

Article

Comparing Different Methods That Measure Bilingual Children's Language Environment: A Closer Look at Audio Recordings and Questionnaires

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

Faculty of Social and Behavioural Sciences, Utrecht University, 3584 CS Utrecht, The Netherlands

* Correspondence: e.g.e.c.verhoeven@uu.nl

Abstract: The quantity of language input is a relevant predictor of children's language development and is frequently used as a variable in child bilingualism research. Studies use various methods to measure bilingual language input quantity, but it is currently unknown what the optimal method is. We investigated the bilingual language input estimates of 31 Turkish–Dutch and 21 Polish–Dutch 3- to 5-year-old bilingual children, obtained via the questionnaire for Quantifying Bilingual Experience (Q-BEx) and day-long audio recordings made with Language Environment Analysis (LENA), and proposed a combined method that could overcome several shortcomings of the individual methods. The three methods are compared to each other in their correlation strength with receptive and expressive vocabulary scores. Each individual method correlated significantly with vocabulary scores, regardless of modality or language. Contrary to our hypothesis, the combined method did not correlate stronger with vocabulary outcomes than the Q-BEx and LENA individually did. The latter two did not differ significantly from each other in their correlational strength with vocabulary outcomes. These findings show that both the Q-BEx, LENA, and combined method can be deemed reliable to measure bilingual language input quantity. Future studies can make more informed decisions about their methodology in children's bilingualism research.

Keywords: language input; methods; bilingualism; questionnaires; Q-BEx; day-long audio recordings; LENA



Citation: Verhoeven, Emma, Merel van Witteloostuijn, Ora Oudgenoeg-Paz, and Elma Blom. 2024. Comparing Different Methods That Measure Bilingual Children's Language Environment: A Closer Look at Audio Recordings and Questionnaires. *Languages* 9: 231. <https://doi.org/10.3390/languages9070231>

Academic Editors: Maria Kihlstedt and Cathy Cohen

Received: 15 April 2024

Revised: 14 June 2024

Accepted: 18 June 2024

Published: 26 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The amount of language input that children receive is a relevant predictor of their language development, in addition to the type of input they receive and the characteristics of their interactions (e.g., Bergelson et al. 2023; Hart and Risley 1995; Huttenlocher et al. 1991; for a recent meta-analysis, see Anderson et al. 2021). Input here refers to all spoken language that the child is exposed to and will be interchangeably used with exposure, much like in other research (Cychosz et al. 2021; Orena et al. 2020; Unsworth 2016; but see Carroll (2017) for a different view). Children vary extensively in their input quantity, and for bilingual children, this variation is even greater than for monolinguals because their input is distributed over different languages (Hoff et al. 2012). Factors that modulate such variation include not only how talkative parents are and how much time they spend with their child, but also which languages are used at home by who and when, whether the child is going to school, and whether the child receives input from siblings, amongst other factors (Hoff 2006; Hoff and Core 2013; Weisleder and Fernald 2013). Estimating language input in each individual language, provided by all these different sources (e.g., parents, siblings, teachers, peers, community, etc.), is challenging. Yet, as the quantity of language input is a key predictor of a child's overall skill in that language, accurate measurement is crucial (De Houwer 2009; Gathercole and Thomas 2009; Pearson et al. 1997; Place and Hoff 2011, 2016).

The present study compared the two methods that are most frequently used to quantify bilingual language input, namely parental questionnaires and day-long recordings. Further, as they could complement each other, we introduce a new method that combines these two methods by using their individual strengths (Orena et al. 2020). By doing so, our study aims to contribute to the optimization of quantitative language input measures in child bilingualism research. To compare the three methods (i.e., questionnaires, recordings, and a combination), correlations with children's vocabulary scores were calculated (see Marchman et al. 2017 for a similar approach). Children's vocabulary scores can be divided into two modalities: expressive vocabulary refers to the number of different words that the child produces, and receptive vocabulary refers to the number of words that the child understands. Expressive vocabulary is typically measured using picture naming tasks, while receptive vocabulary is often measured using picture selection tasks. As the quantity of language input has been proven to be related to children's vocabulary (Bijeljac-Babic et al. 2012; Byers-Heinlein 2013; Carbajal and Peperkamp 2020; Hart and Risley 1995; Hoff 2003; Huttenlocher et al. 1991; Lieven et al. 2019; Marchman et al. 2017; Pearson et al. 1997; Place and Hoff 2011, 2016; Potter et al. 2019; Rowe 2012; Unsworth et al. 2018), we assume that a stronger correlation indicates a better estimate of the language input. We expect the strongest correlations for the combined method, which would indicate that this method estimates the quantitative bilingual language input more accurately than parental questionnaires and audio recordings separately. Furthermore, we hypothesize that language input might be easier to estimate in the minority language than in the majority language. If this is indeed the case, correlations between each method and the vocabulary scores will be higher in the minority language than in the majority language.

1.1. Parental Questionnaires

The most frequently used method to measure the quantity of language input is parental questionnaires. Questionnaires can examine the language input of the child over a longer period of time, e.g., a week, month, year, or even the entire lifespan of the child. Examining longer periods of time allows parents to report on changes that have occurred in the language environment of the child (Byers-Heinlein et al. 2020). Furthermore, questionnaires are able to take into account language exposure that takes place outside of the home environment, for example, in school. A more pragmatic advantage of using parental questionnaires is that they are relatively fast and easy to administer (Byers-Heinlein et al. 2020), and many are readily available for researchers to use (see Kaščelan et al. 2022, for a review). A disadvantage of using questionnaires, however, is that parents do not always seem to accurately report their child's language input (Bail et al. 2015; Cychosz et al. 2021; Marchman et al. 2017; Richards et al. 2017). For example, parents can underestimate or overestimate their language use due to cultural biases (Heine et al. 2002; Ramírez-Esparza et al. 2008). Cultural differences can influence the response style of participants, such that members of some cultures might be more modest in their response than others (Heine et al. 2000), or some cultures tend to answer more towards the center of a scale than others (Chen et al. 1995). This can pose a problem when studies include participants from diverse cultural backgrounds, which is often the case in bilingualism research. Cultural differences also play a role in the extent to which parents might give socially desirable responses (Lalwani et al. 2006). Parents might under- or overestimate their language use to a socially desired response. Another disadvantage is that questionnaires are unable to measure the absolute amount of child-directed speech, which has been proven to explain additional variance on top of language proportion (Marchman et al. 2017).

1.2. Audio Recordings

Another commonly used method is recording the language environment at home (Bruyneel et al. 2021; Casillas et al. 2020; Levin-Asher et al. 2023; Orena et al. 2020). Measures of the language input via naturalistic recordings can provide precise and ecologically valid information (Bergelson et al. 2019; Cristia et al. 2020; Cychosz et al. 2021; Green-

wood et al. 2011). However, these audio recordings might not fully capture the exposure to language that occurs outside of the home environment. This may specifically impact the accurate measurement of exposure to the majority language, which is the predominant language in society and omnipresent outside of children's home environment. Additionally, the duration of recordings is often limited, capturing only a snapshot of the general language environment of the child (Bergelson et al. 2019). Recently, new technologies have allowed researchers to capture full days of naturalistic audio recordings within the homes of bilingual families (Cristia et al. 2020; Greenwood et al. 2011). However, one full day may not be representative (e.g., if one of the parents is not at home on that day) and does not capture the variability of language exposure between different days. For that reason, Orena et al. (2020) advise recording at least three full days. However, increasing the number of days that parents need to record is more invasive for participants, more labor-intensive for the researcher, and remains a mere snapshot. Moreover, it is conceivable that the increased effort and invasiveness for families can lead to smaller sample sizes (e.g., Casillas et al. 2020; Cychosz et al. 2021; Marchman et al. 2017).

1.3. Comparing the Methods

Studies show that parental reports and audio recordings yield different language input estimates of the same language environment (Cychosz et al. 2021; Marchman et al. 2017). To better understand and improve the methods that are used to measure language input, three studies correlated the language input obtained via parental questionnaires with day-long audio recordings (Cychosz et al. 2021; Marchman et al. 2017; Orena et al. 2020). Below we go over these three studies in more detail.

Cychosz et al. (2021) correlated the amount of child-directed speech (CDS) estimated by parents in a bilingual language use questionnaire with random samples of naturalistic recordings made with LENA from ten Spanish–English bilingual infants. They found a weak and non-significant correlation between the language exposure estimated by parents and the observed exposure in the audio recording.

Marchman et al. (2017) conducted a language background environment interview and gathered audio recordings with LENA from 18 Spanish–English bilingual children. They found a moderately strong positive correlation between the two methods, which neared statistical significance (probably due to a small sample size and low power). Besides correlating the two methods with each other, they also looked at the association between each method individually and standardized language outcomes. Both methods correlated moderately with standardized language outcomes. However, the absolute amount of child-directed speech measured by the day-long audio recordings explained additional variance on top of the variance explained by the relative amount of language exposure as estimated in the questionnaire. This indicates that the absolute quantity of speech plays an important role in language development and should ideally be taken into account when constructing a language input estimate.

Orena et al. (2020) collected three day-long audio recordings made with LENA from 21 French–English bilingual infants and conducted a language environment interview with their caregivers. The estimates from the parent reports were correlated with the observed proportion of bilingual language exposure from the three day-long audio recordings. They found a positive relationship that indicated that parents can reliably indicate their child's proportion of language exposure.

These three studies all correlated the language input estimates from parental reports with the observed language exposure from the audio recordings, but did not compare the methods with each other. The strength of the correlation between parental estimates and observed naturalistic data varied between all three studies. These discrepancies call for a more in-depth comparison, not only in terms of whether and how strongly the two methods correlate with each other but also in terms of how they each correlate with child language outcomes. Marchman et al. (2017) also mention the need to explore new methods that capture the variation in bilingual children's language environments, which is one of the

goals of this study. The finding that language input measured by either method is related to children's language outcomes might suggest that both methods measure relevant aspects of language input. However, these relevant parts might not overlap or only partially overlap. Thus, a combination of both methods, similar to the one suggested in the current study, seems promising.

1.4. The Present Study

Previous research has shown that parental questionnaires and audio recordings yield different estimates of the same language environment (Bail et al. 2015; Cychosz et al. 2021; Marchman et al. 2017), but also that both predict children's language outcomes (Marchman et al. 2017). The aim of the present study is to compare the two individual existing methods and explore a new, combined method. The combined method constructs its estimate of language exposure similarly to other parental questionnaires, i.e., it multiplies the number of hours each person was reported to spend with the child by the proportion of language use by that person. However, these proportions of language use are not estimated by parents but are the proportions observed in a day-long audio recording. This method adds to the ecological validity and reduces possible (cultural) bias (Heine et al. 2002; Ramírez-Esparza et al. 2008). Additionally, the absolute amount of speech was incorporated in the novel language input estimate, as the absolute amount of speech has been claimed to hold higher predictive power for language development than the relative amount (De Houwer 2011; Marchman et al. 2017). Furthermore, many questionnaires calculate the language input quantity as time spent with the child, but the construct *time* has been argued to be a lacking measure of exposure (Carroll 2017). The results of this study contribute to better insights into bilingual language input measures.

To guide this study, we formulated the following research questions:

1. Which language input measure correlates best with the vocabulary scores of bilingual children: parental questionnaires, day-long audio recordings, or a combination of both?
2. Is there a difference in measuring the quantitative language input in the minority language compared to the majority language?

As the quantity of language input relates to children's vocabulary (e.g., Hart and Risley 1995; Huttenlocher et al. 1991; Rowe 2012), we correlated the estimates of language input of all three methods with the child's vocabulary scores in both their languages to investigate which method has the strongest correlation. We hypothesize that the combined method correlates more strongly with vocabulary than audio recordings and parental reports individually do.

Furthermore, children receive input in the majority language outside of the home environment that is difficult to estimate or not captured at all in the language input measures. Although minority language exposure can also occur outside the home environment, children will predominantly receive input in this language at home, in particular in families with a migration background. Thus, the language input measure of the minority language could be a more precise estimate than the one in the majority language. Therefore, we hypothesize that language input measures predict the vocabulary scores in the minority language better than they do in the majority language (Dijkstra et al. 2016; Duursma et al. 2007; Hammer et al. 2009).

2. Materials and Methods

2.1. Participants

The participants were 54 multilingual children, aged between 36 and 72 months, and their families. Families were recruited via schools, (local) events, online calls on social media platforms, and personal networks. At the time of data collection, no child 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 Dutch, the majority language in the Netherlands. The reason these two communities were selected is twofold.

Firstly, the Turkish and Polish communities in the Netherlands are well represented, making it feasible to reach a decent sample size. Secondly, these data form part of a larger research project, for which the typological aspects of Turkish and Polish were specifically relevant.

Two children were excluded from the analysis because they received more than 50% exposure to a third language. Six children were reported to be exposed to a small amount of English as a third language (range: 2–14% exposure, $M = 4%$, $SD = 5%$). These children were not excluded because English is frequently used in games and other media in the Netherlands; for some parents, this may have led to reporting it in the questionnaire. Although two children did not meet our preregistered criterion of 15% exposure to the minority language (see procedure for preregistration details), we decided to include them in the analysis because they demonstrated considerable knowledge of the minority language. Thus, the final sample consisted of 52 multilingual children and their families.

In terms of language dominance, the sample is varied and consists of Dutch-dominant, balanced, and Polish/Turkish-dominant children. Dominance was determined based on current overall language exposure according to the Q-BEx questionnaire. In most families ($n = 31$), children were first exposed to Dutch in the home environment. The other children were first exposed to Dutch at pre-school ($n = 23$), school ($n = 3$), or another location ($n = 1$). All children were first exposed to Turkish or Polish in their home environment. While most families used both Dutch and the heritage language in the home environment at the time of testing ($n = 42$), there were some families that used only the heritage language at home ($n = 10$). The education level of parents was measured as the highest level of education attained between parents and was relatively high in our sample. For the demographic information of the sample, see Table 1.

Table 1. Descriptive statistics of the sample included in the analysis.

	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Minority language				
Polish	21			
Turkish	31			
Age (months)	52	55	10	36–72
Gender				
Female	27			
Male	25			
Language Dominance				
>60% Dutch	24			
40–60% Dutch	11			
<40% Dutch	17			
Parents' educational level				
University degree	44			
Post-secondary school	7			
Secondary school	1			

Note. A child was considered dominant in a language if it receives >60% input in that language (Cattani et al. 2014; Siow et al. 2023).

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 children between 3 and 7 years old (Haman et al. 2017) and a valid measure of vocabulary (Van Wonderen and Unsworth 2021). The task consists of 4 parts, each comprising 32 items testing: (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 receptive and expressive parts was counterbalanced. We decided to calculate separate scores for expressive and receptive vocabulary because these represent two different modalities and skills. The

younger children in our sample may perform at the lower end for expressive vocabulary because expressive vocabulary is more challenging than receptive vocabulary (Gershkoff-Stowe and Hahn 2013). The older children, on the other hand, may perform at the ceiling for receptive vocabulary. Having both an expressive and a receptive vocabulary enables us to cover a wider age range. In addition, it allows us to explore potentially differential relations between receptive and expressive modality and input. Nouns and verbs were collapsed to increase the number of items and increase variation within each modality. These sum scores represent the number of correct items (range: 0–64). Accuracy was based on a list of target words, but we decided to make a few adaptations to the target list and include more responses as correct. After consultation with at least two native speakers per language (Dutch, Polish, and Turkish, as spoken in the Netherlands (Doğruöz and 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 *scheppen* (to shovel) in addition to *graven* (to dig) for the image of a man digging a hole/shoveling dirt).

2.2.2. Language Environment Measures

Parental Questionnaires: Quantifying Bilingual Experience (Q-BEx)

We measured the language input with the questionnaire for Quantifying Bilingual Experience (Q-BEx; De Cat et al. 2022). The Q-BEx is a modular questionnaire that has been developed to function as an all-encompassing tool in the field of bilingualism. Two modules are fixed, namely *background information* and *risk factors*, while other modules are optional for the researcher, including, but not limited to, *language proficiency* or *language mixing*. The questionnaire is currently available in 25 different languages, allowing it to be broadly used in many different bilingual communities. Its construction was informed through a Delphi study (Kaščelan et al. 2022), where researchers, parents, and professionals in education and health care indicated what a questionnaire about children's language environment should contain.

For this study, we used the module *language exposure and use* that provides current proportions of exposure for the home situation, at school, in the community, on holidays, and in total. The questionnaire obtains current exposure estimates by asking parents questions such as "Think about a typical week in the current year. At home, how often does [person] use each language when speaking to the child?" for each interlocutor in the household and other contexts (at school, in the community, with friends). Further, parents fill out a schedule with whom the child spends their time during regular days, irregular days, weekend days, and holidays. The proportions of exposure are weighted by multiplying the hours each interlocutor spends with the child by the estimated language proportions of that person. If time is spent with multiple people at once, the time is distributed evenly (e.g., if the child spends 8 h with both parents, it is divided into 4 h with one parent and 4 h with the other). Language exposure during holidays is questioned separately, as the language exposure often shifts when families visit their home country. The questionnaire's algorithm sums the weighted input during regular weeks and holiday weeks to the amount of input in a full year. Finally, the total amount of hours per language in a year is translated into a proportion. Summarizing, the weighted proportions of language exposure from the Q-BEx in our sample represent the general input children received in Dutch ($n = 52$), Turkish ($n = 31$), Polish ($n = 21$), and English ($n = 6$ as a third language) during the past year up to the moment of filling out the questionnaire. The sum of the weighted proportional language estimates for each participant is equal to one (e.g., 0.73 Dutch, 0.20 Turkish, and 0.07 English).

Day-Long Audio Recording: Language Environment Analysis (LENA)

The language environment was also recorded with the Language Environment Analysis (LENA) software (v3.5.0). The LENA recorder is a small device that can record up to 16 h and is worn by the child inside the special pocket of a custom-made shirt. The audio recordings gathered with the LENA system are long and in a naturalistic setting, which

is beneficial for finding stronger effects between language input estimates and language outcomes (Anderson et al. 2021).

All parents were instructed to use the LENA recorder during a weekend day because during the weekdays some children may go to daycare or school. Parents of children who did not attend daycare or school also used the LENA on a weekend day. Since parents were shown to be consistent in their language use across weekdays and weekends within a short period of time (Orena et al. 2020), we assumed that recordings during weekend days would be representative of their general language use. Parents were also instructed to start recording when their child wakes up, put the recorder in the designated t-shirt, and resume the rest of the day as if it were a typical day. The recorder automatically switched off when it was full (after 16 h). In case parents did not wish for some conversations to be heard, they were given the option to have parts of the audio removed before the data were processed and analyzed by emailing the research team which times should be deleted. One family used this option and requested to have two hours of audio removed. The recordings were 14 h on average (range: 4.2–16, *SD* = 3.76).

For reasons of feasibility, the data were sampled. Based on common practice (Orena et al. 2020; Marasli and Montag 2023; Ramírez-Esparza et al. 2014), we first removed silent fragments. Then, to portray the language input during an entire day, we sampled segments that represented periods of high, medium, and low interaction by using the conversational turns for 5-min fragments generated by the LENA software. We selected 18 5-min segments (1.5 h) that contain the most conversational turns, 18 segments that contain the lowest number of conversational turns (but are not silent), and 18 segments that are in the middle (p.c. LENA Foundation). From those 5-min segments, we analyzed every other 30-s segment (Marasli and Montag 2023; Ramírez-Esparza et al. 2014, 2017), reducing the number of audio segments to 270 30-s segments per participant. This is considered more than sufficient to reliably reflect the full day-long audio recording (Cychosz et al. 2021; Marasli and Montag 2023).

Subsequently, we coded the 30-s segments manually for the speaker(s), language(s) spoken, activity, and whether there is speech directed to the target child (CDS). If a segment contained more than one language, it was determined by the coder in a separate column which of these languages occurred most frequently, or whether both languages occurred equally.

The coding is exemplified in Table 2. The full coding manual can be found on the Open Science Framework (OSF; <https://osf.io/x953/>). Coders were two bilingual Turkish–Dutch and three bilingual Polish–Dutch research assistants. The inter-rater reliability between assistants of the same language was determined based on the average Kappa scores over the columns *Speaker*, *Language*, *Dominance*, and *CDS* of one participant (270 segments). For both Polish–Dutch ($\kappa = 0.81$) and Turkish–Dutch ($\kappa = 0.82$), very strong inter-rater reliability was obtained.

Table 2. The coding system of the 30-s segments.

Subject	Segment	Speaker	Language	Dominance	CDS	Activity
00001	1	SIB, FAT	Dutch	Dutch	ODS	Mealtime
00001	2	CHI, FAT	Dutch, Polish	Polish	CDS-ADULT	Mealtime
00001	3	CHI, FAT	Dutch	Dutch	CDS-ADULT	Mealtime
00001	4	CHI, FAT, MOT	Dutch, Polish	Both	CDS-ADULT	Playtime

Note. Abbreviations: FAT = father, MOT = mother, SIB = sibling, CHI = target child, CDS-ADULT = child-directed speech by an adult, and ODS = other directed speech.

The proportion of language input in each language was calculated based on the segments that contain speech that is directed to the target child (CDS) (Marchman et al. 2017; Ramírez-Esparza et al. 2014, 2017; Weisleder and Fernald 2013). The number of segments in a language (*Dominance* column) was divided by the total number of segments that contain CDS, resulting in a proportional score for each language, similar to the Q-BEx. If a segment contains equal amounts of speech in both languages, half a segment will be

assigned to each language. For example, Table 2 contains 1.5 segments of child-directed speech in Dutch and 1.5 segments of child-directed speech in Polish.

Combined Method: LENA and Q-BEx

Table 3 provides a schematic representation of how the three different methods calculate language input. The combined method differs from the Q-BEx in three ways.

Table 3. Schematic representation of how the different methods calculate language input.

Q-BEx	LENA	Combined Method
Step 1 (Q-BEx)	Step 1 (LENA)	Step 1 (Q-BEx Data)
Determine who spends time with the child during a regular week and holiday week.	Code the sampled recording (270 30-s segments) for speaker, language, and child-directed speech.	Determine who spends time with the child during a regular week.
Step 2 (Q-BEx)	Step 2 (LENA)	Step 2 (LENA data)
Parents estimate the proportion of languages used by each interlocutor with the child.	Remove segments that do not contain child-directed speech.	Code the sampled recording (270 30-s segments) for speaker, language, and child-directed speech, and calculate the proportion of observed language use in each unique speaker context.
Step 3 (Q-BEx)	Step 3 (LENA)	Step 3 (LENA data)
Multiply the time spent by each interlocutor with their estimated language use. This is achieved automatically by the Q-BEx interface.	Divide the number of segments in one language by the total number of segments.	Calculate the mean adult word count (AWC) per waking hour to quantify the amount of speech the child receives during the recorded day.
Step 4 (Q-BEx and LENA)		
Multiply the time spent in each context with the quantity of speech (mean AWC) and the observed language use.		

First, instead of using estimated proportions of language use, this method uses observed language use from the LENA recording to prevent bias (Heine et al. 2002; Ramírez-Esparza et al. 2008). We calculated current language exposure by multiplying the number of hours each person was reported to spend with the child by the proportion of language use observed by that person in day-long recordings. If any interlocutor was missing from the recording, their information was extracted from the questionnaire.

Second, instead of dividing time equally between interlocutors when time is spent with multiple people at once, the combined method considers these environments as separate contexts with unique patterns of language exposure. The following example will illustrate why this is relevant. Consider a family in which the mother is a native Dutch speaker but has a high proficiency in Turkish and the father is a native Turkish speaker. When alone with the child, the mother speaks Dutch, and the father speaks Turkish. Their shared language of communication within the family is Turkish. The Q-BEx would divide a full day spent with both parents equally between mother and father, which would result in an estimate of 50% Dutch and 50% Turkish, whilst their language use in these contexts likely contains a lot more Turkish given the fact that Turkish is their shared language of communication. Using the *Speaker* codes, the combined method included the observed language use for segments with multiple speakers. In total, the combined method distinguishes nine possible contexts of unique combinations of speakers (mother; father; mother and sibling; father and sibling; mother and father; mother and father and sibling; sibling; school; community). More information about these contexts can be found in the detailed calculations of the combined method on OSF (<https://osf.io/t9mnb>). See step 2 in Table 3.

Third, unlike questionnaires, the combined method incorporates the absolute amount of speech, which is known to vary greatly between families (Weisleder and Fernald 2013). The absolute amount has been claimed to hold higher predictive power for language outcomes than a relative amount (De Houwer 2011; Marchman et al. 2017; Orena et al. 2020) and may differ quite crucially from the relative amount. Two children might, for example, both receive 60% of Dutch language input, but for one child, this might equal 5000 words, whilst for the other, this could equal 10,000 words. The LENA software provides us with automatic output on the adult word count (AWC), which has proven to correlate with manually coded AWC scores in Dutch (Bruyneel et al. 2021) and also in bilingual settings (Orena et al. 2019). When computing AWC scores, segments containing sleep were filtered out based on the automated classifier for periods of sleep (Bang et al. 2022). The mean AWC per hour was calculated from the remaining hours that the child is awake (step 3 in Table 3); this value is used to quantify language input in the home environment. The mean AWC per hour was multiplied by the hours of input the child receives in each language during a full week (step 4 in Table 3).

Where the Q-BEx and LENA methods both end up with one proportional value per language, the combined method results in a quantitative measure of how many words the child hears during a regular week in each language.

2.3. Procedure

The data used for this study is part of the larger project Children and Language Mixing: developmental, psycholinguistic, and sociolinguistic aspects (CALM). The project has been approved by the Ethics Committee of Utrecht University (FETC20-0291). Data were collected between July 2022 and July 2023. The study was preregistered on the Open Science Framework in September 2023 (<https://osf.io/ajgh6>). Data and scripts can be found in the Supplementary Material on the project page on the Open Science Framework (<https://osf.io/xc953/>).

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. Parents received the LENA recorder with instructions to use it on a weekend day before the second test appointment. To safeguard the privacy of the participants, only fragments of 30 s were listened to. Consequently, the researcher was unable to gain an understanding of 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 given the opportunity to have parts of the audio removed before it would ever be listened to. One family made use of this option and 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 task to the child, retrieved the LENA recorder, and filled out the Q-BEx questionnaire together with a parent in their preferred language (Dutch, Polish, Turkish, or English). As the data from this study are part of a larger project, additional tests were administered during home visits, but these are not discussed in this paper.

2.4. Data Analysis

All variables were first regressed on age to ensure that any variance in the correlation between quantitative language input and vocabulary scores cannot be ascribed to age. All further calculations were made with age-residualized measures. We then correlated the language input estimates from each method (questionnaires, recordings, and combined method) with the expressive and receptive vocabulary scores in both the majority language (Dutch) and minority language (Polish/Turkish). These correspond to correlations 1 to 6 in

Figure 1. For convenience, the correlations are displayed only once, but correlations 1 to 6 are present in both the majority and minority languages.

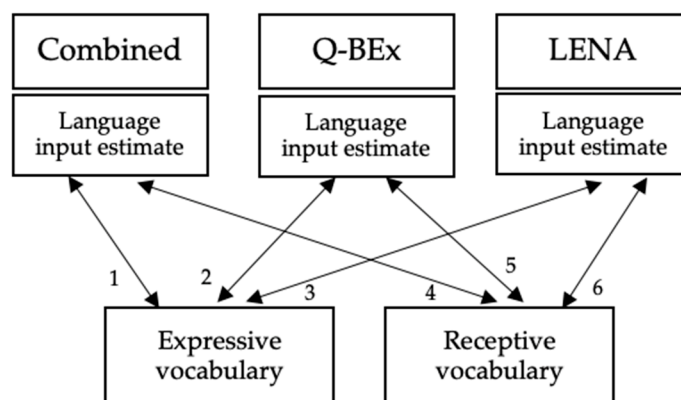


Figure 1. Schematic overview of the correlations between variables. All depicted variables have been regressed by age. All variables are separately available for the majority language (Dutch) and the minority language (Polish/Turkish).

Next, the correlations between each method and the expressive and receptive vocabulary scores were compared with each other in pairs (combined versus Q-BEx, combined versus LENA, and Q-BEx versus LENA) using the *cocor* package (Diedenhofen and Musch 2015) in R (R Core Team 2020). Thus, for both the majority language and the minority language, correlations 1, 2, and 3 (as seen in Figure 1) were compared with each other, as were correlations 4, 5, and 6. As we hypothesized that the combined method would outperform the other two methods, the comparison of these correlations was one-sided ($1 > 2$, $1 > 3$, $4 > 5$, and $4 > 6$). Because we had no hypothesis with respect to a difference between the methods of audio recordings (LENA) and questionnaires (Q-BEx) (Cychosz et al. 2021; Marchman et al. 2017; Orena et al. 2020), their comparison was tested two-sided ($2 \neq 3$, $5 \neq 6$).

Finally, we hypothesized that all methods would be better at estimating the input in the minority language than in the majority language. This was tested with Pearson and Filon's *z*-test (Pearson and Filon 1898) from the *cocor* package. We compared the correlations of each method with the expressive and receptive vocabulary scores in the minority language to the ones in the majority language, e.g., correlation 1 (as shown in Figure 1) in the minority language with correlation 1 in the majority language, etc. All comparisons were tested one-sided, with the hypothesis that the correlations in the minority language would be stronger than those in the majority language. Additionally, given the possible differences between receptive and expressive vocabulary (Gershkoff-Stowe and Hahn 2013), we exploratively compared correlations between the input measures and expressive and receptive vocabulary scores. The alpha level for all analyses was set at 0.05.

3. Results

The results' section is organized as follows: First, descriptive results are provided for the language input measures and vocabulary scores. Second, correlations between language input estimates and vocabulary outcomes of each individual method are presented, followed by comparisons of the three methods. Third, we compare correlations in the minority versus the majority language. Lastly, we compare correlations between the language input estimates and expressive and receptive vocabulary scores as an exploratory analysis.

3.1. Descriptive Results

In Table 4, it can be observed that the Q-BEx and LENA methods yield different proportional estimates, with the LENA method showing higher estimates for the minority language. The combined method shows somewhat balanced estimates for the majority

and minority languages, with more variation in the minority languages. Children’s receptive vocabulary scores are similar in the majority and minority languages, but children’s expressive vocabulary scores are slightly higher in the minority language.

Table 4. Descriptive results for the different language input measures and vocabulary scores in both the majority language (Dutch) and minority language (Polish/Turkish).

Variable	Majority Language (Dutch)				Minority Language (Polish/Turkish)			
	M	SD	min	max	M	SD	min	max
Language input								
Q-BEx	0.52	0.21	0.14	0.96	0.47	0.21	0.04	0.86
LENA	0.36	0.34	0.00	0.97	0.63	0.34	0.03	1.00
Combined method	54,202	34,916	7377	145,115	62,651	43,786	1321	182,797
Combined proportion	0.48	0.28	0.09	0.98	0.52	0.28	0.02	0.91
Vocabulary scores								
Expressive	22.6	11.2	1	44	25.8	17.0	0	55
Receptive	50.4	10.2	16	62	50.1	11.0	16	64

Note. Abbreviations: *M* = mean, *SD* = standard deviation, *min* = minimum, *max* = maximum, Q-BEx = Quantifying Bilingual Experience (De Cat et al. 2022), and LENA = Language Environment Analysis. The language input estimates of the Q-BEx and LENA are proportions; the language input estimate of the Combined Method reflects the absolute amount of words per week. These word counts were converted to a proportion to facilitate interpretation and comparison to the other two methods, but further calculations were conducted with the absolute amount of words per week (row 3).

3.2. Individual Methods

The correlations for both individual methods and the combined method are shown in Table 5. Their corresponding plots are in Appendix A.

Table 5. Correlations between each method and the expressive and receptive vocabulary scores.

Method	Majority Language (Dutch)				Minority Language (Polish/Turkish)			
	Expressive		Receptive		Expressive		Receptive	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Q-BEx	0.66	<0.001	0.35	0.01	0.74	<0.001	0.56	<0.001
LENA	0.62	<0.001	0.43	0.001	0.62	<0.001	0.44	0.001
Combined method	0.49	<0.001	0.28	0.04	0.59	<0.001	0.48	<0.001

Note. *r* = Pearson’s correlational strength, *p* = *p*-value.

All correlations between the three methods and the expressive and receptive vocabulary scores in both the majority and minority languages were significant. The three methods correlated moderately to strongly with the expressive vocabulary scores in the majority language, weakly to moderately with the receptive vocabulary scores in the majority language, moderately to strongly with the expressive vocabulary scores in the minority language, and moderately with the receptive scores in the minority language. Thus, all measures correlate significantly and positively with children’s vocabulary scores.

3.3. Comparing the Methods

The results of the comparisons are presented in Table 6. All one-sided comparisons between the combined method and Q-BEx and LENA are non-significant. Comparisons between LENA and Q-BEx also did not reach significance. The comparison for expressive vocabulary scores in the minority language between the Q-BEx questionnaire (*r* = 0.74) and LENA recording (*r* = 0.62) neared significance, *t*(49) = 1.76, *p* = 0.08. Nonetheless, the effect size for the difference between the two correlations was small, *d* = 0.17. In sum, no significant differences were found between the methods in terms of their correlation with vocabulary scores.

Table 6. *p*-values of the comparisons between the correlations of each method.

Hypotheses	Majority Language (Dutch)		Minority Language (Polish/Turkish)	
	Expressive	Receptive	Expressive	Receptive
Combined > Q-BEx	0.96	0.71	0.96	0.79
Combined > LENA	0.95	0.94	0.63	0.35
LENA ≠ Q-BEx	0.60	0.37	0.08	0.16

3.4. Comparing Majority and Minority Languages

Table 7 presents the comparison of the correlations between the methods and vocabulary scores in the majority and minority languages. Even though none of the comparisons reached significance, there is a trend showing that the correlations with receptive vocabulary are stronger in the minority language than in the majority language for the Q-BEx method ($z = 1.41, p = 0.08$) and for the combined method ($z = 1.26, p = 0.10$). A z -value larger than one implies that the correlation with receptive vocabulary in the minority language is more than one standard deviation higher than the correlation with receptive vocabulary in the majority language. Statistical significance would have been reached at a z -value larger than 1.645. Thus, these two methods might be slightly better at predicting receptive vocabulary in the minority language as compared to the majority language.

Table 7. z - and p -values of the comparisons of the correlations of each method in the majority language (Dutch) and minority language (Polish/Turkish).

Method	Expressive Vocabulary		Receptive Vocabulary	
	z	p	z	p
Q-BEx	0.92	0.18	1.41	0.08
LENA	-0.04	0.52	0.05	0.48
Combined	0.78	0.21	1.26	0.10

Note. All comparisons were tested one-sided, with the hypothesis that r minority > r majority. As the groups in the comparison are independent (majority vs. minority), Pearson and Filon’s z (Pearson and Filon 1898) was used.

3.5. Exploratory Analysis

The correlations in Table 5 suggest that the strength of the relationship between the different language input measures is stronger for expressive vocabulary compared to receptive vocabulary. To examine whether this is indeed the case, an exploratory analysis has been conducted. Again, the *cocor* package was used to compare the overlapping correlations based on dependent groups. The results are presented in Table 8 and confirm that correlations are significantly stronger for expressive vocabulary than for receptive vocabulary, except for the combined method in the minority language, where the comparison did not reach statistical significance. All effect sizes are small, ranging between $d = 0.12$ and $d = 0.33$.

Table 8. The comparisons between the correlations of each method with the expressive and receptive vocabulary scores in the majority (Dutch) and minority language (Polish/Turkish).

Method	Majority Language (Dutch)			Minority Language (Polish/Turkish)		
	t	p	Cohen’s d	t	p	Cohen’s d
Q-BEx	3.31	<0.001	0.33	2.50	0.008	0.26
LENA	1.94	0.03	0.19	2.21	0.02	0.22
Combined	1.91	0.03	0.19	1.23	0.11	0.12

Note. All comparisons were tested one-sided, with the hypothesis that r expressive > r receptive. As the groups in the comparison are dependent, William’s t (Williams 1959) was used.

4. Discussion

In this study, we set out to investigate how different methods that measure bilingual children’s language input correlated with expressive and receptive vocabulary scores

in the majority language (Dutch) and minority language (Polish/Turkish). The results showed that language input estimates correlated significantly with children's vocabulary outcomes, regardless of how input was estimated, which language was measured (majority or minority language), or the modality in which vocabulary size was measured (receptive or expressive). Thus, parental questionnaires, day-long audio recordings, and a combination of these two methods all produce valid measures of (bilingual) language input quantity.

4.1. Comparisons between Language Input Quantity Measures

The Q-BEx questionnaire and LENA recordings produced somewhat different estimates of majority versus minority language input quantity (see Table 4). This difference might be explained by the fact that LENA only measures language input in the home environment. As a consequence, input that occurs outside of the home environment, which is likely more often in the majority language, is not taken into account. The Q-BEx does include language input outside of the home environment in its estimation. This may explain why the estimated proportion of majority language input quantity is larger. Although the estimates of the different measures varied, their correlations with children's vocabulary scores did not significantly differ from each other. Contrary to our expectations, the combined method did not correlate more strongly with expressive nor receptive vocabulary scores than the Q-BEx questionnaire or LENA recording did. One of the hypothesized advantages of the combined method was the incorporation of the absolute amount of language input, which has been found to better predict language outcomes than a relative amount (De Houwer 2011; Marchman et al. 2017; Orena et al. 2020). In our study, we used the average adult word count (AWC) during waking hours, which was calculated from the automatically generated metric based on the day-long LENA recording, as a measure of input quantity. However, this measure might not have been representative of the average input during an entire week. Even though parents have been reported to be consistent in their proportional language use across days (Orena et al. 2020), this might not hold for their absolute quantity of speech. Weekdays might differ from weekend days in many ways; parents can, for example, be more distracted, tired, and busy with work-related tasks during the week and more relaxed and available for engagement with their child during the weekend. Therefore, generalizing the average adult word count recorded in one day-long recording might have created a distorted overview of the quantity of language input during an entire week.

4.2. Minority Language versus Majority Language Vocabulary

We hypothesized that estimates of the majority language input would be less accurate than estimates of the minority language input. This is due to the fact that children receive a relatively large amount of input in the majority language from sources outside of their home environment, which complicates tracking the input quantity. Correlations between receptive vocabulary scores and measures of input quantity indeed suggested somewhat stronger correlations for the minority language compared to the majority language, confirming findings in previous research (Dijkstra et al. 2016; Duursma et al. 2007; Hammer et al. 2009). Caution is warranted, however, because comparisons of correlations between input and child vocabulary outcomes in the minority language versus the majority language did not reach statistical significance. The Q-BEx and combined method did show a trend in the expected direction with receptive vocabulary scores. Previous studies that did find significant differences between the effect of minority language input and majority language input (Dijkstra et al. 2016; Duursma et al. 2007; Hammer et al. 2009) had larger sample sizes ($n = 72-96$) and thus more power. It could be speculated that our trends might reach significance with a larger sample.

For the reasons mentioned above, we expected parents to be better at estimating the language input in the minority language than in the majority language, and that is indeed what we observe in the Q-BEx and combined method. A possible reason why this relation is not found for the LENA method may be due to LENA's limited measuring environment.

The children who participated in our study received minority language input mainly at home. By only measuring the home environment and ignoring language input outside of the home environment, LENA might lead to an overrepresentation of minority language input. Indeed, Table 4 shows that the LENA method yielded higher proportions for the minority language than the Q-BEx and combined method, which both do take into account language exposure outside of the home environment. In addition, the LENA does not capture variation between children in their minority language exposure outside the home, in contrast to the Q-BEx and combined method.

Summarizing, audio recordings made with LENA provide ecologically valid and detailed data about the home environment. They enable investigating specific input properties, such as the absolute amount of input (Marchman et al. 2017) and comparisons across cultures (Bergelson et al. 2023). However, they may be less optimal for representing children's general language environment.

4.3. Expressive versus Receptive Vocabulary

In the exploratory analysis, we found that language input estimates generally correlate more strongly with expressive vocabulary scores than with receptive vocabulary scores, except for the combined method in the minority language, which found no difference. A similar pattern was found by Dijkstra et al. (2016) for bilingual Frisian–Dutch children. In their study, expressive vocabulary scores in both the majority language (Dutch) and the minority language (Frisian) were predicted by the home language input, whilst receptive vocabulary scores in the majority language were not.

Gibson et al. (2014) also studied the effects of language exposure on the gap between expressive and receptive semantic knowledge in bilingual Spanish–English pre-kindergarten children. They refer to the *Weaker Links hypothesis* (Gollan et al. 2008) to explain why exposure has a larger impact on children's expressive language skills compared to their receptive language skills. This hypothesis suggests that phonological representations of words (i.e., the mental representations of the combinations of sounds that comprise words in a specific language) become stronger as a function of more experience, i.e., more input and opportunities to speak. More experience would enable children to gradually break down whole-word representations into segmental representations, resulting in increasingly detailed phonological representations (Gibson et al. 2014). By implication, limited experience in a certain language leads to relatively weak phonological representations in that language. Importantly, while weak phonological representations may be sufficient for recognizing words, they will hamper the production of words because of distinct processes of lexical access in comprehension and production (Gollan et al. 2011). Therefore, the quantity of language input might impact the level of expressive vocabulary more than the level of the receptive vocabulary.

4.4. Limitations and Further Research

The fairly small sample size might have resulted in a lack of power to detect significant relationships. Additionally, we were unable to explore potential differences between the Polish and Turkish bilingual communities in the Netherlands, which may reflect cultural biases and differences (Heine et al. 2002; Ramírez-Esparza et al. 2008).

Moreover, we only used LENA recordings on weekend days. Due to European privacy regulations, we could not record outside of the home environment. This might have resulted in a biased estimate of AWC and an overestimation of the proportion of use of the minority language. Most studies have used the LENA to investigate the home language environment of infants (Marchman et al. 2017; Orena et al. 2019; Ramírez-Esparza et al. 2014, 2017). For many infants, this can be either a weekday or a weekend day, and studies often ask parents to record one of each. As the children in our sample are older and most go to daycare or school during the week, this was not feasible. To stay consistent between families, we asked all parents in our sample to record a weekend day. Future studies might try to also use recordings outside the home environment to evaluate whether they

significantly improve the estimates of language input. Despite these limitations, the current study provides us with rich data regarding the multilingual input of the children. The combination of methods enabled us to thoroughly study the usefulness of each method and the combination of both.

Future studies could investigate whether the average AWC found in one day-long recording corresponds to the average AWC found in an entire week of day-long recordings.

5. Conclusions

In this study, we have taken a closer look into different methods that measure the bilingual language input of 3- to 5-year-old children. We examined the characteristics of questionnaires and day-long audio recordings and proposed a combined method to overcome several shortcomings of the individual methods. Contrary to our hypothesis, we did not find that the combined method correlated more strongly with vocabulary outcomes than questionnaires or recordings alone. No significant difference was found between the questionnaire and the recordings. Importantly, all methods correlated significantly with expressive and receptive vocabulary scores in the majority and minority languages and can thus be deemed reliable methods to measure bilingual language input quantity. In all cases, except for the combined method in the minority language, language input quantity correlated more strongly with expressive vocabulary scores than receptive vocabulary scores, suggesting that the quantity of input in this age range is more important for language production than language comprehension. Finally, although all three methods are reliable methods to measure bilingual language input, it is important to bear in mind that both questionnaires and recordings have their own advantages and shortcomings and may serve different purposes. The findings of this study provide important insights for work on bilingual language input and can be used to guide future methodological choices as well as inspire future studies seeking to optimize measures of bilingual language input.

Supplementary Materials: The following supporting information can be downloaded at: <https://osf.io/xc953/> and <https://www.mdpi.com/article/10.3390/languages9070231/s1>, The project page on the Open Science Framework contains: detailed description combined method, coding manual, and data analysis scripts.

Author Contributions: Conceptualization, E.V., M.v.W., O.O.-P. and E.B.; methodology, E.V.; formal analysis, E.V.; investigation, E.V. and research assistants.; data curation, E.V.; writing—original draft preparation E.V.; writing—review and editing, E.V., M.v.W., O.O.-P. and E.B.; visualization, E.V.; supervision, M.v.W., O.O.-P. and E.B.; funding acquisition, E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Dutch Research Council (NWO), grant number VI.C.191.042.

Institutional Review Board Statement: The study was approved by the Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University (filed under number 23-0362, valid through 30 November 2027).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset used for the analyses in this study is openly available on the Open Science Framework at: <https://osf.io/xc953/>. The raw data behind the dataset is not readily available because the data are part of an ongoing study. Requests to access the original data should be directed to the corresponding author.

Acknowledgments: We would like to thank all families who participated and Laura Koelma, Vera Snijders, Gülşah Yazıcı, Hatice Bulut, Fatma Nur Öztürk, Klaudia Latkowska, Patricia Dworak, and Zuzanna Kruber for their help with collecting and coding the data.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Correlations between Q-BEx language input estimates and vocabulary scores

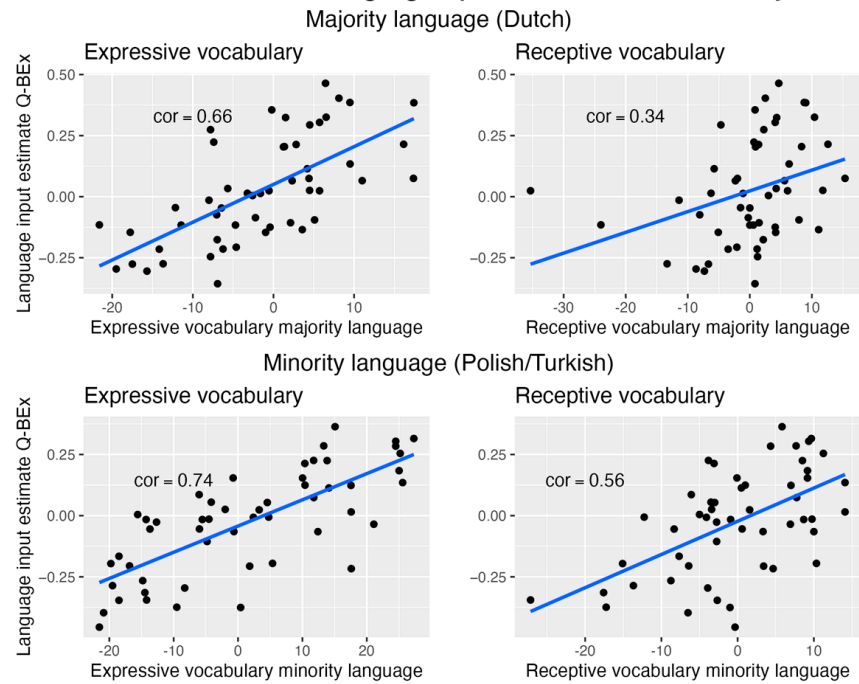


Figure A1. Scatter plots of the correlations between the language input estimates from the Quantifying Bilingual Experience (Q-BEx) questionnaire and receptive and expressive vocabulary in the majority language (Dutch) and minority language (Polish/Turkish).

Correlations between LENA language input estimates and vocabulary scores

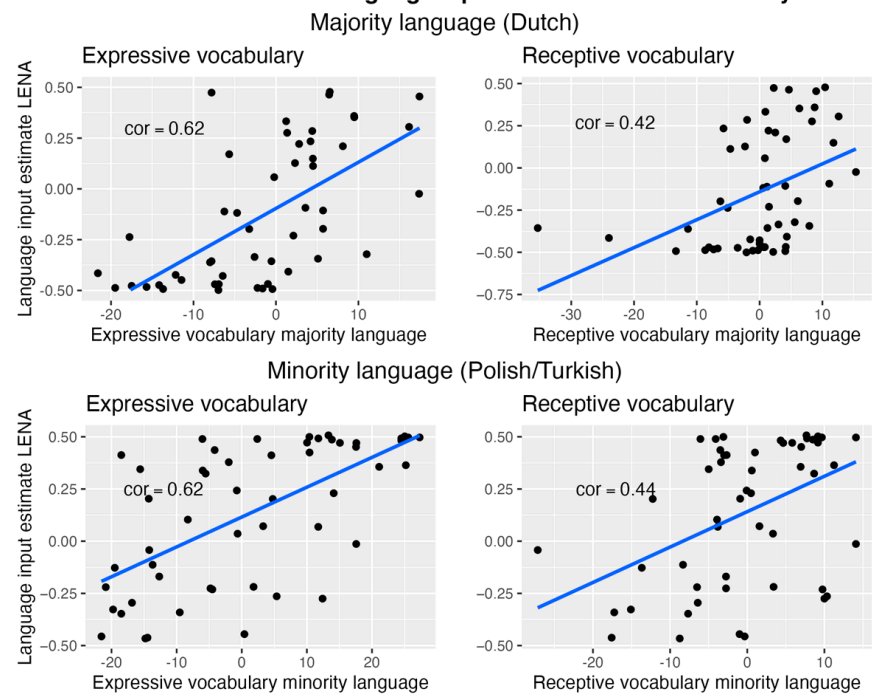


Figure A2. Scatter plots of the correlations between the language input estimates from the Language Environment Analysis (LENA) questionnaire and receptive and expressive vocabulary in the majority language (Dutch) and minority language (Polish/Turkish).

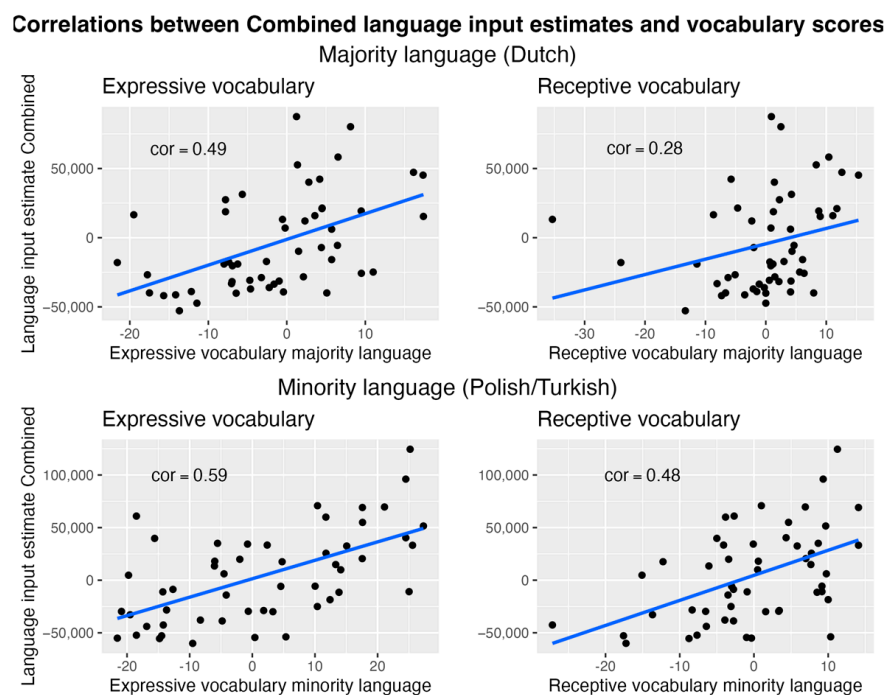


Figure A3. Scatter plots of the correlations between the language input estimates from the combined method and receptive and expressive vocabulary in the majority language (Dutch) and minority language (Polish/Turkish).

References

- Anderson, Nina. J., Susan A. Graham, Heather Prime, Jennifer M. Jenkins, and Sheri Madigan. 2021. Linking quality and quantity of parental linguistic input to child language skills: A meta-analysis. *Child Development* 92: 484–501. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bail, Amelie, Giovanna Morini, and Rochelle S. Newman. 2015. Look at the gato! Code-switching in speech to toddlers. *Journal of Child Language* 42: 1073–101. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bang, Janet Y., George Kachergis, Adriana Weisleder, and Virginia A. Marchman. 2022. *An Automated Classifier for Child-Directed Speech from LENA Recordings*. Somerville: Cascadilla Press. [\[CrossRef\]](#)
- Bergelson, Erika, Marisa Casillas, Melanie Soderstrom, Amanda Seidl, Anne S. Warlaumont, and Andrei Amatuni. 2019. What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science* 22: e12724. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bergelson, Erika, Melanie Soderstrom, Iris-Corinna Schwarz, Caroline F. Rowland, Nairán Ramírez-Esparza, Lisa R. Hamrick, Ellen Marklund, Marina Kalashnikova, Ava Guez, Marisa Casillas, and et al. 2023. Everyday language input and production in 1001 children from six continents. *Proceedings of the National Academy of Sciences of the United States of America* 120: 52. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bijeljac-Babic, Ranka, Josette Serres, Barbara Höhle, and Thierry Nazzi. 2012. Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS ONE* 7: e30843. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bruyneel, Eva, Ellen Demurie, Sofie Boterberg, Petra Warreyn, and Herbert Roeyers. 2021. Validation of the Language ENvironment Analysis (LENA) system for Dutch. *Journal of Child Language* 48: 765–91. [\[CrossRef\]](#) [\[PubMed\]](#)
- Byers-Heinlein, Krista. 2013. Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition* 16: 32–48. [\[CrossRef\]](#)
- Byers-Heinlein, Krista, Esther Schott, Ana Maria Gonzalez-Barrero, Melanie Brouillard, Daphnée Dubé, Amel Jardak, Alexandra Laoun-Rubenstein, Meghan Mastroberardino, Elizabeth Morin-Lessard, Maria P. Tamayo, and et al. 2020. MAPLE: A multilingual approach to parent language estimates. *Bilingualism: Language and Cognition* 23: 951–57. [\[CrossRef\]](#)
- Carbajal, Maria Julia, and Sharon Peperkamp. 2020. Dual language input and the impact of language separation on early lexical development. *Infancy* 25: 22–45. [\[CrossRef\]](#)
- Carroll, Susanne E. 2017. Exposure and input in bilingual development. *Bilingualism: Language and Cognition* 20: 3–16. [\[CrossRef\]](#)
- Casillas, Marisa, Penelope Brown, and Stephen C. Levinson. 2020. Early language experience in a Tzeltal Mayan village. *Child Development* 91: 1819–35. [\[CrossRef\]](#)
- Cattani, Allegra, Kirsten Abbot-Smith, Rafalla Farag, Andrea Krott, Frédérique Arreckx, Ian Dennis, and Caroline Floccia. 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: 649–71. [\[CrossRef\]](#)
- Chen, Chuansheng, Shin-Ying Lee, and Harold W. Stevenson. 1995. Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science* 6: 170–75. [\[CrossRef\]](#)

- Cristia, Alejandrina, Federica Bulgarelli, and Elika Bergelson. 2020. Accuracy of the language environment analysis system segmentation and metrics: A systematic review. *Journal of Speech, Language, and Hearing Research* 63: 1093–105. [CrossRef] [PubMed]
- Cychosz, Margaret, Anele Villanueva, and Adriana Weisleder. 2021. Efficient estimation of children's language exposure in two bilingual communities. *Journal of Speech, Language, and Hearing Research* 64: 3843–66. [CrossRef] [PubMed]
- De Cat, Cécile, Draško Kaščelan, Philippe Prévost, Ludovica Serratrice, Laurie Tuller, and Sharon Unsworth. 2022. *Quantifying Bilingual Experience (Q-BEx): Questionnaire Manual and Documentation*. Charlottesville: Open Science Framework. [CrossRef]
- De Houwer, Annick. 2009. *Bilingual First Language Acquisition*. Bristol: Multilingual Matters, vol. 2. [CrossRef]
- De Houwer, Annick. 2011. Language input environments and language development in bilingual acquisition. *Applied Linguistics Review* 2: 221–40. [CrossRef]
- Diedenhofen, Birk, and Jochen Musch. 2015. cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE* 10: e0121945. [CrossRef]
- Dijkstra, Jelske, Folkert Kuiken, René J. Jorna, and Edwin L. Klinkenberg. 2016. The role of majority and minority language input in the early development of a bilingual vocabulary. *Bilingualism: Language and Cognition* 19: 191–205. [CrossRef]
- Doğruöz, A. Seza, and Ad Backus. 2010. Turkish in the Netherlands: Development of a new variety? In *Language Contact: New Perspectives*. Edited by Muriel Norde, Bob de Jonge and Cornelius Hasselblatt. Amsterdam: John Benjamins Publishing, vol. 28, pp. 87–102. [CrossRef]
- Duursma, Elisabeth, Silvia Romero-Contreras, Anna Szuber, Patrick Proctor, Catherine Snow, Diane August, and Margarita Calderón. 2007. The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Applied Psycholinguistics* 28: 171–90. [CrossRef]
- Gathercole, Virginia C. Mueller, and Enlli Môn Thomas. 2009. Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition* 12: 213–37. [CrossRef]
- Gershkoff-Stowe, Lisa, and Erin R. Hahn. 2013. Word comprehension and production asymmetries in children and adults. *Journal of Experimental Child Psychology* 114: 489–509. [CrossRef] [PubMed]
- Gibson, Todd A., Elizabeth D. Peña, and Lisa M. Bedore. 2014. The relation between language experience and receptive-expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism* 17: 90–110. [CrossRef] [PubMed]
- Gollan, Tamar H., Rosa I. Montoya, Cynthia Cera, and Tiffany C. Sandoval. 2008. More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language* 58: 787–814. [CrossRef] [PubMed]
- Gollan, Tamar H., Timothy J. Slattery, Diane Goldenberg, Eva Van Assche, Wouter Duyck, and Keith Rayner. 2011. Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General* 140: 186. [CrossRef] [PubMed]
- Greenwood, Charles R., Kathy Thiemann-Bourque, Dale Walker, Jay Buzhardt, and Jill Gilkerson. 2011. Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly* 32: 83–92. [CrossRef]
- Haman, Ewa, Magdalena Łuniewska, and Barbara Pomiechowska. 2015. Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters, pp. 196–240. [CrossRef]
- Haman, Ewa, Magdalena Łuniewska, Pernille Hansen, Hanne Gram Simonsen, Shula Chiat, Jovana Bjekić, Agnè Blažienė, Katarzyna Chyl, Ineta Dabašinskiė, Sharon Armon-Lotem, and et al. 2017. Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics* 31: 818–43. [CrossRef]
- Hammer, Carol Scheffner, Megan Dunn Davison, Frank R. Lawrence, and Adele W. Miccio. 2009. The effect of maternal language on bilingual children's vocabulary and emergent literacy development during Head Start and kindergarten. *Scientific Studies of Reading* 13: 99–121. [CrossRef] [PubMed]
- Hart, Betty, and Todd R. Risley. 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Towson: Paul H Brookes Publishing.
- Heine, Steven J., Darrin R. Lehman, Kaiping Peng, and Joe Greenholtz. 2002. What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology* 82: 903. [CrossRef] [PubMed]
- Heine, Steven J., Toshitake Takata, and Darrin R. Lehman. 2000. Beyond self-presentation: Evidence for self-criticism among Japanese. *Personality and Social Psychology Bulletin* 26: 71–78. [CrossRef]
- Hoff, Erika. 2003. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development* 74: 1368–78. [CrossRef] [PubMed]
- Hoff, Erika. 2006. How social contexts support and shape language development. *Developmental Review* 26: 55–88. [CrossRef]
- Hoff, Erika, and Cynthia Core. 2013. Input and language development in bilingually developing children. In *Seminars in Speech and Language*. New York: Thieme Medical Publishers, vol. 34, No. 04. pp. 215–26. [CrossRef]
- Hoff, Erika, Cynthia Core, Sylvia Place, Rosario Rumiche, Melissa Señor, and Marisol Parra. 2012. Dual language exposure and early bilingual development. *Journal of Child Language* 39: 1–27. [CrossRef] [PubMed]
- Huttenlocher, J., W. Haight, A. Bryk, M. Seltzer, and T. Lyons. 1991. Early vocabulary growth: Relation to language input and gender. *Developmental Psychology* 27: 236. [CrossRef]
- Kaščelan, Draško, Philippe Prévost, Ludovica Serratrice, Laurie Tuller, Sharon Unsworth, and Cécile De Cat. 2022. A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition* 25: 29–41. [CrossRef]

- Lalwani, Ashok K., Sharon Shavitt, and Timothy Johnson. 2006. What is the relation between cultural orientation and socially desirable responding? *Journal of Personality and Social Psychology* 90: 165. [CrossRef] [PubMed]
- Levin-Asher, Bonnie, Osnat Segal, and Liat Kishon-Rabin. 2023. The validity of LENA technology for assessing the linguistic environment and interactions of infants learning Hebrew and Arabic. *Behavior Research Methods* 55: 1480–95. [CrossRef] [PubMed]
- Lieven, Elena, Vibeke Grover, Paola Uccelli, and Meredith L. Rowe. 2019. Input, interaction and learning in early language development. In *Learning through Language: Towards an Educationally Informed Theory of Language Learning*. Cambridge: Cambridge University Press, pp. 19–30. [CrossRef]
- Marasli, Zeynep, and Jessica L. Montag. 2023. Optimizing random sampling of daylong audio. Paper presented at the Annual Meeting of the Cognitive Science Society, Sydney, Australia, July 26–29, vol. 45. No. 45.
- Marchman, Virginia A., Lucía Z. Martínez, Nereyda Hurtado, Theres Grüter, and Anne Fernald. 2017. Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science* 20: e12425. [CrossRef] [PubMed]
- Orena, Adriel John, Krista Byers-Heinlein, and Linda Polka. 2019. Reliability of the language environment analysis recording system in analyzing French–English bilingual speech. *Journal of Speech, Language, and Hearing Research* 62: 2491–500. [CrossRef] [PubMed]
- Orena, Adriel John, Krista Byers-Heinlein, and Linda Polka. 2020. What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental Science* 23: e12901. [CrossRef] [PubMed]
- Pearson, Barbara Z., Sylvia C. Fernandez, Vanessa Lewedeg, and D. Kimbrough Oller. 1997. The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics* 18: 41–58. [CrossRef]
- Pearson, Karl, and Louis Napoleon George Filon. 1898. VII. Mathematical contributions to the theory of evolution.—IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 191: 229–311. [CrossRef]
- Place, Sylvia, and Erika Hoff. 2011. Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development* 82: 1834–49. [CrossRef]
- Place, Sylvia, and Erika Hoff. 2016. Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition* 19: 1023–41. [CrossRef]
- Potter, Christine E., Eva Fourakis, Elizabeth Morin-Lessard, Krista Byers-Heinlein, and Casey Lew-Williams. 2019. Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science* 22: e12794. [CrossRef] [PubMed]
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available online: <https://www.r-project.org/> (accessed on 1 August 2023).
- Ramírez-Esparza, Nairán, Adrián García-Sierra, and Patricia K. Kuhl. 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: 880–91. [CrossRef] [PubMed]
- Ramírez-Esparza, Nairán, Adrián García-Sierra, and Patricia K. Kuhl. 2017. The impact of early social interactions on later language development in Spanish–English bilingual infants. *Child Development* 88: 1216–34. [CrossRef] [PubMed]
- Ramírez-Esparza, Nairán, Samuel D. Gosling, and James W. Pennebaker. 2008. Paradox lost: Unraveling the puzzle of Simpatía. *Journal of Cross-Cultural Psychology* 39: 703–15. [CrossRef]
- Richards, Jeffrey A., Jill Gilkerson, Dongxin Xu, and Keith Topping. 2017. How much do parents think they talk to their child? *Journal of Early Intervention* 39: 163–79. [CrossRef]
- Rowe, Meredith L. 2012. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development* 83: 1762–74. [CrossRef]
- Siow, Serene, Nicola A. Gillen, Irina Lepädatu, and Kim Plunkett. 2023. Double it up: Vocabulary size comparisons between UK bilingual and monolingual toddlers. *Infancy* 28: 1030–51. [CrossRef]
- Unsworth, Sharon. 2016. *Quantity and Quality of Language Input in Bilingual Language Development*. Washington, DC: American Psychological Association. [CrossRef]
- Unsworth, Sharon, Vicky Chondrogianni, and Barbora Skarabela. 2018. Experiential measures can be used as a proxy for language dominance in bilingual language acquisition research. *Frontiers in Psychology* 9: 1809. [CrossRef]
- Van Wonderen, Elise, and Sharon Unsworth. 2021. Testing the validity of the Cross-Linguistic Lexical Task as a measure of language proficiency in bilingual children. *Journal of Child Language* 48: 1101–25. [CrossRef]
- Weisleder, Adriana, and Anne Fernald. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science* 24: 2143–52. [CrossRef] [PubMed]
- Williams, Evan J. 1959. The comparison of regression variables. *Journal of the Royal Statistical Society: Series B (Methodological)* 21: 396–99. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.