DOI: 10.1111/infa.12562

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

Double it up: Vocabulary size comparisons between UK bilingual and monolingual toddlers

Serene Siow[1,2] | **Nicola A. Gillen**[1] | **Irina Lepădatu**[1] | **Kim Plunkett**[1]

[1]Department of Experimental Psychology, University of Oxford, Oxford, UK

[2]Department of English, Linguistics & Theatre Studies, National University of Singapore, Singapore, Singapore

Correspondence
Serene Siow, Department of English, Linguistics & Theatre Studies, National University of Singapore, Singapore 119077, Singapore.
Email: s.siow@nus.edu.sg

Funding information
Economic and Social Research Council, Grant/Award Number: CQR01830

Abstract

We compared vocabulary sizes in comprehension and production between bilingual toddlers growing up in the United Kingdom (UK) and age-matched UK English monolinguals (12–36 months old) using parent-report vocabulary questionnaires. We found that bilingual toddlers' vocabulary sizes in English were smaller than the vocabulary sizes of their monolingual peers. Notably, this vocabulary gap was not found when groups were compared on conceptual vocabulary in comprehension. Conceptual scoring also reduced the vocabulary gap in production but group differences were still significant. Bilingual toddlers knew more words than monolinguals when words across their two languages were added together, for both comprehension and production. This large total vocabulary size could be attributed to a high proportion of doublets (cross-linguistic word pairs with the same meaning) in bilinguals' vocabularies. These findings are discussed in relation to language exposure, facilitation from cross-linguistic overlap and maturation constraints on vocabulary size.

1 | INTRODUCTION

For bilingual toddlers living in communities that predominantly speak one language (e.g., the United Kingdom [UK], which uses English), it is important for them to acquire the community language which would be widely used outside the home. As schools in such countries typically use the majority

language as the language of instruction, proficiency in the majority language can have repercussions for school achievement. A study by Howard et al. (2014) showed that Spanish-speaking bilingual children's English vocabulary size in spoken production (as tested using a picture naming task) was positively associated with their English reading proficiency, even after accounting for the effect of socioeconomic status and amount of English exposure. A report by Strand et al. (2015) analyzing the England National Pupil Database in 2013 indicated that the percentage of students in England classified to be learning English as an additional language (EAL) was 16.2%. As a group, EAL students were identified by Strand et al. to have lower rates of academic achievement compared to students with English as their first language when tested at the end of their first year of schooling. This lag decreased over the years of schooling, with EAL students catching up to their peers by age 16. This narrowing of the gap with years of schooling gives rise to the possibility that bilingual children's school-readiness could be boosted by consistent language exposure starting from a younger age. In monolinguals, larger vocabulary size and faster speed of word recognition tested at 25 months of age have been linked to better expressive vocabulary, IQ and working memory at 8 years old (Marchman & Fernald, 2008). Studying the early vocabulary development of children in their first 3 years of life, when their early language skills are rapidly developing, can help us better understand the potential sources of divergences for bilingual students who do or do not lag behind their monolingual peers.

1.1 | Single-language vocabulary, conceptual vocabulary and total vocabulary

The literature on bilinguals' vocabulary size has generally found a bilingual delay when comparing vocabulary in a single language between monolinguals and bilinguals. This group difference has been found in receptive vocabulary tasks across toddlers (Cattani et al., 2014), school-age children (Bialystok et al., 2010) and adults (Bialystok & Luk, 2012). It has also been found for productive vocabulary as tested using a picture naming task with toddlers (Cattani et al., 2014; Hoff & Ribot, 2017) and school-age children (Yan & Nicoladis, 2009). Differences of toddlers' vocabulary size between groups have also been found using parent-report vocabulary questionnaires for both comprehension and production (Cattani et al., 2014; Floccia et al., 2018; Vagh et al., 2009). Nevertheless, in some cases, comprehension and production may not show identical results. Yan and Nicoladis (2009) found that while word comprehension performance was comparable, school-age bilinguals performed significantly poorer than monolingual peers in production. Additionally, there may also be an age-related change in the size of the vocabulary gap. De Houwer et al. (2014) found no significant difference in the receptive vocabulary sizes in Dutch of monolingual toddlers learning Dutch and bilingual toddlers learning Dutch and French at 13 months, but monolinguals knew significantly more Dutch words at 20 months.

The size of vocabulary differences between monolinguals and bilinguals is partially dependent on the amount of exposure the bilinguals receive for each language. Single language vocabulary size has been found to be positively correlated with the relative amount of exposure the child has to that language (Cattani et al., 2014; Hoff et al., 2012; Pearson et al., 1997). English-dominant bilingual toddlers (i.e., toddlers who hear more English than their other language in their day-to-day lives) have been found to display larger English vocabulary than Spanish-dominant bilinguals (Pearson et al., 1993; Vagh et al., 2009). Further supporting the effect of language exposure, Pearson et al. (1993) also found that while Spanish-dominant bilinguals had smaller English vocabulary, they had larger Spanish vocabulary than English-dominant peers. Notably, Pearson et al. (1993) found that the vocabulary sizes of bilingual children were comparable to that of monolinguals in their dominant language. This effect was also found by Cattani et al. (2014)—2.5-year-old bilingual toddlers with at

least 60% English exposure performed equally well as English monolingual peers on various language measures in English. Measures of single language vocabulary are therefore easily affected by language exposure, and paint a limited picture if only one language is tested.

As such, proficiency in the majority language alone is insufficient for characterizing bilingual development. When studying vocabulary knowledge in bilinguals, the method of calculating vocabulary size is important as it can produce varied results. Alternative measures that have been used by researchers are *conceptual vocabulary size* and *total vocabulary size*. Conceptual vocabulary size counts any concepts that bilinguals know, regardless of the language they are known in. A child is said to know a concept if they understand the word in one language or both. Total vocabulary size sums the vocabulary sizes in both languages. Both measures present estimates of bilinguals' vocabulary that are more representative of bilinguals' language development than single-language vocabulary size, as they take into account vocabulary in both languages. Conceptual scoring has been noted to bring school-age bilingual's vocabulary into normal monolingual range for both comprehension and production (Gross et al., 2014). Pearson et al. (1993) found that the conceptual vocabulary sizes and total vocabulary sizes of Spanish–English bilingual toddlers were similar to the vocabulary sizes of English monolinguals for production, and in fact were significantly larger than monolinguals' vocabulary sizes for comprehension. Comparable conceptual vocabulary sizes between groups have also been found by Junker and Stockman (2002). Core et al. (2013) found conceptual vocabulary sizes that were still smaller than monolinguals, but total vocabulary sizes that were comparable to those of monolinguals. The gap between monolinguals and bilinguals increased for conceptual vocabulary between the age of 22–30 months, but total vocabulary size remained similar between groups. Total vocabulary sizes are also commonly used to measure bilingual vocabulary while minimizing the variability of language exposure. Total vocabulary sizes of bilinguals have typically been found to be larger than those of monolinguals (Bosch & Ramon-Casas, 2014; De Houwer et al., 2014; Junker & Stockman, 2002), or at least comparable (Core et al., 2013).

1.2 | Translation equivalents

In addition to calculations of vocabulary sizes, we can also explore the interconnectivity of bilinguals' two languages by studying the overlap in bilinguals' vocabulary in their two languages. Words that share the same meaning across languages are known as *translation equivalents*. When a child knows both words in a translation equivalent pair, they are considered to know a *doublet*. This is in contrast to translation equivalent pairs where the child knows only one word of the pair, otherwise referred to as a *singlet*. Research with bilingual children's early vocabulary has found doublets in the receptive vocabulary of toddlers as young as 13 months old (De Houwer et al., 2006) and in productive vocabulary by 16 months old (Legacy et al., 2017). Investigations into the acquisition of translation equivalents can shed light on the mechanisms underlying bilingual vocabulary acquisition.

Various proposals have been suggested for the acquisition pattern of translation equivalents in comparison to new concepts. A Neutral Account posits that translation equivalents are not specially advantaged—the acquisition of the translation equivalent to a known word is not different from the acquisition of a word for a new concept. Pearson et al. (1995) provided support for a Neutral Account, showing that the number of doublets known by bilingual children in their sample was comparable to the number of doublets expected when random children were matched together. Another account is the Preference Account, where toddlers find it easier to learn translation equivalents over new concepts. In contrast to Pearson et al.'s findings, Bilson et al. (2015) found that bilinguals in their sample showed significantly more translation equivalents in their vocabulary than

would be expected if the two languages were learnt independently, thus supporting the Preference Account. A recent study by Tsui et al. (2022) suggested that there may be developmental differences in the acquisition of translation equivalents based on the child's vocabulary size. While toddlers with larger vocabulary sizes (300 words or more) showed patterns that followed predictions of the Neutral Account, those with smaller vocabulary sizes showed patterns supporting a Preference Account.

Language exposure can also have an effect on the proportion of doublets observed in bilinguals' vocabulary. Bilinguals who receive balanced exposure to both languages are expected to have similar vocabulary sizes in both languages. Words that are learnt earlier in one language are also often learnt earlier in other languages (Łuniewska et al., 2016). Therefore, the words known in each language by bilinguals who have balanced vocabulary sizes are likely to have high conceptual overlap, regardless of whether this acquisition of doublets is incidental or due to a preference toward learning translation equivalents (Pearson et al., 1995; Tsui et al., 2022).

1.3 | The present study

This study investigated whether bilingual toddlers growing up in the UK have similar or smaller vocabulary sizes compared to monolinguals of the same age. This research complements previous work with UK bilingual samples by Cattani et al. (2014) and Floccia et al. (2018), with the advantage of a wider tested age range. We compared vocabulary acquisition trajectories between English-learning monolinguals and bilinguals growing up in the UK aged 12–36 months old, comparing cross-sectional data for comprehension and production collected using vocabulary questionnaires. As the UK is a predominantly English-speaking community, the development of English proficiency is important for both monolingual and bilingual toddlers' long-term communicative and academic outcomes. We were also interested in the extent to which the degree of English exposure a child receives influences their English vocabulary size. Additionally, we investigated whether bilinguals and monolinguals have comparable vocabulary sizes when measured using conceptual vocabulary and total vocabulary. We also calculated the proportion of doublets, to study the interactions between the two languages. A key strength of the present study is the use of a standardized set of words across both comprehension and production and across the full tested age range, allowing us to make more robust inferences about the relationship between comprehension and production. Finally, we ran an exploratory analysis comparing the proportion of doublets in bilinguals' vocabulary against the proportion found when monolinguals with comparable vocabulary sizes were randomly paired. If bilinguals know more doublets than expected by chance (operationalized as randomly-paired monolinguals), this would provide support for the Preference Account of translation equivalent acquisition. In contrast, the Neutral Account predicts that translation equivalents are not privileged over new concepts, and so the number of doublets known by bilinguals should be no different from chance.

Bilingual children growing up in the UK face a situation where their Additional Language (AL) is not widely spoken in the community. For bilinguals growing up in communities where both languages are widely spoken and their languages have equal social prestige, bilinguals may perform more similarly to monolingual peers (Smithson et al., 2014). However, when there is a dominant community language, parents can find it difficult to maintain consistent input in the minority language for their child (see O'Toole & Hickey, 2017 for a report on Irish–English bilinguals). This has impact on both the quantity and quality of language exposure that bilingual toddlers receive. By studying the language development of bilinguals growing up in the UK, we can learn more about the parallel acquisition of two languages with unequal usage in the community.

2 | METHODS

2.1 | Participants

The present study was conducted according to guidelines laid down in the Declaration of Helsinki. Written informed consent was obtained from a parent or caregiver for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Medical Sciences Research Ethics Board at the University of Oxford, with reference number R60939/RE009.

Participants were recruited via advertisements on social media (restricted to England) and via email to families in the lab's database. We excluded participants reported to be premature (gestation weeks of less than 34 weeks), have hearing problems or have diagnosed language delay. The demographics of bilingual and monolingual participants are summarized in Table 1.

2.1.1 | Bilingual

The bilingual toddlers in our sample had at least 20% exposure to English and 20% exposure to their AL. We excluded participants who reported hearing 10% or more of a third language. We required at least one parent to be a native speaker of the AL. We also required at least one parent to have fluent English proficiency (self-rated proficiency of at least 7 out of 10). Finally, to obtain reliable estimates of vocabulary size, we excluded an additional 56 children whose parents expressed uncertainty about their ability to report their child's English vocabulary (e.g., due to not speaking English at home). The ALs spoken by families in our bilingual sample (group sizes in brackets) were Dutch ($N = 29$), French ($N = 81$), German ($N = 46$), Italian ($N = 90$), Polish ($N = 92$), Portuguese ($N = 48$) and Spanish ($N = 147$). Bilingual families were offered the option of a £5 Amazon voucher or a child-sized t-shirt as remuneration.

On average, the sampled bilingual toddlers were slightly skewed to have more exposure to English than to their AL, with mean English exposure of 54.7% (SD = 17.6). This is unsurprizing given that English is the predominant language spoken in the community. These toddlers typically had a higher proportion of English exposure overall than at home. The mean home English language exposure was 37.0% (SD = 22.5). The majority ($N = 465$) of the sample were simultaneous bilinguals who had at least 10% (and no more than 90%) English language exposure at home. There were 39 toddlers with no English exposure at home, and another 29 who received some English exposure but less than 10%.

In the final sample, 173 families reported that both parents were native speaker of the AL; 301 families reported that one parent was a native speaker of the AL and one parent was a native speaker of

TABLE 1 Participant groups.

	Bilingual	Monolingual
N	533	257
N female	275	101
N male	258	156
Age in months	12.0–36.0 (mean 24.1)	12.5–35.9 (mean 25.0)
Country of residence	England	England
Language requirements	≥20% English and ≥20% AL	Only English
Data collection period	2020–2022	2020–2022

English; 2 families had parents who were both bilingual; 21 families had one bilingual parent and one English-native parent; 11 families had one bilingual parent and one AL-native parent; and 25 families had one AL-native parent and one parent who was native in a third language. Additionally, 46.9% of sampled toddlers had spent at least 1 month immersed in a country where their AL is spoken widely in the community. Of the toddlers who had language immersion experience, the average time spent was 3.43 months (SD = 3.88).

2.1.2 | Monolingual

Monolingual participants completed the questionnaires as part of other studies. A subset ($N = 180$) participated in an online study testing the utility of a touchscreen receptive vocabulary task (Gillen et al., 2021). These families participated without remuneration. The remaining monolingual participants ($N = 77$) completed the questionnaires as part of their participation in a lab-based experiment. These participants were offered £5 for travel expenses and a child-sized t-shirt as appreciation for their participation. Questionnaires were sent to parents via email to be completed online at home. The questionnaires were completed prior to any additional tasks. As such, we do not expect that participation in either of the two aforementioned studies would have significant impact on the analyses conducted in this paper.

2.2 | Procedure

Questionnaires were administered online using Qualtrics (2022). Parents received a link to the questionnaires that they could complete at home. The questionnaires could be paused and resumed, but would time out if not completed within 1 week. The questionnaires, anonymized data and analysis scripts can be found in our OSF repository (https://osf.io/w8ke5/).

2.2.1 | Demographics questionnaire

All parents first answered some questions relating to their demographics (toddlers' date of birth, each parent's highest level of education, any languages spoken at home besides English, and whether the toddler had any diagnosed language delay, hearing problems or was born premature).

We used mother's highest education level as a proxy for socioeconomic status. Only entries where information on mother's education level was available were included in the analysis. Education level was converted into a numerical score, with 0—No qualifications; 1—Left school at 16 with GCSE or equivalent; 2—Left school at 18 with A-Levels or equivalent; 3—University degree or equivalent. Overall, mothers' educational level in the sample was high, with 89.5% of mothers in the bilingual sample and 93.4% of mothers in the monolingual sample having a university degree or equivalent.

2.2.2 | Language exposure questionnaire

Bilingual parents (who reported that their child had regular exposure to English and one additional language) went on to complete a simplified version of the Language Exposure Questionnaire (LEQ) developed by Bosch and Sebastián-Gallés (2001). The LEQ is a parental report questionnaire used to

obtain a summary of each toddlers' language environment. It includes questions about each parent's native language, their language proficiency, which languages(s) they usually use when speaking to the toddler, whether or not the toddler is attending nursery (and the language used as nursery), and whether the toddler has spent time immersed in a country where their AL is widely spoken. We also asked parents to give an estimate of the percentage of English, AL and any third language that their child is exposed to in their daily life overall. Parents were also asked about the percentages of these languages heard specifically in the home. Monolingual parents (who reported that their child had no consistent exposure to any other languages besides English) were not administered the LEQ.

Following the demographics questions and LEQ (if applicable), parents were directed to the appropriate version of the vocabulary questionnaires according to their toddler's language background.

2.2.3 | Vocabulary questionnaire

Data on vocabulary knowledge in English, for both the bilingual and monolingual groups, was collected using the Oxford Communicative Development Inventory (CDI) (Hamilton et al., 2000), a questionnaire containing 418 words commonly known to British toddlers. Parents indicated for each word whether their child understands and says, understands but does not say, or does not understand the word. The utility of CDIs to evaluate vocabulary development in toddlers has been supported by studies showing good congruence between parent-reported vocabulary and toddlers' performance on vocabulary tasks for both monolinguals (Gillen et al., 2021; Styles & Plunkett, 2009) and bilinguals (Marchman & Martínez-Sussmann, 2002; Vagh et al., 2009). Parents of bilingual toddlers also completed an adaptation of the Oxford CDI in their AL (also 418 words). Adaptations were created for each of the ALs collected in this study (Dutch, French, German, Italian, Polish, Portuguese and Spanish). While normed versions of the CDI exist in these languages, they vary considerably in length and also have variable amounts of overlapping concepts with the Oxford CDI. We chose to use adaptations of the Oxford CDI as this allowed us to have a high level of conceptual overlap for analyses of conceptual vocabulary size. We worked with native speakers of each AL, who translated the Oxford CDI and replaced words that were not relevant to the target language—for example, "penny" was replaced with "coin" in most languages. If an English word had multiple translations, we listed the different translations as one entry. We also compared the translations to normed adaptations of the MacArthur-Bates CDI in those languages, using the same words where possible—to give an example, "lorry/truck" was listed as "Lastwagen/Laster," following the German MCDI adaptation.

English vocabulary size, conceptual vocabulary size and total vocabulary size were calculated using the concepts that overlap across all the adaptations (365 out of 418 words), excluding words for onomatopoeia. A monolingual child was coded as knowing a concept if they knew the English word for the concept. A bilingual child was coded as knowing a concept if they knew the English word, the word in their AL, or both.

2.3 | Sample subset for vocabulary in comprehension

The Oxford CDI, like other vocabulary questionnaires that use a standardized list across different ages, is prone toward floor and ceiling effects. Toddlers, especially those with larger vocabularies, are likely to also know words that are not included in the CDI. Ceiling effects are problematic as it generates a plateau of scores at the older ages that is not reflective of the actual developmental trajectory. It is therefore crucial to specify a cut-off that minimizes the possibility of scores reaching ceiling. Floor effects constitute a similar problem, creating a plateau at the youngest ages.

To avoid floor and ceiling effects, analyses on vocabulary size in comprehension used a subset of participants from the original 12–36-month-old sample. We defined the ceiling as 90% of the maximum conceptual vocabulary size as calculated from overlapping concepts in our questionnaires, and the floor as 10% of the maximum conceptual vocabulary size. We excluded age groups that had median vocabulary sizes in comprehension smaller than the floor value and those larger than the ceiling value. The bilingual group and the monolingual group had their thresholds at different ages—the bilingual group's accepted range was 12–26 months; the monolingual group's accepted range was 14–25 months. We applied the more conservative age range for both groups, leaving a sample subset of 14–25-month-old toddlers for analyses of vocabulary size in comprehension. This subset included 141 monolingual toddlers (55 female; mean age 21.2 months) and 242 bilingual toddlers (125 female; mean age 19.6 months).

2.4 | Sample subset for vocabulary in production

We similarly aimed to avoid floor and ceiling effects for vocabulary size in production. As toddlers typically learn to produce words later than they learn to understand them, the age at which toddlers' vocabulary sizes hit the floor and the ceiling are likely to differ for comprehension and production, thus requiring different cut-offs. The ceiling for production was defined as 90% of the maximum conceptual vocabulary size, and the floor as 10% of the maximum conceptual vocabulary size. The bilingual group's accepted range was 20–36 months; the monolingual group's accepted range was 19–29 months. Again, we applied the more conservative age range. The sample subset for analyses of vocabulary size in production was 20–29-months-old. This subset included 151 monolingual toddlers (60 female; mean age 23.9 months) and 223 bilingual toddlers (113 female; mean age 24.8 months).

3 | RESULTS

We first ran linear regressions (separately for comprehension and production) with vocabulary size as the dependent variable. Language group (monolingual or bilingual), age, gender and mother's highest education level (numerical score) were predictors. Age was centered on the mean and scaled by standard deviation for comprehension and production separately. The reference level for language group was "Monolingual." The reference level for gender was "Male." An interaction between age and language group was also included, to test if differences between monolingual and bilingual groups changed with age. This analysis was done for English vocabulary size, conceptual vocabulary size and total vocabulary size. The model is defined in R (R Core Team, 2013) as below:

$$\mathbf{lm}(\text{vocabulary\_size} \sim \text{age} + \text{gender} + \text{mother\_education} + \text{group} + \text{age} : \text{group})$$

To test the gradient effect of language exposure within the bilingual group, we ran a second series of linear regressions with vocabulary size as the dependent variable for bilingual participants only. Language exposure, age, gender and mother's highest education level were predictors. Language exposure was a continuous variable derived from the percentage of English exposure the child received. This percentage was centered on 50% and scaled by standard deviation. Monolingual participants were not included in this analysis as their inclusion would result in a skewed distribution where a large proportion of the sample has 100% English exposure.

As bilinguals who receive balanced input in their two languages were predicted to have a higher proportion of doublets than those with English-dominant or AL-dominant input, the quadratic term for language exposure was also included as a predictor. The quadratic term is the squared values of

scaled English exposure. As scaled English exposure has a normal distribution with both positive and negative values (range −1.7 to 1.7), the quadratic conversion would result in a U-shaped curve. A significant linear term would indicate that vocabulary size increases or decreases linearly with an increase of English exposure. Meanwhile, a significant quadratic term would indicate that the effect of English exposure can be represented as a U-shaped curve in the data.

The second model is defined as below:

$$\begin{aligned}\textbf{lm}(\text{vocabulary\_size} \sim\ & \text{age} + \text{gender} + \text{mother\_education} \\ & + \text{English\_exposure} + \textbf{I}(\text{English\_exposure}^2))\end{aligned}$$

Model residuals showed that assumptions for linear regression were met for all models. There was no evidence of non-linear relationships between predictors and the outcome variable that had not already been included.

3.1 | Comprehension (14–25-month-old)

For visualization purposes, we split the bilingual sample into three groups by their language exposure: 83 English-dominant bilinguals (61%–80% English exposure, mean = 72.0, SD = 4.66), 84 Balanced bilinguals (41%–60% English exposure, mean = 54.8, SD = 5.43) and 75 AL-dominant bilinguals (20%–40% English exposure, mean = 31.2, SD = 7.35).

3.1.1 | English vocabulary

The relationship between age, language group and English vocabulary size in comprehension is visualized in Figure 1. As predicted, we observe a difference between the vocabulary trajectories of monolinguals and bilinguals, with bilinguals of all three levels of language exposure having smaller vocabulary sizes in English compared to monolinguals of the same age. The bilingual sample had on average smaller vocabulary sizes than same-age monolinguals for English vocabulary ($t = -6.260$, $p < .001$). The interaction between group and age was not significant ($t = -0.265$, $p = .791$).

When testing the effect of English language exposure as a continuous variable within the bilingual group, we see the expected significant positive linear effect of English exposure on English vocabulary size (Table 2). The quadratic term was also significant, indicating that balanced bilinguals were more similar to English-dominant bilinguals than to AL-dominant bilinguals.

3.1.2 | Conceptual vocabulary

We then studied the relationship between age, language group and conceptual vocabulary size. In Figure 2, which visualizes the relationship for conceptual vocabulary in comprehension, we see that the difference between bilinguals and monolinguals has largely disappeared. The bilingual sample had comparable vocabulary sizes in comprehension with same-age monolinguals when vocabulary size was calculated using conceptual vocabulary ($t = -1.392$, $p = .165$). There was also no significant interaction between group and age ($t = -0.318$, $p = .750$).

The effect of language exposure was not significant for predicting conceptual vocabulary size in comprehension within the bilingual group (Table 3).

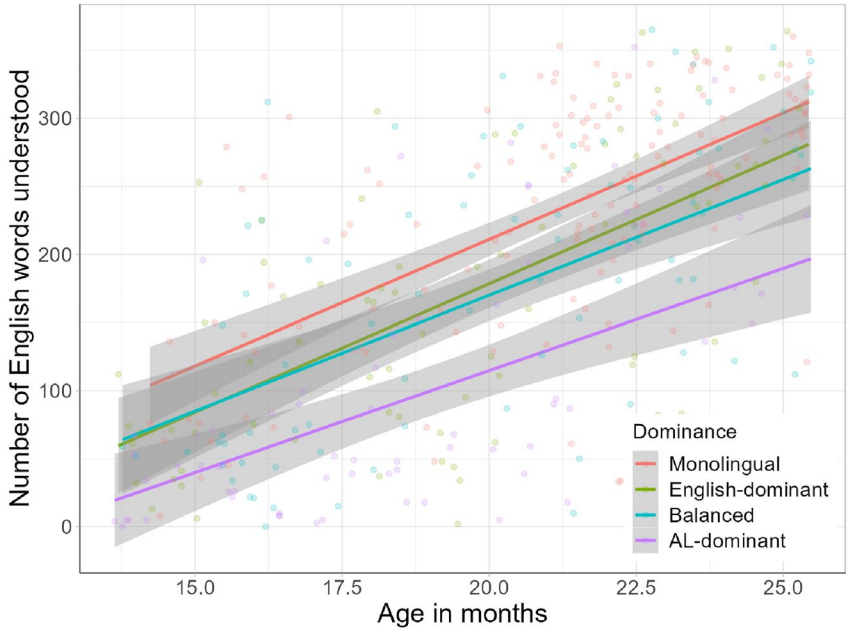

FIGURE 1 Scatterplot of English vocabulary size in comprehension against toddler's age, split by language dominance groups.

TABLE 2 Linear model for bilinguals' English vocabulary size in comprehension, with age, gender, mother's education and English language exposure as predictors.

Predictor	Estimate	Std error	*t*	*p*
(Intercept)	172.7	37.6	4.59	<.001
Age	57.7	5.25	11	<.001
Gender	1.92	10.6	0.182	.856
Mother's education	−1.23	12.6	−0.097	.922
English exposure	27.2	5.45	4.99	<.001
English exposure (quadratic)	−15.4	5.89	−2.61	.00973

3.1.3 | Total vocabulary and doublets

Analyses of the proportion of doublets in bilingual toddlers' vocabulary showed that bilingual toddlers have a large number of words that they understand in both languages. While the exact ratio varies greatly (range 0%–99.7%), on average participants understood both translation equivalents of 48.6% (SD = 24.4%) of the concepts they knew. As a result of this high rate of doublets, bilingual toddlers know a large number of words when their two languages are added together, exceeding the vocabulary size of monolinguals of the same age. This difference was significant ($t = 7.642$, $p < .001$). Additionally, there was also a significant interaction between group and age ($t = 3.443$, $p < .001$), where the difference between groups increased with age.

Additionally, the effect of language exposure was significant for predicting total vocabulary size in comprehension within the bilingual group, both for the linear term and the quadratic term (Table 4). This indicates that English-dominant bilinguals had larger total vocabulary sizes than AL-dominant

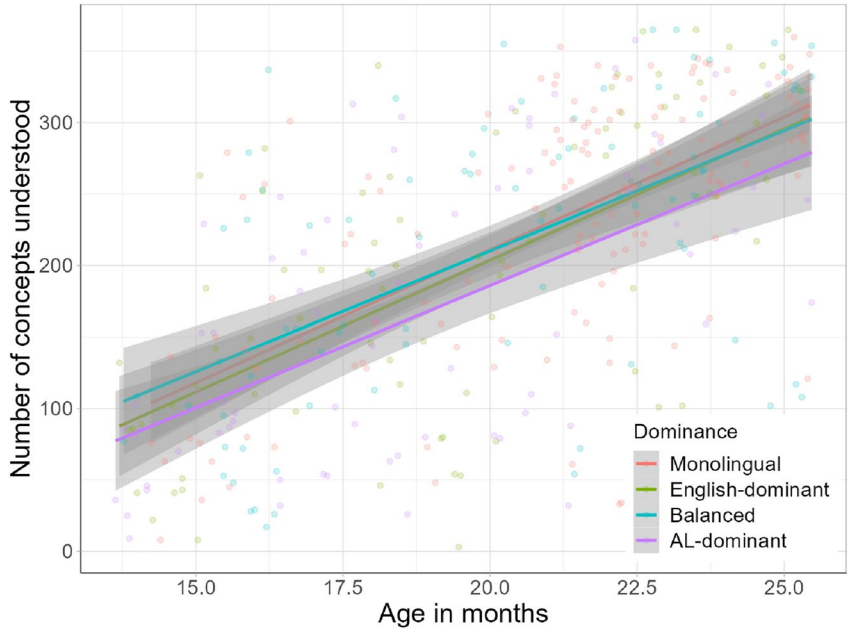

FIGURE 2 Scatterplot of conceptual vocabulary size in comprehension against toddler's age, split by language dominance groups.

TABLE 3 Linear model for bilinguals' conceptual vocabulary size in comprehension, with age, gender, mother's education and English language exposure as predictors.

Predictor	Estimate	Std error	*t*	*p*
(Intercept)	190.3	37.3	5.10	<.001
Age	59.6	5.21	11.4	<.001
Gender	5.5	10.5	0.523	.601
Mother's education	6.28	12.5	0.501	.617
English exposure	5.16	5.41	0.954	.341
English exposure (quadratic)	−7.77	5.84	−1.33	.185

TABLE 4 Linear model for bilinguals' total vocabulary size in comprehension, with age, gender, mother's education and English language exposure as predictors.

Predictor	Estimate	Std error	*t*	*p*
(Intercept)	301.5	69.9	4.31	<.001
Age	111.0	9.76	11.4	<.001
Gender	9.63	19.7	0.489	.626
Mother's education	13.0	23.5	0.555	.579
English exposure	20.3	10.1	2.00	.0463
English exposure (quadratic)	−22.2	11.0	−2.03	.0436

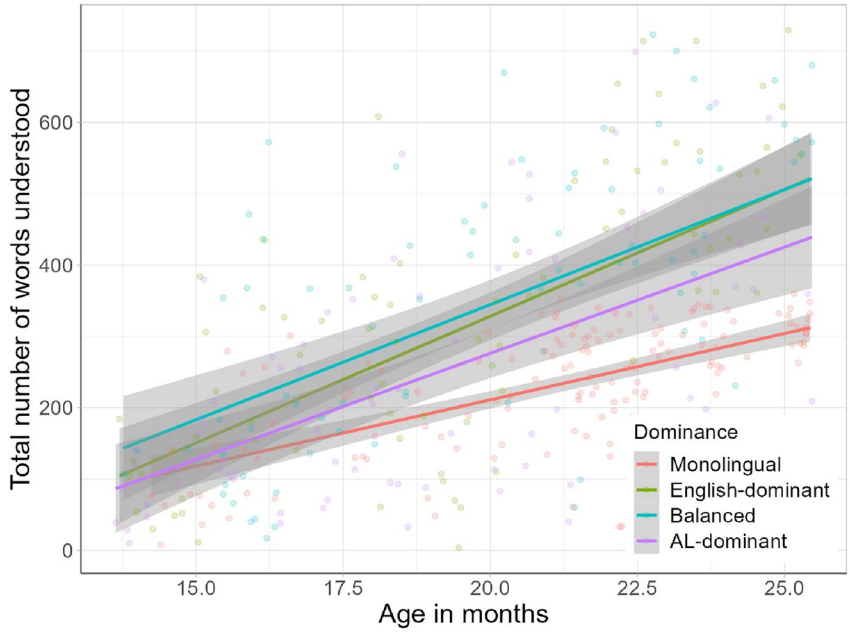

FIGURE 3 Scatterplot of total vocabulary size in comprehension against toddler's age, split by language dominance groups.

bilinguals, and that balanced bilinguals also had larger vocabulary sizes than AL-dominant bilinguals, as can also be seen in Figure 3.

3.2 | Production (20 to 29-month-old)

For visualization purposes, we split the bilingual sample for production into three groups by their language exposure: 79 English-dominant bilinguals (61%–80% English exposure, mean = 73.0, SD = 5.00), 86 Balanced bilinguals (41%–60% English exposure, mean = 54.7, SD = 5.46) and 58 AL-dominant bilinguals (20%–40% English exposure, mean = 31.7, SD = 6.97).

3.2.1 | English vocabulary

As with comprehension, differences between monolinguals and bilinguals were found in vocabulary size for production. Bilinguals produced less English words than monolinguals ($t = -6.579$, $p < .001$). There was no significant interaction between group and age ($t = -1.019$, $p = .309$).

When testing the effect of English language exposure as a continuous variable, we see the expected significant positive effect of the proportion of English language exposure on English vocabulary size in production (Table 5). The quadratic term was again significant, indicating that balanced bilinguals were more similar to English-dominant bilinguals than to AL-dominant bilinguals (Figure 4).

3.2.2 | Conceptual vocabulary

As with comprehension, conceptual vocabulary scoring reduced the difference between monolinguals and bilinguals. However, in contrast to the results with comprehension, there remained a difference between groups for conceptual vocabulary production. This difference between groups was significant

TABLE 5 Linear model for bilinguals' English vocabulary size in production, with age, gender, mother's education and English language exposure as predictors.

Predictor	Estimate	Std error	*t*	*p*
(Intercept)	87.7	43.3	2.03	.0439
Age	49.3	4.99	9.88	<.001
Gender	21.4	10.6	2.02	.0451
Mother's education	7.50	14.4	0.519	.604
English exposure	33.9	5.64	6.01	<.001
English exposure (quadratic)	−13.9	5.82	−2.38	.018

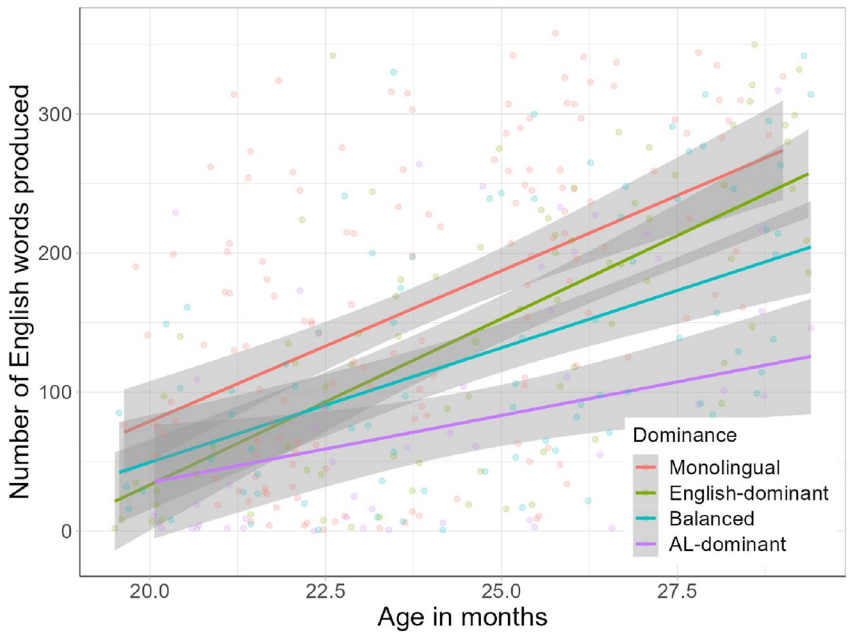

FIGURE 4 Scatterplot of English vocabulary size in production against toddler's age, split by language dominance groups.

($t = -3.446$, $p < .001$). There was no significant interaction between group and age ($t = -0.184$, $p = .854$).

In contrast to conceptual vocabulary in comprehension, language exposure was significant for predicting conceptual vocabulary size in production within the bilingual group, both in the linear term and quadratic term (Table 6). While English-dominant and Balanced bilinguals had vocabulary sizes that were close to that of monolinguals, AL-dominant bilinguals had lower vocabulary size (Figure 5).

3.2.3 | Total vocabulary and doublets

Bilingual toddlers have a large number of doublets in their expressive vocabulary, albeit lower than the values seen in comprehension. Again, the proportion varies (range 0%–89.8%), but on average toddlers in the sample produced the translation equivalents in both languages of 32.7% (SD = 22.4%) of the concepts they produced.

TABLE 6 Linear model for bilinguals' conceptual vocabulary size in production, with age, gender, mother's education and English language exposure as predictors.

Predictor	Estimate	Std error	*t*	*p*
(Intercept)	110.2	46.3	2.38	.0182
Age	57.5	5.34	10.8	<.001
Gender	26.5	11.4	2.33	.0206
Mother's education	9.99	15.5	0.65	.519
English exposure	18.9	6.03	3.13	.00197
English exposure (quadratic)	−12.5	6.22	−2.02	.0450

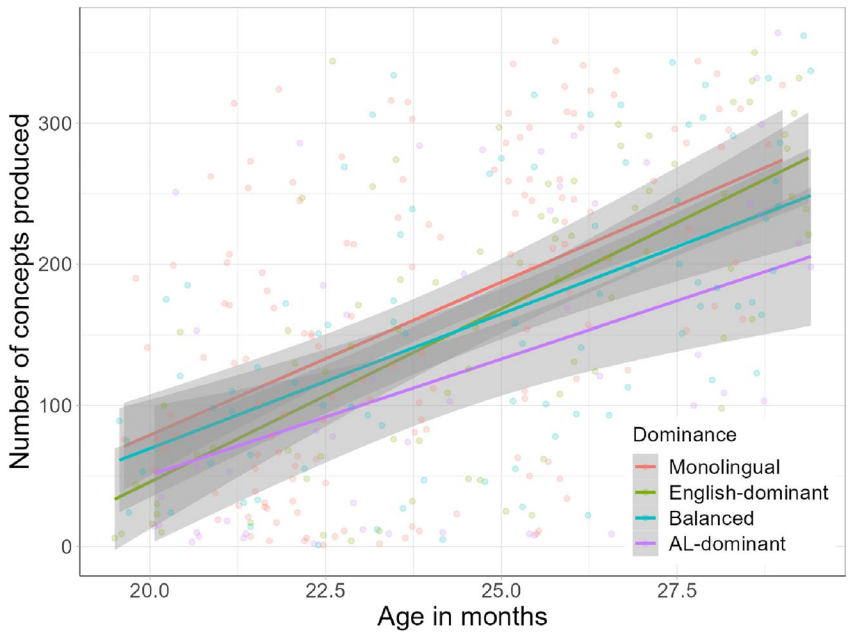

FIGURE 5 Scatterplot of conceptual vocabulary size in production against toddler's age, split by language dominance groups.

In comparison to total vocabulary size in comprehension, where bilinguals had larger vocabulary sizes than monolinguals, the difference between groups for total vocabulary size in production was smaller, albeit still significant ($t = 2.103$, $p = .0361$). The interaction between group and age was also significant ($t = 2.051$, $p = .0410$).

The effect of language exposure was significant for predicting total vocabulary size in production within the bilingual group, both for the linear term and the quadratic term, mirroring the pattern observed for comprehension (Table 7). Balanced bilinguals had higher total vocabulary sizes than both English-dominant and AL-dominant bilinguals (Figure 6).

To investigate whether bilinguals had more doublets in their vocabulary than would be expected by chance, we ran an exploratory analysis to compare the proportion of doublets produced by bilingual children to the proportion of doublets generated when two vocabulary-matched monolinguals are paired randomly. We ran this exploratory analysis on English–Spanish bilinguals (the largest subgroup of our sample) and English monolinguals reported in this manuscript, along with Spanish monolinguals from Wordbank (Frank et al., 2017). Data from Spanish monolinguals was collected using the Spanish

TABLE 7 Linear model for bilinguals' total vocabulary size in production, with age, gender, mother's education and English language exposure as predictors.

Predictor	Estimate	Std error	*t*	*p*
(Intercept)	158.3	77.8	2.034	.0432
Age	90.5	8.97	10.1	<.001
Gender	41.5	19.1	2.17	.0310
Mother's education	17.8	26.0	0.684	.494
English exposure	27.6	10.1	2.72	.00697
English exposure (quadratic)	−32.0	10.5	−3.06	.00247

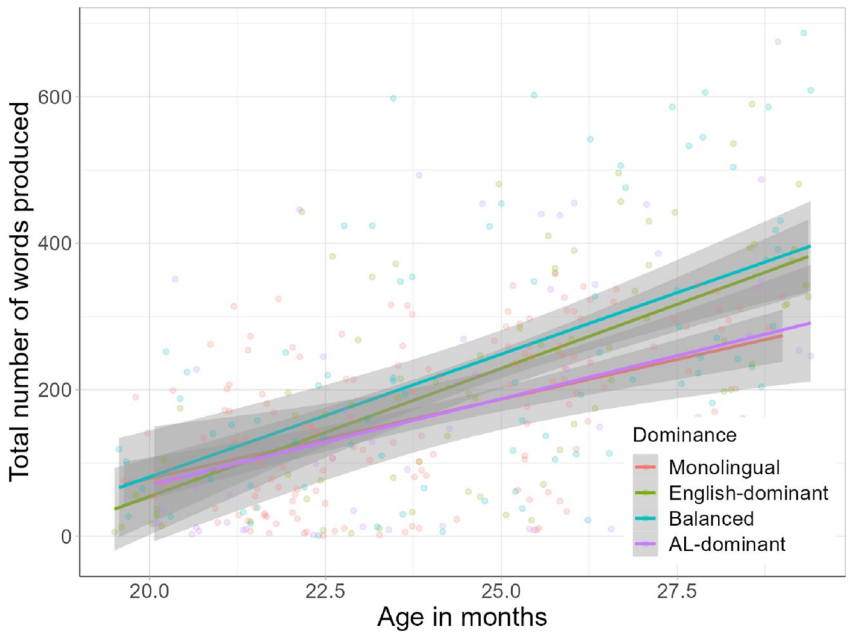

FIGURE 6 Scatterplot of total vocabulary size in production against toddler's age, split by language dominance groups.

(European) Words & Sentences CDI (López-Ornat et al., 2005) ($N = 593$), which measures production vocabulary only. There was an overlap of 136 concepts between the Oxford CDI and the Spanish CDI. To avoid floor effects, we restricted analyses to toddlers with at least 10 words produced in each language.

Monolinguals were paired so that the English monolingual of each pair had the same English vocabulary size as a given bilingual, while the Spanish monolingual had the same Spanish vocabulary size as the bilingual. For example, if a bilingual produced 30 English words and 55 Spanish words, English monolinguals with vocabulary size of 30 will be matched to each Spanish bilingual with vocabulary size of 55. This process resulted in 2238 monolingual pairs, comprised of 104 English monolinguals and 174 Spanish monolinguals. These pairs were compared against 62 bilinguals.

While there was a strong correlation of .95 between the proportion of doublets known by bilinguals and the proportion of doublets by chance, bilinguals knew a significantly larger proportion of doublets than randomly paired monolinguals ($t = 8.54$, $p < .001$). This provides support for the Preference Account of translation equivalent acquisition. There was a strong positive effect of vocabulary size on the proportion of doublets ($t = 12.7$, $p < .001$), but no significant interaction between group

and total vocabulary size ($t = 0.421$, $p = .674$), suggesting that there was no vocabulary size-related change in preference for translation equivalents in the sampled age range.

4 | DISCUSSION

In this paper, we presented findings regarding the vocabulary sizes of a large sample of UK bilingual and monolingual toddlers that are convergent with previous findings in the literature (Cattani et al., 2014; De Houwer et al., 2006, 2014; Hoff & Ribot, 2017; Vagh et al., 2009). The use of the Oxford CDI and its adaptations, which collects information about both comprehension and production using the same set of words, allowed us to make direct comparisons between vocabulary size in comprehension and production for our sample. While other studies have reported both comprehension and production (De Houwer et al., 2014; Pearson et al., 1993), they utilized the MacArthur-Bates CDI and its adaptations, reporting data from the Infant Form for younger participants and the Toddler Form for older participants. In our study, we used the Oxford CDI across the full range of our sample, thus collecting vocabulary size for a common set of words across all ages. Another advantage of this study was the sample size, which exceeded that of previous studies with similar questions and methodology (Bosch & Ramon-Casas, 2014; Core et al., 2013; De Houwer et al., 2014; Pearson et al., 1993). This gave us sufficient power to analyze the data with age and language exposure as continuous variables, facilitating interpretations of these predictors.

4.1 | English vocabulary

For English vocabulary, bilinguals' vocabulary sizes were found to be smaller than that of same-age monolinguals, with the difference significant in both comprehension and production. This difference was modulated by the amount of English exposure received by the child, where bilingual children with a higher proportion of exposure to English showed larger vocabulary sizes in English than bilinguals who received proportionally more exposure to their AL. The relationship between the percentage of English exposure and vocabulary size in English highlights the importance of language input in vocabulary learning. Words that are learnt earlier have been associated with higher frequency in the language input (Braginsky et al., 2019; Goodman et al., 2008; Hills, 2013; Hills et al., 2010). It is therefore unsurprizing that toddlers who hear less of a language understand and produce less words in that language.

The differences in English vocabulary did not show a fully linear pattern. Bilinguals who received similar amounts of input to both English and their AL were more similar to English-dominant bilinguals than to AL-dominant bilinguals. This was found for both comprehension and production. It is possible that vocabulary gains from increased exposure to a particular language may be smaller when language input is already skewed toward that language. This reflects the findings by Pearson et al. (1993) and Cattani et al. (2014) that English-dominant bilingual toddlers performed similarly to same-age English monolinguals.

4.2 | Conceptual vocabulary

The second comparison that we conducted in this paper was between bilinguals' and monolinguals' conceptual vocabulary size. For bilinguals, concepts were considered to be known if either or both of the translation equivalent pair associated with the concept was known. For English monolinguals, conceptual vocabulary size was equal to their English vocabulary size. We found that bilinguals' conceptual vocabulary size in comprehension was comparable to that of monolinguals. This suggests that learning two languages simultaneously did not have a detrimental effect on bilingual toddlers'

acquisition of new concepts in their vocabulary. This pattern is consistent with previous research measuring conceptual vocabulary size in comprehension (Pearson et al., 1993). For production, conceptual vocabulary size was smaller for bilinguals than monolinguals. This mirrored findings by Core et al. (2013) for production.

4.3 | Total vocabulary

When comparing total vocabulary size, bilingual toddlers showed larger vocabulary sizes than monolinguals for both comprehension and production. This result mirrors findings by Junker and Stockman (2002) and De Houwer et al. (2014), but differs from Pearson et al. (1993) and Core et al. (2013). We attribute the significant effect of group for production in our analysis to our large sample size, exceeding that of Core et al. (2013) ($N = 113$) and Pearson et al. (1993) ($N = 60$). The larger total vocabulary sizes of bilingual toddlers relative to monolinguals can be attributed to a high number of words that bilingual toddlers understand in both languages. As a result of receiving language input in two languages, bilingual toddlers acquire words in both languages which overlap in meaning. For our tested languages, the high total vocabulary size may be partially attributable to cognates among the doublets known, though likely not to the extent reported by Bosch and Ramon-Casas (2014) for Spanish-Catalan bilinguals. The languages tested in our study had a modest number of form-similar cognates and very few form-identical cognates. The tested languages which shared the most cognates with English were Dutch and German, with 30% and 27% cognates (defined as translation equivalents that overlapped on more than 40% of their transcribed phonemes) among the words in the CDIs, while the remaining languages (French, Italian, Polish, Portuguese and Spanish) had between 5% and 12% cognates.

Additionally, an interaction between group and age was found for total vocabulary size, where the difference between bilinguals and monolinguals grew larger with increased age. This is similar to the pattern found by Core et al. (2013) where only the oldest age group (30 months) showed significant difference between monolinguals and bilinguals for total vocabulary size.

4.4 | Translation equivalents

Bilingual children showed a high proportion of doublets in their vocabulary. Notably, there was a quadratic effect of language exposure, with balanced bilinguals showing larger total vocabulary sizes and more doublets. The mechanism behind the acquisition of doublets is outside the scope of this paper, but we will make some conjectures based on past research and our exploratory analysis. A Neutral Account for doublet acquisition has been supported by Pearson et al. (1993) and Tsui et al. (2022) for toddlers with larger vocabularies exceeding 300 words. The Neutral Account posits that translation equivalents are neither easier nor more difficult to learn than new concepts. Bilingual children may learn doublets incidentally due to similar trajectories of word acquisition order between languages. The words which are easiest to learn in one language are also the words which are easiest to learn in the other (Łuniewska et al., 2016). Another popular theory is the Preference Account. While some doublets may still be learnt incidentally, the Preference Account posits that it is easier to learn the translation equivalent for a known word than a new concept. The predictions of this account matches findings by Bilson et al. (2015) and by Tsui et al. (2022) for toddlers with smaller vocabulary sizes below 300 words. A preference for doublets may explain the higher total vocabulary sizes of bilinguals compared to monolinguals in our sample, as an ease of learning translation equivalents

may make it easier for bilinguals to acquire more words in their total vocabulary than monolinguals. Our exploratory analyses with English-Spanish bilinguals showed that bilingual toddlers produced a higher proportion of doublets than expected by chance, providing support in line with the Preference Account for the sampled age range of 12–36 months.

4.5 | Maturation constraints for expressive vocabulary

Another important finding of this study is that despite having comparable conceptual vocabulary sizes in comprehension between groups, bilinguals had smaller conceptual vocabulary sizes in production than monolinguals. In total vocabulary, while bilinguals exceeded monolinguals in vocabulary size for both comprehension and production, the difference between groups was smaller in production than in comprehension.

When making direct comparison between comprehension and production, there may be concerns that the samples reported in this manuscript for analyses on comprehension and production were not identical, due to the differing age ranges for avoid ceiling and floor effects. An analysis of the common participants between these two samples (20–25 months old, N monolingual = 107, N bilingual = 124) showed similar patterns. Total vocabulary size in comprehension was significantly higher for bilinguals than monolinguals ($t = 6.641$, $p < .001$), while total vocabulary size in production did not differ between groups ($t = 1.072$, $p = .285$).

This finding that monolingual and bilinguals of the same age produce, on average, a similar total number of words despite bilinguals understanding many more words has interesting implications for the developmental trajectory of word production. These patterns observed in production are likely to be linked to maturation constraints in word production. Toddlers typically understand many more words than they produce. This comprehension-production gap has been observed even in adults when adult participants are taught new words in a novel word learning experiment (Gershkoff-Stowe & Hahn, 2013). Unlike comprehension, which only involves the cognitive process of speech recognition, word production is also dependent on development of motor skills and active rehearsal. Sosa and Stoel-Gammon (2012) found that toddlers' early word productions showed differences in the production accuracy relative to typical adult pronunciations and have variability in pronunciations of the same word. There was more variability found for words that were phonetically more complex and had later-learnt sound combinations. Green et al. (2010) put forward a model of speech development that involves both constraints (milestones in perceptual, cognitive or motor development which limits the learning speed for producing new words) and catalysts (cognitive, neural or environmental supports that help children learn to produce new words). Constraints for speech production by toddlers may include the gradual maturation of articulatory coordination of speech articulators (Green et al., 2000; Iuzzini-Seigel et al., 2015). Given that bilingual toddlers are likely to be equally constrained by developmental milestones in fine motor control, it is unsurprizing that the total number of words produced does not differ significantly between monolingual and bilingual groups.

4.6 | Imbalance in the effect of language dominance

The last point of note to be discussed is that English-dominant bilinguals showed larger vocabulary sizes than AL-dominant bilinguals for both conceptual and total vocabulary, not only for English vocabulary as predicted. This may be a feature of the UK bilingual sample, who hear English widely in the community and have different patterns of language usage in the home. Children who are dominant

in their AL (20%–40% overall English exposure) are less likely to be simultaneous bilinguals (defined as having at least 10% exposure to each language at home). Only 75% of AL-dominant bilinguals can be classified as simultaneous bilinguals. In comparison, 97% of English-dominant bilinguals and 87% balanced bilinguals are simultaneous bilinguals. The later and limited exposure to their non-dominant language may have repercussions on AL-dominant bilinguals' early vocabulary development.

However, there is also a possibility that parents of AL-dominant children find it more difficult to report their child's English vocabulary accurately. In the present study, when given the opportunity to write comments in a free text box, several parents of bilingual toddlers indicated uncertainty in answering the CDI in their non-native language. Vagh et al. (2009) observed a similar issue, with 16 parents of 118 opting out of reporting their child's English vocabulary due to lack of confidence. This opt-out rate of approximately 10% is similar to the rate observed in the present study. For the reported analyses, we excluded the 56 families (out of 589) who explicitly expressed uncertainty in their reporting accuracy. We also required at least one parent to be a native speaker of the AL, and at least one parent to have fluent English proficiency (self-rated proficiency of at least 7 out of 10). Through these criteria, we aimed to reduce the variability in parents' reporting accuracy as a result of low proficiency in one of the target languages.

Another possibility that may affect conceptual vocabulary size is the set of words used for measuring vocabulary size. The CDI is not an exhaustive list of all the words that a child may know, but instead is a subset of commonly-known words aimed to provide an estimate of a child's vocabulary knowledge compared to their peers. Our AL CDIs were adapted from the Oxford CDI, which was normed using data from English-monolingual British toddlers. As such, the subset of words in the Oxford CDI (and subsequently our AL CDIs) may be biased toward concepts that are familiar to the UK English-speaking community. While the toddlers in the bilingual sample were also growing up in the UK, there may be certain concepts less common in their home environment due to cultural differences. We attempted to reduce this bias by using only the subset of concepts that was common across all our CDIs after appropriate substitutions were made by native speakers of those languages. Further research on bilingual children would benefit from questionnaires and tasks specifically designed to reflect the experiences of bilingual communities.

5 | CONCLUSION

In this paper, we presented analyses on the vocabulary size trajectories of bilingual toddlers relative to monolingual toddlers in English vocabulary, conceptual vocabulary and total vocabulary. Observed differences between groups in English vocabulary size highlight the effect of language exposure on vocabulary size in a single language. In contrast, comparable conceptual vocabulary sizes in comprehension suggest that bilinguals and monolinguals acquire new concepts at similar rates. Bilinguals also have high total vocabulary sizes, with many words being doublets. This parallel development of their two languages is important for the communicative needs of bilingual toddlers, who may need different languages to communicate effectively with the people in their lives. We also discussed maturation constraints for the development of vocabulary in production, linking it to comparable total vocabulary sizes of bilingual and monolingual toddlers. The patterns of bilingual vocabulary size observed in this study highlight both similarities and differences between bilinguals and monolinguals. Both groups show similar trajectories for learning new concepts and development of production. However, a unique aspect of bilinguals' language learning is the availability of cross-linguistic overlap. Translation equivalents can facilitate bilinguals' vocabulary learning through shared meaning. Learning a new word for a known concept is posited to be easier than learning a new unique concept. Further

research into how different semantic and phonological aspects of the languages may affect bilinguals' vocabulary trajectories can shed light onto the specific aspects that facilitate parallel development of two languages.

ACKNOWLEDGMENTS

The study reported in this manuscript was supported by the Economic and Social Research Council grant CQR01830 to KP and Nuria Sebastian-Galles. At the time of data collection, SS, NG and IL were employed as research assistants on this grant. At the time of publication, SS is employed as a Postdoctoral Fellow under an Academic Research Fund Tier 1 grant FY2021-FRC3-002 from the Singapore Ministry of Education. The authors declare no conflicts of interest with regard to the funding source for this study. Data collection was conducted by SS, NG and IL. Data analysis was conducted by SS. We would like to thank all families who participated in this study. We also thank the reviewers for their time and insightful feedback that improved the quality of this manuscript. Many thanks to associates who helped with adapting the translations of the Oxford CDI:

- Dutch: Prof Paula Fikkert
- German: Dr Marlene Spangenberg
- French: Prof Caroline Floccia
- Italian: Dr Davide Volpi
- Polish: Zuzanna Patryas
- Portuguese: Andreia Correia Brecha
- Spanish: Prof Nuria Sebastian-Galles, Dr Daniela Avila-Varela and Gonzalo Garcia-Castro

DATA AVAILABILITY STATEMENT

Data and materials are available on OSF: https://osf.io/w8ke5/.

ORCID

Serene Siow https://orcid.org/0000-0001-6482-2191
Nicola A. Gillen https://orcid.org/0000-0001-8731-4827
Kim Plunkett https://orcid.org/0000-0003-0216-7480

REFERENCES

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, *15*(2), 397–401. https://doi.org/10.1017/s136672891100040x

Bialystok, E., Luk, G., Peets, K. F., & Sujin, Y. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/s1366728909990423

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134. https://doi.org/10.1016/j.cognition.2015.03.013

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322. https://doi.org/10.1177/0165025414532559

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49. https://doi.org/10.1207/s15327078in0201_3

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

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 and Communication Disorders*, *49*(6), 649–671. https://doi.org/10.1111/1460-6984.12082

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: Implications for assessment. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1637–1649. https://doi.org/10.1044/1092-4388(2013/11-0044)

De Houwer, A., Bornstein, M. H., & De Coster, S. (2006). Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism*, *10*(3), 331–347. https://doi.org/10.1177/13670069060100030401

De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied PsychoLinguistics*, *35*(6), 1189–1211. https://doi.org/10.1017/s0142716412000744

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (2018). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*, 1–135.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/s0305000916000209

Gershkoff-Stowe, L., & Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of Experimental Child Psychology*, *114*(4), 489–509. https://doi.org/10.1016/j.jecp.2012.11.005

Gillen, N. A., Siow, S., Lepadatu, I., Sucevic, J., Plunkett, K., & Duta, M. (2021). Tapping into the potential of remote developmental research: Introducing the oxfordbabylab app. PsyArXiv.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/s0305000907008641

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, *43*(1), 239–255. https://doi.org/10.1044/jslhr.4301.239

Green, J. R., Nip, I. S., Maassen, B., & Van Lieshout, P. (2010). Some organization principles in early speech development. *Speech Motor Control: New Developments in Basic and Applied Research*, *10*, 171–188.

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23*(4), 574–586. https://doi.org/10.1044/2014_ajslp-13-0026

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*(3), 689–705. https://doi.org/10.1017/s0305000900004414

Hills, T. (2013). The company that words keep: Comparing the statistical structure of child-versus adult-directed language. *Journal of Child Language*, *40*(3), 586–604. https://doi.org/10.1017/s0305000912000165

Hills, T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*, 1–27. https://doi.org/10.1017/s0305000910000759

Hoff, E., & Ribot, K. M. (2017). Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 years. *The Journal of Pediatrics*, *190*, 241–245. https://doi.org/10.1016/j.jpeds.2017.06.071

Howard, E. R., Páez, M. M., August, D. L., Barr, C. D., Kenyon, D., & Malabonga, V. (2014). The importance of SES, home and school language and literacy practices, and oral vocabulary in bilingual children's English reading development. *Bilingual Research Journal*, *37*(2), 120–141. https://doi.org/10.1080/15235882.2014.934485

Iuzzini-Seigel, J., Hogan, T. P., Rong, P., & Green, J. R. (2015). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of Motor Learning and Development*, *3*(1), 53–68. https://doi.org/10.1123/jmld.2014-0054

Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology*, *11*(4), 381–394. https://doi.org/10.1044/1058-0360(2002/042)

Legacy, J., Reider, J., Crivello, C., Kuzyk, O., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2017). Dog or chien? Translation equivalents in the receptive and expressive vocabularies of young French–English bilinguals. *Journal of Child Language*, *44*(4), 881–904. https://doi.org/10.1017/s0305000916000295

López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., & Martínez, M. (2005). *Macarthur: Inventario de desarrollo comunicativo*. TEA Ediciones.

Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Anđelković, D., Blom, E., Boerma, T., Chiat, S., de Abreu, P. E., Gagarina, N., Gavarró, A., Håkansson, G., Hickey, T., de López, K. J., Marinis, T., Popović, M., Thordardottir, E., Blažienė, A., … Ünal-Logacev, Ö. (2016). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods*, *48*(3), 1154–1177. https://doi.org/10.3758/s13428-015-0636-6

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(0), F9–F16. https://doi.org/10.1111/j.1467-7687.2007.00671.x

Marchman, V. A., & Martínez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research*, *45*(5), 983–997. https://doi.org/10.1044/1092-4388(2002/080)

O'Toole, C., & Hickey, T. M. (2017). Bilingual language acquisition in a minority context: Using the Irish–English communicative development inventory to track acquisition of an endangered language. *International Journal of Bilingual Education and Bilingualism*, *20*(2), 146–162. https://doi.org/10.1080/13670050.2016.1179256

Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied PsychoLinguistics*, *18*(1), 41–58. https://doi.org/10.1017/s0142716400009863

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*(1), 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language*, *22*(2), 345–368. https://doi.org/10.1017/s030500090000982x

Qualtrics. (2022). *Qualtrics*. Qualtrics. https://www.qualtrics.com

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Smithson, L., Paradis, J., & Nicoladis, E. (2014). Bilingualism and receptive vocabulary achievement: Could sociocultural context make a difference? *Bilingualism: Language and Cognition*, *17*(4), 810–821. https://doi.org/10.1017/s1366728913000813

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research*, *55*(2), 596–608. https://doi.org/10.1044/1092-4388(2011/10-0113)

Strand, S., Malmberg, L. and Hall, J. (2015) English as an additional language (EAL) and educational achievement in England: An analysis of the national pupil database.

Styles, S., & Plunkett, K. (2009). What is 'word understanding' for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, *36*(4), 895–908. https://doi.org/10.1017/s0305000908009264

Tsui, R. K.-Y., Gonzalez-Barrero, A. M., Schott, E., & Byers-Heinlein, K. (2022). Are translation equivalents special? Evidence from simulations and empirical data from bilingual infants. *Cognition*, *225*, 105084. https://doi.org/10.1016/j.cognition.2022.105084

Vagh, S. B., Pan, B. A., & Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: The utility of combining parent and teacher report. *Child Development*, *80*(5), 1545–1563. https://doi.org/10.1111/j.1467-8624.2009.01350.x

Yan, S., & Nicoladis, E. (2009). Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism: Language and Cognition*, *12*(03), 323–335. https://doi.org/10.1017/s1366728909990101

How to cite this article: Siow, S., Gillen, N. A., Lepădatu, 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