Journal of Child Language (2025), 1–24
doi:10.1017/S0305000924000655

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Bilingual children reach early language milestones at the same age as monolingual peers

Karolina Muszyńska[1], Grzegorz Krajewski[1], Agnieszka Dynak[1], Nina Gram Garmann[2,3], Anna Sara H. Romøren[2], Magdalena Łuniewska[1], Katie Alcock[4], Napoleon Katsos[5], Joanna Kołak[6], Hanne Gram Simonsen[3,6], Pernille Hansen[7], Magdalena Krysztofiak[1], Krzysztof Sobota[1] and Ewa Haman[1]

[1]University of Warsaw, Poland; [2]OsloMet – Oslo Metropolitan University, Norway; [3]University of Oslo, Norway; [4]Lancaster University, UK; [5]University of Cambridge, UK; [6]University College London, UK and [7]Inland Norway University of Applied Sciences, Norway
Corresponding author: Karolina Muszyńska; Email: karolina.muszynska@psych.uw.edu.pl

(Received 13 December 2023; revised 09 September 2024; accepted 22 October 2024)

Abstract

In this longitudinal study, we compare the age of reaching early developmental milestones in bilingual and monolingual children and between the bilinguals' two languages. We present data from 302 Polish bilinguals (living outside of Poland with various majority languages) and 302 Polish monolinguals, aged $M = 12.78$ months on study entry (range: 0–24 months), matched on sex, age at study entry, duration of parental reporting, and parental education. The milestones under investigation include crawling, walking, babbling, first, 10th, 50th word, and first multi-word utterances. The data was collected with a specially designed mobile app, in which parents reported their children's development repeatedly. Using this relatively big sample and looking at a wide range of investigated milestones, we present evidence that typical bilingual development follows a trajectory similar to monolingual development. We also evaluate the feasibility and usefulness of online data collection using mobile apps to study early language development.

Keywords: bilingual development; early language milestones; first words; online data collection

Streszczenie

W badaniu podłużnym za pomocą specjalnie zaprojektowanej aplikacji mobilnej zbieraliśmy dane od rodziców, raportujących rozwój swoich dzieci. W aplikacji rodzice raportowali m.in. wiek osiągania kamieni milowych, takich jak raczkowanie, chodzenie, gaworzenie, wypowiadanie pierwszego, dziesiątego i pięćdziesiątego słowa, oraz łączenie słów w wielowyrazowe wypowiedzi. Na podstawie danych pochodzących od rodziców 302 dzieci dwujęzycznych przyswajających język polski (oraz inny język większościowy, mieszkających poza Polską) i 302 polskich dzieci jednojęzycznych, porównaliśmy wiek osiągania wczesnych kamieni milowych przez dzieci jedno- i dwujęzyczne oraz pomiędzy dwoma językami dzieci dwujęzycznych.

Dzieci jedno- i dwujęzyczne zostały dopasowane do siebie pod względem płci, wieku w momencie dołączenie badania (M = 12.78, zakres od 0 do 24 miesięcy), poziomu wykształcenia rodziców i tego, przez ile dni rodzice raportowali rozwój dzieci w aplikacji. Analizując dane pochodzące z tej stosunkowo dużej próby, wskazujemy na podobieństwa w przebiegu rozwoju dzieci jedno- i dwujęzycznych. Oceniamy również możliwości i ograniczenia zbierania danych za pomocą aplikacji mobilnych w badaniach dotyczących wczesnego rozwoju językowego.

Abstrakt

I denne longitudinelle studien sammenligner vi polske enspråklige og tospråklige barn for å undersøke hvor gamle de er når de når bestemte milepæler i utviklingen, og om det er forskjeller mellom de tospråkliges to språk. De tospråklige er polskspråklige barn som bor utenfor Polen i land med ulike majoritetsspråk. Vi presenterer data fra 302 tospråklige og 302 enspråklige barn, med en gjennomsnittsalder på 12.78 måneder ved innsamlingsstart (intervall: 0–24 måneder). Barna er matchet for kjønn, alder ved innsamlingsstart, varighet av foreldrerapportering og foreldrenes utdanning. Milepælene som undersøkes inkluderer krabbing, gange, babling, det første ordet, det tiende ordet, det femtiende ordet og de første flerordsytringene. Dataene ble samlet inn med en spesialdesignet mobilapp, der foreldre rapporterte om barnas utvikling flere ganger. I dette relativt store utvalget har vi funnet evidens for at tospråklige barn følger tilsvarende kurve som enspråklige barn når det gjelder typiske språklige og utviklingsmessige milepæler. Vi evaluerer også gjennomførbarheten og nytten av online datainnsamling ved hjelp av mobilapper for å studere tidlig språkutvikling.

1. Introduction

Establishing a developmental path in children's language acquisition may inform both psycholinguistic theory and practitioners (e.g., educators and speech therapists) about the individual trajectories of language development that are typical in the general population. The issue of a typical development gains additional importance in the case of bilingual children (e.g. Boerma & Blom, 2017; Paradis et al., 2011, 2023). Traditionally, the starting point for scientific descriptions of language development in children would be a monolingual English-speaking child. The monolingual bias in the literature and research (Kidd & Garcia, 2022) may misrepresent the way many children develop their languages, failing to understand the factors (both within and around the child) which shape language acquisition as many children in the world grow up exposed to more than one language (see for example De Houwer, 2023). Comparing monolingual and bilingual development sheds light on the common and distinctive processes underlying language development in diverse contexts. As pointed out by Hoff et al. (2012) and Höhle et al. (2020), among others, exploring the early language development of bilingual children may provide additional insights into the debate on the (in)dependence of early language acquisition from variation in input. Bilingual children's input is naturally divided between two languages, and as a result, a bilingual child typically hears less of each language compared to a monolingual. Potential differences in timing of reaching early language milestones between bilinguals and monolinguals could reflect the impact of input variation in early

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

language development. On the other hand, similarities between bilinguals and monolinguals (or between the two languages of bilinguals) could suggest that the timing of reaching early language milestones is more independent of input variations and either dependent on other environmental factors, e.g. maternal responsiveness (e.g. Tamis-LeMonda et al. 2001), or – at least in the early stages – on general cognitive and social learning abilities (Gervain & Mehler, 2010).

In the literature, there is substantial evidence that children are capable of successfully acquiring two languages (for a review, see Höhle et al., 2020; Werker & Byers-Heinlein, 2008). When the bilinguals' vocabularies in each language are summed to a measure of total vocabulary, bilinguals are often reported to know as many words as their monolingual peers (e.g. Core et al., 2013; Miękisz et al., 2017, see also Pontecorvo et al. 2023 for evidence on ASL – English bilingual deaf and hard of hearing children) or even more (Byers-Heinlein et al. 2024, Legacy et al., 2016). On the other hand, in single language comparisons, bilinguals often exhibit smaller vocabularies in each language in comparison to their monolingual peers (cf. Genesee 2022). In a study of Spanish-American English bilinguals (age: 30–60 months), Hoff and Ribot (2017) found that bilinguals' (productive) vocabulary growth in English lagged 6 months to 1 year behind their American English monolingual peers (including parent education as covariate). The size of the vocabulary lag depended on the relative exposure to American English, and the impact of exposure was not linear – the higher the exposure to English, the larger was the increase in English vocabulary knowledge. This clearly shows that in bilinguals, as they may have less frequent access to each language compared to monolinguals (Hoff et al., 2012; Unsworth, 2016), the rate of vocabulary growth in each language varies as a (possibly not linear) function of the relative amount of exposure (e.g., Hoff et al. 2012, Elin Thordardottir, 2019). Elin Thordardottir (2019) found that Canadian English-French bilinguals' amount of exposure to each language predicted their receptive and expressive vocabulary size in each language in school age and was a better predictor of vocabulary size than the age of first exposure to each language. In another study in Canada, Paradis and Jia (2017) tracked school-aged bilinguals' performance on standardised measures of vocabulary, grammar, and global comprehension and found that bilinguals needed approximately 5½ years of English exposure to reach monolingual norms on most of the tasks. Thus, bilinguals' distributed exposure across languages has been repeatedly shown to impact their growth of vocabulary or other language skills. However, researchers note that these lags do not necessarily imply clinically relevant delays, as long as children are reported to have some skills in another language.

Previous studies on early language acquisition have focused mainly on vocabulary size and vocabulary growth (as measured by parental questionnaires, i.e., CDIs, MacArthur-Bates Communicative Development Inventories, Fenson et al., 2007). We propose that investigating the timing of reaching fundamental language milestones (such as babbling, producing first words, and combining words into sentences) may provide further evidence regarding the pace of language development in bilingual vs. monolingual children. Investigating developmental milestones is a recommended foundation for practitioners (e.g. speech therapists, social and health workers and others) in screening for developmental delays (Lipkin et al., 2020). Importantly, researchers warn that the available evidence supporting the milestones is still lacking and conflicting, illustrating the critical need for further research in this area (Roberts et al., 2023). Moreover, investigating early language milestones can inform our knowledge of language development by showing to what extent milestone attainment is dependent on (the quantity and quality of) input. Unquestionably, input variability influences the vocabulary size in children. However, it is unclear to what extent the timing of reaching language milestones

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

is susceptible to variations characteristic of the bilingual input. Reaching the earliest milestones, such as producing first words, is also driven by general cognitive and linguistic learning abilities, such as language discrimination, statistical learning, and word segmentation (Gervain & Mehler, 2010). A recent review from Höhle et al. (2020) presents evidence that bilingual and monolingual infants show similar trajectory in the early language acquisition steps: bilingual infants can discriminate between their native language from non-native rhythmically similar languages at the same time as monolinguals (Bosch & Sebastián-Gallés, 2001). Group similarities are found also in word segmentation abilities at 6 and 8 months of age (e.g. Bosch et al., 2013; Singh, 2018), which are closely related to later lexical acquisition (e.g. Junge et al., 2012). As Höhle et al. (2020, p. 66) conclude, these early similarities prove that infants are equipped with robust learning mechanisms that allow them to create stable linguistic representations even in diverse input conditions. There are also other factors, such as maternal responsiveness, that also have great influence on early language milestones. Tamis-LeMonda et al. (2001) showed that children whose mothers responded contingently to their child's vocalisations and play activities reached early language milestones (such as first word production, 50 words production, combining words into sentences) earlier than children of less responsive mothers (the study did not control for general language input children were receiving). Potential differences in timing of reaching early language milestones between bilinguals and monolinguals could reflect the impact of input variation in early language development, while group similarities could suggest that timing of reaching early language milestones is robust against input variations or dependent on other factors or combinations of factors.

So far, few studies of early bilingual development have focused on the age of reaching language milestones. Petitto et al. (2001) performed a longitudinal observational study with three French-English bilingual children and three bimodal bilingual children (i.e., children who learn both signed and spoken language). No monolingual group was studied. They found that both bilingual groups produced their first word/sign around their first birthday (in both languages), reached their 50-word milestone, and started to combine words/signs before their second birthday (in both languages). Similar ages are reported for typically developing monolingual children (Scharf et al., 2016). Oller et al. (1997), in their longitudinal study of over 70 infants, found that the onset of canonical babbling was the same between bilingual (American English-Spanish) and monolingual (American English) children. In a study of Spanish-American English toddlers, Hoff et al. (2012) found that just before their second birthday (at 1;10) a similar percentage of bilingual and monolingual children were combining words in either language. A study of bilingual and monolingual children with autism spectrum disorder (ASD) diagnosis also showed no group differences in the age of producing first words or first phrases (nor in vocabulary growth) (Ohashi et al., 2012). These studies suggest that early developmental trajectories may be similar between bilingual and monolingual children, though the evidence comes from relatively small-scale studies, studies of atypical bilingual populations, or studies with a narrow range of early milestones.

In the present study, we aimed to investigate early bilingual language development not only in comparison to monolingual peers but also between the two languages of the bilinguals. To date, the only evidence on the comparison of bilingual children's milestones between their two languages comes from Petitto et al.'s (2001) bilingual and bimodal case study. Importantly, the children observed by Petitto et al. had their language exposure equally divided between their two languages, as each of their parents spoke one language towards the child (one parent one language, OPOL, strategy). However, bilingual children

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

typically have unequal exposure to their languages, and as a result, their vocabulary in each language may develop at different rates (Hoff, 2017; Marchman et al., 2009). Examining bilingual children's language milestones in both of their languages while considering the context and amount of language exposure is essential for a more accurate assessment of their language development and for grasping the dynamics of this development (Hoff, 2012).

In studies of developmental milestones, the common method of testing is a parental report or diary. Parents are a natural source of knowledge about their child's early language development, as they have the most access to their child's everyday experiences (Dale, 1991; De Houwer, 2011). At their best, parental reports or diaries allow frequent data collection (which is desirable for tracking developmental milestones) and deliver ecologically valid insights based on the child's natural environment. As for the validity of parental diaries, to date, there is little evidence, but Reznick and Goldfield (1994) found moderate to strong correlations (r = 0.64–0.86 on various timepoints) between parental diaries and vocabulary checklists (i.e. CDIs) for parents of children between 1;2 and 1;8. When research staff was to confirm parental reports of attained milestones in a study by Oller et al. (1997), the agreement between parental reports of their children's canonical babbling and researchers' subsequent evaluation was very high (in only two out of 73 cases the staff did not confirm the parental report). On another note, Moore et al. (2019) compared parental reports of a child's onset of talking with researchers' reports during lab visits and found the mean estimates did not differ significantly, but the estimates per child were not correlated. The authors suggest this arises from the difficulty (of both parents and researchers) to determine whether what the child has produced is indeed a word (Vihman & McCune, 1994). In our study we hoped to help the participating parents by establishing the study's interpretation of a new word (see Measures and data coding). When parents are asked to recall (after some time) when their child reached a particular milestone, their recalls are most precise if they refer to culturally salient milestones, such as birth weight and walking, which are accurately recalled even after a few years (Hus et al., 2011; Jaspers et al., 2010; Treharne, 1992). Speech milestones, such as the age of first word production, seem to be more difficult to recall with time, especially after a number of years (e.g. Alcock & Brennan, 2011; Hus et al., 2011; Majnemer & Rosenblatt, 1994). Together, this evidence shows that parents are quite reliable in their continuous reporting in a language diary, as compared to a checklist, and that the parental recall should be collected at the time or shortly after the milestones occur.

In the present study, we aimed to provide robust evidence, grounded in the timing of reaching developmental milestones, characterising typical early language development in bilingual children. We asked parents about their children's basic motor and language milestones. The motor milestones (i.e., crawling, walking unassisted) were included to check the consistency of parental reports, expecting no differences between the bilingual and monolingual groups. However, the development of motor skills, especially the transition from crawling to walking, is directly linked to language development – as walking children can use their free hands to gesture and elicit language from caregivers (Moore et al., 2019; West & Iverson, 2021). We also asked parents to repeatedly report their children's language milestones: babbling, gestures, words, and utterances as they appeared in the children's production. The study made use of a mobile app to reach a large, heterogeneous sample of participants and to minimise the effort associated with language diary collection methods. With a mobile app, within a short reach, parents could report more easily and be regularly reminded to report their child's language productions. However, the biggest and most common challenge of using mobile apps in research has

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

been the low retention of participants over longer periods of time. Some studies have shown 15-day and 30-day retention rates as low as 3–4% (Baumel et al., 2019). In a cross-evaluation from eight remote digital health studies, the median participant retention was only 5.5 days (Pratap et al., 2020). We attempted to mitigate this expected challenge of using mobile apps in research by adding some gamification features to increase parents' engagement in the study (see Mieszkowska et al., 2022 for more details). At the end of the present paper, we explore and report on the usefulness of this relatively novel approach to language data collection, as well as share the challenges and limitations of the method.

2. Research aims

The present paper follows a pre-registration published on OSF before the data collection began (https://osf.io/pzykv/). We aimed to investigate the early bilingual language development in comparison to monolingual peers and between the two languages of the bilinguals. We focused on the reported age of crawling, walking, and specific language milestones: babbling, age of reporting the first word, 10th, 50th word, the first multiword utterance, the first 3-word utterance, and the first 4-word utterance. The research questions considered in the present paper are as follows:

RQ1: Will bilinguals reach specific language milestones at a similar time as their Polish monolingual peers, at least in one of their languages?
RQ2: Will bilinguals reach specific language milestones at a similar time in both their languages, the home language (Polish) and the majority language?

Regarding RQ1, we hypothesised that as long as we compare the bilinguals' and monolinguals' age of reaching the milestones independently of the language in which the milestones were achieved, we might expect similar outcomes between the groups. Similar results were shown in studies comparing total vocabulary growth and early milestones in bilinguals and monolinguals (e.g. Oller et al., 1997; Petitto et al., 2001).

Regarding RQ2, we expected that the bilingual children in our sample would not reach all language milestones at a similar time in each of their languages, given previous studies (e.g. DeAnda et al., 2018; Hoff & Core, 2013; Marchman et al., 2009) showing that language-specific outcomes in bilingual children are dependent on various factors (e.g. input quantity) and therefore are not necessarily growing in parallel.

Originally, we also planned to account for the influence of language-specific input on the bilinguals' language milestones in the two languages, but due to missing data, we were unable to perform analyses as planned. Instead, in our exploratory analysis, we looked at a small subset of the bilingual sample for whom we obtained information on their language input patterns from a single (typical) day. We aimed to link the children's relative input in the home language to the timing of reaching language milestones in their home and majority language.

3. Method

3.1. Participants

In our pre-registration, we set our target sample at 300 parents of children aged 0–24 months: 100 parents of Polish-English bilinguals living in the UK, 100 parents of Polish-Norwegian bilinguals living in Norway, and 100 parents of Polish monolinguals

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

living in Poland. These groups were chosen because Norway and the UK are common destinations for migrating Poles.

There were two modifications to this plan, both stemming at least to some extent from widespread recruitment on social media and the resulting popularity of the app. First, the study attracted a significant number of parents of bilingual children with Polish and another language other than English or Norwegian. This group was not the focus of our recruitment. We decided to include these data in our analyses if the children met the following inclusion criteria: they were between 0 and 24 months at the study entry, they lived outside of Poland, and at least one of the parents was an L1 speaker of Polish. However, we excluded children who were bilingual with another Slavic language (Croatian, Czech, Russian, Slovak, Ukrainian) or Korean since we could not rule out potential impact from strong typological closeness/distance with Polish. This new group included children bilingual with Polish and the following majority languages: German (n = 45), English spoken outside of the UK (n = 26), Dutch (n = 11), French (n = 10), Spanish (n = 9), Danish (n = 8), Swedish (n = 5), Italian (n = 3), Faroese (n = 1), Flemish (n = 1), Greek (n = 1), Portuguese (n = 1).

The second modification concerns the sample size of the groups. We aimed at 100 participants in each group, but since no stopping rule was employed, we gathered data from 2,178 parents of Polish monolinguals, 116 parents of Polish-English bilinguals, 74 parents of Polish-Norwegian bilinguals, and 121 parents from the above-mentioned group of Polish bilinguals with majority languages other than Norwegian or English (311 bilinguals in total).

The 311 bilingual children all heard Polish on an everyday basis (at least one of their parents was Polish and spoke Polish to the child) and lived in the country where another language was the majority language. This does not mean that all of these children heard the majority language on a regular basis – for 7 children, parents reported they had not yet had regular contact with the majority language (see Appendix A). These 7 children all lived in Norway, and their ages ranged from four to 22 months at study entry (M = 11.85, SD = 7.12). The monolingual children lived in Poland, heard only Polish on an everyday basis, but could have some minimal input, e.g. from English-language media, popular among parents in Poland.

According to the pre-registered criteria, we excluded 75 children born before 37 weeks of pregnancy (70 monolinguals and 5 bilinguals), 17 children with low birth weight (Reyes & Mañalich, 2005) (16 monolinguals and 1 bilingual), and 39 children (35 monolinguals, 4 bilinguals) with health problems that could impact the early timing of developmental milestones (e.g. hearing problems, sight problems, muscle tension). For 686 monolinguals and 165 bilinguals (31.5% and 55.4% of the sample, respectively), we did not obtain any information regarding potential health and/or developmental difficulties. We did not predict these missing data points in our pre-registration, and we decided to keep these children in the dataset, as excluding them from the analyses would substantially diminish the analytical power. The sample after employing our exclusion criteria included 302 bilinguals (111 Polish-English bilinguals, 73 Polish-Norwegian bilinguals, and 118 bilinguals with Polish and a language other than English/Norwegian) and 2,055 Polish monolingual children.

Our pre-registration stipulated matching each bilingual child with a monolingual child as similar as possible in regards to the age (in months) upon study entry, sex, and the parental level of education. Because of many missing data for the latter (671 monolinguals and 179 bilinguals), we included a dummy level for the missing parental education (see Table 1 for the education levels used in the matching procedure). We also added another matching variable (not included in the pre-registration), a proxy for parent's engagement

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

in the study, calculated as the number of days between the parent's first and last reported element (i.e., gesture, word, utterance) for their child. This matching variable was added due to high variability in the sample regarding the length of reporting (see Table 2), which could potentially influence the amount (and precision) of data gathered from each individual. Each matching variable was given an equal weight, and for each bilingual, a monolingual with most similar characteristics was found (the matching script is available on OSF; https://osf.io/3ruj9/). The resulting paired groups did not differ on the matching variables, i.e. sex ($\chi^2(1) = 1$, $p = 1$), parental education ($\chi^2(6) = 3$, $p = 0.8$), age of entering the study ($t(602) = -0.17$, $p = 0.87$) or the length of parental reporting ($t(601) = 0.28$, $p = 0.78$) (see Tables 1 and 2). Additionally, we found that the bilingual and monolingual children did not differ in the mean age of crawling, $t(147) = -0.38$, $p = 1.0$, and walking unassisted, $t(74) = -1.25$, $p = 1.0$, as reported by parents.

Since the Polish monolingual sample was unexpectedly large, we decided to use the matching procedure to create two control (monolingual) groups to make maximum use of the gathered data (this was not included in the pre-registered plan). The results were similar between the three groups, and here we will report results from the bilinguals and monolinguals from the first match. These exploratory analyses, including all three groups, are available in the Supplementary Materials (Appendix B).

Upon entering the study, each parent gave their informed consent to their child's study participation. The study was evaluated and approved by the Ethics Committee at the Faculty of Warsaw, University of Warsaw, and Sikt – Norwegian Agency for Shared Services in Education and Research.

Table 1. Frequency table of parental education by group after the matching procedure

Parental education level	Bilinguals *n*	Monolinguals *n*
0 (Not available)	179	181
PhD	8	6
Higher (university degree)	88	94
Unfinished higher	15	11
Vocational (technical)	4	2
Secondary	7	8
Primary	1	0

Table 2. Group characteristics after the matching procedure

Group	*n*	Sex	Age (months) at study entry (*M*)	Age (months) at study entry (*SD*)	Days reporting (*M*)	Days reporting (*SD*)	Days reporting (min)	Days reporting (max)
Bilinguals	302	155f 147m	12.80	7.15	32.91	79.69	1	529
Monolinguals	302	155f 147m	12.90	7.00	31.16	76.25	1	492

Note. "Days reporting" indicates the number of days (mean) parents were reporting their children's development in the app.

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

3.2. Procedure

The detailed procedure is described in the study protocol (Mieszkowska et al., 2022). Parents who wished to join the study downloaded the "StarWords – every word counts" app from Google Play or Apple App Store. Parents were asked to report/recall the earliest developmental milestones already reached (if applicable, i.e., crawling, walking, babbling), and from then on regularly report on their child's new gestures, words, and utterances as they appeared, and answer questions about the child's language environment (see Measures). The app sent automatic notifications to users to provide information on the child's family characteristics, language environment, and early development. Parents were rewarded (with access to popular-scientific podcasts, articles and colouring pages with their children's words, see Appendix C) for responding to those notifications, but not for the number of elements (gestures, words, utterances) reported in the app. Parents entering the study were encouraged to report regularly and for as long as possible, but were allowed to opt out at any point without giving any reason. The notifications encouraging them to report new elements were usually sent a week after a previous element (gesture, word, utterance) was reported. The exact timing would depend also on the child's age (e.g., these notifications were assigned a lower priority and thus could be skipped if the child was younger than 12 months). Apart from the notifications, there was no intended direct contact between the parents and the researchers, but parents could contact the research team, and some did, mostly to report bugs or problems in the app.

3.3. Measures and data coding

3.3.1 The reported age of reaching early developmental milestones

In the app, parents provided their children's age (in months) of crawling, walking unassisted, and babbling, e.g., parents of children who were older than 10 months were asked whether their child had already started to babble (i.e. repeat single sounds, e.g. "ma-ma" or "da-da") and if yes, they were asked to recall at what age. For the remaining milestones, we relied on the words and utterances reported in the app: we calculated the age of the child (in months) at the moment when the parent reported the child's first word, 10th or 50th word, first multiword utterance (regardless of its length), first 3-word and 4-word utterance. The method of using multiple data points (i.e. 10, 50 words) was employed before in studies on cross-linguistic early language trajectories (e.g. de Boysson-Bardies & Vihman, 1991). The milestones we chose are commonly investigated both in language development research and practice: babbling (e.g. de Boysson-Bardies & Vihman, 1991; Oller et al., 1997), first word (Bloom et al., 1993; Bloom, 2001; Clark, 1993; Nelson, 1973; Rudolph & Leonard, 2016; Tamis-LeMonda et al., 2001), first 10 words (Tardif et al., 2008; Visser-Bochane et al., 2020), first 50 words (Bigelow, 1987; Bloom et al., 1993; Tamis-LeMonda et al., 2001), first multiword utterance (Bloom et al., 1993; Boerma & Blom, 2017).

3.3.2 What counted as a word

We operationalised a word spoken by the child as a word that a child has said a few times in a similar context that may have a few close meanings (e.g. "doggie" for a dog or any four-legged animal) and does not have to resemble the adult form. To minimise demands on parental linguistic awareness in order to respond, the app included a tooltip (i.e., hint, an element which, when clicked on, displays additional information) with information on what can be counted as a new word (see Figure C1 in Appendix C). Parents were also

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

encouraged to report the child's neologisms (lexical innovations). These were included in the present analyses as "words." We did so because the vast majority of these "neologisms" were in fact child forms (mispronounced or shortened versions of regular word-forms or onomatopoeia) and moreover, such child forms were equally often reported by parents as "words" and "neologisms," and thus combining the two categories seemed reasonable (see also Vihman and McCune (1994) for criteria of identifying early words). Parents always reported the form of the word, its meaning, and the language that the word/utterance was spoken in. If a word was marked as belonging to two languages at the same time (e.g. "jogart" meaning "jogurt" in Polish and "yoghurt" in English), in our analyses for the first research question (i.e., will bilinguals reach specific language milestones at a similar time as their Polish monolingual peers, at least in one of their languages?), we counted this word once. The same was done for utterances: if an utterance was reported as a mixture of two languages, we counted the utterance once. Note that for the first research question it did not matter which language the words/utterances were reported in, as we were interested in whether bilinguals reach a given milestone in any language at a similar age as their monolingual peers. In analyses for the second research question (i.e., will bilinguals reach specific language milestones at a similar time in both their languages, the home language (Polish) and the majority language?), if a word or utterance was marked as belonging to two languages at the same time, we counted it twice, i.e. for the home language (Polish) milestones and the majority language milestones.

3.3.3 What counted as a multiword utterance

We operationalised a multiword utterance as an utterance or phrase made of multiple words. The app also included a tooltip with some examples of multiword utterances, e.g. "gimme milk'" or "mama here" (see Figure C1 in Appendix C). The parents always reported a form of the utterance and its meaning. There were altogether 325 multiword utterances reported by parents of the final overall sample. We calculated the number of words per multiword utterance automatically by counting words in a string. We went through the reported words and phrases manually to ensure consistency and correct human error. For instance, some parents wrote "bye bye" as a two-word utterance, others as a single word ("byebye"). We coded all instances as the latter. Some multiword phrases were reported as a single (clustered) word, e.g. "mima" (short form of Polish *nie ma* "all gone") or "yowkay" (short of "you okay?"). These were kept as one-word utterances, in agreement with the parents' intuition. Some utterances that included multiple multiword phrases were divided and recoded. For example, the reported six-word phrase *Mama śpi, tata śpi, bobo śpi* ("Mom sleeping, dad sleeping, baby sleeping") was recoded as three two-word utterances. Multiword utterances that were in fact part of a song or rhyme and instances of counting and exclamations were discarded.

3.3.4 Input quantity and quality

Every 9 days, parents were encouraged (via notification) to report on the child's exposure to each language on the day before (see Figure C1 in Appendix C): the amount of time (in hours) the child heard each of their languages (input quantity), the amount of time (in minutes) spent on book reading, playing, singing, playing video/smartphone games, and other screen time in each language (proxy of input quality). Only 104 parents reported on the child's language environment, and seven did so more than once. The

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

data concerning children's language input was thus insufficient to perform the pre-registered analyses, but in the exploratory analyses below, we did attempt to link relative input in the home language to children's timing of reaching language milestones in both home and majority language (for a subgroup with any data available).

3.4. Data analysis

All calculations in the present paper were done in R version 4.2.2 (R Core Team, 2022). The scripts with R code and the anonymised data are available via OSF on GitHub (https://osf.io/3ruj9/). In accordance with the pre-registration, we conducted a series of *t*-tests with the Bonferroni-Holm correction (for multiple comparisons), and – since frequentist inference only provides evidence against the null hypothesis and cannot provide probabilistic evidence in favour of the null hypothesis – we also calculated a series of Bayes factors, following practices recommended by Hoijtink et al. (2019). The Bayes factor provides the ratio of the likelihood of the null hypothesis, BF_{01} (e.g. that bilinguals and monolinguals do not differ on the timing of reaching a particular milestone) to the likelihood of the alternative hypothesis, BF_{10} (e.g. that bilinguals and monolinguals differ on the timing of reaching a particular milestone), given the observed data. The value of the Bayes factor for a given hypothesis, e.g. 10, indicates that the support for this hypothesis in the observed data is 10 times larger than for the other hypothesis. A Bayes factor below 3 indicates "weak" evidence for a hypothesis, a Bayes factor between 3 and 10 indicates "moderate" or "substantial" evidence, a Bayes factor between 10 and 30 indicates "strong" evidence, and a Bayes factor above 30 indicates "very strong" evidence for the particular hypothesis (Jeffreys, 1998). All Bayesian calculations for group comparisons were done in R with package bain, version 0.2.10 (Gu et al., 2023). With the bain package, the default characteristics of the prior distribution and its variance are specified by using a fraction of the information in the data. Since we used a default prior, we also report a sensitivity analysis, showing the effect of varying the prior's settings: using the default minimal fraction (bg), twice the minimal fraction ($2 \times$ bg), and thrice the minimal fraction ($3 \times$ bg). By altering the base prior parameter values, we can check whether the results are robust or the posterior distribution changes dramatically. We also present the posterior probabilities of H0 and H1 after observing the data. The posterior probabilities add up to 1 and indicate how likely it is that the other hypothesis is still true, e.g., the posterior probability of H0 being 0.93 indicates that there is still a 7% chance that H1 (alternative hypothesis) is true. Posterior probabilities can be seen as Bayesian error probabilities.

4. Results

4.1. Early language milestones between bilinguals and monolinguals

Following the pre-registration, we first used a series of *t*-tests with the Bonferroni-Holm correction to compare the mean reported age of reaching the language milestones between bilinguals and monolinguals. We considered seven language milestones, i.e. babbling, the age of reporting the first word, 10th, 50th word, first multiword utterance, first 3-word, and 4-word utterance. For bilinguals, when we considered milestones such as first words or first utterances, we used the age of reporting these language milestones in whichever language came first (Polish /Norwegian /English / other). For milestones of 10th and 50th words, we counted both words reported in their home and majority language.

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

Before running the analyses, we excluded 12 outliers, 2 bilinguals, and 10 monolinguals (i.e. children with age reported values for a given milestone more than 1.5 interquartile range (IQR) above the third quartile or below the first quartile). The *t*-tests revealed no significant differences between the bilingual and monolingual groups in the mean age of babbling, age of reporting the first word, 10th word, 50th word, first multiword utterance, or first 3-word utterance (see Table 3 and Figure 1). We also calculated – in accordance with the pre-registered plan – a series of Bayes factors (see Table 3). We found strong evidence for the hypothesis stating no difference between the groups in reference to the reported age of the first word and the 10th word (with less than 10% chance that the alternative hypothesis is true). We found moderate evidence for the no difference hypothesis in the reported age of babbling and the age of reporting the first multiword utterance and 3-word utterance (with 10%–18% chance that the alternative hypothesis is true). The evidence was inconclusive in relation to the age of reporting the 50th word (with 24% chance that the alternative hypothesis is true). No statistical analysis was performed for the 4-word-utterance due to the very small sample size. Still, we report available data as described in the preregistration (see Table 3).

4.2. Exploratory analyses

4.2.1 Bilinguals' age of reaching milestones in HL vs. ML

Originally, we planned to compare the age of reaching language milestones between bilinguals' home and majority language for Polish-English and Polish-Norwegian bilinguals separately. However, in this sample, we were able to perform the comparisons only for one milestone, i.e., the age of reporting the first word. For the remaining milestones in the pre-registered bilingual group, we either collected data from single participants or none. We show their data in Appendix D, and here we present results from a larger sample, including bilinguals with majority languages other than English (in the UK) and Norwegian.

In our exploratory analyses, we compared the age of reaching milestones in the home vs. the majority language on the whole bilingual sample with various majority languages (not only English and Norwegian). For this unpre-registered sample, we ran a series of paired *t*-tests with Holm correction and a series of Bayes factors. We included only children for whom a given milestone was reported in both languages. We present the calculations only for two milestones: the age of reporting the first word and the 10th word. For the remaining milestones, we had data from very few participants (see Table 4 and Figure 2). We excluded one outlier (with age of reporting the first word 1.5 IQR below the first quartile). The *t*-test showed that the bilinguals did not differ significantly in the mean age of reporting the first word or the 10th word in their home and majority languages (see Table 4). The Bayes factor confirmed these results, yielding strong evidence for the no-difference hypothesis between the languages regarding the age of reporting the first word (with 9% chance that the alternative hypothesis is still true). The evidence for no difference between the languages regarding the age of reporting the 10th word was moderate and indicated 16% of chance that the alternative hypothesis was true.

4.2.2 Input effects on the age of reaching milestones in bilinguals

Originally, we planned to test whether in bilinguals, the language-specific (L1/L2) mean input quantity/quality before reaching a milestone is linked to the age of reaching that

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

Table 3. Age of reaching each milestone in bilinguals (with Polish as a home language and various majority languages) and a matched group of Polish monolinguals

	Bilinguals			Monolinguals								Sensitivity analysis	
Milestone	*n*	Age (*M*)	Age (*SD*)	*n*	Age (*M*)	Age (*SD*)	*t* statistic	*p* adj.	CI 95%	BF_{01} (posterior probab.)	BF_{10} (posterior prob.)	BF_{01} 2 x bg (posterior prob.)	BF_{01} 3 x bg (posterior prob.)
Babbling	58	7.31	2.64	54	6.76	1.80	$t(101) = 1.30$	1.00	[–0.29, 1.39]	7.43 (0.88)	0.13 (0.12)	5.26 (0.84)	4.29 (0.81)
1st word	140	14.92	5.53	137	15.22	5.13	$t(274) = –0.46$	1.00	[–1.56, 0.96]	14.95 (0.94)	0.07 (0.06)	10.57 (0.91)	8.63 (0.90)
10 words	50	16.82	4.34	47	16.83	3.75	$t(94) = –0.01$	1.00	[–1.64, 1.62]	9.52 (0.91)	0.11 (0.09)	6.73 (0.87)	5.50 (0.85)
50 words	15	18.60	3.72	14	20.00	1.92	$t(21) = –1.29$	1.00	[–3.66, 0.86]	3.12 (0.76)	0.32 (0.24)	2.21 (0.69)	1.80 (0.64)
1st multiword	34	19.21	3.62	41	19.29	3.72	$t(71) = –0.10$	1.00	[–1.78, 1.61]	8.58 (0.90)	0.12 (0.10)	6.07 (0.86)	4.96 (0.83)
1st 3-word utterance	14	20.36	3.03	18	21.44	1.98	$t(21) = –1.16$	1.00	[–3.03, 0.85]	4.59 (0.82)	0.22 (0.18)	3.24 (0.76)	2.65 (0.73)
1st 4-word utterance	3	17.67	0.58	6	22.83	1.94	–	–	–	–	–	–	–

Note. BF_{01} indicates support for the null hypothesis in the observed data. BF_{10} indicates support for the alternative hypothesis: $1/BF_{01}$. Posterior probability of a specified hypothesis is given in the parentheses.

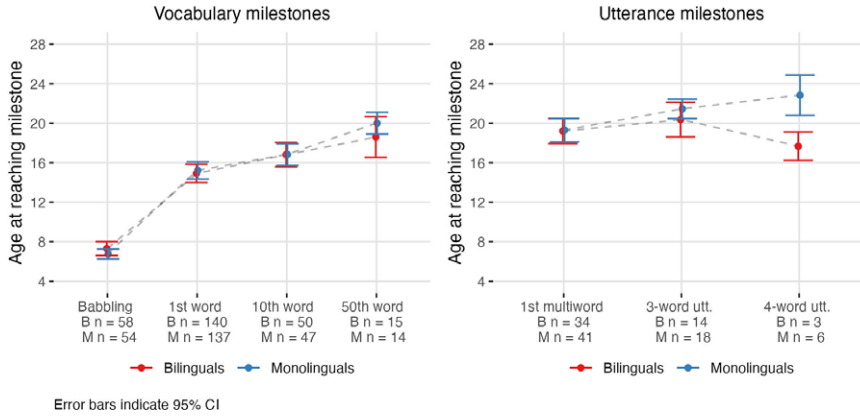

Figure 1. Age of reaching each milestone in bilinguals (with Polish as a home language and various majority languages) and a matched group of Polish monolinguals. Error bars indicate 95% confidence Intervals.

milestone (in L1/L2). Parents were asked to indicate how many hours the child heard each language on the previous day (and whether the previous day was typical to the child's regular experience). For the 302 bilinguals in our sample, we got information about input patterns from 58 parents (19%) who indicated that their estimates were representative of a typical child's day. Most of these estimates were collected only once, which hindered calculating input indices (as planned in the pre-registration). Whenever parents provided the input estimates more frequently than once, we used these values to calculate means. Parents varied in the sum of the waking hours they accounted for in providing input estimates – overall, they accounted for 8 to 24 hours of their child's day ($Mdn = 10$, $M = 11$, $SD = 6.8$ hours). To level out the variability in the total number of hours accounted for by the parents, we calculated the relative exposure to bilinguals' two languages (in percentage) by taking the estimate in the home language and dividing it by the sum of both estimates (i.e. input in the home language and input in the majority language). This way, we obtained a measure of relative input in the home language, Polish (percentage). The children in our bilingual sample varied largely in the relative input they received in Polish, min = 0%, max = 100%, $Mdn = 10\%$, $M = 65\%$, $SD = 34\%$, with 22 children (over ⅓ of all for whom input estimates were provided) hearing only Polish.

We used the relative input in the home language to run correlations with children's ages of reaching the following milestones: first word, 10th word, and first multiword utterance in both languages of the bilinguals. We did not include the 50th word milestone nor the 3-word and 4-word utterance milestones, since for these milestones we obtained data from few participants (see Table A5 in Appendix A). Neither of the correlations between the relative input in the home language and bilinguals' age of reaching milestones in either home or majority language were significant. We also ran a Bayes factor correlation using the BayesFactor package (Morey & Rouder, 2024) since the previously used package, bain, did not contain the relevant functionality. The evidence was inconclusive for the relation between relative exposure to the home language and the age of reaching the first word, the 10th word, and the first multiword utterance in either language (see Table A5 in Appendix A).

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

Table 4. Age of reaching each milestone in bilinguals (with various majority languages): home language (Polish) vs. majority language

	Home language (Polish)			Majority language								Sensitivity analysis	
Milestone	*n*	Age (*M*)	Age (*SD*)	*n*	Age (*M*)	Age (*SD*)	*t* statistic	*p* adj.	CI 95%	BF_{01} (posterior prob.)	BF_{10} (posterior prob.)	BF_{01} 2 x bg (posterior prob.)	BF_{01} 3 x bg (posterior prob.)
1st word	86	15.38	5.62	86	15.94	5.24	*t*(85) = –1.89	0.12	[–1.10, 0.03]	10.46 (0.91)	0.10 (0.09)	7.39 (0.88)	6.04 (0.86)
10 words	15	17.53	3.91	15	18.07	3.58	*t*(14) = –0.77	0.46	[–2.00, 0.96]	5.08 (0.84)	0.20 (0.16)	3.59 (0.78)	2.93 (0.75)
50 words	–	–	–	–	–	–	–	–	–	–	–	–	–
1st multiword utterance	7	20.29	2.56	7	21.14	3.89	–	–	–	–	–	–	–
1st 3-word utterance	1	23	–	1	23	–	–	–	–	–	–	–	–
1st 4-word utterance	–	–	–	–	–	–	–	–	–	–	–	–	–

Note. BF_{01} indicates support for the null hypothesis in the observed data. BF_{10} indicates support for the alternative hypothesis: $1/BF_{01}$. Posterior probability of a specified hypothesis is given in the parentheses

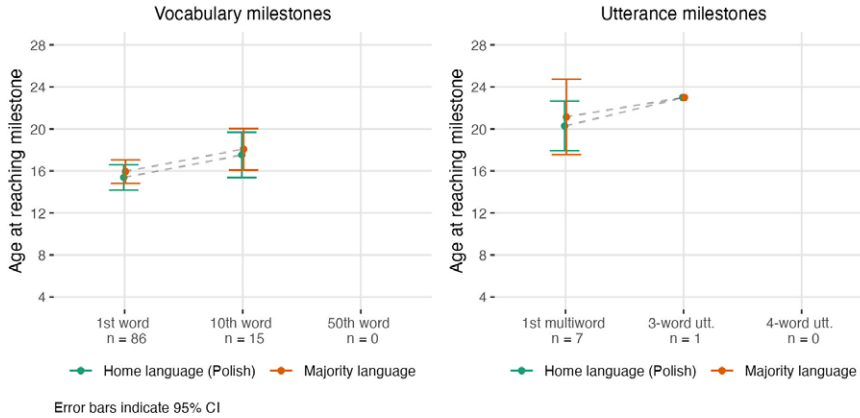

Figure 2. Age of reaching each milestone in bilinguals (with various majority languages): home language (Polish) vs. majority language. Error bars indicate 95% Confidence Intervals.

5. Discussion

This study responds to the scarcity of research that directly compares the timing of reaching early developmental milestones by bilingual and monolingual children. So far, most studies have focused on comparisons of bilingual and monolingual vocabulary growth. They reported similar developmental patterns in both groups (when total vocabulary measures were used) (e.g., Hoff, 2017). We asked parents of bilingual and monolingual children 0–24 months to continuously report their children's gestures, words, and utterances as they appeared in children's production using a mobile app. We also asked them to report or recall (if they had already occurred) some basic motor milestones (i.e., crawling, walking unassisted), and babbling. Our bilingual sample consisted of children living outside of Poland with the same home language (Polish) and differing majority languages; the monolingual sample included Polish monolingual children living in Poland. In our analyses, we compared the age of reaching language milestones between bilinguals and monolinguals and between the two languages of the bilinguals. We employed both frequentist and basic Bayesian statistics.

5.1. Do bilinguals reach specific language milestones at a similar time as their Polish monolingual peers, at least in one of their languages?

We found that the bilingual and monolingual children did not differ in the age of reaching most of the language milestones, i.e. babbling, age of reporting the first word, 10th word, and first multiword utterance production. This was confirmed by the Bayes factor analyses in comparisons between bilinguals and one monolingual group (indicating weak/moderate to strong evidence in favour of no group difference) and in comparisons of three groups: one bilingual group and two Polish monolingual groups matched to the bilingual group and selected from a very large monolingual sample (indicating strong and very strong evidence in favour of no group difference, see Appendix B). It should be noted here that the evidence we have collected was most robust for the first word, the 10th word, and the first multiword utterance. For the other milestones that required more engagement from the parent (i.e., 50th word, 3-word utterance, 4-word utterance), the samples

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

were smaller as fewer and fewer participants remained in the study. As a consequence, we cannot draw strong conclusions on these milestones, i.e. 50th word, 3-word utterance, and 4-word utterance. For the age of reporting the 50th word, evidence from Bayes was either inconclusive (bilinguals vs. one monolingual group) or moderate (bilinguals vs. two monolingual groups, see Appendix B). Given the existent literature on vocabulary growth in bilingual and monolingual children, we could expect that bilinguals would reach this milestone at a similar age as their monolingual peers if words from both languages would be taken into account (e.g. Core et al., 2013; Byers-Heinlein et al. 2024; Hoff et al., 2012; Miękisz et al., 2017) and that bilinguals might be slower in reaching that milestone in each of their languages, compared to monolinguals (e.g. Byers-Heinlein et al. 2024; Hoff et al., 2012). For the first 3-word utterance, evidence from Bayes was either weak (bilinguals vs. one monolingual group) or inconclusive (bilinguals vs. two monolingual groups, see Appendix B). Small sample sizes for the last milestone, i.e. the age of reporting the first 4-word utterance, hindered a group comparison in that aspect. In essence, our evidence does not rule out bilingual differences (relative to monolingual peers) in the area of these milestones: the 50th word, producing 3-word and 4-word utterances.

Nevertheless, we believe our study provides an important contribution to both psycholinguistic theory and practice. As for the psycholinguistic theory, our results corroborate findings from other aspects of bilingual development (e.g. Höhle et al., 2020; Werker & Byers-Heinlein, 2008), showing that children are equally prepared to acquire two languages or one, even if bilinguals are tasked with a bigger challenge than their monolingual peers. This suggests that early language development is rather robust against variation in the amount of input characteristic to bilingual experience (see Höhle et al., 2020 for a similar conclusion). It is not to say that input does not matter for the trajectory of early language development. Rather, it is the combination of language input and linguistic processing skills that drive a child's language development. Until now, most of the evidence in this regard came either from small-scale studies (Petitto et al., 2001) or studies focused on selected milestones (Oller et al., 1997; Hoff et al., 2012). We found the same pattern of results when we compared our bilingual group to two Polish monolingual groups, strengthening our inferences regarding the bilingual and monolingual age of reaching language milestones. Moreover, previous research in this area focused on vocabulary growth (e.g. Hoff, 2017; Hoff & Core, 2013; Pearson et al., 1993), and while such evidence is highly informative, the present evidence directly shows that bilingual children do not display a developmental delay at the beginning of productive language use. This may encourage psycholinguistic research to investigate the potential compensatory strategies that bilingual children employ (e.g. Fennell et al., 2007), or explore other factors, e.g. maternal responsiveness, that have previously been shown to be important in a child's timing of reaching early language milestones.

For practitioners, our results propose that when a screening or diagnosis of a bilingual child is performed, one may expect that the typically developing child would reach at least some of the language milestones, those not referring to large vocabulary growth, i.e. babbling, first word, 10th word, and first multiword utterance, at the same age as conventional for monolingual children, if production in both languages is taken into account. At the same time, more research is needed to establish whether the milestone of 50 words (which is a recognized milestone in monolingual child development, e.g. Stanford's Age-Appropriate Speech and Language Milestones; ALSHA's Communication Milestones) is similarly informative for bilinguals.

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

5.2. Do bilinguals reach specific language milestones at a similar time in both their languages, the home language (Polish) and the majority language?

Second, we set out to compare the development of the home and the majority language in bilingual children. It should be noted here that due to insufficient data, especially in the later milestones referring to the 50th word and the first multiword utterances, our preregistered analyses (on subgroups of children by their majority language) could not be performed. In our exploratory analyses of the whole bilingual sample, the frequentist approach did not reveal significant differences in the mean age of reporting the first word or the 10th word in two languages of bilinguals. The Bayes factor analyses yielded strong evidence for no difference between the languages in the age of the first word, but the evidence was weak for the 10th word milestone. These results are partly in line with Petitto et al.'s (2001) case study in which no major differences were identified between the two languages of bilingual and bimodal children in the first words, first phrases, and first 50 words milestones. However, the children observed by Petitto et al. had their language exposure equally divided between their two languages, while in our case, the bilingual children were experiencing various language exposure patterns at home. The bilinguals varied largely in the relative amount of input they received in Polish – half of the children (for whom input estimates were provided) were hearing Polish for less than 10% of their waking time, and over ⅓ of all children were hearing only Polish. Despite that, the participants in the present study exhibited no cross-language difference in the age of the first reported word. Given the heterogeneity of the bilingual experience captured in the present sample, it is meaningful that the classic milestone of first word production was attained at a similar age in bilinguals' two languages. This again strengthens the conclusion that early language development is resilient to bilingual input variation. However, it is to be expected that input gains more importance in the area of later language milestones, as it has been generally found that exposure to each language (both its quantity and quality) exerts strong influence on language-specific vocabulary growth (e.g. Byers-Heinlein et al. 2024; De Houwer, 2011; Hoff, 2017). Consequently, bilingual children with unequal exposure to their languages may show large variability in their vocabulary growth, as found in previous studies (e.g. Core et al., 2013; Hoff, 2017; Marchman et al., 2009). In our study, the evidence for no between-language difference was weak in the age of the 10th reported word, and as for further milestones, the available data was virtually nonexistent, hindering further comparisons. Thus, we cannot assume that bilinguals would go on to reach the next milestones at similar ages in both languages. We have also attempted to consider bilinguals' relative input in the home language (Polish) and its relation to the age of reaching language-specific milestones. However, this exploratory analysis was done on a very small subset of children – at best, in the milestone of the first word, we had input data from 58 children and samples smaller than 20 in later milestones. Also, while the frequentist analysis did not allow us to reject the no-cross-language-difference hypothesis, the results from the Bayes factor yielded inconclusive evidence. It may be that our aim of collecting regular and frequent information on input patterns in the bilingual family was too effortful for parents who were already asked to regularly report their child's new gestures, words, and phrases.

5.3. Observations from an online study

When setting out, we expected that online data collection with the use of a mobile app (together with wide-spread recruitment via social media and other media outlets) would allow us to reach a large and linguistically heterogeneous sample and gather data

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

longitudinally and continuously. Though we did manage to attract a wide group of parents in the study (over 5,470 families downloaded the app), we found that providing parents with a mobile app means little control over their engagement in the study. First, we observed low long-term engagement in the app, with only 5% of parents reporting their children's progress over 6 months. Notably, similar retention rates are reported for other research apps. Some studies have shown 15-day and 30-day retention rates as low as 3–4% (Baumel et al., 2019). In a cross-evaluation of more than 100,000 participants from eight remote digital health studies, the median participant retention was only 5.5 days out of a cumulative participation of 850,000 days (Pratap et al., 2020). Importantly though, the number of parents reporting children's first word and first 10-word milestones in our study was considerably higher and thus more generalisable. Second, we received a low rate of responses to the questions we sent out to parents, especially concerning children's linguistic environment. Even though we designed a complex notification system and gamification features to increase parental engagement in the study (see Mieszkowska et al., 2022), for 59% of the children, we did not obtain any reports of words or multiword utterances (over the whole course of data collection, i.e. 2021–2023). These were parents that downloaded the app, provided consent to take part in the study, but did not report anything. For 21% of the children, their parents reported their child's language (e.g., words) only upon downloading the app; for 9% of the children; parents were reporting for up to 30 days; for 7% of the children, parents were reporting up to half a year; and 4% of the users used the app for longer than half a year. We believe that data collected with a mobile app, which entails no direct contact between researchers and parents, means that the data that we collect is largely driven by parental interest. Many parents are interested in tracking their children's words, or early development, but fewer parents are interested in reporting their children's language exposure patterns, leading to a significant lack of data in that aspect. Also, the parents who remained in the study for longer periods likely represent a specific (engaged, motivated) profile that is not generalisable to the entire population.

Trying to mitigate the expected small retention rate, we included some gamification features in the app to extend or increase user engagement. These were carefully planned so that we would not alter parental behaviours towards their children and so that we would reward the users' engagement in the app and not their reports of new words, which could lead to falsified data or frustration in parents of late-talkers. There was a separate "Knowledge" tab in the app with various resources (e.g. podcasts with researchers on general child development, popular-scientific articles on raising children, see Figure C1 in Appendix C), but they did not manage to substantially guarantee user engagement. It could be that these materials did not meet the parental interests and that they would be more interested, e.g. in seeing their child's progress against other children or norming data. However, there were a few reasons we decided against that. First, such information in the app could falsely alarm or upset parents of late talkers or even encourage parents to over-report. Second, the app should not be treated as a replacement for a professional diagnosis or include information that could be interpreted in that way. And lastly, in some languages, including Polish, there are no normed standards available for bilingual children that take into account the heterogeneity of the bilingual experience.

6. Limitations

The first limitation is that the data presented here is not fully longitudinal: for some children, we obtained data only for some of the early milestones (which is visible in the way the sample sizes decrease from early to late milestones). The second limitation is

https://doi.org/10.1017/S0305000924000655 Published online by Cambridge University Press

closely linked to the first one, i.e. due to the high drop-out rate and missing data, some of the pre-registered analyses could not be performed, and others were performed on relatively small samples (e.g. the 50th word in the bilingual vs. monolingual comparison), and as such should be approached with caution. Also, neither of the statistical strategies applied (frequentist or Bayesian) are resistant to small sample sizes. However, we provided effect sizes (for the frequentist) and posterior probabilities (for Bayesian analyses) to show the relative certainty of the obtained results. We also compared our bilingual sample to two monolingual samples (see Supplementary Materials) and obtained similar results of no group difference, thus strengthening our inference. Finally, the small percentage of parents who remained in the study for over six months likely represents a specific profile that is not generalisable to the entire population (i.e., parents particularly interested in the topic and highly motivated).

7. Conclusion

Our study provides new evidence that bilingual and monolingual children can achieve their early language milestones at comparable ages. The findings complement earlier research that primarily focused on vocabulary growth and enrich our understanding of bilingual language acquisition. By employing new data collection methods (i.e. mobile app), our study contributes valuable insights on how such methods might be used to offer real-time insights into language acquisition in a digital age. Using a mobile app and a fully online study protocol allowed us to attract a worldwide group of parents in the study (over 5,470 families downloaded the app from all over the world). Even though we have struggled with keeping our participants engaged in the study for longer periods of time, the amount of data was sufficient to create closely matched groups of bilinguals and monolinguals and to obtain robust results for early language milestones, i.e. first word, 10 words, and first multiword utterance. Further research using similar methods should reconsider parental involvement in a longer perspective (months rather than weeks), possibly by adding regular direct contact with the research team and/or providing rewards that would be, on the one hand, attractive to parents but, on the other hand, not alter their behaviours towards their children. Feedback on actual data provided by the parents (e.g. child's scores relative to the population) could be most attractive to parents, but is possible only when norms are available. Additionally, not only written reports but also recordings could be considered (this however would be a challenge in terms of personal data protection). Overall, we see big potential in using mobile apps for gathering large numbers of parental reports on children's language, especially considering mobility and migration.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0305000924000655.

Acknowledgements. The authors would like to express their gratitude to all the families who voluntarily took part in this study. We would also like to thank Alicja Kost (@mataja_blog), Ula Petrycka (@kiedy_do_logopedy) and Gosia Dvořáková (@gazelawlaponii) for their invaluable help and support in conducting this study.

Author contribution. Conceptualization: KM, GK, AD, NGG, ASHR, MŁ, KA, NK, HGS, PH, KS, EH; Data curation: KM, GK, KS; Formal analysis: KM; Funding acquisition: KM, GK, AD, NGG, ASHR, MŁ, KA, NK, HGS, EH; Investigation: KM, GK, AD, NGG, ASHR, MŁ, JK, KS, EH; Methodology: KM, GK, AD, NGG, ASHR, MŁ, KA, NK, JK, HGS, PH, MK, KS, EH; Project administration: KM, GK, EH; Software: KM, GK, ASHR, MK, KS; Supervision: KM, NGG, EH; Validation: KM, GK, KS; Visualization: KM, MK; Writing – original draft: KM, GK, AD, MŁ, JK, EH; Writing – review & editing: KM, GK, AD, NGG, ASHR, MŁ, KA, NK, JK, HGS, PH, MK, KS, EH.

Funding. This study was funded by the National Science Center (grant no. 2018/31/B/HS6/03916) granted to Ewa Haman. Partial funding was also obtained from the Research Council of Norway through its Centres of Excellence funding scheme (project number 223265), and BWPs awarded to KM, EH, and GK in years 2021–2024

Competing interest. The authors declare none.

Disclosure of use of AI tools. No AI tools were used in the preparation of the present manuscript.

References

Age-appropriate speech and language milestones. (n.d.). Retrieved August 7, 2024, from https://www.stanfordchildrens.org/en/topic/default?id=age-appropriate-speech-and-language-milestones-90-P02170

Alcock, K. J., & **Brennan, E.** (2011, September). *Parents' memory for early language – Relationship between 18 month CDI scores and retrospective scores at 7 years.* British Psychological Society, Developmental Section, Newcastle. Center For Open Science. https://osf.io/dv6bu/

Baumel, A., **Muench, F.**, **Edan, S.**, & **Kane, J. M.** (2019). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *Journal of Medical Internet Research*, **21**(9), e14567. https://doi.org/10.2196/14567

Bigelow, A. (1987). Early words of blind children. *Journal of Child Language*, **14**(1), 47–56. https://doi.org/10.1017/S0305000900012721

Bloom, L., **Tinker, E.**, & **Margulis, C.** (1993). The words children learn: Evidence against a noun bias in early vocabularies. *Cognitive Development*, **8**(4), 431–450. https://doi.org/10.1016/S0885-2014(05)80003-6

Bloom, P. (2001). How children learn the meanings of words. *Behavioral and Brain Sciences*, **24**(06), 1095–1103.

Boerma, T., & **Blom, E.** (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, **66**, 65–76. https://doi.org/10.1016/j.jcomdis.2017.04.001

Bosch, L., **Figueras, M.**, **Teixidó, M.**, & **Ramon-Casas, M.** (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: Evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology*, **4**, 106.

Bosch, L., & **Sebastián-Gallés, N.** (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy: The Official Journal of the International Society on Infant Studies*, **2**(1), 29–49.

Byers-Heinlein, K., **Gonzalez-Barrero, A. M.**, **Schott, E.**, & **Killam, H.** (2024). Sometimes larger, sometimes smaller: Measuring vocabulary in monolingual and bilingual infants and toddlers. *First Language*, **44**(1), 74–95.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511554377

Communication Milestones: 19 to 24 Months. (n.d.). American Speech-Language-Hearing Association. Retrieved August 7, 2024, from https://www.asha.org/public/developmental-milestones/communication-milestones-19-to-24-months/

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)

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research*, **34**(3), 565–571. https://doi.org/10.1044/jshr.3403.565

de Boysson-Bardies, B., & **Vihman, M. M.** (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, **67**(2), 297–319. https://doi.org/10.2307/415108

De Houwer, A. (2011). Language input environments and language development in bilingual acquisition. *Applied Linguistics Review*, **2**(2011).

De Houwer, A. (2023). The danger of bilingual–monolingual comparisons in applied psycholinguistic research. *Applied Psycholinguistics*, **44**(3), 343–357.

DeAnda, S., **Hendrickson, K.**, **Zesiger, P.**, **Poulin-Dubois, D.**, & **Friend, M.** (2018). Lexical access in the second year: A study of monolingual and bilingual vocabulary development. *Bilingualism: Language and Cognition*, **21**(2), 314–327. https://doi.org/10.1017/S1366728917000220

Fennell, C. T., **Byers-Heinlein, K.**, & **Werker, J. F.** (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*, **78**(5), 1510–1525. https://doi.org/10.1111/j.1467-8624.2007.01080.x

Fenson, L., **Marchman, V. A.**, **Thal, D. J.**, **Dale, P. S.**, **Reznick, J. S.**, & **Bates, E.** (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual* (2nd ed.). Brookes Publishing.

Genesee, F. (2022). The monolingual bias: A critical analysis. *Journal of Immersion and Content-Based Language Education*, **10**(2), 153–181.

Gervain, J., & **Mehler, J.** (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, **61**(1), 191–218.

Gu, X., **Hoijtink, H.**, **Mulder, J.**, & **van Lissa, C.** (2023). *bain: Bayes factors for informative hypotheses.* (R package version 0.2.10) [Computer software]. https://CRAN.R-project.org/package=bain

Hoff, E. (2017). How bilingual development is the same as and different from monolingual development. *OLBI Working Papers*, *8*. https://uottawa.scholarsportal.info/ojs/index.php/ILOB-OLBI/article/view/2114

Hoff, E., & **Core, C.** (2013). Input and language development in bilingually developing children. *Seminars in Speech and Language*, **34**(4), 215–226. https://doi.org/10.1055/s-0033-1353448

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), 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.e1. https://doi.org/10.1016/j.jpeds.2017.06.071

Höhle, B., **Bijeljac-Babic, R.**, & **Nazzi, T.** (2020). Variability and stability in early language acquisition: Comparing monolingual and bilingual infants' speech perception and word recognition. *Bilingualism: Language and Cognition*, **23**(1), 56–71. https://doi.org/10.1017/S1366728919000348

Hoijtink, H., **Mulder, J.**, **van Lissa, C.**, & **Gu, X.** (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, **24**(5), 539–556. https://doi.org/10.1037/met0000201

Hus, V., **Taylor, A.**, & **Lord, C**. (2011). Telescoping of caregiver report on the autism diagnostic interview – Revised. *Journal of Child Psychology and Psychiatry*, **52**(7), 753–760. https://doi.org/10.1111/j.1469-7610.2011.02398.x

Jaspers, M., **de Meer, G.**, **Verhulst, F. C.**, **Ormel, J.**, & **Reijneveld, S. A.** (2010). Limited validity of parental recall on pregnancy, birth, and early childhood at child age 10 years. *Journal of Clinical Epidemiology*, **63**(2), 185–191. https://doi.org/10.1016/j.jclinepi.2009.05.003

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Junge, C., **Kooijman, V.**, **Hagoort, P.**, & **Cutler, A.** (2012). Rapid recognition at 10 months as a predictor of language development: Rapid recognition. *Developmental Science*, **15**(4), 463–473.

Kidd, E., & **Garcia, R.** (2022). How diverse is child language acquisition research? *First Language*, **42**(6), 703–735. https://doi.org/10.1177/01427237211066405

Legacy, J., **Zesiger, P.**, **Friend, M.**, & **Poulin-Dubois, D.** (2016). Vocabulary size, translation equivalents, and efficiency in word recognition in very young bilinguals. *Journal of Child Language*, **43**(4), 760–783.

Lipkin, P. H., **Macias, M. M.**, **Council on Children With Disabilities, Section on Developmental and Behavioral Pediatrics**. (2020). Promoting optimal development: Identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics*, **145**(1), e20193449. https://doi.org/10.1542/peds.2019-3449

Majnemer, A., & **Rosenblatt, B.** (1994). Reliability of parental recall of developmental milestones. *Pediatric Neurology*, **10**(4), 304–308. https://doi.org/10.1016/0887-8994(94)90126-0

Marchman, V. A., **Fernald, A.**, & **Hurtado, N.** (2009). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish–English bilinguals. *Journal of Child Language*, **37**(04), 817–840. https://doi.org/10.1017/S0305000909990055

Mieszkowska, K., **Krajewski, G.**, **Sobota, K.**, **Dynak, A.**, **Kolak, J.**, **Krysztofiak, M.**, **Łukomska, B.**, **Łuniewska, M.**, **Garmann, N. G.**, **Hansen, P.**, **Romøren, A. S. H.**, **Simonsen, H. G.**, **Alcock, K.**, **Katsos, N.**, & **Haman, E.** (2022). Parental report via a mobile app in the context of early language trajectories: Starwords study protocol. *International Journal of Environmental Research and Public Health*, **19**(5), 3067. https://doi.org/10.3390/ijerph19053067

Moore, C., **Dailey, S.**, **Garrison, H.**, **Amatuni, A.**, & **Bergelson, E.** (2019). Point, walk, talk: Links between three early milestones, from observation and parental report. *Developmental Psychology*, **55**(8), 1579–1593.

Morey, R., **Rouder, J.**, (2024). BayesFactor: Computation of bayes factors for common designs. R package version 0.9.12-4.7, https://CRAN.R-project.org/package=BayesFactor

Miękisz, A., **Haman, E.**, **Łuniewska, M.**, **Kuś, K.**, **O'Toole, C.**, & **Katsos, N.** (2017). The impact of a first-generation immigrant environment on the heritage language: productive vocabularies of Polish toddlers living in the UK and Ireland. *International Journal of Bilingual Education and Bilingualism*, **20**(2), 183–200. https://doi.org/10.1080/13670050.2016.1179259

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, **38**(1/2), 1–135. https://doi.org/10.2307/1165788

Ohashi, J. K., **Mirenda, P.**, **Marinova-Todd, S.**, **Hambly, C.**, **Fombonne, E.**, **Szatmari, P.**, **Bryson, S.**, **Roberts, W.**, **Smith, I.**, **Vaillancourt, T.**, **Volden, J.**, **Waddell, C.**, **Zwaigenbaum, L.**, **Georgiades, S.**, **Duku, E.**, & **Thompson, A.** (2012). Comparing early language development in monolingual- and bilingual- exposed young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, **6**(2), 890–897. https://doi.org/10.1016/j.rasd.2011.12.002

Oller, D. K., **Eilers, R. E.**, **Urbano, R.**, & **Cobo-Lewis, A. B.** (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language*, **24**(2), 407–425. https://doi.org/10.1017/S0305000997003097

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, **50**(4), 793–817.

Paradis, J., **Genesee, F.**, & **Crago, M. B.** (Eds.). (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning* (2nd ed). Paul H. Brookes Pub. Co.

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

Petitto, L. A., **Katerelos, M.**, **Levy, B. G.**, **Gauna, K.**, **Tétreault, K.**, & **Ferraro, V.** (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language*, **28**(2), 453–496. https://doi.org/10.1017/S0305000901004718

Pontecorvo, E., **Higgins, M.**, **Mora, J.**, **Lieberman, A. M.**, **Pyers, J.**, & **Caselli, N. K.** (2023). Learning a sign language does not hinder acquisition of a spoken language. *Journal of Speech, Language, and Hearing Research*, **66**(4), 1291–1308.

Pratap, A., **Neto, E. C.**, **Snyder, P.**, **Stepnowsky, C.**, **Elhadad, N.**, **Grant, D.**, **Mohebbi, M. H.**, **Mooney, S.**, **Suver, C.**, **Wilbanks, J.**, **Mangravite, L.**, **Heagerty, P. J.**, **Areán, P.**, & **Omberg, L.** (2020). Indicators of retention in remote digital health studies: A cross-study evaluation of 100,000 participants. *NPJ Digital Medicine*, **3**, 21.

Paradis, J., & **Jia, R.** (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, **20**(1), e12433. https://doi.org/10.1111/desc.12433

R Core Team. (2022). *R: A language and environment for statistical computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Reznick, J. S., & **Goldfield, B. A.** (1994). Diary vs. representative checklist assessment of productive vocabulary. *Journal of Child Language*, **21**(2), 465–472. https://doi.org/10.1017/S0305000900009351

Roberts, M. Y., **Sone, B. J.**, **Jones, M. K.**, **Standley, M.**, **Conner, T.**, **Lee, E. D.**, **Norton, E. S.**, **Roman, J.**, **Speights, M.**, **Young, R.**, & **Weisleder, A.** (2023). What the evidence does (and does not) show for the Centers for Disease Control and Prevention child development milestones: An illustrative example using expressive vocabulary. *Journal of Speech, Language, and Hearing Research*, **66**(9), 3622–3632. https://doi.org/10.1044/2023_JSLHR-23-00020

Rudolph, J. M., & **Leonard, L. B.** (2016). Early language milestones and specific language impairment. *Journal of Early Intervention*, **38**(1), 41–58. https://doi.org/10.1177/1053815116633861

Reyes, L., & **Manalich, R.** (2005). Long-term consequences of low birth weight. *Kidney International.* Supplement, **97**, S107–11. https://doi.org/10.1111/j.1523-1755.2005.09718.x

Scharf, R. J., **Scharf, G. J.**, & **Stroustrup, A.** (2016). Developmental milestones. *Pediatrics in Review*, **37**(1), 25–37.

Singh, L. (2018). He said, she said: Effects of bilingualism on cross-talker word recognition in infancy. *Journal of Child Language*, **45**(2), 498–510.

Tamis-LeMonda, C. S., **Bornstein, M. H.**, & **Baumwell, L.** (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, **72**(3), 748–767. https://doi.org/10.1111/1467-8624.00313

Tardif, T., **Fletcher, P.**, **Liang, W.**, **Zhang, Z.**, **Kaciroti, N.**, & **Marchman, V. A.** (2008). Baby's first 10 words. *Developmental Psychology*, **44**(4), 929–938. https://doi.org/10.1037/0012-1649.44.4.929

Thordardottir, Elin (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, **23**(1), 236–255.

Treharne, D. A. (1992). Parental recall of children's early development. *European Journal of Disorders of Communication*, **27**(3), 221–230. https://doi.org/10.3109/13682829209029422

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 103–121). American Psychological Association. https://doi.org/10.1037/14939-007

Vihman, M. M., & **McCune, L.** (1994). When is a word a word? *Journal of Child Language*, **21**(3), 517–542. https://doi.org/10.1017/S0305000900009442

Visser-Bochane, M. I., **Reijneveld, S. A.**, **Krijnen, W. P.**, **van der Schans, C. P.**, & **Luinge, M. R.** (2020). Identifying milestones in language development for young children ages 1 to 6 years. *Academic Pediatrics*, **20**(3), 421–429. https://doi.org/10.1016/j.acap.2019.07.003

Werker, J. F., & **Byers-Heinlein, K.** (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in Cognitive Sciences*, **12**(4), 144–151. https://doi.org/10.1016/j.tics.2008.01.008

West, K. L., & **Iverson, J. M.** (2021). Communication changes when infants begin to walk. *Developmental Science*, **24**(5), e13102.

Cite this article: Muszyńska, K., Krajewski, G., Dynak, A., Garmann, N.G., Romøren, A.S.H., Łuniewska, M., Alcock, K., Katsos, N., Kołak, J., Simonsen, H.G., Hansen, P., Krysztofiak, M., Sobota, K., & Haman, E. (2025). Bilingual children reach early language milestones at the same age as monolingual peers. *Journal of Child Language* 1–24, https://doi.org/10.1017/S0305000924000655