. (2003). Phonological memory, phonological awareness, and foreign language word learning. *Language Learning, 53*, 429–462.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685–701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 1217–1232.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38*, 241–253.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science, 10*, 853–873.

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

Kovacs, G., & Racsmany, M. (2008). Handling L2 input in phonological short-term memory: The effect of non-L1 phonetic segments and non-L1 phonotactics on nonword repetition. *Language Learning, 58*, 597–624.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.

Majerus, S., & Van der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology, 21*, 303–310.

Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language, 51*, 297–306.

Martin, N. (2009). The roles of semantic and phonological processing in short-term memory and learning: Evidence from aphasia. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 220–243). Hove, UK: Psychology Press.

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

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61–78.

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus Users Guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology, 19*, 233–270.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Section 2: The Coloured Progressive Matrices. UK: Oxford Psychologists Press.

Roodenrys, S., Hulme, C., & Brown, G. (1993). The development of short-term memory span: Separable effects of speech rate and long-term memory. *Journal of Experimental Child Psychology, 56*, 431–442.

Rypma, B., Prabhakaran, V., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Load-dependent roles of frontal brain regions in the maintenance of working memory. *NeuroImage, 9*, 216–226.

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

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

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*, 1657–1661.

Standing, L., Bond, B., Smith, P., & Isely, C. (1980). Is the immediate memory span determined by subvocalization rate? *British Journal of Psychology, 71*, 525–539.

Swanson, H. L., Saez, L. M., & Gerber, M. (2006). Growth in literacy and cognition in bilingual children at risk or not at risk for reading disabilities. *Journal of Educational Psychology, 98*, 247–264.

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 729–735.

Thorn, A., Frankish, C. R., & Gathercole, S. E. (2009). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 198–219). Hove, UK: Psychology Press.

Thorn, A. S. C., 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) multinominal processing tree model. *Cognitive Psychology, 50*, 133–158.

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