Effects of bilingualism on statistical learning in preschoolers

Verhagen, J.; de Bree, E.

DOI
10.1075/lab.18102.ver

Publication date
2021

Document Version
Final published version

Published in
Linguistic Approaches to Bilingualism

Effects of bilingualism on statistical learning in preschoolers

Josje Verhagen and Elise de Bree
University of Amsterdam

Earlier work indicates that bilingualism may positively affect statistical learning, but leaves open whether a bilingual benefit is (1) found during learning rather than in a post-hoc test following a learning phase and (2) explained by enhanced verbal short-term memory skill in the bilinguals. Forty-one bilingual and 56 monolingual preschoolers completed a serial reaction time task and a nonword repetition task (NWR). Linear mixed-effect regressions indicated that the bilinguals showed a stronger decrease in reaction times over the regular blocks of the task than the monolinguals. No group differences in accuracy-based measures were found. NWR performance, which did not differ between the groups, did not account for the attested effect of bilingualism. These results provide partial support for effects of bilingualism on statistical learning, which appear during learning and are not due to enhanced verbal short-term memory. Taken together, these findings add to a growing body of research on effects of bilingualism on statistical learning, and constitute a first step towards investigating the factors which may underlie such effects.

Keywords: bilingualism, statistical learning, serial reaction time task, non-adjacent dependency learning, verbal short-term memory

1. Introduction

Previous studies have found that bilingual speakers show an increased ability in statistical learning as compared to monolingual speakers, that is, an increased ability to track frequency information about the co-occurrence of elements (Bartolotti et al., 2011; de Bree, Verhagen, Kerkhoff, Doedens, & Unsworth, 2017; Wang & Saffran, 2014). The available literature leaves open two questions that will be investigated in this study on statistical learning in monolingual and bilingual preschoolers. The first is whether a bilingual advantage in statistical learning is visible *during* the learning task, rather than in a test phase following a learning

https://doi.org/10.1075/lab.18102.ver | Published online: 11 February 2020
Linguistic Approaches to Bilingualism 11:5 (2021), pp. 611–639. ISSN 1879-9264 | E-ISSN 1879-9272

phase. This will help determine whether bilinguals' increased performance stems from better abilities to track structural regularities in the input as it is encountered rather than to better retrieval of these regularities in a post-hoc test. The second question is whether bilinguals' advantage in statistical learning is due to enhancements in verbal short-term memory, as has been proposed – but not yet tested – in earlier research (Bartolotti et al., 2011; Wang & Saffran, 2014).

Over the past few years, evidence has accumulated that bilingualism may impact positively on statistical learning (for reviews, see Bulgarelli, Lebkuecher, & Weiss, 2018; Hirosh & Degani, 2017). Such positive effects of bilingualism on statistical learning have been observed in participants of different ages, including young infants and adults, and for both auditory and visual statistical learning. For infants, Kovács and Mehler (2009) found that bilingual seven-month-old infants learned two three-syllable structures simultaneously, while monolingual infants learned only one. De Bree and colleagues (2017) found that bilingual 24-month-olds tracked non-adjacent dependencies more readily than monolingual peers, at least in a condition in which the input contained exceptions from a predominant rule. For adults, Wang and Saffran (2014) found that bilingual Mandarin-English and Spanish-English speakers outperformed English monolinguals in learning statistical regularities from a tonal language. Since only the Mandarin-English group had experience with a tonal language, bilinguals' higher performance could not be attributed to prior experience with a tonal system.

However, mixed results have been reported. Bartolotti and colleagues (2011) found that bilingual adults outperformed monolinguals in learning novel word forms from Morse Code in only one out of two experiments in their study. Specifically, these authors found that bilinguals who spoke English and another language performed significantly better than English monolinguals in an experiment in which participants were presented with Morse Code for the first time and only one cue was presented to detect word boundaries. In a second experiment in which participants were presented with Morse Code again and two cues (rather than one cue) were provided, these same bilinguals did not outperform monolinguals. Similarly, Poepsel and Weiss (2016) found that bilingual adults outperformed monolinguals in a statistical word learning task testing participants' learning of one-to-one mappings between word labels and referents, but not in an experiment testing one-to-many mappings. Finally, Yim and Rudoy (2013) did not find effects of bilingualism in two experiments assessing auditory and visual rule learning in 5- to 13-year-old children. As a possible explanation of their null findings, Yim and Rudoy proposed that bilinguals' advantage is only found when two sets of statistical rules are presented, rather than one set of rules, since the presence of multiple rules mirrors the dual language input situations bilinguals have experience with. However, whereas this explanation could explain the results obtained in some of the previous studies (de Bree et al., 2017; Kovács & Mehler, 2009; Poepsel & Weiss, 2016), it is at odds with the results of studies reporting a bilingual advantage in statistical learning tasks containing one set of rules only (Bartolotti et al., 2011; Bonifacci, Giombini, Bellochi, & Contento, 2011; Wang & Saffran, 2014).

Previous studies with children and adults have typically compared monolinguals and bilinguals statistical learning abilities in training-test designs (Bartolotti, et al., 2011; de Bree et al., 2017; Kovács & Mehler, 2009; Wang & Saffran, 2014; Yim & Rudoy, 2013). In such experiments, participants first listen to a continuous stream of stimuli governed by a rule or set of rules. Subsequently, they perform a test assessing their knowledge of the rule(s). For adults, this test is commonly a forced-choice selection task, in which participants indicate which one out of two strings conforms to the novel language just heard. For infants and toddlers, head-turn preference paradigms have been used, assessing children's listening times to strings that either were or were not presented during training (de Bree et al., 2017; Kovács & Mehler, 2009). Importantly, results obtained with such training-test designs leave open whether bilinguals' increased performance stems from better abilities to track structural regularities from the input (Kuo & Anderson, 2010, 2012), or to increased abilities to *retrieve* the knowledge of these regularities in a post-hoc test.

Preliminary evidence that bilinguals are better in statistical learning *online*, at least in a non-verbal task, comes from a study by Bonifacci and colleagues (2011). In this study, monolingual and bilingual 6- to 12-year-olds and 14- to 22-year-olds performed a visual statistical learning task in which the ability to learn co-occurrence patterns between shapes was assessed. Specifically, in this task, participants were trained on a sequence of colored shapes. On some trials, the sequence stopped and participants were asked to report the color of the shape. Across age groups, the bilingual participants showed faster response times on the anticipation trials than the monolinguals. The authors proposed that bilinguals' increased ability to track statistical regularities in the input is due to their extensive practice in anticipating linguistic elements in processing two different linguistic systems. The results of this study contrast with those of Park and colleagues (2018), however, who found no effects of bilingualism on a visual serial reaction timed task in which bilingual and monolingual 8- to 12-year-olds with or without specific language impairment responded to sequences of visual cues in specific locations. Since no earlier studies using online assessments have targeted *verbal* statistical learning, it is currently unknown whether bilinguals have enhanced abilities to track statistical relationships from verbal stimuli, as these are presented to them.

An important question, moreover, is how previously found bilingual advantages in statistical learning should be explained. One proposal is the structural sensitivity hypothesis (Kuo & Anderson, 2010, 2012), which holds that bilinguals show heightened sensitivity to structural properties of language because dual language exposure renders the structural properties of languages more salient. As a result of their heightened sensitivity to linguistic structure, bilinguals would be better in learning structural regularities from novel input (de Bree et al., 2017; Kuo & Anderson, 2012). Proponents of the structural sensitivity hypothesis also point to the possibility that enhanced executive functioning may play a role in bilinguals' enhanced statistical learning. Specifically, to overcome interlingual interference, bilinguals would need to attend to structural features of language and flexibly inhibit attention to less relevant linguistic information, which would help them impute linguistic structure more readily.

A second explanation that has been proposed to account for bilinguals' improved statistical learning is enhanced verbal short-term memory (Bartolotti et al., 2011; Wang & Saffran, 2014). Verbal short-term memory refers to the ability to hold auditory information in memory for a brief period of time, and is generally assessed with digit span or nonword repetition tasks (Gathercole, 2006). Previous studies examining effects of bilingualism on verbal short-term memory tasks have provided mixed findings. On the one hand, studies have found positive effects of bilingualism (Biedroń & Szczepaniak, 2012; Delcenserie & Genesee, 2017; Kaushanskaya, 2012), and attributed these to bilinguals' broader linguistic knowledge, which would facilitate short-term storage of nonwords, or to the high demands that bilingual processing and use place on verbal short-term memory. Other studies found either no or negative effects of bilingualism on verbal short-term memory tasks (Boerma et al., 2015; Fernandes, Craik, Bialystok, & Kreuger, 2007). A possible explanation is that results vary as a function of the specific properties of the task stimuli used (digits or nonwords, degree of language-specificity of the nonwords) as well as of the bilingual participants. Specifically, bilinguals may perform more poorly than monolinguals on verbal short-term memory tasks if tasks are based on a language that the bilinguals are less proficient in (Armon-Lotem & Meir, 2016; Boerma et al., 2015; Messer et al., 2010).

Although the previous evidence regarding effects of bilingualism on verbal short-term memory is mixed, the idea that increased verbal short-term memory underlies bilinguals' advantage in (auditory) statistical learning is plausible. Previous research shows that individual differences in performance on tasks that rely on verbal short-term memory tasks at least to some degree are positively associated with statistical learning in monolingual children (Kapa & Colombo, 2014) and adults (Misyak & Christiansen, 2012). Two processes have been proposed to underlie statistical learning: extraction and integration (Erickson & Thiessen, 2015). Extraction refers to the process of holding statistically congruent clusters in memory (Perruchet & Tillmann, 2010); integration to the process of combining information across the stored clusters. In particular extraction is assumed to rely on verbal working memory, including short-term storage of information. In fact, in a recent framework in which statistical learning in various tasks (word segmentation, category learning, artificial language learning, SRT tasks) was considered, statistical learning was argued to arise from a set of memory processes, with a key role for the storage of elements (Thiessen, 2017). Specifically, the idea is that participants store exemplars in memory, and subsequently, integrate information from these exemplars, such that features that are consistent across them are strengthened, and features that are inconsistent across them are weakened, leading to knowledge of statistical regularities.

Despite previous claims that bilinguals' enhanced statistical learning is explained by enhanced verbal short-term memory (Bartolotti et al., 2011; Wang & Saffran, 2014), no previous studies have yet tested this proposal. Some tentative evidence comes from de Bree and colleagues (2017), who found a positive and significant correlation between monolingual and bilingual toddlers' performance in a non-adjacent dependency learning experiment and children's scores on a nonword repetition task. However, in this study, correlations between verbal short-term memory and statistical learning were calculated with monolingual and bilingual children collapsed. Hence, it is as yet an open question whether bilingual advantages in statistical learning are explained by bilinguals' improved verbal short-term memory skill.

1.1 This study

The main aim of the current study was to investigate if bilingual children outperform monolingual peers in learning structural relationships from a novel language when the learning process itself is assessed, rather than knowledge of the learned material in a post-test. Furthermore, our study aimed to test if any effects of bilingualism found were due to improved verbal short-term memory skill in the bilinguals.

To address these aims, we employed a task which was based on the Serial Reaction Time (SRT) task paradigm (Nissen & Bullemer, 1987). In a typical SRT task, participants are presented with visual stimuli and asked to press a button as quickly as possible upon appearance of a specific stimulus (Hunt & Aslin, 2001; Lum et al., 2014; Vicari et al., 2003). Typically, there are a number of regular blocks in which stimuli conform to a rule or set of rules, and there is a final block in which presentation is irregular. If participants learn the rule(s), their performance should improve over the regular blocks, as indicated by more accurate and faster responses to specific stimuli over blocks, followed by a drop in performance or stabilized performance in the final, random block. The task we used resembled an earlier online statistical learning task used with monolingual children (Lammertink, van Witteloostuijn, Boersma, Wijnen & Rispens, 2019), and tested children's learning of non-adjacent dependencies. Non-adjacent dependencies have been well-researched in both children and adults (Gómez, 2002), and have ecological relevance, as they commonly occur in natural language (e.g., subject-verb agreement as in '*He* walk-*s*'). The children in our study also completed a nonword repetition (NWR) task, to assess their verbal short-term memory abilities. Our questions were the following:

1. Do bilingual children outperform monolingual children during statistical learning of non-adjacent dependencies in the SRT-based task?
2. Do bilingual children outperform monolingual children on the NWR task? If so, can this bilingual advantage in statistical learning performance be explained by bilinguals' advanced verbal short-term memory skills?

As for the first question, we hypothesized that bilingual children would outperform monolingual peers on the SRT task. This bilingual advantage could show up as the presence of a typical SRT-curved pattern in the bilinguals, but not in the monolinguals, or a more prominent SRT-curved pattern (e.g., steeper increase in performance over the first three blocks and/or stronger decrease in performance in the final block) in the bilinguals than monolinguals. Our prediction was based on earlier research showing a bilingual advantage in verbal statistical learning studies with bilinguals of different ages using a training-test design (Bartolotti et al., 2011; Kovács & Mehler, 2009), including earlier work on non-adjacent dependency learning in young children (de Bree et al., 2017) and online sequence learning in bilingual children (Bonifacci et al., 2011).

As for the second question, we predicted that the bilinguals would outperform the monolinguals on NWR. This prediction was based on earlier work showing increased performance on verbal memory measures in bilinguals, using NWR tasks, at least for adults (Delcenserie & Genesee, 2016; Kaushanskaya, 2012). However, as described earlier, lower performance in bilinguals has also been reported (Boerma et al., 2015; Fernandes et al., 2007), especially in situations where the bilinguals were less proficient in the language that the NWR task was based on (Boerma et al., 2015; Messer et al., 2010; Parra et al., 2011). Therefore, in the current study, we took into account the degree to which items were based on Dutch, by considering the phonotactic probability of the NWR items. In so doing, we explored the possibility that any advantage on the task for the bilingual children would surface for the low-probability items, but not high-probability items (as these latter items were more specific to Dutch). Finally, we predicted that, in case the bilinguals would outperform the monolinguals on (part of) the NWR task, enhanced verbal short-term memory would explain their improved statistical learning abilities. This prediction was based on earlier proposals that verbal storage is implicated in statistical learning (Erickson & Thiessen, 2015; Thiessen, 2017) as well as earlier findings that verbal short-term memory skill is associated with statistical learning in monolingual and bilingual children (de Bree et al., 2017; Kapa & Colombo, 2014).

2. Method

2.1 Participants

Participants were 56 monolingual and 41 bilingual kindergarteners. All children came from families that resided in the Netherlands. Mean age was 5;1 years in the monolingual group ($SD = 0;7$) and 5;2 years in the bilingual group ($SD = 0;9$). The monolingual group contained 31 girls (55%) and the bilingual group 19 girls (46%). Age and gender did not differ significantly between the groups ($t(95) = 1.073$ $p = .286$ for age; $\chi^2(1, N = 97) = 0.770$, $p = .380$ for gender). Six additional children were tested (four bilinguals; two monolinguals), but not included in the final sample, because they did not seem to understand the task, as indicated by very low response rates (i.e., no more than two hits per block of the task). The monolingual children all came from families in which no language other than Dutch was spoken, as indicated by parental report in an electronic questionnaire. The bilingual children were from families in which one out of a diverse set of other languages was spoken instead of or next to Dutch ($N = 7$ Turkish, $N = 5$ German, $N = 5$ Polish, $N = 4$ English, $N = 3$ Kurdish, $N = 2$ Surinamese, $N = 2$ Italian, $N = 2$ Moroccan Arabic, $N = 2$ Portuguese, $N = 1$ Chinese, $N = 1$ Russian, $N = 1$ Thai, $N = 1$ Danish, $N = 1$ Spanish, $N = 1$ Frisian, $N = 1$ French, $N = 1$ Farsi, $N = 1$ Somali). Most children heard both languages at home ($N = 29$ (71%)); the others were exposed to the other language at home, and were in contact with the majority language (Dutch) only at kindergarten and through contacts outside of their homes. Parents indicated which language(s) they spoke to their child for each parent separately. They also indicated which language their child spoke best: Dutch ($N = 18$ (44%)), the other language ($N = 12$ (29%)), or both languages equally well ($N = 11$ (27%)). In both groups, the majority of the children came from families in which parents were highly educated, as indicated by parents' responses in the parental questionnaire. Specifically, 47 (84%) of the monolinguals had at least one parent who had attained a college or university degree. Twenty-eight (70%) of the bilinguals had at least one parent who had attained a college or university degree. For one bilingual child, this information was missing. Parental education did not differ significantly between the groups ($\chi^2(1, N = 96) = 2.649$, $p = .104$. Informed consent was obtained from children's parents prior to data collection.

2.2 Materials

2.2.1 *Statistical learning experiment*

In our SRT-based task, children were presented with auditory speech strings that consisted of three elements. These triplets took the form a-X-b and c-X-d, in which X was a variable element and a, b, c, and d were fixed. Two counterbalanced languages were created that children were randomly assigned to: a language containing triplets of the type a-X-b (*rak-X-toef*) and c-X-d (*sot-X-lut)* and a language containing counterbalanced a-X-d (*rak-X-lut*) and c-X-b (*sot-X-toef*). Even though *toef* is an existing word in Dutch, we included it in our stimuli to have the exact same stimuli as in earlier work on statistical learning in Dutch children (de Bree et al., 2017) and adults (Grama, Kerkhoff, & Wijnen, 2016). Importantly, *toef* is a highly infrequent word, occurring only 0.05 in every million (SUBTLEX, Keuleers, Brysbaert & New, 2010). As X-elements, the following 18 two-syllable pseudowords were used: *fidan, bensim, sulep, hiftam, wiffel, domo, vami, banip, kengel, naspoe, rogges, noeba, kasi, snigger, poemer, plizet, wadim,* and *mofig.* The task contained four blocks. In the first three blocks, all stimuli were regular, such that they only contained triplets conforming to the rules of the languages presented. In Language 1, triplets were a-X-b, c-X-d (*rak*-X-*toef, sot*-X-*lut*); in Language 2, these were a-X-d and c-X-b (*rak*-X-*lut* and *sot*-X-*toef*). In the fourth block, stimuli did not conform to these rules, such that the final elements *lut* and *toef* were combined with the incorrect first element (e.g., *rak-X-toef* for the language that contained *sot-X-toef*) as well as with other a- and c-elements that had not been presented during the previous blocks (e.g., *<u>jik</u>-X-toef, <u>tep</u>-X-toef*). Specifically, in this random block, 18 triplets containing the target word were presented, six of which were paired with the first element of the other triplet (e.g., *rak-X-toef* for the language that contained *sot-X-toef*), and twelve were paired with either one of two new elements that had not been presented before (*jik* or *tep*). For an overview of all stimuli, see Table 1.

All triplets were presented to the children over headphones with a 250-ms interval in between the pseudo-words in a triplet and a 750-ms interval in between triplets. Triplets had been constructed by cross-splicing the pseudowords from pre-recorded speech from a native speaker of Dutch, to ensure uniformity of the stimuli and avoid co-articulation effects (see de Bree et al., 2017 for details). To make the task more engaging for the children, it was embedded in a narrative: prior to the task, children were told that they were going to see pictures of moles on the laptop screen and, at the same time, hear the name of each mole. All moles looked exactly the same (see Figure 1), so children had to listen carefully to the names of the moles. Children were then told that there was going to be a party for moles and that their task was to invite to this party as many moles as they could. Crucially, they were instructed to invite only the mole named *lut* (or *toef* in the counterbalanced version), by pressing a button on a push-button box as quickly as possible when they heard a mole called *lut* (or *toef* in the other version), or when they expected a mole called *lut* (or *toef*) to appear. This procedure allowed investigating whether children became faster and more accurate in responding to the final word of the triplet over time, and actually could predict it, which would be indicative of children's learning of the non-adjacent dependency. To keep children engaged over blocks, they were shown a picture after each block which depicted all the moles they had 'collected' during the immediately preceding block. They then counted the number of moles they had collected together with the experimenter, and were given positive feedback irrespective of whether they had performed well, to keep them motivated for the next block. Button presses were recorded from the offset of the X-element to 750 ms after the offset of the final word of the triplet. If children pressed the button correctly, that is, if the triplet indeed contained the target word, the respective ('target') mole was given a party hat, to signal a correct response. If children did not press

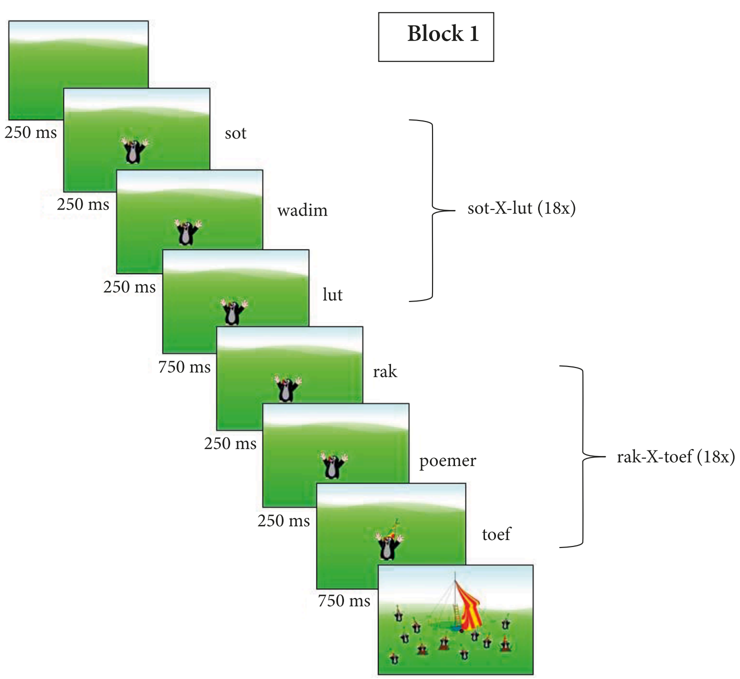


Figure 1. Illustration of block 1 in the SRT-based non-adjacent dependency learning experiment

Note. In the experiment, triplets were presented in randomized order. Button presses were recorded from the offset of the X-element. A mole with a party hat was shown immediately after children had pushed the button upon identifying a correct target (here *toef*); in all other instances, a mole without a hat was shown.

the button when the target word was presented or pressed the button for strings that did not contain the target word, no feedback was provided. Positive feedback to correct answers only was chosen over other types of feedback, based on pilot data showing that children understood the task and were very motivated to perform well when given positive feedback only. Prior to the first test block, a practice block was presented that contained 28 singlets, four of which were target words. Children were instructed to press a button as fast as they could when they heard a target word (*lut* or *toef*). The items presented in the practice phase were the same elements used in the experimental blocks, in random order. If children pressed the button correctly, the experimenter made a brief 'thumps up' gesture, as a means of positive feedback to the child. The aim of the practice block was to familiarize children with the procedure of pressing the button *only* if they heard a target word and do so as quickly as possible. Each of the first three (regular) test blocks was composed of 18 triplets containing a target word and 18 triplets of the other dependency (without a target word). Each block thus contained 18 target words. Counterbalanced across experiment versions were the specific target word (i.e., *toef* vs. *lut*) as well as the language that children were presented with (Language 1: *rak-X-toef* and *sot-X-lut* vs. Language 2: *rak-X-lut* and *sot-X-toef*), resulting in four experiment versions. The final, random block also contained 18 targets, but these occurred in unpredictable positions. The task was programmed using the experimental software Zep (http://beexy.org/zep/), and administered on a laptop. A button box was used to record children's responses. For each trial, accuracy (hit/false alarm) and response time (in the case of a button press) were recorded.

Table 1. Stimuli of the SRT-based non-adjacent dependency learning experiment

		Language1		**Language2**	
Experiment	**Nr. of strings**	**Non-adjacent dependencies**		**Non-adjacent dependencies**	
Block 1	2 × 18	a-$X_{(1-18)}$-b (rak X toef)	c-$X_{(1-18)}$-d (sot X lut)	a-$X_{(1-18)}$-d (rak X lut)	c-$X_{(1-18)}$-b (sot X toef)
Block 2	2 × 18	a-$X_{(1-18)}$-b (rak X toef)	c-$X_{(1-18)}$-d (sot X lut)	a-$X_{(1-18)}$-d (rak X lut)	c-$X_{(1-18)}$-b (sot X toef)
Block 3	2 × 18	a-$X_{(1-18)}$-b (rak X toef)	c-$X_{(1-18)}$-d (sot X lut)	a-$X_{(1-18)}$-d (rak X lut)	c-$X_{(1-18)}$-b (sot X toef)
Block 4	2 × 18	c-$X_{(1-6)}$-b (sot X toef)	c-$X_{(1-18)}$-d (sot X lut)	c-$X_{(1-6)}$-d (sot X lut)	c-$X_{(1-18)}$-b (sot X toef)
		e-$X_{(1-6)}$-b (tep X toef)		e-$X_{(1-6)}$-d (tep X toef)	
		f-$X_{(1-6)}$-b (jik X toef)		f-$X_{(1-6)}$-d (jik X toef)	

Note. $X_{(\text{subscript number})}$ refers to the different X-items used

2.2.2 *Nonword repetition*

The nonword repetition (NWR) task by Rispens and Baker (2012) was used to assess children's verbal short-term memory. In this task, children listen to nonwords and are then asked to repeat these nonwords. The nonwords in the task are based on Dutch, and manipulated for phonotactic probability such that half of the nonwords contain phoneme combinations that are frequent in Dutch (e.g., 'tanoolon' /ta:no:lɔn/) and the other half combinations that are infrequent in Dutch (e.g., 'muihuuguf' /mœyhuxʏf/). The task consists of two blocks, but for the current study, only the first block was administered in order to reduce testing time (le Clercq et al., 2017). The task contained six three-syllable, eight four-syllable, and eight five-syllable nonwords. Children's responses were recorded with a voice recorder, and coded later as correct/incorrect. Eight (8.2%) percent of randomly selected data was coded by an additional researcher, following earlier work in which similar percentages of NWR data were double-coded (Duncan & Paradis, 2016 (10%); McKean, Letts, & Howard, 2013 (10%)). Interrater agreement for the scores was 85% and task reliability was .71 (Cronbach's alpha).

2.3 Procedure

Children were tested individually by a research assistant in a quiet room at their schools or in the Babylab at Utrecht University. Task order was fixed with the SRT task preceding the NWR task. Together, these tasks lasted about twenty minutes. At the end of the session, children received a small gift. Parents filled out an electronic questionnaire about their child and their home language(s), and provided written informed consent for their child's participation.

2.4 Variables and analyses

Linear mixed-effects regression models were run in R version 3.4.1 (R Core Team, 2015) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015), with *p*-values obtained through the LmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). In these models, we were interested in the effects of 'group' and 'block' as well as interaction effects between these factors, which is why we tested for the effects of these factors directly, rather than through model comparisons. Interaction effects were plotted using the interaction.plot package (Chambers, Freeny, & Heiberger, 1992). In all models, orthogonal sum-to-zero contrast coding was applied to our fixed effects (group, block) (Baguley, 2012). Experiment version (Language 1 or Language 2) was not included as a factor, as it did not yield an effect in any of the models. To solve issues of non-converging models, we increased the number of iterations to 100,000 (Powell, 2009).

To address our first research question on possible effects of bilingualism on SRT performance, four separate models were run, with four different dependent variables (see below). In these models, we entered group (monolinguals/bilinguals), block, and group*block as fixed-effect factors. For block, we set contrasts between (i) block 1 versus block 2, (ii) blocks 1 and 2 versus block 3 (to see whether performance increased over these regular blocks) and (iii) block 3 versus block 4 (to see whether performance decreased from the final regular to the random block). We included by-subject and by-item random intercepts, by-subject random slopes for block, and by-item random slopes for block, group, and block*group.

To address our second question of whether the bilinguals outperformed the monolinguals on the NWR task, we ran a model on children's correct/incorrect scores in the NWR task with group (monolinguals/bilinguals), item length (three, four, or five syllables) and phonotactic probability (high vs. low) as fixed-effect factors. By-subject and by-item random intercepts were included, as well as by-subject random slopes for item length and phonotactic probability and by-item random slopes for group. Item length was included based on earlier work showing a (negative) effect of bilingualism on nonword repetition for long nonwords, but not for shorter ones (Boerma et al., 2015). Furthermore, to see whether any effects of bilingualism on children's performance in the SRT-based task remained once NWR was controlled, we ran additional analyses in which children's NWR sum scores were added as a fixed-effect factor.

As for the dependent variables in the models, four variables were analyzed. First, two variables were constructed that were based on children's response accuracy: hits and *d'*. The variable 'hits' was a categorical variable: 0 for strings that contained a target word, but for which there was no button press; 1 for strings that contained a target word and for which there was a button press. *D'* is a statistic from signal detection theory, which reflects the percentage of button presses for targets relative to the percentage of button presses for non-targets (false alarms). Taking into account both hits and false alarms, *d'* controls for potential response bias, such as a child pressing the button in response to each stimulus. *D'* is typically calculated with the following formula: $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$ (MacMillan & Creelman, 2005), with a higher *d'* signaling more accurate signal detection. For our data, a correction was applied in order to deal with zero scores (Hautus, 1995; Stanislaw & Todorov, 1999), such that a score of 0.5 was added to the number of hits and false alarms and a score of 1 to the number of targets/non-targets, as follows: $d' = Z(\text{number of hits} + 0.5 / \text{total targets presented} + 1) - Z(\text{number of false alarms} + 0.5 / \text{total non-targets presented} + 1)$.

The third and fourth dependent variables in our analyses were based on children's reaction times. First, to control for differences in length of the target words (*lut, toef*), residual reaction times for children's hit responses were calculated by subtracting the duration of the target word from the total reaction time for each hit. The second reaction time-based measure reflected how often children actually pressed the button prior to hearing the target, signaling how often they correctly predicted the target. This variable was calculated by recoding each hit as to whether it occurred prior to the onset of the target word, resulting in a binary variable: 1 for reaction times below or equal to 250 ms; 0 for reaction times above 250 ms (as a 250-ms pause occurred between the X-element and the target).

The inclusion of these four different variables was based on pilot results which showed considerable variation among children in both accuracy and response speed. Moreover, since each variable provided a different piece of information, we included them all, to obtain a complete picture of children's results. Hits were our most basic measure and could be analyzed at the item level. However, this variable did not take into account false alarms, resulting in very high scores for children who pressed the button at each trial. Therefore, *d'* was included. Regarding response speed, residual reaction times to hits reflected how quickly children reacted to or predicted the final word in the triplet. However, a potential flaw in our task was that children would just always wait until they heard the target word before pressing the button. In this case, any decrease in reaction times over successive blocks might be due to children's faster processing of the target or faster motor responses, rather than being indicative of learning the non-adjacent dependency. Therefore, children's anticipations to targets were also analyzed.

Previous research has shown that reaction times may be affected by performance on previous trials, such that reaction times slow down after incorrect responses, a phenomenon known as post-error slowing (Dutilh et al., 2012). Therefore, in the models with reaction times and anticipations, an additional fixed-effect factor was included: accuracy on the previous trial, which reflect whether performance on the previous trial was correct.

3. Results

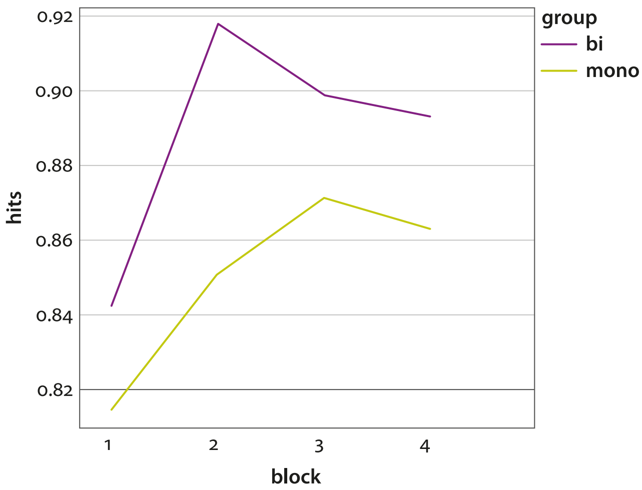
3.1 Comparing SRT performance between the monolingual and bilingual children

Mean scores and standard deviations per block are presented in Table 2 for the monolingual and bilingual children separately. Figure 2 presents these same scores, plotted in graphs, to enable visual inspection of the results. Note that the y-axis in these figures is truncated, differs across response measures, and sometimes has a narrow range. Although this improves visibility of the curves, a drawback is that differences across blocks or groups may seem larger than they actually are.

Table 2. Descriptive statistics per block of the SRT-based task for the monolingual and bilingual children

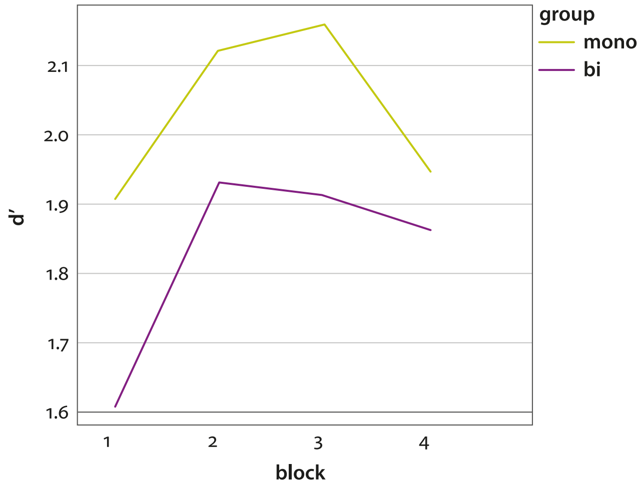
	Monolinguals		Bilinguals	
	M	(SD)	M	(SD)
Accuracy-based variables				
Hits (proportions)				
Block 1	0.81	(0.39)	0.84	(0.36)
Block 2	0.85	(0.36)	0.92	(0.27)
Block 3	0.87	(0.34)	0.90	(0.30)
Block 4	0.86	(0.34)	0.89	(0.31)
d'				
Block 1	1.91	(1.06)	1.61	(1.12)
Block 2	2.12	(1.09)	1.93	(1.14)
Block 3	2.16	(1.06)	1.91	(1.11)
Block 4	1.94	(0.86)	1.86	(1.13)
Reaction-time based variables				
Residualized RTs to hits				
Block 1	405.13	(286.66)	387.58	(280.35)
Block 2	343.46	(248.55)	331.95	(250.10)
Block 3	340.02	(269.85)	268.21	(253.07)
Block 4	347.54	(257.26)	294.03	(261.49)
Anticipations (proportions)				
Block 1	0.027	(0.161)	0.026	(0.158)
Block 2	0.029	(0.168)	0.023	(0.151)
Block 3	0.028	(0.166)	0.046	(0.211)
Block 4	0.030	(0.170)	0.030	(0.172)

Note. For anticipations, three decimals are given, because subtle differences between blocks otherwise remain unnoticed.



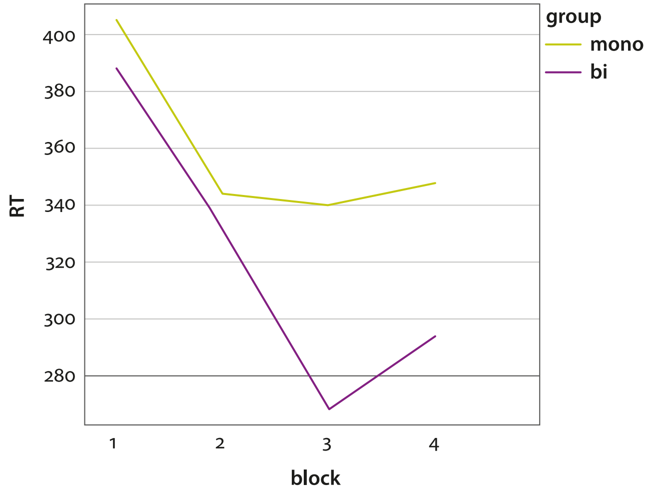
group
bi
mono
hits
0.92
0.90
0.88
0.86
0.84
0.82
1
2
3
4
block

a.

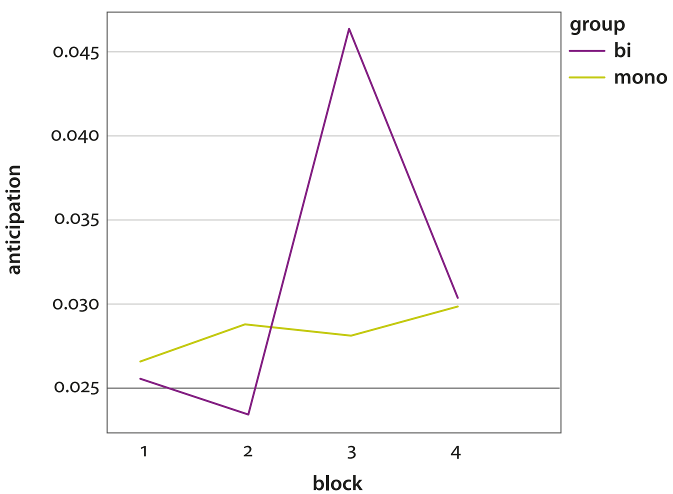


group
mono
bi
d’
2.1
2.0
1.9
1.8
1.7
1.6
1
2
3
4
block

b.



c.



d.

Figure 2. Plotted results per block in the SRT-based task for the monolinguals and bilinguals separately: (a) hit responses (proportions), (b) mean *d'*, (c) residualized reaction times for hits, and (d) anticipations (proportions)

Visual inspection of Figure 2a indicates that children's hit scores became higher over the regular blocks, followed by a drop in scores from the third to the final (random) block, with the exception of a slight decrease in (the otherwise very high) scores from the second to third block in the bilinguals. Although scores were higher overall in the bilingual group, a linear logistic mixed-effect model with group and block as fixed factors (and random intercepts and slopes for subjects and items, see under 'Analyses') showed no effect of group ($\beta = 0.344$, $SE = 0.258$, $z = 1.333$, $p = .182$). Regarding effects of block, the increase in hits between block 1 and 2 was significant ($\beta = 0.532$, $SE = 0.231$, $z = 2.300$, $p = .021$), but the other comparisons were not ($\beta = 0.077$, $SE = 0.176$, z = 0.439, $p = .660$ for blocks 1 and 2 vs. block 3; $\beta = 0.011$, $SE = 0.246$, $z = 0.044$, $p = .964$ for block 3 vs. block 4). The interactions between group and block were not significant either (ps > .1). See Table A1 in the Appendix for the full model results.

A typical SRT-curved pattern also emerged from children's mean *d'* scores per block, shown in Figure 2b. Although scores seemed higher for the monolingual children than for the bilingual children, there was no effect of group ($\beta = 0.206$, $SE = 0.193$, $t = 1.067$, $p = .288$). Rather, there was a significant increase in performance from blocks 1 to 2, regardless of group ($\beta = 0.297$, $SE = 0.099$, $t = 2.993$, $p = .003$) as well as a significant increase from blocks 1 and 2 to block 3 ($\beta = 0.160$, $SE = 0.078$, $t = 2.061$, $p = .040$). The difference in mean *d'* between blocks 3 and 4 was not significant ($\beta = -0.057$, $SE = 0.103$, $t = -0.546$, $p = .585$) and there were no interactions between block and group (ps > .1, see Table A1 in the Appendix).

Residualized reaction times for children's hit responses showed the expected, reversed pattern of the accuracy scores (hits and *d'*), as there was a decline in reaction times during the regular blocks, followed by an increase from the third to the final block (see Figure 2c). A linear mixed-effect model with group and block as fixed effect factors showed a main effect of group ($\beta = 39.730$, $SE = 6.930$, $t = 5.773$, $p < .001$), which indicated that the bilinguals responded more quickly overall than the monolinguals. There also was a significant interaction between group and 'blocks 1 and 2 versus block 3', which indicated that the decrease in response rate from blocks 1 and 2 to block 3 was significantly stronger in the bilinguals than in the monolinguals ($\beta = 49.130$, $SE = 17.810$, $t = 2.759$, $p = .006$). There were no further main effects or interaction effects (ps > .1, see Table A1 in the Appendix).

Finally, concerning children's anticipations to target words, that is, how often they pressed the button prior to or exactly at the onset of the target word, it is noteworthy that such responses were infrequent overall. A linear logistic mixed-effect model showed no effects (see Table 2). However, there were two trends in the data. First, there was a tendency for an increase in anticipations from blocks 1 and 2 to block 3 ($\beta = 0.337$, $SE = 0.186$, $z = 1.810$, $p = .070$). Second, a trend towards an interaction effect between group and 'blocks 1 and 2 versus block 3' indicated that this tendency to anticipate more to the upcoming target in block 3 as compared to blocks 1 and 2 was stronger in the bilingual group than in the monolingual group ($\beta = -0.678$, $SE = 0.372$, $z = -1.820$, $p = .069$). Fort the full model results, see Table A1 in the Appendix.

3.2 Comparing verbal short-term memory between the bilingual and monolingual groups

Mean nonword repetition scores were available for 95/97 children, with missing data for two monolingual children due to technical error. Table 3 presents proportions correct on items of high- and low-probability items and different lengths for the monolingual and bilingual children, separately.

Table 3. Mean scores (proportions correct) and standard deviations on the NWR task for the monolingual and bilingual children

	Monolinguals (*n* = 54)		Bilinguals (*n* = 41)	
	M	(SD)	M	(SD)
All items	0.36	(0.48)	0.31	(0.46)
High-probability items (all)	0.46	(0.50)	0.39	(0.49)
3-syllable items	0.78	(0.41)	0.63	(0.48)
4-syllable items	0.49	(0.49)	0.39	(0.49)
5-syllable items	0.22	(0.41)	0.15	(0.36)
Low-probability items (all)	0.28	(0.45)	0.25	(0.43)
3-syllable items	0.57	(0.50)	0.61	(0.49)
4-syllable items	0.19	(0.40)	0.18	(0.38)
5-syllable items	0.18	(0.38)	0.08	(0.28)

A linear mixed-effect model with group, item length, and phonotactic probability as fixed effects showed no effect of group ($\beta = -0.372$, $SE = 0.745$, $z = -0.499$, $p = .618$). There was a main effect of item length, such that longer items were repeated less accurately than shorter ones ($\beta = -0.733$, $SE = 0.089$, $z = -8.191$, $p < .001$). Moreover, there were two significant interactions: (i) a group*phonotactic probability interaction indicated that the effect of phonotactic probability was stronger for the monolingual than for the bilingual children ($\beta = -3.635$, $SE = 1.364$, $z = -2.665$, $p = .008$) and (ii) a group*phonotactic probability*item length interaction indicated that that the monolinguals' greater sensitivity to phonotactic probability as compared to that of the bilinguals was stronger for the shorter versus longer items ($\beta = 0.400$, $SE = 1.364$, $z = -2.665$, $p = .008$). There were no other main effects or significant interactions ($ps > .1$).

Even though the bilingual children did not outperform the monolingual children on the NWR task, we examined whether our results for the SRT task remained unchanged when differences in NWR were controlled. To this aim, the above models were re-run with NWR sum scores as an additional fixed-effect factor. The results of these models showed that NWR positively predicted children's overall performance for three of the four response measures: *d'* ($\beta=0.092$, $SE=0.027$, $z=3.426$, $p<.001$), reaction times ($\beta=-10.248$, $SE=5.048$, $z=-2.030$, $p=.045$), and anticipations ($\beta=-0.286$, $SE=0.097$, $z=2.950$, $p=.003$) (see Table A2 in the Appendix). Moreover, the results showed that the group*block interaction effect that was reported above remained, which indicated that the bilinguals showed a steeper increase in reaction times from blocks 1 and 2 to block 3 than the monolinguals ($\beta=34.020$, $SE=14.319$, $z=2.376$, $p=.017$). The results showed, furthermore, that NWR interacted significantly with children's hit scores across successive blocks, such that children with higher NWR scores showed a steeper increase in hits from blocks 1 and 2 to block 3 ($\beta=0.103$, $SE=0.041$, $z=2.503$, $p=.012$) and a steeper decrease in hits from block 3 to 4 than children with lower NWR scores ($\beta=-0.155$, $SE=0.061$, $z=-2.527$, $p=.012$). Likewise, children with higher NWR scores showed a steeper decrease in anticipations from blocks 3 to 4 than children with lower NWR scores ($\beta=-0.466$, $SE=0.201$, $z=-2.319$, $p=.020$). However, caution in interpreting these results is needed, due to the very large number of statistical comparisons made, and hence, the increased likelihood of finding false positives.

4. Discussion

The main aim of our study was to see whether bilingual children showed an advantage over monolingual children in statistical learning when, instead of the product of statistical learning, the process itself was studied. Our second aim was to see whether, if a bilingual advantage was found, this could be explained by enhanced verbal short-term memory in the bilinguals. To address these aims, an experiment was conducted based on the Serial Reaction Timed (SRT) task paradigm, which assessed children's sensitivity to non-adjacent dependencies. Scores on a nonword repetition (NWR) task, a measure of verbal short-term memory, were also analyzed to see if the bilinguals outperformed the monolinguals, as well as how NWR performance related to differences in statistical learning. The results suggested learning curves that are characteristic of SRT tasks, showing an increase in performance over the regular blocks, followed by a decrease or stabilized performance from the final, regular to the random block. Results of linear mixed-effect models showed effects of block for two out of four responses measures: hits (increase from blocks 1 to 2) and *d'* (increase from blocks 1 to 2, and from blocks 1 and 2 to 3). These results are suggestive of learning, as they, at least in part, signal increases in performance over the regular blocks, followed by stabilized performance from the final regular to the random block (e.g., Vicari et al., 2003). In addition, an interaction effect between group and block was found for children's reaction times, which indicated that the increase in performance during the regular blocks of the task was stronger for the bilinguals than for the monolinguals. Trend effects emerged for children's anticipations: the bilinguals showed a trend towards a stronger increase in anticipations from blocks 1 and 2 to block 3 than the monolinguals. These findings indicate that in particular the bilingual children were sensitive to the regularity tested in the current experiment that, in triplets of the type *sot-X-toef* or *rak-X-lut* (or vice versa), the third element had a dependency relation with the first element. While caution is needed in interpreting these results for anticipations (as these were based on few data points), the trend for anticipations is important, as it suggests that bilinguals did not simply become more accurate in *responding* to a particular target word, pressing the button when they heard (part of) this target, but became slightly faster over time in *predicting* the target on the basis of the first element of the triplet.

Regarding NWR, there were no significant differences between the bilinguals and monolinguals. When NWR was controlled, the interaction effect between group and block on children's reaction times remained, such that bilinguals showed a stronger decrease in reaction times from blocks 1 and 2 to block 3 than the monolinguals. Also, there were several significant interactions between NWR and children's performance over consecutive blocks, albeit not for all blocks and for all response measures. This suggests that NWR was associated with learning the non-adjacent dependency relation in our experiment at least to some degree, which accords with the earlier literature on the role of verbal memory in statistical learning (Erickson & Thiessen, 2015) as well as earlier research findings showing associations between NWR and statistical learning (de Bree et al. 2017; Misyak & Christiansen, 2012). The lack of an effect of bilingualism on NWR performance is in keeping with previous studies reporting no effects of bilingualism on measures of verbal short-term memory, including nonword repetition as well as digit and word span tasks (de Bree et al., 2017; Fernandes, Craik, Bialystok, & Kreuger, 2007), but contrasts with studies showing enhanced performance on such tasks in bilingual speakers (Biedroń & Szczepaniak, 2012; Delcenserie & Genesee, 2017; Kaushanskaya, 2012). In our study, the stronger effect of phonotactic probability in the monolingual as opposed to bilingual children suggests that language-specificity of the items played a role, which might explain why no positive effect of bilingualism was found. Indeed, earlier work has shown that bilinguals perform more poorly than monolinguals on verbal short-term memory measures containing items that are based on a language that the bilinguals are less proficient in (Boerma et al., 2015; Messer et al., 2010). Future work could investigate in more detail how the specific properties of the nonword stimuli and the bilingual participants may modulate effects of bilingualism on NWR performance. Also, an NWR measure that is less subject to effects of existing language knowledge (e.g., quasi-universal NWR task, Boerma et al., 2015) could be used to see how verbal short-term memory relates to statistical learning in monolingual and bilingual children.

Positive effects of bilingualism did not emerge for the accuracy-based measures in our study – hits and *d'* (a measure of signal sensitivity). One possibility that awaits further investigation is that effects of bilingualism on statistical learning are modulated by bilinguals' language pairings, an idea that could be interesting to pursue in the light of the structural sensitivity hypothesis (Kuo & Anderson, 2012). Perhaps, a certain degree of typological distance is needed for bilinguals to develop enhanced sensitivity to linguistic structure. Another possibility to be investigated is that bilinguals' children language proficiency plays a role. For adults, Onnis et al. (2017) found that bilinguals with more balanced proficiency levels had improved statistical learning skills as compared to bilinguals with less balanced profiles. Future work could address whether more balanced proficiency profiles in bilinguals lead to enhanced statistical learning directly, through enhanced knowledge of the properties of two languages, or indirectly, through enhanced executive processing skills (Bulgarelli, Lebkuecher, & Weiss, 2018; Hirosh & Degani, 2017).

Our study has a number of limitations. The first is that we only assessed children's verbal short-term memory and not verbal working memory. While verbal-short term memory is implicated in statistical learning through the process of extraction, verbal working memory is thought to be involved in the integration of information across clusters stored in memory (Erickson & Thiessen, 2015; Thiessen, 2017). Given that previous research has shown that bilinguals may outperform monolinguals on working memory (Blom et al., 2014), it is worth investigating whether, and if so to what extent, bilinguals' enhanced statistical learning is due to improved working memory abilities.

A second limitation of our study was that the time window in our task in which children could predict the upcoming target word was relatively short (i.e., 250 ms). Consequently, anticipations were infrequent overall, perhaps also because young children are typically slow to respond. While the trend effects for anticipations support the validity of our task, any potential stronger effects of bilingualism on this variable may have been hard to detect, because of the low number of anticipations and high variation within the groups. Furthermore, we cannot exclude the possibility that anticipatory responses, as operationalized in the current study, were due to low impulse control. Future assessments could include longer time intervals, enabling a better measurement of children's ability to predict upcoming targets, perhaps combined with an assessment of impulse control as a control measure. In addition, future studies could assess children's online learning of multiple rules that mirror more closely the dual language input that bilingual children are exposed to (Kovács & Mehler, 2009; Weiss, Gerfen & Mitchell, 2009).

A third limitation is that, in our study, we did not add an offline task, and thus were unable to see how children's outcomes in the online task related to more traditional measures of statistical learning used in previous work. Online measures might, however, be able to capture children's learning better than offline measures. A recent study by Lammertink and colleagues (2019) assessed online learning of auditory non-adjacent dependencies in five- to eight-year-old monolingual children in an SRT task, and found that performance on this task did not correlate with performance on an offline grammaticality judgement test. The authors proposed that offline measures may be more complex for children, as they rely on explicit decisions (Franco et al., 2015; Lammertink et al., 2019). Nevertheless, in future studies, attempts could be made to supplement online measures with offline measures, to see how online performance relates to children's knowledge as assessed in post-hoc tests.

The current study contributes to earlier work in at least three ways. First, it adds to the growing body of research on bilingual advantages in statistical learning, as it looked into the process of statistical learning (i.e., during the learning task) rather than at learning outcomes in a post-hoc test. Second, our study is the first to test the proposal that bilinguals' improved statistical learning is due to improved verbal short-term memory skill, a proposal that was not supported by the present data from monolingual Dutch and bilingual Dutch + other language preschoolers. Finally, the findings indicate that a bilingual advantage can be found in online statistical learning, but that it is contingent on the type of response measures looked at, as, in our study, it emerged for reaction time-based measures, but not for accuracy-based measures.

Funding

Research funded by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (275-89-01) to Josje Verhagen.

Acknowledgements

We would like to thank Willemijn Doedens for recruiting and testing participants and for valuable comments on the procedure of the experiment, Iris Welvaarts and Milou Damen for recruiting and testing participants, Chris van Run for programming the experiment, Stijn Verhagen for creating the visual stimuli, Sharon Unsworth for her contribution to subject recruitment, and Annemarie Kerkhoff for valuable input on the procedure of the experiment. Furthermore, we are grateful to the Babylab Utrecht for enabling us to recruit children through their database, and to all schools, parents and children who participated in our study. This research was partly financed with a VENI grant (number 275–89–01) from the Netherlands Organization for Scientific Research awarded to the first author.

References

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language and Communication Disorders*, *51*, 715–731. https://doi.org/10.1111/1460-6984.12242

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*, 158–175. https://doi.org/10.3758/s13428-011-0123-7

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology*, *2*, 1–10. https://doi.org/10.3389/fpsyg.2011.00324

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effect models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Biedroń, A., & Szczepaniak, A. (2012). Working memory and short-term memory abilities in accomplished multilinguals. *The Modern Language Journal*, *96*, 290–306. https://doi.org/10.1111/j.1540-4781.2012.01332.x

Blom, E., Küntay, A. C., 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. https://doi.org/10.1016/j.jecp.2014.06.007

Boerma, T., Chiat, S., Leseman, P. P. M., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal 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

Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science*, *14*, 256–269. https://doi.org/10.1111/j.1467-7687.2010.00974.x

de Bree, E. H., Verhagen, J., Kerkhoff, A. O., Doedens, W. J., & Unsworth, S. (2017). Language learning from inconsistent input: Bilingual and monolingual toddlers compared. *Infant and Child Development*, *26*, e1996. https://doi.org/10.1002/icd.1996

Bulgarelli, F., Lebkuecher, A. L. & Weiss, D. J. (2018). Statistical learning in bilingualism. *Language, Speech, and Hearing Services in Schools*, *49*, 740–753. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0139

Chambers, J. M., Freeny, A. R., & Heiberger, R. M. (1992). *Analysis of variance; designed experiments*. In Chambers, J. M., & Hastie, T. J. (Eds.), *Statistical Models in S*. Wadsworth & Brooks/Cole.

le Clercq, C. M. P., van der Schroeff, M. P., Rispens, J. E., Ruytjens, L., Goedegebure, A., van Ingen, G., & Franken, M-C. (2017). Shortened nonword repetition task (NWR-S): A simple, quick and less expensive outcome to identify children with combined specific language and reading impairment. *Journal of Speech, Language, and Hearing Research*, *60*, 2241–2248. https://doi.org/10.1044/2017_JSLHR-L-16-0060

Delcenserie, A., & Genesee, F. (2017). The effects of age of acquisition on verbal memory in bilinguals. *International Journal of Bilingualism*, *21*, 600–616. https://doi.org/10.1177/1367006916639158

Duncan, T.S., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L1 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

Dutilh, G., Vandekerckhove, J., Forstmann, B.U., Keuleerst, E., Brysbaert, M., & Wagenmakers, E-J. (2012). Testing theories of post-error slowing. *Attention, Perception and Psychophysics*, *74*, 454–465. https://doi.org/10.3758/s13414-011-0243-2

Erickson, L.C., & Thiessen, E.D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid serial auditory presentation: A new measure of statistical learning in speech segmentation. *Experimental Psychology*, *62*, 346–351. https://doi.org/10.1027/1618-3169/a000295

Fernandes, M.A., Craik, F.I.M., Bialystok, E., & Kreuger, S. (2007). Effects of bilingualism, aging, and semantic relatedness on memory under divided attention. *Canadian Journal of Experimental Psychology*, *61*, 128–141. https://doi.org/10.1037/cjep2007014

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

Grama, I.C., Kerkhoff, A.O., & Wijnen, F.N.K. (2016). Gleaning structure from sound: The role of prosodic contrast in learning non-adjacent dependencies. *Journal of Psycholinguistic Research*, *45*, 1427–1449. https://doi.org/10.1007/s10936-016-9412-8

Hautus, M.J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods*, *27*, 46–51.

Hirosh, Z., & Degani, T. (2017). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin Review*, *25*, 892–916. https://doi.org/10.3758/s13423-017-1315-7

Hunt, R.H., & Aslin, R.N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology*, *130*, 658–680. https://doi.org/10.1037/0096-3445.130.4.658

Jaeger, T.F. (2010). Post to HLP/Jaeger lab blog. https://hlplab.wordpress.com/2009/05/14/random-effect-structure/, accessed September 2019.

Kapa, L.L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, *76*, 237–252. https://doi.org/10.1016/j.jml.2014.07.004

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

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, *42*, 643–650. https://doi.org/10.3758/BRM.42.3.643

Kovács, A. & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*, 611–612. https://doi.org/10.1126/science.1173947

Kuo, L.J., & Anderson, R.C. (2010). Beyond cross-language transfer: Reconceptualizing the impact of early bilingualism on phonological awareness. *Scientific Studies of Reading*, *14*, 365–385. https://doi.org/10.1080/10888431003623470

Kuo, L. J., & Anderson, R. C. (2012). Effects of early bilingualism on learning phonological regularities in a new language. *Journal of Experimental Child Psychology*, *111*, 455–467. https://doi.org/10.1016/j.jecp.2011.08.013

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lammertink, I., van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2019). Auditory statistical learning in children: Novel insights from an online measure. *Applied Psycholinguistics*, *40*, 279–302. https://doi.org/10.1017/S0142716418000577

Lum, J. A. G., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedure learning deficits in specific language impairment (SLI): a meta-analysis of serial reaction time task performance. *Cortex*, *51*, 1–10. https://doi.org/10.1016/j.cortex.2013.10.011

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd edition). Mahwah, New Jersey: Lawrence Erlbaum Associates.

McKean, C., Letts, C., & Howard, D. (2013). Developmental change is key to understanding primary language impairment: The case of phonotactic probability and nonword repetition. *Journal of Speech, Language, and Hearing Research*, *56*, 1579–1594. https://doi.org/10.1044/1092-4388(2013/12-0066)

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. https://doi.org/10.1016/j.jecp.2009.12.006

Misyak, J. B., & Christiansen, M. H. (2012). Statistical language and language: An individual differences study. *Language Learning*, *62*, 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, *19*, 1e32. https://doi.org/10.1016/0010-0285(87)90002-8

Onnis, L., Chun, W. E., & Lou-Magnuson, M. (2017). Improved statistical learning abilities in adult bilinguals. *Bilingualism: Language and Cognition*, *21*, 427–433. https://doi.org/10.1017/S1366728917000529

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

Park, J. S., Miller, C., Rosenbaum, D. A., Sanjeevan, T., van Hell, J. G., Weiss, D. J., & Mainela-Arnold, E. (2018). Bilingualism and procedural learning in typically developing children and children with language impairment. *Journal of Speech, Language, and Hearing Research*, *61*, 634–644. https://doi.org/10.1044/2017_JSLHR-L-16-0409

Perruchet, P. & Tillmann, B. (2010). Exploiting multiple sources of information in learning an artificial language: Human data and modeling. *Cognitive Science*, *34*, 255–285. https://doi.org/10.1111/j.1551-6709.2009.01074.x

Poepsel, T. J., & Weiss, D. (2016). The influence of bilingualism on statistical word learning. *Cognition*, *152*, 9–19. https://doi.org/10.1016/j.cognition.2016.03.001

Powell, M.J.D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives (Report)*. Department of Applied Mathematics and Theoretical Physics, Cambridge University. http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf, accessed September 2019.
R Core Team (2015). R: A language environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.
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)
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection measures. *Behavior Research Methods, Instruments and Computers: A Journal of the Psychonomic Society*, 31, 137–149. https://doi.org/10.3758/BF03207704
Thiessen, E.D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical transactions of the Royal Society B*, *372*, 1711. https://doi.org/10.1098/rstb.2016.0056
Vicari, S., Marotti, L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologica*, *41*, 108–114. https://doi.org/10.1016/S0028-3932(02)00082-9
Wang, T., & Saffran, J.R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, 5, 953. https://doi.org/10.3389/fpsyg.2014.00953
Weiss, D., Gerfen, C., & Mitchel, A. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development*, 5, 30–49. https://doi.org/10.1080/15475440802340101
Yim, D., & Rudoy, J. (2013). Implicit statistical learning and language skills in bilingual children. *Journal of Speech, Language, and Hearing Research*, 56, 310–323. https://doi.org/10.1044/1092-4388(2012/11-0243)

Appendix. Results of linear Mixed-Effect Regression Models

Table A1. Results of linear mixed-effect models testing with group (bilingual vs. monolingual children) and block as fixed effects

	Fixed factors	β	**S.E.**	z/t^a	p
Hit responses	Intercept	2.493	0.134	18.611	<.001
	Group	−0.344	0.258	−1.333	.182
	Block: 1 vs. 2	0.532	0.231	2.300	.021
	Block: 1, 2 vs. 3	−0.077	0.176	−0.439	.661
	Block: 3 vs. 4	−0.011	0.246	−0.044	.964
	Group*Block 1 vs. 2	−0.517	0.383	−1.350	.177
	Group*Block 1, 2 vs. 3	0.145	0.289	0.502	.616
	Group*Block 3 vs. 4	−0.003	0.426	−0.008	.994

	Fixed factors	β	S.E.	z/t[a]	p
d'	Intercept	1.930	0.097	19.954	<.001
	Group	0.206	0.193	1.067	.288
	Block: 1 vs. 2	0.297	0.099	2.993	.003
	Block: 1, 2 vs. 3	0.160	0.078	2.061	.040
	Block: 3 vs. 4	−0.057	0.103	−0.546	.585
	Group*Block 1 vs. 2	0.016	0.198	0.080	.936
	Group*Block 1,2 vs. 3	0.068	0.155	0.441	.660
	Group*Block 3 vs. 4	−0.246	0.207	−1.186	.237
RT	Intercept	335.88	351.04	0.957	.613
	Group	39.73	6.93	5.733	<.001
	Accuracy on previous trial	−271.71	656.97	−0.414	.971
	Block: 1 vs. 2	−26.43	28.01	−0.944	.516
	Block: 1, 2 vs. 3	−47.04	20.29	−2.318	.256
	Block: 3 vs. 4	−34.50	20.27	−2.146	.255
	Group*Block 1 vs. 2	−23.57	22.90	−1.029	.303
	Group*Block 1,2 vs. 3	49.13	17.81	2.759	.006
	Group*Block 3 vs. 4	34.83	23.89	1.458	.145
Anticipations[b]	Intercept	−3.498	0.083	−42.26	<.001
	Group	−0.067	0.155	−0.43	.669
	Accuracy on previous trial	−0.092	0.162	−0.57	.570
	Block: 1 vs. 2	0.015	0.267	0.06	.955
	Block: 1, 2 vs. 3	0.337	0.186	1.81	.070
	Block: 3 vs. 4	−0.024	0.268	−0.09	.928
	Group*Block 1 vs. 2	0.112	0.534	0.21	.834
	Group*Block 1,2 vs. 3	−0.678	0.372	−1.82	.069
	Group*Block 3 vs. 4	0.137	0.536	0.26	.798

Notes. [a] Z-values are presented for models with hits and anticipations, as binary dependent variables; *t*-values are presented for model with *d'* and reaction times, as continuous dependent variables. [b] In the model with anticipations, by-item random slopes were added for block but not for group, yielding the maximal random effect structure supported by the data, in line with recommendations made by Jaeger (2010).

Table A2. Results of linear mixed-effect models testing with nonword repetition (sum scores) as an additional fixed-effect factor

	Fixed factors	β	S.E.	z/t[a]	p
Hit responses	Intercept	2.476	0.133	18.569	<.001
	Group	−0.392	0.258	−1.519	.129
	Nonword repetition	0.028	0.038	0.729	.466
	Block: 1 vs. 2	0.545	0.228	2.389	.017
	Block: 1, 2 vs. 3	−0.075	0.173	−0.434	.665
	Block: 3 vs. 4	−0.023	0.242	−0.094	.925
	Group*Block 1 vs. 2	−0.567	0.377	−1.505	.132
	Group*Blocks 1,2 vs. 3	0.231	0.280	0.824	.410
	Group*Block 3 vs. 4	0.099	0.418	0.238	.812
	NWR*Block 1 vs. 2	0.097	0.057	1.710	.087
	NWR*Blocks 1,2 vs. 3	0.103	0.041	2.503	.012
	NWR*Block 3 vs. 4	−0.155	0.061	−2.527	.012
d'	Intercept	1.923	0.092	20.981	<.001
	Group	0.124	0.184	0.671	.504
	Nonword repetition	0.092	0.027	3.426	<.001
	Block: 1 vs. 2	0.295	0.099	2.975	.003
	Block: 1, 2 vs. 3	0.166	0.078	2.142	.033
	Block: 3 vs. 4	−0.062	0.104	−0.602	.548
	Group*Block 1 vs. 2	−0.015	0.199	−0.075	.940
	Group*Block 1,2 vs. 3	0.097	0.156	0.620	.536
	Group*Block 3 vs. 4	−0.213	0.208	−1.022	.307
	NWR*Block 1 vs. 2	0.036	0.029	1.241	.215
	NWR*Blocks 1,2 vs. 3	−0.020	0.023	−0.896	.371
	NWR*Block 3 vs. 4	−0.058	0.030	−1.916	.056
RT	Intercept	272.153	41.155	6.613	<.001
	Group	46.365	34.629	1.339	.184
	Nonword repetition	10.248	5.048	2.030	.045
	Accuracy previous trial	−12.420	5.806	−2.139	.032
	Block: 1 vs. 2	−13.141	22.265	−0.590	.555
	Block: 1, 2 vs. 3	−74.527	17.052	−4.371	<.001
	Block: 3 vs. 4	−78.839	23.195	−3.399	.001
	Group*Block 1 vs. 2	24.574	19.210	1.279	.201
	Group*Block 1,2 vs. 3	34.020	14.319	2.376	.017
	Group*Block 3 vs. 4	24.574	19.210	1.279	.201
	NWR*Block 1 vs. 2	−3.178	2.275	−1.166	.244
	NWR*Blocks 1,2 vs. 3	2.645	2.106	1.256	.209
	NWR*Block 3 vs. 4	5.015	2.828	1.773	.076

	Fixed factors	β	S.E.	z/t[a]	p
Anticipations	Intercept	−6.096	0.455	−13.391	<.001
	Group	−0.677	0.608	−1.114	.265
	Nonword repetition	−0.286	0.097	−2.950	.003
	Accuracy previous trial	−0.134	0.203	−0.660	.509
	Block: 1 vs. 2	−1.083	1.212	−0.894	.371
	Block: 1, 2 vs. 3	0.669	0.727	0.920	.358
	Block: 3 vs. 4	−1.425	1.258	−1.133	.257
	Group*Block 1 vs. 2	0.892	1.075	0.830	.407
	Group*Block 1,2 vs. 3	−0.724	0.650	−1.113	.266
	Group*Block 3 vs. 4	−0.899	1.024	−0.878	.380
	NWR*Block 1 vs. 2	0.070	0.187	0.373	.709
	NWR*Blocks 1,2 vs. 3	0.080	0.106	0.757	.449
	NWR*Block 3 vs. 4	−0.466	0.201	−2.319	.020

Note. [a] *Z*-values are presented for models with the binary variables hits and anticipations; *t*-values are presented for model with the continuous variables *d'* and reaction times.

Address for correspondence

Josje Verhagen
University of Amsterdam
Postbus 1637
1000 BP Amsterdam
The Netherlands
J.Verhagen@uva.nl

Co-author information

Elise de Bree
Research Institute of Child Development and Education
University of Amsterdam
e.h.debree@uva.nl

Publication history

Date received: 14 November 2018
Date accepted: 4 November 2019
Published online: 11 February 2020