



# To mix or not to mix? The relation between parental language mixing and bilingual children’s language outcomes

Emma Verhoeven , Merel van Witteloostuijn , Ora Oudgenoeg-Paz and Elma Blom

Department of Education and Pedagogy, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, The Netherlands

## Research Article

**Cite this article:** Verhoeven, E., van Witteloostuijn, M., Oudgenoeg-Paz, O. and Blom, E. (2025). To mix or not to mix? The relation between parental language mixing and bilingual children’s language outcomes. *Bilingualism: Language and Cognition* 1–15. <https://doi.org/10.1017/S1366728925100175>

Received: 8 November 2024  
Revised: 29 April 2025  
Accepted: 13 May 2025

**Keywords:**  
bilingualism; language mixing; language input; language development; parenting

**Corresponding author:**  
Emma Verhoeven;  
Email: [e.g.e.c.verhoeven@uu.nl](mailto:e.g.e.c.verhoeven@uu.nl)

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

### Abstract

Language mixing is a common phenomenon in the language input of bilingual children. However, the relation between the frequency of parental language mixing and children’s language development remains unclear. The present study investigates the relation between language mixing as observed in daylong audio recordings (LENA) and as reported by parents in the questionnaire for Quantifying Bilingual Experience (Q-BEx) and children’s language outcomes in the majority and minority language. Participants were 52 3-to-5-year-old Polish-Dutch and Turkish-Dutch children in the Netherlands and Bayesian informative hypothesis evaluations were applied. In 14 out of 15 regression analyses, the LENA and Q-BEx measures yielded similar associations with children’s language outcomes. Parental language mixing was not related to majority language outcomes, but a negative relation was found with expressive vocabulary in the minority language. Longitudinal studies are needed to pinpoint the directionality of this negative relation.

### Highlights

- Polish-Dutch and Turkish-Dutch parents do not vary significantly in mixing behavior.
- Parental language mixing is not related to any majority language outcome.
- Parental mixing negatively relates to expressive minority language vocabulary.
- Method to measure mixing does not moderate the relation with language outcomes.

## 1. Introduction

Language mixing, also called code-switching, is the use of more than one language within the same conversation (Lam & Matthews, 2020; Poeste et al., 2019; Poplack, 2001; Yow et al., 2018). Language mixing is common among bilingual speakers, with large individual differences between speakers and communities (Hoff & Core, 2015; Myers-Scotton, 2017; Yow et al., 2018). Different forms of language mixing occur regularly in bilingual children’s language input (Kremin et al., 2022). Bilingual families are sometimes advised to avoid language mixing and to use the “One-Parent-One-Language” (OPOL) approach (Blom et al., 2018; De Houwer, 2007). This advice comes from the belief that children would acquire their languages better when the languages are separated in their input (Aronsson, 2020) and may stem from fears that mixing confuses children, even though empirical support for this idea is scarce. To date, merely five studies have investigated the relation between the degree of parental language mixing and children’s language outcomes, and their results are equivocal (Bail et al., 2015; Byers-Heinlein, 2013; Carbajal & Peperkamp, 2020; Place & Hoff, 2011, 2016). Their various findings will be discussed throughout this introduction. The present study aims to provide further insight by investigating the relation between language mixing by caregivers and children’s language outcomes. It focuses on an underinvestigated age range (3–5 years) and includes data from two different bilingual groups, Polish-Dutch and Turkish-Dutch. Its findings are of both theoretical and practical relevance. They will provide insights into the relation between specific language environment factors and children’s language development, and they will contribute to improved advice provided to bilingual families.

### 1.1. Parental language mixing in relation to children’s language outcomes

Investigating the quantity and characteristics of child-directed language mixing is crucial to understanding how it might affect language acquisition (Kremin et al., 2022). As previously mentioned, only a few studies have examined the relation between parental language mixing and

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



bilingual children's language skills. Two studies found a negative relation such that more parental language mixing was related to smaller vocabularies in children (Byers-Heinlein, 2013; Carbajal & Peperkamp, 2020). Two other studies did not find a relation between the frequency of mixed language input children heard and their vocabulary or grammar outcomes (Place & Hoff, 2011, 2016). One final study reported a positive relation where more mixing was associated with children having larger vocabularies (Bail *et al.*, 2015). Overall, it remains unknown whether the relation between parental language mixing and children's language outcomes may be negative, positive or non-existent, and whether it differs for various language aspects (e.g., receptive and expressive vocabulary outcomes or grammar). Moreover, hypothetically, all three empirical outcomes could theoretically be expected, as we will further elaborate on in this section.

### 1.1.1. More mixing relates to lower child language outcomes

The first possibility is that more parental language mixing is related to lower language outcomes in children. According to the theoretical framework of the Bilingual Interactive Activation Model (Dijkstra & Van Heuven, 2002; Grainger & Dijkstra, 1992; Grainger *et al.*, 2010), both languages of bilinguals are simultaneously active and accessible. When children process language input, they implicitly expect an upcoming word to be in the same language as the previous word. Thus, when language mixing occurs, this is counter to their expectations. They need to inhibit the previously activated language and retrieve knowledge from the other language, resulting in a processing cost that may temporarily hinder language comprehension (Byers-Heinlein *et al.*, 2022). Ruan *et al.* (2023) note, moreover, that children's comprehension could be diminished due to language mixing disrupting the statistical regularities in the speech stream. Children use mechanisms such as statistical learning to implicitly detect patterns in their language input (Erickson & Thiesen, 2015; Romberg & Saffran, 2010). Parental language mixing may cause disruptions in these regularities, which could temporarily complicate the processing and comprehension of mixed language utterances (Place & Hoff, 2016; Potter *et al.*, 2019). It is not expected that a limited amount of mixed language will affect children's language outcomes, but frequent mixing might have a detrimental effect on children's language outcomes (Ruan *et al.*, 2023).

Two studies found a negative relation between parental language mixing and children's language skills. Byers-Heinlein (2013) investigated 129 18-month-old and 39 24-month-old bilingual infants with various language backgrounds and found a negative relation between parents' language mixing score and children's receptive vocabulary score in English in the 18-month-olds. No negative relation was found for the 24-month-old children. The frequency of parental language mixing was measured with the Language Mixing Scale (LMS; Byers-Heinlein, 2013), which comprises five statements. Four out of the five statements refer to language mixing within sentences (i.e., intra-sentential), and only one statement refers to general language mixing. Therefore, the LMS could be seen as a measure of intra-sentential language mixing instead of overall language mixing (see also Place & Hoff, 2016).

The second study by Carbajal and Peperkamp (2020) examined the dual language exposure of 58 11-month-old bilingual infants in Paris who were exposed to a variety of languages besides the majority language French. The study did not assess the frequency of language mixing; rather, it focused on the inverted form, namely language purity. More language mixing corresponded to a lower language purity. Carbajal and Peperkamp reported a positive relation between receptive vocabulary in French and within-speaker

language purity (i.e., the less a speaker mixed, the higher the children's receptive vocabulary scores in French). No relation was found for within-block language purity (i.e., the absence of language mixing within a 30-minute block). It should be noted that this method lacks fine-grained information regarding the types of mixing that occur within a block or speaker. The occurrence of multiple languages within a 30-minute time frame does not indicate how these languages are being mixed. We assume that intra-sentential language mixing is better reflected by within-speaker language purity than within-block language purity, as language mixing within a sentence implicitly takes place within a speaker. Taken together, the results suggest that hearing multiple languages within a 30-minute time frame is not negatively related to children's early receptive vocabulary outcomes but hearing a speaker mix their languages may be.

Summarizing, two studies found negative relations between parental language mixing and children's receptive vocabulary, which is in line with the hypothesis that children might have more difficulties comprehending mixed language input. The variables that are negatively related to vocabulary outcomes more closely represent intra-sentential language mixing. This may support the hypothesis that intra-sentential language mixing induces more processing costs and may hinder language comprehension, whereas inter-sentential language mixing may not (Byers-Heinlein *et al.*, 2017; Gullifer *et al.*, 2013; Kremin *et al.*, 2022).

### 1.1.2. More mixing relates to higher child language outcomes

There are also reasons why more language mixing could relate to higher language outcomes in bilingual children. First, language mixing might support the acquisition of new vocabulary by increasing the contextual variability (i.e., the range of unique contexts in which words appear). Some studies have shown that more variability in the linguistic context can enhance learning (Denby *et al.*, 2018; Gómez, 2002). Language mixing increases this variability by creating infrequent contexts that may draw the child's attention and highlight novel words, making it easier for children to learn novel words in a mixed context (Kaushanskaya *et al.*, 2023).

Second, parents indicate that they mainly mix their languages to bolster their children's comprehension or to teach new words (Kremin *et al.*, 2022). Additionally, considering the often unbalanced language levels of bilingual children (Baker & Jones, 1998; Yip & Matthews, 2006), learning new vocabulary in their weaker language might be aided by presenting these words within sentences of their stronger language (Ruan *et al.*, 2023; but see Potter *et al.*, 2019). These instances of language mixing could facilitate language learning and could have a positive effect on children's language outcomes.

Third, parents may alter their language use based on children's increasing language abilities (Soderstrom, 2007). If children's language abilities are more developed, parents tend to form longer sentences, thereby creating more possibilities for switching within sentences (Bail *et al.*, 2015). Additionally, parents may perceive an increase in their child's language abilities to be indicative of an enhanced capacity to process more complex linguistic input. Indeed, it has been found that parents mixed more when speaking to their infant at 18 months of age than at 10 months of age (Kremin *et al.*, 2022). Importantly, the direction of the relation remains unknown. It may be that language mixing enhances language abilities, or that enhanced language abilities cause parents to mix more.

A positive relation between intra-sentential language mixing and productive vocabulary outcomes was found in one study that

involved toddlers (Bail et al., 2015). In the study, 24 18- to 24-month-old Spanish-English bilingual children were observed during a 13-minute play session with their parents in a laboratory setting. Parents were given some toys and were instructed to play with their children as they usually would. Language mixing was measured as the number of observed intra- and inter-sentential mixes during the play session. For vocabulary outcomes, the authors created total and conceptual vocabulary scores, comprising both English and Spanish expressive vocabulary scores. The frequency of intra-sentential language mixing was positively correlated to both types of vocabulary outcomes, while no relation was found for the number of inter-sentential switches. The observed positive relation does not indicate whether language mixing may be beneficial for vocabulary development, or whether parents mix more when children show enhanced language skills.

### 1.1.3. Mixing frequency is unrelated to children's language outcomes

The final possibility is that language mixing frequency is not associated with children's language outcomes. Again, there are several potential explanations for null findings. First, a non-existent relation may be explained by the type of mixed input that children generally receive. The most common type of language mixing occurs between sentences (Bail et al., 2015; Byers-Heinlein et al., 2017; Kremin et al., 2022), which is not associated with processing costs (Byers-Heinlein et al., 2017; Gullifer et al., 2013; Kremin et al., 2022). It could also be the case that children do not experience much processing costs when they are used to language mixing in their natural language environment, regardless of type of language mixing (Adamou & Shen, 2019; Blanco-Elorrieta & Pyllkänen, 2017).

Second, language mixing is a natural phenomenon (Hoff & Core, 2015; Yow et al., 2018) and is considered fluid and effortless for highly proficient bilingual speakers (Poplack, 1980, 2001). If language mixing were detrimental to language comprehension, we would not expect it to be such a common behavior among bilingual speakers in naturalistic settings (Backus, 2005; Myers-Scotton, 1993; Zentella, 1998).

Third, although language mixing occurs frequently in bilingual speech, mixed language input makes up a relatively small part of a bilingual child's total language input (Kremin et al., 2022). It could be that, regardless of any possible processing costs, the frequency of mixed language input is too low to impact children's language outcomes in either direction.

Place and Hoff have investigated the relation between parental language mixing and children's expressive vocabulary and grammar outcomes in two studies. Their first study included 29 2-year-old Spanish-English bilingual children whose parents kept detailed language diaries of their children's language exposure over seven-days (Place & Hoff, 2011). Language mixing was measured as the percentage of 30-minute blocks in which the children were exposed to two languages. They found no relation between language mixing and expressive vocabulary or grammar outcomes. The authors noted that the percentage of mixed 30-minute blocks lacked fine-grained information about the actual frequency of language mixing occurring within these time blocks. In a follow-up study (Place & Hoff, 2016), 90 parents kept a language diary on the language exposure of their 30-month-old Spanish-English bilingual children, and 58 of them also filled in the LMS (Byers-Heinlein, 2013). The results from Place and Hoff (2011) were replicated, as the total frequency of mixed blocks was again unrelated to any measure of English or Spanish development. Furthermore, the score from the language mixing scale also did not relate to children's expressive

vocabulary or grammar outcomes (marginally to one out of five Spanish language outcomes).

### 1.2. The present study

The present study examines the relation between parental language mixing and language outcomes in 3- to 5-year-old children. The participants come from two distinct bilingual groups in the Netherlands: Turkish-Dutch and Polish-Dutch families. It is important to investigate language mixing in diverse groups of bilingual speakers to increase our understanding of the factors that underlie the influence of parental language mixing on children's language acquisition, such as the type of mixing or the amount of experience with mixed language input (Adamou & Shen, 2019; Byers-Heinlein et al., 2022). We selected Polish-Dutch and Turkish-Dutch families because both groups are well represented in the Netherlands, where the study is situated. Moreover, the Polish and Turkish immigrants in the Netherlands exhibit notable demographic differences, particularly with regard to their migration history (Central Bureau of Statistics (CBS), n.d.). Therefore, before turning to studying the relation between parental language mixing and children's language outcomes, we first address the question:

RQ1: *How do the language mixing behaviors of Turkish-Dutch parents and Polish-Dutch parents differ in terms of frequency and type of mixing?*

Language mixing is a frequent phenomenon in Turkish families in the Netherlands (Backus, 2012; Backus & Demirçay, 2021; Yagmur, 2009). This is because most Turkish parents in the Netherlands are second- or third-generation immigrants (CBS, n.d.) and have a good command of both languages (Backus & Demirçay, 2021; Yagmur, 2009). In comparison, the Polish group in the Netherlands largely consists of first-generation immigrants (78.3%; CBS, n.d.). Furthermore, the Turkish community in the Netherlands is nearly twice as large as the Polish community, resulting in Turkish children having more (bilingual) speakers in their environment with whom they can speak both the heritage language and the societal language in comparison to Polish children. As further-generation immigrants show more frequent language mixing than first-generation immigrants (Backus & Demirçay, 2021), and Turkish children may have more bilingual interlocutors in their immediate environment, we hypothesize that the Turkish-Dutch group mixes languages more frequently than the Polish-Dutch group. We do not have any hypotheses regarding differences between the two groups in terms of intra- and inter-sentential language mixing.

The other research questions look at the relation between parental language mixing and bilingual children's language outcomes. We investigate both overall language mixing behavior and intra- and inter-sentential language mixing separately (Poplack, 2001), and distinguish between two types of language outcomes: vocabulary skills (research question 2) and sentence repetition scores (research question 3). For research question 2, vocabulary outcomes in the majority language (Dutch) and the minority language (Polish/Turkish) are separated into expressive and receptive vocabulary outcomes. Previous studies have found negative effects on receptive vocabulary (Byers-Heinlein, 2013; Carbajal & Peperkamp, 2020), but not on expressive vocabulary (Bail et al., 2015; Byers-Heinlein, 2013; Place & Hoff, 2011, 2016) and only two studies addressed grammar and found no relation (Place & Hoff, 2011, 2016). In summary, to investigate the relation between parental language mixing and language outcomes, we set up the following two research questions:

- RQ2: *To what extent does parental language mixing relate to children's receptive and expressive vocabulary outcomes?*
- RQ3: *To what extent does parental language mixing relate to children's sentence repetition scores?*

The three possible hypotheses regarding the relation between parental language mixing and children's language outcomes have been substantiated in the introduction. With the use of Bayesian statistics, we evaluate the evidence for a positive, a negative or a non-existent relation (Hoijsink, Gu, *et al.*, 2019).

Finally, we investigate whether the method used to measure language mixing affects the observed relations. The present study measured language mixing using both daylong audio recordings and questionnaires. Using two methods allows us to investigate whether the relation between parental language mixing and children's language outcomes varies as a result of the type of measure used. This leads us to our final research question:

- RQ4: *Does the method used to measure language mixing affect the observed relations between parental language mixing and children's language outcomes?*

Previous studies that used different methods to measure parental language mixing yielded mixed evidence (Bail *et al.*, 2015; Byers-Heinlein, 2013; Carbajal & Peperkamp, 2020; Place & Hoff, 2011, 2016). Whether their disparate findings can be (partly) ascribed to their differences in methodology is currently unknown.

## 2. Materials and methods

The data in this study are part of the larger project "Children and Language Mixing: developmental, psycholinguistic and sociolinguistic aspects" (CALM; <https://osf.io/p9gje/>). The project has been approved by the Ethics Committee of Utrecht University (FETC20–0291). The data collection took place from July 2022 to July 2023 in the Netherlands. At that time, the overarching project had gathered data from Turkish-Dutch ( $n = 33$ ), Polish-Dutch ( $n = 23$ ) and English-Dutch ( $n = 4$ ) bilingual children and their families. For this study, a subsample of only the Turkish and Polish groups was used. The study was preregistered on the Open Science Framework on August 1, 2024 (<https://doi.org/10.17605/OSF.IO/849RZ>). Deviations from the preregistration are made explicit throughout the methods section. Data and scripts are available on the project page for this study on the Open Science Framework (<https://osf.io/qudr7/>).

### 2.1. Participants

The participants were 56 multilingual children, aged between 36 and 72 months ( $M = 55$ ,  $SD = 10$ ), and their families. Data from four children was excluded from the analysis because they either received more than 50% exposure to a third language ( $n = 2$ ) or did not finish the daylong audio recording and second test appointment ( $n = 2$ ). Our final sample consisted of 21 Polish-Dutch (10 girls) and 31 Turkish-Dutch (17 girls) bilingual children. Families were recruited via schools, (local) events, online calls on social media platforms and personal networks. At the time of data collection, none of the children had received a diagnosis of a (suspected) language disorder. All children lived in the Netherlands and heard either the minority language Polish or Turkish in addition to the majority language Dutch.

Table 1 presents the descriptive statistics of the sample per group. Both groups are varied in terms of language exposure and language dominance. Six children were reported to be exposed to a limited amount of English as a third language ( $M = 4\%$  exposure,  $SD = 5\%$ , range: 2–14%). The educational level of parents was measured as the highest level of education attained between parents via the questionnaire for Quantifying Bilingual Experience (Q-BEx; De Cat *et al.*, 2022) and was relatively high in both groups.

## 2.2. Measures

### 2.2.1. Vocabulary

Children's vocabulary was measured in both their languages (Dutch and Polish/Turkish) via the Cross-linguistic Lexical Task (CLT; Haman *et al.*, 2015). The CLT is appropriate for our age range (Haman *et al.*, 2017) and is considered a valid measure of vocabulary (Van Wonderen & Unsworth, 2021). The task consists of four parts, each comprising 32 items: (1) receptive knowledge of nouns; (2) receptive knowledge of verbs; (3) expressive knowledge of nouns and (4) expressive knowledge of verbs. The order of the four parts was counterbalanced.

In the receptive part, children were asked about the target word via a prerecorded phrase (e.g., "Where is the candle?"). The child was prompted to point to one of the four images on the screen that corresponded to the word. In the expressive part, children saw an image and were asked to describe it in a single word (e.g., "What is this?"). Each correct response was awarded one point. Nouns and verbs were collapsed to increase the number of items tapping into receptive and expressive vocabulary, thereby increasing variation within each modality (sum score range: 0–64). The analyses were conducted using the raw sum scores. Accuracy in the CLT is based on a list of target words, but we made a few adaptations to the target list and included more responses as correct. After consultation with at least two native speakers per language (Dutch, Polish and Turkish as spoken in the Netherlands (Doğruöz & Backus, 2010)), we decided to include fourteen Dutch, four Polish and five Turkish additional synonyms in the expressive vocabulary test (e.g., for Dutch, we included both *strijkijzer* and *strijkbout* as correct responses to the picture of an iron, as these were considered synonyms, but *strijkbout* is more old-fashioned).

### 2.2.2. Sentence repetition

The standardized sentence repetition test from the Clinical Evaluation of Language Fundamentals Preschool-II-Dutch (CELF-P-2-NL; Wiig *et al.*, 2012) was used. The test is appropriate for children between the ages of 3 and 7 years. The sentence repetition measure was administered in Dutch only, because, currently, there is no equivalent for Polish or Turkish. The CELF sentence repetition comprised thirteen prerecorded sentences of increasing length and complexity, preceded by three practice sentences. The experimenter demonstrated the task by repeating the practice sentences together with the child. Afterward, the child repeated the thirteen test sentences on their own. The experimenter scored how many errors the child made. Errors consisted of repeating, omitting, replacing or adding words or saying words in the wrong order. The task was recorded so the experimenter could listen to the child's productions again and score accurately. Children's productions received a score of 3 if no errors were made, 2 if one error was made, 1 if two to three errors were made and 0 if they made four or more errors. The task was discontinued after three consecutive 0-scores. The analyses were conducted using the raw scores.

**Table 1.** Descriptive statistics of the sample per group included in the analysis

Group	Polish-Dutch				Turkish-Dutch				
	<i>n</i>	<i>M</i>	<i>SD</i>	Range	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SD</i>	Range
Age (months)	21	53	11	36–72	31	57	9		38–72
Exposure									
Dutch	21	.51	.19	.22–.85	31	.53	.23		.14–.96
Polish/Turkish	21	.48	.19	.15–.78	31	.47	.23		.04–.86
English	2	.04	.07	.03–.14	4	.04	.05		.02–.13
Language Dominance									
> 60% Dutch	11				13				
40–60% Dutch	3				8				
< 40% Dutch	7				10				
Parents' educational level									
Secondary school	0				1				
Post-secondary	3				4				
University degree	18				26				
Migration generation									
First generation	1				4				
Second generation	18				16				
Third generation	0				6				
NA	2				5				

Note. Language exposure is expressed as a proportion. Language dominance was determined via the current overall language exposure from the Q-BEx questionnaire, with >60% exposure being considered the threshold for dominance (Cattani et al., 2014; Siow et al., 2023).

### 2.2.3. Language mixing – Naturalistic daylong audio recordings (LENA)

The present study employed two methods to assess parental language mixing. The first method involves naturalistic, daylong audio recordings in the home environment, gathered with the Language Environment Analysis (LENA; Greenwood et al., 2011) recording device. Parents turned the device on and put it in the pocket of a T-shirt the child was wearing during a weekend day. No researchers were present during the recording day, and parents were instructed to go about their day as usual. From the full daylong audio recording, 270 30-second segments were sampled based on the conversational turn count (CTC) provided by the automatic LENA output. The 270 segments were manually coded for speaker(s), language(s) spoken, activity and (target) child-directed speech. The segments that contained more than one language were transcribed in CHAT (MacWhinney, 2014) by bilingual research assistants. All instances of intra- and inter-sentential language mixing uttered by a parent were used to create the language mixing variables. More details can be found in the Appendix and Supplementary Materials. The language mixing variables represent the frequency of intra- and inter-sentential language mixing per hour produced by parents to the target child. An overall language mixing score was calculated by summing intra- and inter-sentential language mixing.

### 2.2.4. Language mixing – Questionnaire (Q-BEx)

The second measure of language mixing was the questionnaire for Quantifying Bilingual Experience (Q-BEx; De Cat et al., 2022). The Q-BEx is a modular questionnaire with two fixed modules, namely “background information” and “risk factors”, and several optional modules, one of which being “language mixing”. This module asks

parents about different types of language mixing. The frequency with which parents switch is questioned via three questions, formulated as “At home, when people (including yourself) speak to the child, how often do you do any of the following...” The questions ask about one-word switches, two- or three-word switches or inter-sentential switches. Parents answer on an ordinal 5-point scale ranging from *almost never* to *more than 5 conversations per day* with the option to select *I do not know* or *not relevant*. We transformed these responses into an interval scaled variable by converting them into number of mixes per week (e.g., *one or two conversations per week* becomes 1.5, *one or two conversations per day* becomes 10.5 and we set *more than five conversations per day* to 42). Both one-word switches and two- or three-word switches take place within the sentence and were summed to create one score for intra-sentential language mixing. In conclusion, the Q-BEx provided separate scores for the number of intra- and inter-sentential switches that parents report to utter per week. Overall language mixing was calculated as the sum of the intra- and inter-sentential switches per week.

### 2.3. Procedure

The research took place in the home environment of the participant and consisted of two home visits. Parents provided informed consent during their first test appointment. The first visit was a bilingual session with a bilingual research assistant (either Turkish-Dutch or Polish-Dutch) to assess children’s vocabulary skills in the minority language. During the first appointment, parents also received the LENA recorder with instructions to use it on a day on the weekend before the second test appointment. To safeguard the privacy of the participants, only fragments of 30 seconds were listened

to. Consequently, the researcher was unable to understand the full context of the conversations. It was made clear to the participants that the focus of the study was on the languages being spoken, rather than on the content of the conversations. Furthermore, parents were allowed to have parts of the audio removed before it would ever be listened to. One family had two hours of audio deleted.

The average time between two test appointments was approximately three weeks (range: 2–9 weeks). During the second visit, a Dutch speaker administered the Dutch vocabulary and sentence repetition tasks, retrieved the LENA recorder and filled out the Q-BEx questionnaire together with a parent. The data from this study are part of a larger project and additional tests were administered during home visits, but these are not discussed in this paper.

#### 2.4. Data analysis

Analyses and hypotheses were preregistered on the Open Science Framework. All analyses have been carried out in *R* (version 4.3.0; R Core Team, 2024). To address our research questions, Bayesian informative hypothesis evaluations were used. Where null-hypothesis significance testing (NHST) can only investigate one specific hypothesis against its null hypothesis, Bayesian informative hypothesis testing allows for the simultaneous comparison of multiple hypotheses with each other. In other words, NHST would require multiple analyses to compare the negative hypothesis to the null hypothesis and the positive hypothesis to the null hypothesis, which increases the type I error. The Bayesian analysis calculates the support for each hypothesis that is included in the set based on its relative fit and relative complexity. The hypothesis that describes the data best, without being unnecessarily complex, receives the most support. As we wanted to evaluate multiple hypotheses at once (i.e., a negative, positive or non-existent relation), Bayesian analyses were conducted.

We used two values to evaluate the hypotheses, the posterior model probabilities (PMP) and Bayes Factor (BF). We did not set predetermined thresholds of these values for accepting or rejecting a hypothesis, as this defies the purpose of Bayesian hypothesis testing (Hojtink, Mulder, et al., 2019). Instead, we looked at the  $PMP_a$  to quantify the evidence for each hypothesis compared to the other pre-specified hypotheses. The PMP can be used to compute Bayesian error probabilities. When  $H_1$  is selected as the preferred hypothesis,  $1 - PMP_{1a}$  denotes the probability of choosing the wrong hypothesis (Hojtink, Mulder, et al., 2019). After selecting the hypothesis with the highest  $PMP_a$  as the best of the set, we looked at the Bayes Factor ( $BF_C$ ) to evaluate how good each hypothesis was on its own, without comparing it to the other hypotheses in the set. The Bayes Factor  $BF_C$  quantifies how much more likely the data are to be observed under that hypothesis than under its complement (i.e., any hypothesis other than the hypothesis under consideration). The values can be interpreted as follows: a higher  $PMP_a$  indicates more certainty that the hypothesis is the right one from the set and a higher BF implies stronger evidence for that hypothesis compared to its complement.

*RQ1: How do the language mixing behaviors of Turkish-Dutch parents and Polish-Dutch parents differ in terms of frequency and type of mixing?*

First, the difference in the frequency of language mixing between the two groups was tested with Bayesian independent sample *t*-tests using the *bain* package (Gu, 2016; Hoijtink, Gu, et al., 2019). The two informative hypotheses were that the Turkish group mixed more than the Polish group, or that there was no difference in overall language mixing behavior between the two groups. We

compared the two groups on the frequency of overall language mixing as observed in the daylong audio recording and as reported in the Q-BEx questionnaire. In our preregistration, we stated that we would not investigate the difference between groups on overall language mixing behavior reported in the Q-BEx questionnaire, as these might reflect parents' attitudes toward mixing rather than actual frequencies (Treffers-Daller et al., 2020). However, considering all other analyses were conducted for both methods, we deemed it appropriate to use the Q-BEx data for this question as well.

Second, the difference in the types of language mixing used by the two groups was tested by comparing the ratio of inter- versus intra-sentential switches between the two groups via a Bayesian independent sample *t*-test. A ratio score was calculated by dividing the number of inter-sentential switches by the number of intra-sentential switches for each family. The comparison of the ratios was exploratory since no hypothesis was formulated.

*RQ2: To what extent does parental language mixing relate to children's vocabulary outcomes?*

To test the influence of language mixing on vocabulary outcomes, we ran eight multivariate regressions based on a  $2 \times 2 \times 2$  matrix. The first factor in the matrix was the vocabulary modality. Separate multivariate regression models were created for expressive and receptive vocabulary scores. The second factor was the method used to measure language mixing. Individual models were created for the different methods of how language mixing was measured (questionnaire and daylong audio recordings). The third factor was the type of language mixing. One set of models included the overall frequency of language mixing whereas the other models investigated the specific types of language mixing (intra- and inter-sentential). We did not have different hypotheses for the individual languages. Therefore, each model has two dependent variables: vocabulary scores in the majority language (Dutch) and in the minority language (Polish/Turkish). Summarizing, we ran eight multivariate regression models to investigate the relation between parental language mixing and children's receptive and expressive vocabulary outcomes. All regression models were controlled for age, Dutch language exposure, parental education, and group (Polish or Turkish). The models can be found in the preregistration.

*RQ3: To what extent does parental language mixing relate to children's sentence repetition scores?*

To test the relation between overall language mixing and Dutch sentence repetition scores, we ran multiple linear regression models containing the same set of control variables as used in the models for vocabulary outcomes. The regression models were based on a  $2 \times 2$  matrix. The two factors were the method of measuring language mixing (questionnaire and daylong audio recordings) and the type of language mixing (overall language mixing and types of language mixing). Summarizing, four multiple linear regression models were run to investigate the relation between parental language mixing and children's sentence repetition scores.

### 3. Results

#### 3.1. *RQ1: How do the language mixing behaviors of Turkish-Dutch parents and Polish-Dutch parents differ in terms of frequency and type of mixing?*

All descriptive statistics are presented in Table 2. The overall language mixing frequency per group and method is displayed in Figure 1 and the corresponding BFs and PMPs are presented in

**Table 2.** Descriptive statistics of the sample per group included in the analysis

	Polish-Dutch (n = 21)			Turkish-Dutch (n = 31)		
	M	SD	Range	M	SD	Range
<b>Receptive vocabulary</b>						
Dutch	48.0	11.5	16–62	51.9	8.9	25–62
Polish/Turkish	46.9	12.1	16–63	52.3	9.8	23–64
<b>Expressive vocabulary</b>						
Dutch	19.7	11.6	3–38	24.6	10.6	1–44
Polish/Turkish	22.7	17.1	0–53	28.3	17.0	2–55
Sentence repetition	12.5	7.4	0–25	15.8	7.0	3–30
LENA recording (hours)	14.6	3.1	5.4–16	13.8	3.5	4.2–16
<b>Language mixing (LENA)</b>						
Overall	29.5	34.5	0–121	34.5	31.2	0–103
Inter-sentential	26.0	31.6	0–105	27.6	26.4	0–96
Intra-sentential	3.4	3.8	0–16	7.0	6.4	0–20
<b>Language mixing (Q-BEx)</b>						
Overall	39.8	44.1	0–168	66.6	65.9	0–168
Inter-sentential	28.7	29.9	0–84	37.9	34.0	0–84
Intra-sentential	11.1	22.7	0–84	28.8	36.8	0–84

Note. Vocabulary outcomes are number of items correct. Sentence repetition scores are raw scores from the CELF-P-2-NL. The LENA represents the observed instances of language mixing per hour. The Q-BEx represents the number of times that parents report to mix per week.

**Table 3.** The results from the Bayesian independent samples *t*-tests are indecisive regarding differences in the overall frequency of language mixing observed between the Turkish-Dutch group and the Polish-Dutch group. The BF<sub>C</sub>s show that the LENA data suggest that the Turkish-Dutch and Polish-Dutch groups do not differ in terms of overall language mixing and that the Q-BEx data strongly suggest that the Turkish-Dutch group mixes more than the Polish-Dutch group. However, similar to non-significant *p*-values in classical hypothesis testing, the relatively high Bayesian error probabilities prevent us from fully discarding the other hypotheses.

Summarizing, we have no conclusive evidence for a difference in overall frequency of language mixing between the two groups.

Regarding the ratio of inter- and intra-sentential switches (Figure 2) both groups mix more between sentences than within sentences. The independent samples *t*-tests revealed that Turkish-Dutch parents have a lower ratio of inter- versus intra-sentential language mixing than Polish-Dutch parents. Given the large BF<sub>C</sub> values and the low Bayesian error probabilities (5% and 10% respectively), we conclude that, compared to the Polish-Dutch group, the Turkish-Dutch group makes relatively more use of intra-sentential language mixing than the Polish-Dutch group.

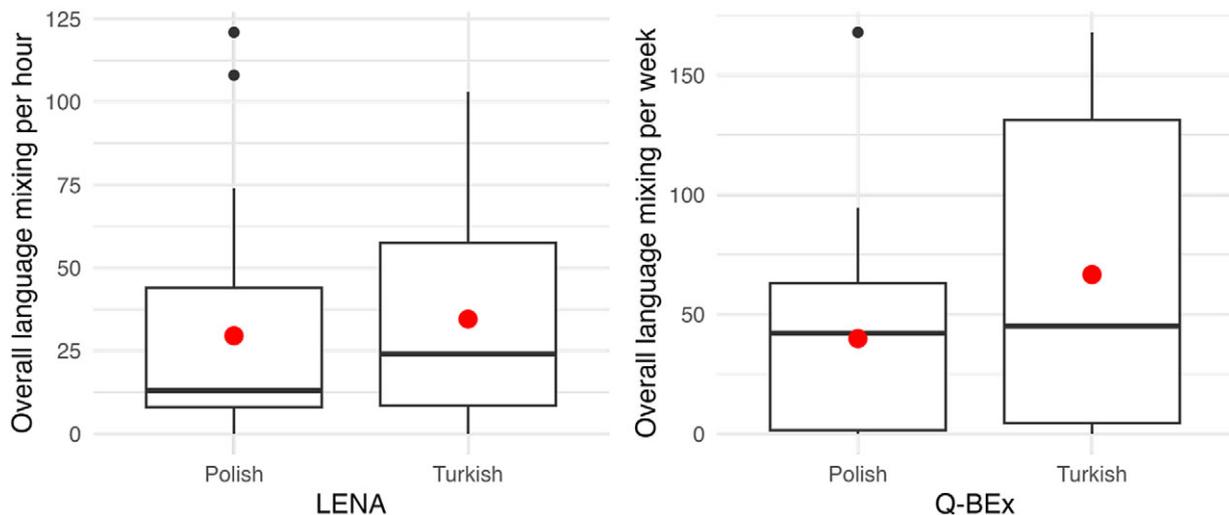
**3.2. RQ2: To what extent does parental language mixing relate to children’s vocabulary outcomes?**

All regression coefficients of language mixing on language outcomes and their 95% confidence intervals can be found in Table 4 and Figure 3. The corresponding BF<sub>C</sub> and PMP<sub>a</sub> for our hypotheses are presented in Table 5. In Figures 3A–F, a general trend can be observed that language mixing seems to have little to no relation with the majority language (Dutch), regardless of type of language mixing, method of measurement or vocabulary modality. Regarding relations with the minority language (Polish/Turkish), overall language mixing and intra-sentential language mixing seem to be negatively related to vocabulary outcomes. We will discuss the model outcomes in more detail below.

**3.2.1. Language mixing and receptive vocabulary outcomes**

By inspecting the regression coefficients (Table 4), we see that overall, intra-sentential and inter-sentential language mixing do not relate to receptive vocabulary outcomes in Dutch or the minority language. A combination of relatively high BFs and posterior model probabilities confirm that both the Q-BEx and the LENA find positive evidence that there is no relation between overall language mixing and children’s receptive vocabulary outcomes.

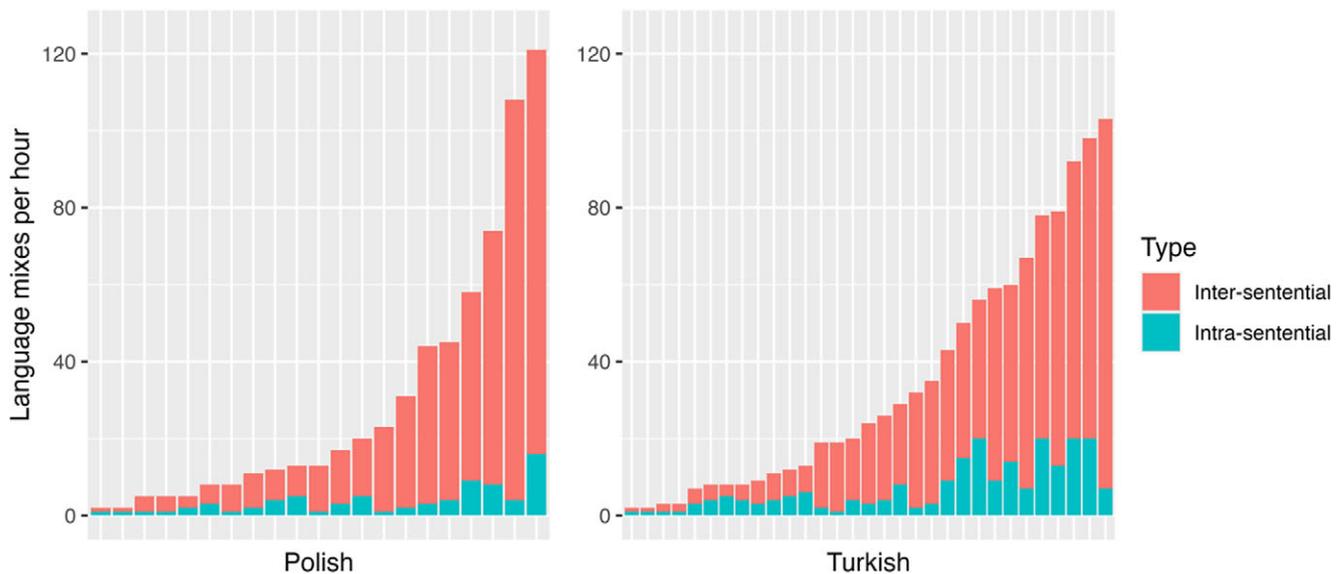
The results of the multivariate regressions are not decisive regarding the relation between the types of mixing and children’s receptive vocabulary outcomes. The Q-BEx finds strong evidence for a non-existent relation, but the Bayesian error probability of .20 prevents us from fully discarding the other hypotheses. This can be interpreted similarly to how a non-significant *p*-value does not lead



**Figure 1.** Boxplots of overall language mixing per group and method.

**Table 3.** Bayes Factors and Posterior Model Probabilities for group differences in mixing behavior

Hypothesis	Overall language mixing				Ratio inter-/intra-sentential			
	LENA		Q-BEx		LENA		Q-BEx	
	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>
H <sub>0</sub> : $\mu_{\text{Turkish-Dutch}} = \mu_{\text{Polish-Dutch}}$	6.06	.81	1.56	.45				
H <sub>1</sub> : $\mu_{\text{Turkish-Dutch}} > \mu_{\text{Polish-Dutch}}$	1.41	.19	24.54	.55	0.06	.05	0.11	.10
H <sub>2</sub> : $\mu_{\text{Turkish-Dutch}} < \mu_{\text{Polish-Dutch}}$					18.31	.95	9.43	.90

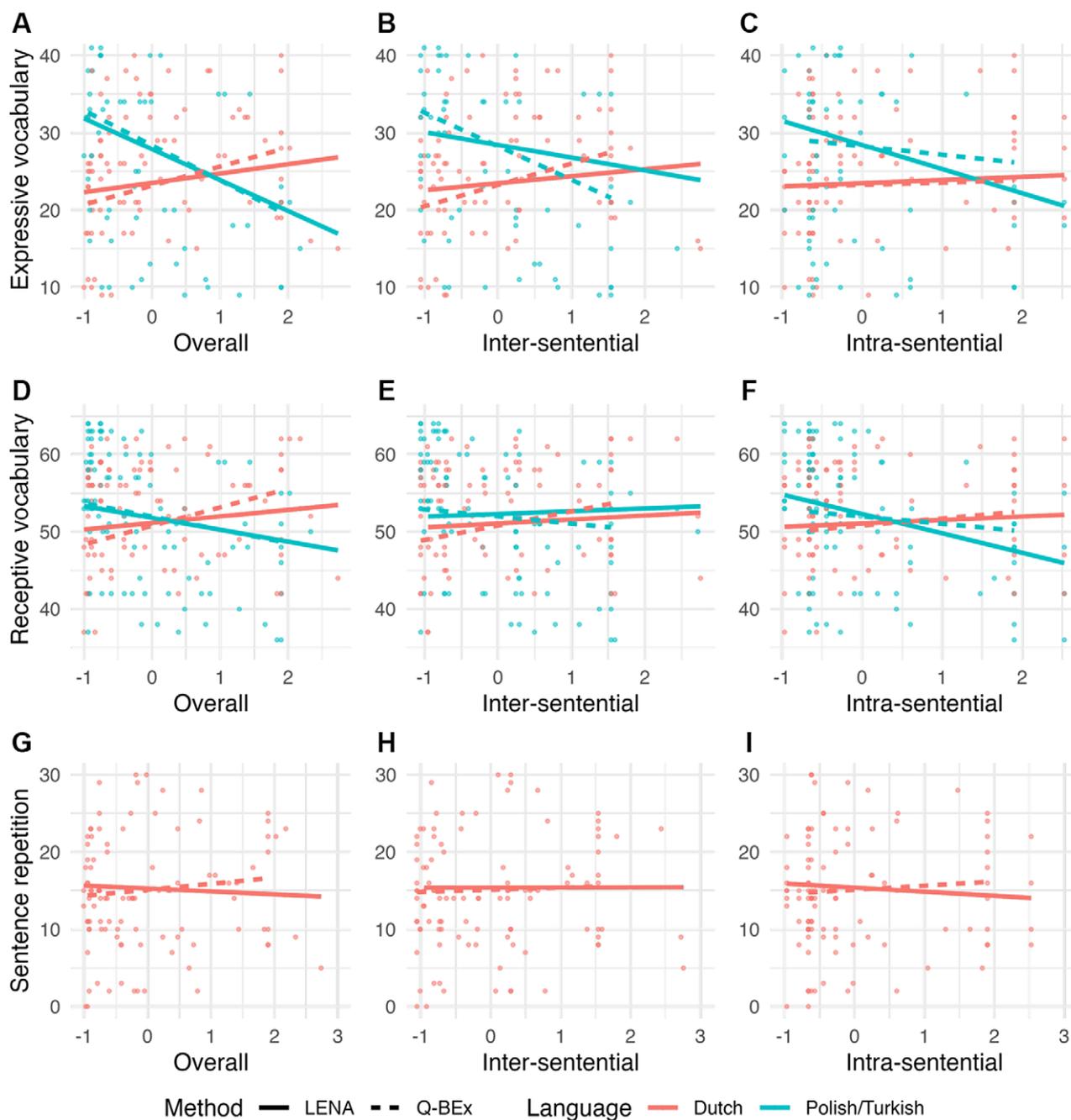
**Figure 2.** Ratio of parental inter- and intra-sentential language mixing per family, as observed in the LENA recordings. Each bar represents the parental language mixing in one household.**Table 4.** Regression coefficients and 95% confidence intervals of parental language mixing on children's language outcomes

		Receptive vocabulary				Expressive vocabulary				Sentence repetition	
		Majority language (Dutch)		Minority language (Polish/Turkish)		Majority language (Dutch)		Minority language (Polish/Turkish)		Majority language (Dutch)	
		B	95% CI	B	95% CI	B	95% CI	B	95% CI	B	95% CI
Overall	LENA	0.85	[-1.76, 3.47]	-1.54	[-3.71, 0.64]	1.20	[-0.98, 3.39]	-3.99	[-6.78, -1.19]	-0.38	[-2.02, 1.26]
	Q-BEx	2.41	[-1.02, 5.84]	-1.72	[-4.64, 1.20]	2.46	[-0.41, 5.32]	-4.52	[-8.34, -0.70]	0.79	[-1.42, 3.01]
Inter-sentential	LENA	0.52	[-2.88, 3.92]	0.36	[-2.39, 3.11]	0.90	[-1.94, 3.75]	-1.66	[-5.23, 1.92]	-0.06	[-2.07, 1.96]
	Q-BEx	1.86	[-1.94, 5.66]	-0.94	[-4.18, 2.30]	2.76	[-0.39, 5.90]	-4.34	[-8.53, -0.15]	0.29	[-2.17, 2.74]
Intra-sentential	LENA	0.45	[-2.88, 3.78]	-2.52	[-5.21, 0.17]	0.41	[-2.37, 3.20]	-3.12	[-6.62, 0.37]	-0.50	[-2.50, 1.51]
	Q-BEx	0.93	[-2.21, 4.07]	-0.96	[-3.64, 1.71]	0.30	[-2.30, 2.90]	-1.10	[-4.57, 2.36]	0.55	[-1.48, 2.58]

Note. The meaning of each color: blue: no relation, red: negative relation, yellow: negative trend, green: positive trend. Regressions were considered significant relations when the confidence interval did not contain 0. Regressions were considered trends when the confidence interval crossed the 0 border with less than .5.

to acceptance of the H<sub>0</sub>, but merely that the data fail to fully discard H<sub>0</sub>. When language mixing is measured with the LENA, equal evidence is found for H<sub>0</sub>: there is no relation between each type of language mixing and receptive vocabulary, and for H<sub>1</sub>: there is a negative relation between intra-sentential language mixing and receptive vocabulary. The shared evidence likely stems from both hypotheses partly predicting the true effect. The regression coefficients in Table 4 show that inter-sentential language mixing is

indeed unrelated to Dutch and Polish/Turkish receptive vocabulary (as predicted by both H<sub>0</sub> and H<sub>1</sub>), intra-sentential language mixing is unrelated to Dutch receptive vocabulary (as predicted by H<sub>0</sub>) but negatively related to Polish/Turkish receptive vocabulary (as predicted by H<sub>1</sub>). Thus, our data suggest that intra-sentential language mixing is not related to receptive Dutch vocabulary, but negatively related to receptive Polish/Turkish vocabulary. This newly generated hypothesis should be tested in future studies with new data.



**Figure 3.** The relation between all types of language mixing and expressive vocabulary (A–C), receptive vocabulary (D–F) and sentence repetition scores (G–I). Sentence repetition scores were only available in Dutch.

**3.2.2. Language mixing and expressive vocabulary outcomes**

We specified in our hypotheses that the relations between language mixing and vocabulary outcomes would be the same in both languages. In Table 4 we observe that overall language mixing is unrelated to Dutch expressive vocabulary, but negatively related to Polish/Turkish expressive vocabulary. It can thus be concluded that our hypotheses are not in line with our data. Therefore, none of our informative hypotheses about the relation between overall language mixing and expressive vocabulary receive any support compared to their complement (All  $BF_{cS} < 0$ ; Table 5).

When looking at the types of language mixing, both LENA and Q-BEx do not find any relation between inter-sentential or intra-sentential language mixing and Dutch expressive vocabulary. The

results of the multivariate regressions are not decisive regarding the relation between types of language mixing and expressive vocabulary in the minority language. A negative relation is observed between inter-sentential language mixing as measured by the Q-BEx questionnaire and Polish/Turkish expressive vocabulary. No negative relations with inter-sentential language mixing were predicted by any of our informative hypotheses. Therefore, the Q-BEx data do not receive support for any hypothesis (Table 5). With the LENA method, a negative trend between intra-sentential language mixing and Polish/Turkish expressive vocabulary is observed (Table 4). This negative trend contributes to the evidence that LENA finds for  $H_1$ : expressive vocabulary is not related to inter-sentential language mixing, but negatively related to intra-sentential language mixing.

**Table 5.** Bayes factors and posterior model probabilities for our hypotheses regarding the relation between language mixing and language outcome

Hypothesis	Receptive vocabulary				Expressive vocabulary				Sentence repetition			
	LENA		Q-BEx		LENA		Q-BEx		LENA		Q-BEx	
	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>	BF <sub>c</sub>	PMP <sub>a</sub>
Overall language mixing												
H <sub>0</sub> : $B_{1,2} = 0$	7.31	.86	4.19	.85	0.31	.35	0.42	.65	6.50	.77	5.63	.74
H <sub>1</sub> : $B_{1,2} < 0$	0.89	.11	0.23	.06	0.49	.64	0.14	.29	2.08	.16	0.48	.06
H <sub>2</sub> : $B_{1,2} > 0$	0.20	.03	0.36	.09	0.01	.01	0.03	.06	0.48	.08	3.15	.20
Types of language mixing												
H <sub>0</sub> : $B_{inter1,2} = 0$ & $B_{intra1,2} = 0$	10.88	.46	26.83	.80	0.72	.12	0.86	.75	21.27	.68	20.23	.67
H <sub>1</sub> : $B_{inter1,2} = 0$ & $B_{intra1,2} < 0$	12.48	.52	3.28	.10	5.42	.88	0.21	.19	7.50	.24	2.50	.08
H <sub>2</sub> : $B_{inter1,2} = 0$ & $B_{intra1,2} > 0$	0.57	.02	3.42	.10	0.03	.01	0.07	.06	2.69	.09	7.44	.25

Note.  $B_{inter}$  = inter-sentential language mixing,  $B_{intra}$  = intra-sentential language mixing,  $_1$  = Dutch vocabulary and  $_2$  = Polish/Turkish vocabulary. Note that the hypotheses for sentence repetition were only available for the majority language.

### 3.3. RQ3: To what extent does parental language mixing relate to children's sentence repetition scores?

As can be seen in Figures 3G–I and Table 4, all regression coefficients are weak and indicate a non-existent relation between all types of language mixing and Dutch sentence repetition scores, but the results from our Bayesian regression analyses show high error probabilities (Table 5). Both the Q-BEx and LENA find the most support for the H<sub>0</sub> in the case of overall language mixing and types of language mixing. In conclusion, based on regression coefficients and BF<sub>s</sub>, our data strongly suggest that all types of language mixing are unrelated to Dutch sentence repetition scores, although high error probabilities prevent us from ruling out other possible hypotheses.

### 3.4. RQ4: Does the method used to measure language mixing affect the observed relations between parental language mixing and children's language outcomes?

As can be seen in Figure 3, the full lines (LENA) and the dotted lines (Q-BEx) are very close to each other, indicating that both methods of measurement generally predict the same relation between language mixing and language outcomes. Indeed, it can be seen that, except trends, data obtained with LENA and Q-BEx yield the same conclusion in 14 out of 15 regressions (either no relation or a negative relation; Table 4). The only relation where the Q-BEx and LENA differ is that between inter-sentential language mixing and Polish/Turkish expressive vocabulary; Q-BEx points to a negative relation, whereas for LENA no relation emerged. This discrepancy results in the LENA method providing evidence for the H<sub>1</sub>, whereas the Q-BEx did not find evidence for any hypothesis in the set (Table 5). In nearly all other cases, both methods agreed on the same informative hypothesis of the set receiving the most support, resulting in identical conclusions.

## 4. Discussion

The present study investigated the relation between parental language mixing and bilingual children's language outcomes. Previous studies have investigated this relation, and their results were equivocal. Their findings varied from a negative relation to a positive relation to no relation at all (Bail et al., 2015; Byers-Heinlein, 2013;

Carbajal & Peperkamp, 2020; Place & Hoff, 2011, 2016). These inconsistent findings may be related to which aspect of language development was examined, which languages were examined, which group of participants was included in the study, which type of language mixing was investigated and/or which method was used to measure language mixing. To ensure a thorough approach, the present study used two different methods to measure language mixing and distinguished between intra- and inter-sentential language mixing. Moreover, it included receptive and expressive vocabulary outcomes in both the majority and minority languages as well as sentence repetition scores in two different groups of participants. This discussion will address the relation between parental language mixing and language outcomes in greater detail, focusing on how our results align with the existing body of literature on the potentially negative, positive or non-existent relations. Moreover, we will discuss the role of the method that is used to measure language mixing. However, before going into this, we will discuss the observed frequency and type of language mixing of our two groups of parents. This is important because the disparate findings between previous studies may be partly attributed to having investigated participants from different communities, who may exhibit different language-mixing practices (Torres & Potowski, 2016). The inclusion of two groups and the comparison of their language-mixing behavior is something that has not been addressed so far in other studies.

### 4.1. Frequency and type of language mixing by Polish-Dutch and Turkish-Dutch parents

We hypothesized that the Turkish-Dutch parents would engage more in language mixing than the Polish-Dutch parents because in general Turkish immigrants in the Netherlands have a longer migration history than the Polish immigrants and have developed a practice of intense language mixing (Backus & Demirçay, 2021; CBS, n.d.). Additionally, it was speculated that the Turkish group has access to more bilingual interlocutors, facilitating bilingual language use such as language-mixing practices. However, we could not conclude that the Turkish parents mixed more than the Polish parents based on our data.

There are two potential explanations for the lack of evidence. First, our sample may not be representative of the general Turkish community in the Netherlands. According to the Statistics Netherlands (CBS, n.d.), 52% of Dutch Turks are born in the Netherlands. In our

sample, however, only 38% of Turkish parents were born in the Netherlands. This means that a relatively large percentage of Turkish parents in our sample did not grow up in the context of the intense mixing that characterizes the Turkish community in the Netherlands (Backus & Demirçay, 2021). As a result, the Turkish families in our sample may mix less than the average Turkish family in the Netherlands. Second, the *bain* package uses the largest prior variance by default, resulting in a slight preference for the null hypothesis. This is because “In an era of heightened awareness of publication bias, sloppy science, and irreplaceability of research results, researchers should be conservative, that is, convincing evidence is needed before another hypothesis is preferred over  $H_0$ ” (Hojtink et al., 2019, p. 30). This may have led to an increase in support for the null hypothesis.

Regarding the different types of language mixing, both groups mixed more inter-sententially than intra-sententially, which is the common pattern among bilingual speakers (Bail et al., 2015; Kremin et al., 2022; Torres & Potowski, 2016). However, the discrepancy between inter- and intra-sentential language mixing was smaller in the Turkish-Dutch group. One possible explanation for this finding is that we considered conversational language mixing (e.g., when the parent would respond in a different language than being spoken to by the child; Gross et al., 2022; Ribot & Hoff, 2014) to be a form of inter-sentential language mixing. It may be that parents who do not (natively) speak Dutch are more likely to respond in the minority language than parents who do speak Dutch, resulting in relatively more cross-speaker inter-sentential language mixing. In our Polish sample, 86% of parents were born outside of the Netherlands (in contrast to 69% of Turkish parents); as a result, the Polish group may engage more in cross-speaker inter-sentential language mixing compared to the Turkish group. In addition, several Turkish parents may show more intra-sentential language mixing as a result of the intense language mixing practices that are typical for the Turkish community in the Netherlands (Backus & Demirçay, 2021).

## 4.2. Relations between caregiver language mixing and children’s language outcomes

In the introduction, we reviewed empirical and conceptual support for a negative, positive or non-existent relation between parental language mixing and children’s language outcomes. The present study found (a) evidence for a negative relation between overall language mixing and children’s expressive vocabulary in the minority language, regardless of the method with which language mixing was measured (LENA or Q-BEx); (b) no evidence for any positive relation with children’s language outcomes and (c) in particular, evidence for the absence of relation between parental language mixing and children’s language outcomes when it comes to receptive and expressive vocabulary and sentence repetition scores in the majority language and receptive vocabulary in the minority language.

### 4.2.1. Negative relation

The negative relation between overall language mixing and expressive vocabulary in the minority language can be explained by two reasons. First, the minority language may be susceptible to effects of language mixing in the home environment because mixed language input constitutes a relatively large proportion of the total language input in this language. This contrasts with the majority language to which children are frequently exposed outside of the home environment and in largely monolingual settings (e.g., in daycare/school). It may thus be that the proportion of mixed language input

is too small to impact any language outcomes in the majority language (Kremin et al., 2022), whereas it may be substantial enough to affect minority language outcomes. This could also explain why we found evidence for a relation between overall language mixing and child language outcomes, while no relations emerged for intra- and inter-sentential language mixing individually. The latter two types of language mixing individually make up too small a part of the total language input (Bail et al., 2015; Kremin et al., 2022). Additionally, it has been argued that language mixing might diminish the prestige of the minority language (Cooper & Fishman, 1971; Place & Hoff, 2016), which may discourage children to learn the minority language and negatively affect their language outcomes.

Second, (mixed) language input may specifically impact expressive language skills (Dijkstra et al., 2016; Verhoeven et al., 2024). The Weaker Links Hypothesis (Gollan et al., 2008) explains this phenomenon based on phonological representations (i.e., the mental representations of the combination of sounds that make up words) that become stronger with experience. When children hear sentences or words more frequently, their phonological representations of those words become stronger (Gibson et al., 2014). By mixing two languages, the monolingual equivalents occur less frequently and become less entrenched. This weakens their phonological representations. Importantly, these weaker phonological representations may be sufficient for recognizing words, but not for producing them (Gollan et al., 2011). Then, in line with the findings reported here, language mixing would not result in any issues with regard to the recognition of words (receptive vocabulary), but they would impede the retrieval of words from the mental lexicon during production (expressive vocabulary).

Place and Hoff (2011, 2016) are the only other studies that have related the frequency of parental language mixing to expressive vocabulary outcomes in the minority language, but they found no relation. This may be explained by their measures of language mixing not accurately representing the overall language mixing that takes place within the home environment. Language diaries lack information on how frequently the languages are being mixed within the 30-minute blocks and the Language Mixing Scale (LMS) is a closer representation of intra-sentential language mixing than overall language mixing (Place & Hoff, 2016). The two studies that did find negative relations did so for receptive vocabulary in the majority language in children younger than 18 months (Byers-Heinlein, 2013; Carbajal & Peperkamp, 2020). These effects were not found in older children, suggesting that children may become more proficient at processing mixed language input with age, thus overcoming the previously identified negative effect of language mixing.

### 4.2.2. Positive relation

We did not find any evidence regarding a positive relation between language mixing and children’s vocabulary outcomes as was found by Bail et al. (2015). Importantly, Bail and colleagues used correlational analyses without controlling for other predictor variables. It may be that the positive relation they found between intra-sentential language mixing and children’s total vocabulary score disappears when age, gender, language input and parental education are controlled for.

### 4.2.3. No relation

In this study, we investigated 15 potential relations between language mixing by parents and children’s language outcomes. Out of those 15 relations, evidence for only one negative relation emerged,

as discussed above. Importantly, we found evidence that parental language mixing was unrelated to children's language outcomes in the majority language, regardless of type of mixing and language outcome measure. In addition, parental language mixing was unrelated to receptive vocabulary in the minority language. This suggests that the null findings of Bail *et al.* (2015) and Place & Hoff (2011, 2016), who looked at Spanish-English bilinguals between 17–30 months old, can be generalized to other populations (i.e., Polish-Dutch and Turkish-Dutch language combinations and children aged between 36–72 months). Most previous research focused on children's vocabulary outcomes, except for Place and Hoff (2011, 2016) who also investigated the relation between parental language mixing and grammatical aspects of children's language. Using a sentence repetition measure, we found, in line with Place and Hoff (2011, 2016), that language mixing in the input was unrelated to children's grammar outcomes, at least in the majority language.

#### 4.2.4. The role of method used to measure language mixing

To determine whether the method of how language mixing was measured underlies the disparate findings from previous studies, we measured language mixing with the use of daylong audio recordings made with LENA and the Q-BEx questionnaire. Previous studies have used language diaries, the LMS or short recordings in a laboratory setting, but called for more naturalistic measures of language mixing in the home environment (Bail *et al.*, 2015; Byers-Heinlein, 2013; Carbajal & Peperkamp, 2020; Place & Hoff, 2011, 2016). Place and Hoff (2016) measured language mixing with both a language diary and the LMS and found no relations with any language outcome with either method. The present study found that the LENA recording and the Q-BEx questionnaire mostly yielded comparable results with regard to children's language outcomes, suggesting that they are comparable instruments to measure language mixing (Verhoeven *et al.*, 2024). However, the Q-BEx revealed a negative relation between inter-sentential language mixing and expressive vocabulary in the minority language, while no relation emerged when language mixing was measured with the LENA. This may be due to the difference in what an inter-sentential language mix entails for the two methods. The LENA included conversational language mixes (i.e., cross-speaker) as cases of inter-sentential language mixing, whereas the Q-BEx only contained within-speaker inter-sentential language mixes. It was mentioned that the disparate findings from previous studies may in part be ascribed to their different methodologies (e.g., Byers-Heinlein (2013) finding a negative relation with the LMS and Bail *et al.* (2015) finding a positive relation with audio recordings). However, the current study shows that the language mixing measure has limited impact, as long as the measures target the same construct (e.g., whether inter-sentential language mixing contains conversational language mixing or not). Other factors that may underlie these disparate results, such as group of participants, ages, the languages they speak and how accustomed the participants are to language mixing, should be addressed in future work.

#### 4.3. Limitations and future research

The present study is subject to some limitations. The first limitation is our relatively small sample size. Inclusion criteria such as specific language combinations and age ranges, in combination with extensive research methods involving daylong audio recordings make it challenging to obtain a large sample. Furthermore, the (manual) processing of naturalistic data leads to intensive workloads (see also suggestions from Cychosz *et al.*, 2021). The rapid development of

technology will hopefully facilitate automatic processing of bilingual daylong audio recordings in the future, thereby facilitating the collection and analysis of larger sample sizes using this method. Smaller sample sizes result in higher Bayesian error probabilities, preventing us in some cases from discarding the other hypotheses in the set. Thus, despite our data sometimes showing strong evidence in support of a hypothesis with a high BF, the high error probabilities forced us to conclude that the results remained inconclusive (e.g., the hypothesis that intra- and inter-sentential language mixing did not relate to sentence repetition outcomes had a BF of 21.27, but a Bayesian error probability of .32). Even though our results may not allow for very strong conclusions, our analytic approach is strong and nuanced. Moreover, communicating weaker results to the field is important to help mitigate publication bias (Song *et al.*, 2013).

Second, only one (weekend) day was recorded. As the children in our sample go to daycare or school, it was not feasible to record weekdays. Even though parents exhibit similar language use during weekdays and weekend days (Orena *et al.*, 2020), recording a single day provides merely a snapshot and may lead to inaccurate representations of the typical language use in the home environment (e.g., when a parent who mixes their languages frequently is not at home during the recording day). However, Verhoeven *et al.* (2024) showed that a single daylong audio recording can be considered a good measure of the bilingual language environment.

Third, we measured grammatical abilities only in Dutch, the majority language, and not in Polish and Turkish, the two minority languages. It thus remains an open question whether the observed negative relation between language mixing and minority language outcomes is limited to expressive vocabulary.

Finally, the cross-sectional design of the study does not allow us to investigate the direction of the relation between overall language mixing and expressive vocabulary in the minority language (Bail *et al.*, 2015; Soderstrom, 2007). It remains unknown whether parents alter their language use based on children's expressive skills, or whether parental language mixing affects children's ability or desire to express themselves in the minority language. Future studies should conduct a longitudinal investigation to ascertain whether there is a causal relation, and in which direction.

#### 4.4. Conclusion

The current study provides rich and naturalistic data regarding the language-mixing behavior of multilingual families in the Netherlands and its relation to children's language outcomes. Our data do not provide strong evidence for differences in overall language mixing frequency between the Turkish-Dutch and Polish-Dutch families, but they do suggest that Turkish-Dutch parents mix more intra-sententially than Polish-Dutch parents. Parental language mixing does not relate to any language outcome in the majority language, while overall language mixing was negatively related to children's expressive vocabulary outcomes in the minority language. We abstain from advising parents to separate their languages because (a) the direction of the negative relation is unclear, (b) overall, our study provided evidence for the unrelatedness of parental language mixing and child language outcomes and (c) language mixing is such common practice (Kremin *et al.*, 2022). Parents may, however, consider specific rules about minority language use in the home environment, as these have been found to support minority language maintenance (Hollebeke *et al.*, 2023), but these do not need to involve a strict separation of the two languages.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S1366728925100175>.

**Data availability statement.** The data that support the findings of this study will be openly available on the Open Science Framework at <https://osf.io/qudr7/>.

**Acknowledgments.** We would like to thank all participating children and parents. Furthermore, we thank Laura Koelma, Vera Snijders, Gülşah Yazıcı, Hatice Bulut, Fatma Nur Öztürk, Klaudia Latkowska, Patricia Dworak, Zuzanna Kruber and Jakob Kaiser for their help with collecting and/or processing the data. Lastly, we want to thank Herbert Hoijtink for his help with the Bayesian statistics.

**Funding statement.** This research is funded by the VICI grant awarded to E.B. by the Netherlands Organization for Scientific Research (NWO; Grant Number VIC.191.042).

**Competing interests.** The authors declare that there are no competing interests.

**Disclosure of AI tools.** The authors declare that no images, text, data or ideas have been AI-generated. The translational tool DeepL was used for translation purposes, as English is not the authors' native language.

**Ethical standard.** The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

## References

- Adamou, E., & Shen, X. R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, 23(1), 53–70. <https://doi.org/10.1177/1367006917709094>.
- Aronsson, K. (2020). Myths about bilingual learning in family life settings: Werner Leopold's child language biographies and contemporary work on children's play practices. *Learning, Culture and Social Interaction*, 26, 100226.
- Backus, A. (2005). Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism*, 9(3–4), 307–340. <https://doi.org/10.1177/13670069050090030101>.
- Backus, A. (2012). Turkish as an immigrant language in Europe. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (1st ed., pp. 770–790). Wiley. <https://doi.org/10.1002/9781118332382.ch31>
- Backus, A., & Demirçay, D. (2021). Intense Turkish-Dutch bilingualism leads to intense Turkish-Dutch mixing: A usage-based account of increasing integration of two typologically different languages. *Belgian Journal of Linguistics*, 35(1), 13–33. <https://doi.org/10.1075/bjl.00062.bac>.
- Bail, A., Morini, G., & Newman, R. S. (2015). Look at the gato! Code-switching in speech to toddlers. *Journal of Child Language*, 42(5), 1073–1101. <https://doi.org/10.1017/S0305000914000695>.
- Baker, C., & Jones, S. P. (1998). *Encyclopedia of bilingualism and bilingual education*. Multilingual Matters.
- Blanco-Elorrieta, E., & Pyllkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, 37(37), 9022–9036.
- Blom, E., Oudgenoeg-Paz, O., & Verhagen, J. (2018). Wat als ouders talen mixen? Een onderzoek naar de relatie tussen mixen van talen door ouders en de taalvaardigheid van meertalige kinderen. [what if parents mix languages? A study into the relationship between parents' language mixing and multilingual children's language proficiency]. *Nederlands Tijdschrift Voor Logopedie*, 90(7), 18–25.
- Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, 16(1), 32–48. <https://doi.org/10.1017/S1366728912000120>.
- Byers-Heinlein, K., Jardak, A., Fourakis, E., & Lew-Williams, C. (2022). Effects of language mixing on bilingual children's word learning. *Bilingualism: Language and Cognition*, 25(1), 55–69. <https://doi.org/10.1017/S1366728921000699>.
- Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences*, 114(34), 9032–9037. <https://doi.org/10.1073/pnas.1703220114>.
- Carbajal, M. J., & Peperkamp, S. (2020). Dual language input and the impact of language separation on early lexical development. *Infancy*, 25(1), 22–45. <https://doi.org/10.1111/infa.12315>.
- Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders*, 49(6), 649–671. <https://doi.org/10.1111/1460-6984.12082>.
- CBS. (n.d.). *Bevolking—Integratie en Samenleven* | CBS [Webpagina]. Bevolking—Integratie en Samenleven | CBS. Retrieved May 14, 2024, from <https://longreads.cbs.nl/integratie-en-samenleven-2022/bevolking>
- Cooper, R. L., & Fishman, J. A. (1971). *Bilingualism in the barrio*. Indiana University.
- Cychosz, M., Villanueva, A., & Weisleder, A. (2021). Efficient estimation of children's language exposure in two bilingual communities. *Journal of Speech, Language, and Hearing Research*, 64(10), 3843–3866. [https://doi.org/10.1044/2021\\_JSLHR-20-00755](https://doi.org/10.1044/2021_JSLHR-20-00755).
- De Cat, C., Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2022). Quantifying Bilingual Experience (Q-BEX): Questionnaire manual and documentation. *Charlottesville: Open Science Framework*
- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(3), 411–424.
- Denby, T., Schecter, J., Arn, S., Dimov, S., & Goldrick, M. (2018). Contextual variability and exemplar strength in phonotactic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(2), 280.
- Dijkstra, J., Kuiken, F., Jorna, R. J., & Klinkenberg, E. L. (2016). The role of majority and minority language input in the early development of a bilingual vocabulary. *Bilingualism: Language and Cognition*, 19(1), 191–205.
- Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197.
- Doğruöz, A. S., & Backus, A. (2010). Turkish in the Netherlands: Development of a new variety. *Language Contact. New Perspectives* (pp. 87–102). John Benjamins Publishing Company.
- 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.
- Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014). The relation between language experience and receptive-expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism*, 17(1), 90–110. <https://doi.org/10.1080/13670050.2012.743960>.
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787–814.
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436. <https://doi.org/10.1111/1467-9280.00476>.
- Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (Ed.), *Advances in psychology* (Vol. 83, pp. 207–220). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)61496-X](https://doi.org/10.1016/S0166-4115(08)61496-X).
- Grainger, J., Midgley, K., & Holcomb, P. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 267–283). John Benjamins Publishing Company. <https://doi.org/10.1075/lald.52.18gra>
- Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, 32(2), 83–92. <https://doi.org/10.1177/1525740110367826>.
- Gross, M. C., López González, A. C., Girardin, M. G., & Almeida, A. M. (2022). Code-switching by Spanish-English bilingual children in a code-switching conversation sample: Roles of language proficiency, interlocutor behavior, and parent-reported code-switching experience. *Language*, 7(4), 246.

- Gu, X. (2016). *Bayesian evaluation of informative hypotheses* [PhD Thesis]. Utrecht University.
- Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00278>.
- Haman, E., Łuniewska, M., Hansen, P., Simonsen, H. G., Chiat, S., Bjekić, J., Blažienė, A., Chyl, K., Dabašinskiė, I., De Abreu, E., Gagarina, N., Gavarró, A., Håkansson, G., Harel, E., Holm, Kapalková, S., Kunnari, S., Levorato, C., Lindgren, J., & Armon-Lotem, S. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from cross-linguistic lexical tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, 31(11–12), 818–843. <https://doi.org/10.1080/02699206.2017.1308553>.
- Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). 8. Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. De Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 196–240). Multilingual Matters. <https://doi.org/10.21832/9781783093137-010>.
- Hoff, E., & Core, C. (2015). What clinicians need to know about bilingual development. *Seminars in Speech and Language*, 36(2), 89–99. <https://doi.org/10.1055/s-0035-1549104>.
- Hojtink, H., Gu, X., & Mulder, J. (2019). Bayesian evaluation of informative hypotheses for multiple populations. *British Journal of Mathematical and Statistical Psychology*, 72(2), 219–243. <https://doi.org/10.1111/bmsp.12145>.
- Hojtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, 24(5), 539–556. <https://doi.org/10.1037/met0000201>.
- Hollebeke, I., Dekeyser, G. N. M., Caira, T., Agirdag, O., & Struys, E. (2023). Cherishing the heritage language: Predictors of parental heritage language maintenance efforts. *International Journal of Bilingualism*, 27(6), 925–941. <https://doi.org/10.1177/13670069221132630>.
- Kaushanskaya, M., Crespo, K., & Neveu, A. (2023). Does code-switching influence novel word learning? *Developmental Science*, 26(2), e13292. <https://doi.org/10.1111/desc.13292>.
- Kremin, L. V., Alves, J., Orena, A. J., Polka, L., & Byers-Heinlein, K. (2022). Code-switching in parents' everyday speech to bilingual infants. *Journal of Child Language*, 49(4), 714–740. <https://doi.org/10.1017/S0305000921000118>.
- Lam, C.-F., & Matthews, S. (2020). Inter-sentential code-switching and language dominance in Cantonese–English bilingual children. *Journal of Monolingual and Bilingual Speech*, 2, 73–105. <https://doi.org/10.1558/jmbs.13308>.
- Marasli, Z., & Montag, J. L. (2023). Optimizing random sampling of daylong audio. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 45(45). <https://escholarship.org/uc/item/0478r9pb>
- MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs (3rd ed.)*. Psychology Press. <https://doi.org/10.4324/9781315805672>
- Muysken, P. (1997). Code-switching processes: Alternation, insertion, congruent lexicalization. In M. Pütz (Ed.), *IMPACT: Studies in language and society* (Vol. 1, p. 361). John Benjamins Publishing Company. <https://doi.org/10.1075/impact.1.25muy>.
- Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Oxford University Press.
- Myers-Scotton, C. (2017). Code-switching. In *The handbook of sociolinguistics* (pp. 217–237). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405166256.ch13>
- Orena, A. J., Byers-Heinlein, K., & Polka, L. (2020). What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental Science*, 23(2), e12901. <https://doi.org/10.1111/desc.12901>.
- Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, 82(6), 1834–1849. <https://doi.org/10.1111/j.1467-8624.2011.01660.x>.
- Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, 19(5), 1023–1041. <https://doi.org/10.1017/S1366728915000322>.
- Poeste, M., Müller, N., & Arnaus Gil, L. (2019). Code-mixing and language dominance: Bilingual, trilingual and multilingual children compared. *International Journal of Multilingualism*, 16(4), 459–491. <https://doi.org/10.1080/14790718.2019.1569017>.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching I. *Linguistics*, 18(7–8), 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>
- Poplack, S. (2001). Code-switching (linguistic). In N. J. Smelser & B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 2062–2065). Elsevier.
- Potter, C. E., Fourakis, E., Morin-Lessard, E., Byers-Heinlein, K., & Lew-Williams, C. (2019). Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science*, 22(4), e12794. <https://doi.org/10.1111/desc.12794>.
- R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.3.0) [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, 17(6), 880–891. <https://doi.org/10.1111/desc.12172>.
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interactions on later language development in Spanish–English bilingual infants. *Child Development*, 88(4), 1216–1234. <https://doi.org/10.1111/cdev.12648>.
- Ribot, K. M., & Hoff, E. (2014). ¿Cómo estas? “I'm good.” conversational code-switching is related to profiles of expressive and receptive proficiency in Spanish–English bilingual toddlers. *International Journal of Behavioral Development*, 38(4), 333–341. <https://doi.org/10.1177/0165025414533225>.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science*, 1(6), 906–914. <https://doi.org/10.1002/wcs.78>.
- Ruan, Y., Byers-Heinlein, K., Orena, A. J., & Polka, L. (2023). Mixed-language input and infant volubility: Friend or foe? *Bilingualism: Language and Cognition*, 26(5), 1051–1066. <https://doi.org/10.1017/S1366728923000287>.
- Siow, S., Gillen, N. A., Lepädatu, I., & Plunkett, K. (2023). Double it up: Vocabulary size comparisons between UK bilingual and monolingual toddlers. *Infancy*, 28(6), 1030–1051. <https://doi.org/10.1111/infa.12562>.
- Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27(4), 501–532. <https://doi.org/10.1016/j.dr.2007.06.002>.
- Song, F. H., & Loke, Y. (2013). Publication bias: What is it? How do we measure it? How do we avoid it? *Open Access Journal of Clinical Trials*, 71. <https://doi.org/10.2147/OAJCT.S34419>.
- Torres, L., & Potowski, K. (2016). Hablamos los dos in the Windy City. *Spanish-English Codeswitching in the Caribbean and the US*, 11, 83.
- Treffers-Daller, J., Ongun, Z., Hofweber, J., & Korenar, M. (2020). Explaining individual differences in executive functions performance in multilinguals: The impact of code-switching and alternating between multicultural identity styles. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.561088>.
- Van Wonderen, E., & Unsworth, S. (2021). Testing the validity of the cross-linguistic lexical task as a measure of language proficiency in bilingual children. *Journal of Child Language*, 48(6), 1101–1125. <https://doi.org/10.1017/S030500092000063X>.
- Verhoeven, E., van Witteloostuijn, M., Oudgenoeg-Paz, O., & Blom, E. (2024). Comparing different methods that measure bilingual children's language environment: A closer look at audio recordings and questionnaires. *Language*, 9(7), 231.
- Wiig, E. H., Secord, W. A., Semel, E., & de Jong, J. (2012). CELF preschool-2-NL: Clinical evaluation of language fundamentals: Preschool – Nederlandstalige versie. Handleiding. <https://dare.uva.nl/search?metis.record.id=376553>
- Yagmur, K. (2009). Language use and ethnolinguistic vitality of Turkish compared with the Dutch in the Netherlands. *Journal of Multilingual and Multicultural Development*, 30(3), 219–233. <https://doi.org/10.1080/01434630802369445>.
- Yip, V., & Matthews, S. (2006). Assessing language dominance in bilingual acquisition: A case for mean length utterance differentials. *Language Assessment Quarterly*, 3(2), 97–116. [https://doi.org/10.1207/s15434311aq0302\\_2](https://doi.org/10.1207/s15434311aq0302_2).
- Yow, W. Q., Tan, J. S. H., & Flynn, S. (2018). Code-switching as a marker of linguistic competence in bilingual children. *Bilingualism: Language and Cognition*, 21(5), 1075–1090. <https://doi.org/10.1017/S1366728917000335>.

Zentella, A. C. (1998). *Growing up bilingual: Puerto Rican children in New York*. Blackwell. <https://ehrafworldcultures.yale.edu/cultures/nk05/documents/056>

### Appendix: Steps to calculate the number of occurrences of parental language mixing from daylong audio recordings.

1. Based on the automatic output provided by the LENA software, we first removed silent segments. Subsequently, 54 5-minute segments were sampled based on their conversational turn count (CTC). We selected 18 segments that contained the highest number of conversational turns, 18 segments with the lowest number of conversational turns (that were not silent) and 18 segments that were in the middle. More detailed information regarding the sampling method can be found on the Open Science Framework page of the study.
2. From those 5-minute segments, we analyzed every other 30-second segment (Marasli & Montag, 2023; Ramírez-Esparza et al., 2014, 2017), resulting in 270 30-second segments per participant.
3. All 270 30-second segments were manually coded for speaker(s), language(s) spoken, activity and whether there was speech directed to the target child (CDS).
4. Every 30-second segment that was coded as containing more than one language was fully transcribed in CHAT (MacWhinney, 2014) by bilingual research assistants. In the transcript, each time the language changed, it was coded as either intra- or inter-sentential language mixing.
5. All individual occurrences of language mixing were further coded for speaker, addressee, switch direction (i.e., whether the switch was from the majority language to the minority language or vice versa) and switch type. More specifically, we coded whether a switch was within a speaker, between speakers or between two different conversations. The latter was coded when a change of language occurred in the transcript, but the languages were not part of the same conversation. In this case, the switch was not considered an instance of language mixing and was not analyzed further. For within-speaker switches, we further coded the following subtypes: insertions, alternations, congruent lexicalizations (Muysken, 1997) and inter-sentential switches. As we are only interested in the difference between intra- and inter-sentential switches in this study, all instances of insertions, alternations and congruent lexicalizations are collapsed into one category of intra-sentential switches. Switches that occur between speakers are by definition inter-sentential.
6. As we investigate the role of parental language mixing on children's language outcomes, we retain only those switches that were uttered by a parent to the target child. The numbers of intra- and inter-sentential switches represent how many times these types of switches occur in the sampled LENA audio, which had a total duration of 135 minutes (i.e., 270 30-second segments). The outcome variable used in the analysis is the frequency of each type of switch per hour of recording, which is calculated by dividing these total mixing frequencies by 2.25.