Journal of Child Language (2025), **52**, 709–728
doi:10.1017/S0305000924000199

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Examining Dutch children's vocabularies across infancy and toddlerhood: Demographic effects are age-specific and task-specific

Anika van der Klis[1,2], Caroline Junge[2], Frans Adriaans[1] and René Kager[1]

[1]Institute for Language Sciences, Department of Languages, Literature and Communication, Utrecht University and [2]Experimental Psychology, Department of Developmental and Experimental Psychology, Utrecht University

Corresponding author: Anika van der Klis; Email: a.vanderklis@uu.nl

(Received 20 July 2023; revised 08 April 2024; accepted 21 April 2024)

Abstract

Limited studies have examined demographic differences in children's vocabulary in longitudinal samples, while there are questions regarding the duration, direction, and magnitude of these effects across development. In this longitudinal study, we included over 400 Dutch children. Caregivers filled out N-CDIs when children were 9–11 months (measuring word comprehension, word production, and gestures) and around 2–5 years of age (measuring word production). At 2–5 years, we also administered a receptive vocabulary task in the lab. We examined demographic effects on vocabulary size across infancy and toddlerhood. We found a disadvantage for males in infants' gestures and toddlers' vocabulary production. We found a negative effect of maternal education on infants' caregiver-reported vocabulary, but a positive effect on toddlers' lab-administered receptive vocabulary. Lastly, we found a negative effect of multilingualism – but only for the lab-administered task. Examining predictors in large, longitudinal samples ensures their robustness and generalisability across development.

Keywords: language assessment; vocabulary development; caregiver reports; demographic effects; YOUth Cohort Study

Introduction

Many studies have shown that the Communicative Development Inventories (CDIs) are reliable and valid checklists for measuring early vocabulary across a wide range of participants (Feldman et al., 2005; Fenson et al., 2007; Frank et al., 2021). Administering CDIs is a standardised, fast, and cost-effective approach that does not require trained lab assistants, lab visits, or the labour-intensive transcription of speech that is required for analysing naturalistic speech samples. This allows for larger sample sizes which is beneficial, especially when examining demographic effects on vocabulary development. Demographic effects typically only capture a small proportion of the large variance in children's vocabulary size, and most significantly in the early years (e.g., Eriksson et al., 2012; Fenson et al., 2007; Kidd & Donnelly, 2020). Previous studies have identified key demographic predictors – including maternal education, children's gender, gestational age, birth weight, and multilingualism – but there are uncertainties regarding the duration, direction, and magnitude of these effects on different vocabulary outcome measures across development. One advantage of cohort studies is that we can re-evaluate previous research findings in a large sample with repeated measurements to ensure their robustness and generalisability. Longitudinal data also provide insights into how these predictors unfold over time. In the current study, we make use of the large, longitudinal YOUth cohort study that measured vocabulary between infancy and toddlerhood of over four hundred Dutch children. This cohort with multiple different vocabulary measurements over time provided us with an excellent opportunity to study the direction and magnitude of demographic predictors across early development.

Maternal education

Maternal education is often used as a proxy for socio-economic status (SES). Previous studies often reported positive effects of maternal education on children's vocabularies measured by CDIs for toddlers (e.g., Feldman et al., 2000; Fenson et al., 2007; but cf. Reese & Read, 2000; Kuvač-Kraljević et al., 2021). Mothers with higher educational backgrounds produce a higher quantity (i.e., they speak more) and quality (i.e., they use more diverse language) of speech towards their children, mediating the positive relationship between maternal SES and children's language development (Hoff, 2003; Huttenlocher et al., 2010). However, studies employing CDIs have frequently observed negative correlations between maternal education and children's vocabularies during infancy (e.g., Bavin et al., 2008; Feldman et al., 2000; Fenson et al., 2007; Reese & Read, 2000). This early negative effect of maternal education on CDIs is likely driven by a caregiver reporting bias: a negative effect of SES is more often reported for vocabulary comprehension which requires more interpretation by the caregiver than vocabulary production, although a negative effect is sometimes reported for production as well (Bavin et al., 2008; Reese & Read, 2000). In contrast, studies rarely report a negative effect of SES on the gesture scale (Bavin et al., 2008; Feldman et al., 2000; Rowland et al., 2022). Determining whether a child produces a word or gesture does not require the caregiver to draw inferences about the child's understanding. In addition, there are fewer expectations from caregivers surrounding children's gesture development compared to their vocabulary development. On the one hand, caregivers could believe that larger vocabularies are more desirable – leading to over-reporting of their infants' vocabularies, or because some caregivers have more liberal criteria for word comprehension than others (see Feldman et al., 2000; Tomasello & Mervis, 1994 for discussions). On the other hand, caregivers may underestimate what their children already know when their children do not produce many words yet (see Houston-Price et al., 2007). These findings make it relevant to study the effects of maternal education in large, longitudinal samples throughout the first years of development on a variety of vocabulary measures.

Children's gender

Many studies have identified small effects of children's gender[1]. More specifically, girls tend to outperform boys on many vocabulary scales (e.g., Eriksson et al., 2012; Feldman et al., 2005; Frank et al., 2021; Reese & Read, 2000; Reilly et al., 2009; Zink & Lejaegere, 2002, but cf. Bavin et al., 2008). Simonsen et al. (2014) showed that boys are characterised by a less steep increase in receptive vocabulary growth than girls – at least until 20 months of age. Feldman et al. (2000) examined over 2,000 American English children using CDIs and reported lower scores for boys in vocabulary production and vocabulary comprehension across children aged 10–13 months. These differences persisted for older children, except for vocabulary comprehension. Girls have also been found to have larger gesture repertoires than boys based on CDIs (Feldman et al., 2000; Germain et al., 2022; Simonsen et al., 2014; Zink & Lejaegere, 2002). These studies suggest that overall, girls have faster developmental trajectories than boys. In contrast, previous studies using naturalistic speech samples or lab-administered tasks of children's receptive vocabularies typically do not report gender differences in diverse samples (e.g., Huttenlocher et al., 2010; Pan et al., 2004; Washington & Craig, 1999), although these findings are inconsistent, particularly for children's expressive language skills where girls tend to outperform boys (e.g., Bornstein et al., 1998; Frank et al., 2021; Qi et al., 2003; but cf. Bergelson et al., 2023). The effect of gender could be small and variable across children's ages and vocabulary measures, causing inconsistent results across studies.

Gestational duration and birth weight

Some studies suggest that preterm children are at a larger risk of having smaller vocabularies than full-term children (e.g., Foster-Cohen et al., 2007; Guarini et al., 2009; Sansavini et al., 2011, but cf. Ogneva & Pérez-Pereira, 2023). There may be negative effects only in extremely or very preterm children. Kern and Gayraud (2007) found that very preterm (28–32 weeks) and extremely preterm (under 28 weeks) children had smaller vocabulary sizes based on CDIs than moderately preterm (33–36 weeks) and full-term children when they were assessed at 24–26 months of age. However, Pérez-Pereira and Cruz (2018) found that gestational age did not affect vocabulary growth in a sample of low-risk preterm children with a wide range of gestational ages and birth weights without other medical complications. Still, a meta-analysis showed that very preterm (under 32 weeks) and/or very low birth weight (under 1500 g) children have persistent language delays (Barre et al., 2011). Moreover, differences between preterm and full-term children in gestural and lexical development become increasingly more evident during the first two years of life (Sansavini et al., 2011; van Baar et al., 2006). Previous studies have to our knowledge not concurrently examined the effects of gestational duration and birth weight, and it remains a question whether these factors influence children's vocabulary development in a non-clinical sample. It also remains largely understudied whether vocabulary differences between preterm and full-term children are apparent during the first year of life. Therefore, it is relevant to study the effects of gestational age and birth weight in a large, longitudinal sample starting from infancy.

[1]In this research, we use the term "gender" because we are primarily interested in the psychological, behavioural, social, and cultural differences across genders, rather than differences based on biological sex alone.

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

Multilingualism

In many studies examining children's vocabularies using the CDIs, multilingual children are excluded. CDI norming samples also typically exclude multilingual children, while being multilingual is the norm in most places across the world. Therefore, it is important to assess how multilingualism affects children's performance on a variety of widely used vocabulary tasks. When assessing only one language, multilingual children have smaller vocabularies than their monolingual peers (Blom et al., 2020; De Houwer et al., 2014; Hoff et al., 2012). De Houwer et al. (2014) showed using CDIs that monolingual toddlers knew more Dutch words than bilingual toddlers (20 months), but both groups understood and produced the same number of lexicalised meanings. They did not find any differences between monolinguals and bilinguals in vocabulary comprehension or vocabulary production for infants (13 months). A recent study showed that multilingualism does not affect infants' gesture repertoires either (Germain et al., 2022). Other studies suggest that multilingual toddlers do not have smaller vocabularies than their monolingual peers when they receive at least 60% exposure to the assessed language (Cattani et al., 2014). In our study, we included multilingual children to examine whether their vocabularies are negatively affected when examining only one of their languages using the N-CDIs and PPVT-III-NL.

Research aim

Well-known demographic predictors of language – including maternal education, child gender, gestational age and birth weight, and multilingualism – have been documented extensively, and many researchers accept their influences on language development without further question. However, particularly in the 90s when the American CDIs were first created, these predictors were often studied in smaller samples at one point in time. Longitudinal data provide insights into how these predictors of children's vocabulary unfold over time. Given the replication crisis in psychology, it is valuable to re-examine the findings in a large, longitudinal sample using different vocabulary measures to ensure their robustness and generalisability. In the present study, we aimed to examine whether key predictors that explain variation in children's early vocabularies are age-specific and task-specific in a large, longitudinal sample of Dutch children. A limited number of studies have examined these predictors within large, longitudinal samples, while there are uncertainties regarding the duration, directionality, and magnitude of these effects on different vocabulary measures across development. By examining the effects on multiple vocabulary measures in a large sample from infancy to early childhood, we analyse whether the effects of well-known predictors are age-specific and task-specific while keeping the characteristics of the sample constant. This helps us to identify whether widely discussed demographic predictors of vocabulary are robust and generalisable across development – ultimately advancing our understanding of children's language development.

Methods

Participants

The data for this study are derived from the YOUth cohort study following Dutch children prenatally up to early childhood (Onland-Moret et al., 2020). The cohort involved repeated measurements at regular intervals. From the cohort, 444 Dutch infants

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

Table 1. Sample characteristics including the mean (and standard deviation) for continuous variables or frequency counts (and percentage of sample) for categorical variables

	N	*Mean (SD) or n (% of sample)*
Age in weeks		
Wave 1	338	46.11 (3.79)
Wave 2	444	175.20 (41.48)
Male	444	214 (48%)
Highest maternal education	426	
Primary school		1 (<1%)
High school		16 (4%)
Vocational education		60 (14%)
Higher education		143 (34%)
University education		215 (50%)
Gestational duration in days	444	278 (12)
Birth weight in grams	399	3514 (476)
Multilingual	369	29 (8%)

around 10 months of age (230 females, 214 males; age M = 10.6 months; age range = 9.0 – 13.1 months; age SD = 0.9) (hereafter Wave 1) were included in this study. These were all the children in the YOUth cohort study who had participated in the next wave by March 2022. During this wave, the same children were on average 3.4 years of age (range = 2.0 – 6.0 years; SD = 0.8) (hereafter Wave 2). There were approximately one to five years (M = 2.5; SD = 0.8) in between measurement waves, randomly varying per participant. We followed the Code of Ethics of the World Medical Association (Declaration of Helsinki), and all caregivers signed informed consent prior to participating. During Wave 1, children received a Miffy picture book for their participation. During Wave 2, children received a frog umbrella.

In total, 426 mothers filled out the demographics questionnaire including questions on the caregivers' education. All caregivers provided us with their child's due date and birth date which we used to calculate the child's gestational duration. Of this sample, 399 caregivers also provided us with their child's birth weight in grams. Lastly, 369 caregivers filled out the questionnaire including languages spoken at home. The summary of sample characteristics is shown in Table 1. In this sample, at least 29 children were not growing up as monolingual Dutch speakers. We considered a child monolingual when only Dutch was spoken at home. Given the small number of multilingual children, we did not differentiate the group further based on the children's estimated time of exposure to Dutch.

Materials and procedure

N_{YOUth}-CDIs

We administered the N_{YOUth}-CDI 1 – measuring vocabulary production, vocabulary comprehension, and gestures – during Wave 1. The N_{YOUth}-CDI 1 contains the short

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

form of words (Zink & Lejaegere, 2003). We used the short form because it contains only 103 compared to 434 items, which makes this form far less time-consuming to complete. This was desired since caregivers already had to fill out a broad range of questionnaires in the YOUth cohort study. Caregivers were asked to check for each item whether their child UNDERSTANDS or SPEAKS the word – also when the child produces synonyms or pronunciation errors. In the N_{YOUth}-CDI 1, we replaced or removed 12 typical Flemish words with synonyms that are more common in Standard Dutch spoken in the Netherlands (e.g., we removed *mantel* from *jas(je) / mantel* ("coat")) to make the lists more suitable for children included in the YOUth cohort study. We included the list containing 65 gestures and actions from the full-length N-CDI-WG (Zink & Lejaegere, 2002) which is typically not administered with short forms. This scale contains "early gestures" including the first communicative gestures (e.g., pointing) and games and routines (e.g., playing peekaboo) and "late gestures" including actions with objects (e.g., eating with a spoon or fork) and pretending to be a caregiver (e.g., pretending to feed a doll). The gesture scale could be more suitable compared to the vocabulary scales for this young age group, as it does not suffer from floor effects and is related to children's later vocabulary size (Zink & Lejaegere, 2002). The N_{YOUth}-CDIs were emailed to the primary caregiver. The N_{YOUth}-CDIs are fully digitised so caregivers could fill them out online. We scored the lists following the instructions of the manuals (Zink & Lejaegere, 2002, 2003).

The N_{YOUth}-CDI 2 is a combination of the short forms N-CDI 2A (16-30 months) and N-CDI 3 (30-37 months) (Zink & Lejaegere, 2003). We combined the two forms because there was only one measurement wave during the toddler and preschool years in the YOUth cohort study (Wave 2). The combined version resulted in a total number of 207 vocabulary items after removing the overlapping ones. Caregivers were asked to check the items that the child SPEAKS – also in case the child produces synonyms or pronunciation errors. In the N_{YOUth}-CDI 2, we also replaced or removed 26 typical Flemish words with similar words that are more common in Standard Dutch spoken in the Netherlands (e.g., *bank* instead of *zetel/sofa* ("couch")). The CDIs for toddlers (including adaptations in other languages) do not measure vocabulary comprehension or gestures. Most toddlers and older children have already acquired all the gestures resulting in a ceiling effect. Children of this age group are also old enough to participate in a lab-administered task of vocabulary comprehension. Caregivers were instructed to fill the N_{YOUth}-CDI 2 out within four weeks after the administration of the PPVT-III-NL in the lab during Wave 2.

Peabody Picture Vocabulary Task

During Wave 2, we also administered the third version of the Dutch Peabody Picture Vocabulary Task (PPVT-III-NL) which is a lab-administered task of receptive vocabulary (Schlichting, 2005). The task measures whether a person can match a spoken word to one of the four pictures (i.e., multiple choice). It is designed as a behavioural task in which the participant points to one of the images and the experimenter produces the target words and scores manually. For the YOUth cohort study, we developed a computerised version of the PPVT-III-NL. The experimenter runs a script on a computer with a touch screen where children are provided with recordings of the test items and four pictures on the screen. This controls for differences in speaker pronunciations and minimises the role of the experimenter. Children can use the touch screen to select one of the pictures after the target item has been presented. During the task, items become increasingly more

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

complex. The PPVT-III-NL has a total of 204 items, divided into 17 sets of 12 items. The task terminates when the child makes nine or more errors in one set ("final set") (see Schlichting, 2005). The programme automatically subtracts the number of errors from the maximum score (which is the number of the final set * 12 items), resulting in the child's raw score. During the task, the child's caregiver was present in the back of the room out of the child's view. Caregivers were explicitly instructed not to help or communicate with the child.

Validity evidence

Reliability

Due to the modifications we made to the N_{YOUth}-CDIs, we first assessed whether the adapted checklists measure a valid approximation of children's vocabulary size. First, we examined whether we could find evidence for the reliability of the N_{YOUth}-CDIs. For the N_{YOUth}-CDI 1, we calculated Cronbach's alpha separately for comprehension (α = .97), production (α = .91), and gestures (α = .89) which represents the consistency of items within each scale. We also calculated Cronbach's alpha for the N_{YOUth}-CDI 2 word production (α = .99) indicating that the items on the scale measured the same construct. Overall, this indicates that the different items included in the caregiver reports show excellent reliability.

Validity

We also present several types of validity evidence. First, we assessed correlations between the different scales included in the N_{YOUth}-CDI 1 for infants. The results of the correlation tests indicate that for the N_{YOUth}-CDI 1, comprehension was positively correlated with both production, $r_s(336) = .50$, $p < .001$, and gestures, $r(335) = .65$, $p < .001$. Production was also correlated positively with gestures, $r_s(335) = .47$, $p < .001$. We also examined whether vocabulary production obtained by the N_{YOUth}-CDI 2 shows a concurrent relationship with vocabulary comprehension measured by the lab-administered PPVT-III-NL. The correlation test indicates there is a strong, positive correlation between N_{YOUth}-CDI 2 production and concurrent PPVT-III-NL comprehension scores, $r_s(292) = .64$, $p < .001$. The relationship is depicted in Figure 1.

The last step was to assess longitudinal relations between the different vocabulary measures. We examined whether vocabulary production, vocabulary comprehension, and gestures measured at Wave 1 were correlated with N_{YOUth}-CDI 2 production and PPVT-III-NL comprehension measured at Wave 2. In total, 266 participants completed the N_{YOUth}-CDIs during Wave 1 and Wave 2, and 325 participants completed both the N_{YOUth}-CDI 1 at Wave 1 and the PPVT-III-NL at Wave 2. We ran partial correlations correcting for the varying time interval between the two waves using the *ppcor* package in *R* (Kim, 2015). We therefore used PPVT-III-NL raw scores which are not yet corrected for age. The results of all (partial) correlation tests are summarised in Table 2.

The results show that all measures of the N_{YOUth}-CDI 1 were positively correlated with later N_{YOUth}-CDI 2 production scores. Overall, the strengths of the correlations were weak to moderate. We also found that comprehension at Wave 2 (i.e., PPVT-III-NL) only correlated with the gesture scale in Wave 1.

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

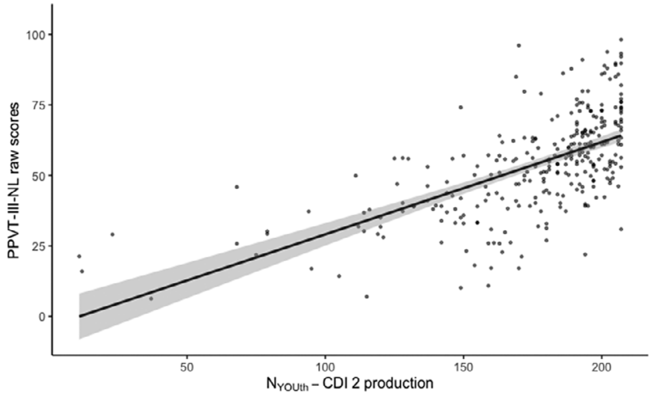

Figure 1. The concurrent relationship between N_{YOUth}-CDI 2 production and PPVT-III-NL comprehension.

Table 2. (Partial) correlation table showing the links between the different N_{YOUth}-CDI scales at Wave 1 and Wave 2 and the PPVT-III-NL at Wave 2 while controlling for the varying time interval in between measurement waves

	1	2	3	4
1. N_{YOUth}–CDI 1 comprehension	–			
2. N_{YOUth}–CDI 1 production	.50***	–		
3. N_{YOUth}–CDI 1 gestures	.65***	.47***	–	
4. N_{YOUth}–CDI 2 production	.31***	.17**	.15*	–
5. PPVT–III–NL comprehension	.08	.08	.15**	.64***

Note: * $p < .05$; ** $p < .01$; ***$p < .001$

Questionnaires

We collected the previously described characteristics of the sample via digital questionnaires. These included questionnaires on the mother's demographics (e.g., educational background), which we collected when the mother was 20 weeks pregnant; the child's birth (e.g., due date, birth date, and birth weight), which we collected shortly after the child's birth; and the languages spoken at home (including questions about the caregivers' native language(s) and the language(s) spoken at home), which we collected during Wave 1 concurrently with the N_{YOUth}-CDI 1.

Coding and analyses

All analyses were carried out in *R* version 4.2.0 (R Core Team, 2022). For the N_{YOUth}-CDI 1, we calculated "vocabulary production" by summing all vocabulary items for which caregivers ticked the box SPEAKS, "vocabulary comprehension" by summing all vocabulary items for which caregivers ticked the box UNDERSTANDS or SPEAKS, and "total gestures" by summing all YES, SOMETIMES, and OFTEN responses on the gesture scale. Gestures can be subdivided into two categories: "early gestures" and "late gestures" (Zink

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

& Lejaegere, 2002). The sum of both scales results in the score "total gestures". We used these raw scores to analyse the data. For the N_{YOUth}-CDI 2, we calculated "vocabulary production" by summing all items that were marked by the caregivers indicating that the child produces the word. For the PPVT-III-NL, we obtained "vocabulary comprehension" through the raw scores which were automatically calculated by the computer script. We coded the highest educational degree obtained on a nine-point scale ranging from 1 = no education to 9 = university degree. We calculated gestational duration in days using the discrepancy between children's due dates and birth dates and adding or subtracting this from 280 days (i.e., full-term gestation). Caregivers reported their children's birth weight in grams. Lastly, we determined whether a child was growing up multilingual (i.e., at least one caregiver does not only speak Dutch at home).

We fitted robust generalised linear models using the package *robustbase* version 0.95-0 (Maechler et al., 2022) following Frank et al. (2021). We used "vocabulary comprehension", "vocabulary production", and "gestures" measured by the N_{YOUth}-CDI 1, "vocabulary production" measured by the N_{YOUth}-CDI 2, and "vocabulary comprehension" measured by the PPVT-III-NL as continuous outcome measures. We added children's ages in weeks, gender (female or male), gestational, birth weight, maternal education, and language status (monolingual or multilingual) as predictors to the models. For categorical predictors, we used dummy coding with the categories containing the largest number of observations (gender: female; language status: monolingual) as reference levels. We centred and scaled children's age, gestational duration, birth weight, and maternal education. We modelled raw scores instead of normed scores or percentiles. By adding age in weeks as a predictor to the models, all other predictors are independent of the effects of age.

Results

Descriptive statistics

We included 444 participants from the YOUth cohort study. During Wave 1, 338 of these participants completed the N_{YOUth}-CDI 1. There was one participant who did not complete the gestures list; this participant is only excluded from analyses involving gestures. During Wave 2, we had to exclude four participants from the PPVT-III-NL because the children did not participate (n = 2) or the test day had ended prematurely before administering the PPVT-III-NL (n = 2) resulting in no data. We excluded an additional 11 children from any analyses involving the PPVT-III-NL because they did not fully complete the task, resulting in a total of 429 participants. There were 303 participants whose caregivers completed the N_{YOUth}-CDI 2 for Wave 2. The descriptive results of the vocabulary tests are presented in Table 3. The high standard deviations indicate that vocabulary scores are spread out over a wide range, revealing a large amount of individual variability.

Demographic effects

During Wave 1, vocabulary comprehension, vocabulary production, and gestures were measured with the N_{YOUth}-CDI 1. The results of the robust regression models for vocabulary outcomes at Wave 1 are presented in Table 4. During Wave 2, vocabulary production was measured with the N_{YOUth}-CDI 2 and comprehension was measured with the PPVT-III-NL. The results of the robust generalised linear regression models for vocabulary outcomes at Wave 2 are presented in Table 5. When examining the effects on

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

Table 3. Descriptive results of the vocabulary measures

Wave	*n*	Comprehension[a] *M* (*SD*)	Production *M* (*SD*)	Gestures *M* (*SD*)
1	337–338	36.75 (21.79)	2.86 (4.22)	18.47 (7.83)
2	303–429	52.82 (18.37)	173.9 (35.78)	

[a]At Wave 1, vocabulary comprehension is measured with the N_{YOUth}-CDI 1. At Wave 2, vocabulary comprehension is measured with the PPVT-III-NL (raw scores).

Table 4. Robust regression results for vocabulary outcomes at Wave 1

	Outcome variable (95% CI)		
	N_{YOUth}-CDI 1 comprehension	N_{YOUth}-CDI 1 production	N_{YOUth}-CDI 1 gestures
(Intercept)	38.66***	1.79***	19.01***
	(35.42, 41.89)	(1.15, 2.44)	(17.98, 20.04)
Age in weeks	8.16***	0.71**	3.60***
	(5.81, 10.51)	(0.26, 1.16)	(2.91, 4.29)
Maternal education	–4.32**	–0.37*	–0.26
	(–7.21, –1.43)	(–0.67, –0.08)	(–0.99, 0.46)
Gender (male)	–4.42	–0.30	–2.60***
	(–9.33, 0.49)	(–0.81, 0.21)	(–4.03, –1.18)
Gestational duration	–0.12	0.04	0.79
	(–3.09, 2.84)	(–0.32, 0.41)	(–0.09, 1.67)
Birth weight	0.53	–0.06	0.25
	(–1.81, 2.86)	(–0.35, 0.24)	(–0.47, 0.98)
Language status (multilingual)	–2.30	0.36	–0.10
	(–10.90, 6.30)	(–0.76, 1.47)	(–2.55, 2.35)
Observations	312	312	311
R^2	0.19	0.14	0.32
Adjusted R^2	0.17	0.12	0.31
Residual std. error	19.47	1.68	5.54
	(df = 305)	(df = 305)	(df = 304)

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

all vocabulary outcomes of infants and toddlers, we find one consistent predictor: age in weeks has a positive effect on all collected outcome measures. We expected a robust age-related effect as children's vocabularies grow fast during the first years of development. Figure 2 shows the effect of children's age and gender on the different N_{YOUth}-CDI 1 scales measured during infancy.

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

Table 5. Robust regression results for vocabulary outcomes at Wave 2

	Outcome variable (95% CI)	
	N_{YOUth}-CDI 2 production[a]	PPVT-III-NL comprehension
(Intercept)	183.41***	53.82***
	(179.96, 186.86)	(52.38, 55.25)
Age in weeks	14.85***	14.80***
	(11.55, 18.14)	(13.50, 16.09)
Maternal education	0.18	1.70**
	(–3.53, 3.89)	(0.42, 2.97)
Gender (male)	–5.65*	–1.06
	(–10.59, –0.70)	(–3.28, 1.15)
Gestational duration	–1.42	–0.33
	(–4.65, 1.80)	(–1.87, 1.21)
Birth weight	1.54	1.31
	(–1.60, 4.67)	(–0.51, 3.14)
Language status (multilingual)	–8.11	–5.23*
	(–20.03, 3.80)	(–10.16, –0.30)
Observations	264	325
R^2	0.39	0.68
Adjusted R^2	0.38	0.68
Residual std. error	16.95	9.58
	(df = 257)	(df = 318)

Note: * $p < .05$; ** $p < .01$; ***$p < .001$

[a]The N_{YOUth}-CDI 2 was administered across all children included in Wave 2 (up until 72 months), while the N-CDI 2 is designed for children until 37 months. We also fitted the model including only the children aged until 37 months ($n = 251$). All results remain unchanged, but the magnitude of the negative effect for males becomes much larger ($b = -15.11$, $SE = 6.75$, p =. 027). See the *R* Markdown file on OSF for the results.

Except for age in weeks, all other predictors show inconsistent patterns across the different measurement waves and vocabulary outcomes. For infants, we found a disadvantage for boys on gestures measured with the N_{YOUth} -CDI 1 ($b = -2.60$, $SE = 0.72$, $p < .001$), but not on word production ($b = -0.30$, $SE = 0.26$, $p = .25$) or word comprehension ($b = -4.42$, $SE = 2.49$, $p = .07$) during Wave 1. During Wave 2, we found a disadvantage for boys on N_{YOUth}-CDI 2 production ($b = -5.65$, $SE = 2.51$, $p = .03$), but not on the lab-administered PPVT-III-NL at this age ($b = -1.06$, $SE = 1.13$, $p = .35$). Figure 3 shows the effect of children's age and gender on the N_{YOUth}-CDI 2 and the PPVT-III-NL at Wave 2. Although both measures show a similar increase with age, there is a clear ceiling effect for production measured using the N_{YOUth}-CDI 2.

Second, we found a negative effect of maternal education on caregiver-reported vocabulary comprehension for infants ($b = -4.31$, $SE = 1.47$, $p < .01$) and vocabulary production for infants ($b = -0.37$, $SE = 0.15$, $p < .05$), but not on gestures ($b = -0.26$, SE =

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

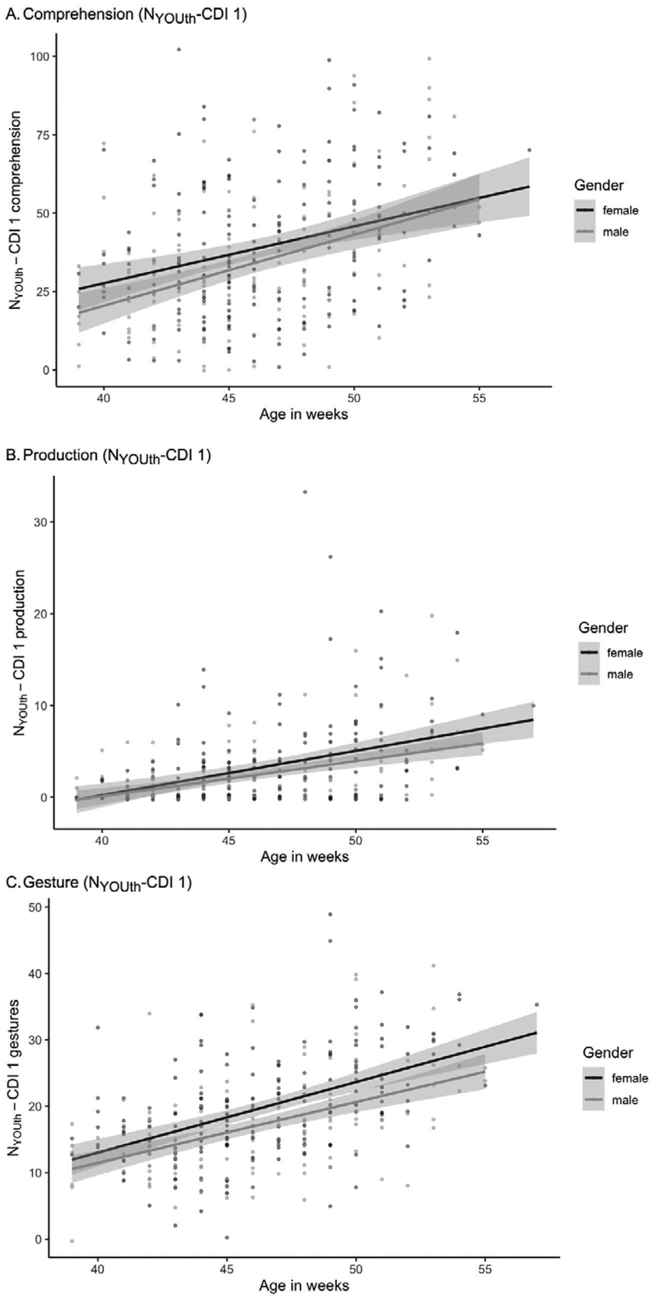

Figure 2. Effects of children's age and gender on vocabulary comprehension (A), production (B), and gestures (C) measured with the N_{YOUth}-CDI 1.

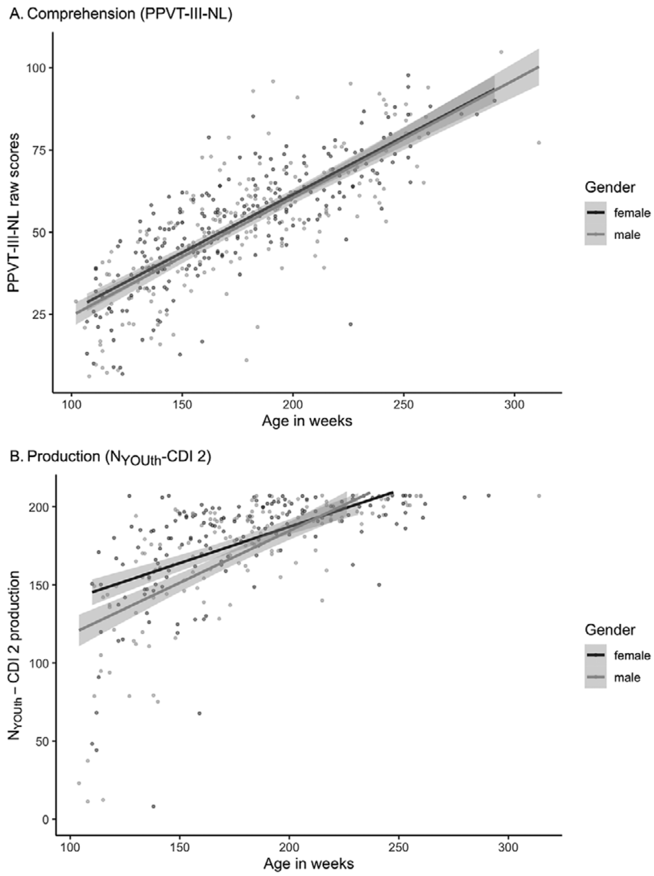

Figure 3. Effects of children's age and gender on vocabulary comprehension measured with the PPVT-III-NL (A) and production measured with the N_{YOUth}-CDI 2 (B).

0.37, p = .48). The negative effect of maternal education has shifted to a positive effect on the lab-administered PPVT-III-NL task during Wave 2 (b = 1.70, SE = 0.65, p < .01), but we found no effect of maternal education on caregiver-reported vocabulary production during this wave (b = 0.18, SE = 1.89, p = .92).

Third, we did not find any significant effects of children's gestational age or birth weight on any of the vocabulary outcomes during Wave 1 or Wave in our sample.

Lastly, we only found a negative effect of multilingualism on the PPVT-III-NL (b = -5.23, SE = 2.50, p < .05), but not on any of the caregiver-reported N_{YOUth}-CDIs.

Discussion

We aimed to examine whether key demographic predictors that explain variation in children's early vocabularies – maternal education, children's gender, gestational age and birth weight, and multilingualism – were age-specific and task-specific in this large, longitudinal sample of Dutch children. Apart from a consistent positive effect of

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

children's ages on all outcomes, we found that none of the other predictors remained constant across the different vocabulary outcomes measured in this study. Below, we address all other factors one by one.

Effect of maternal education shifts over time

We examined the effects of maternal education as a proxy for SES on children's vocabulary outcomes. First, we found negative effects of maternal education on vocabulary production and vocabulary comprehension measured by the N_{YOUth}-CDI 1. This is in line with previous studies that have also reported negative effects of SES on CDIs filled out for infants – usually for vocabulary comprehension and to a lesser extent for word production (Feldman et al., 2000; Fenson et al., 1994; Reese & Read, 2000). This is possibly caused by a caregiver reporting bias. The latter interpretation is strengthened by the finding that there is no effect of maternal education on N_{YOUth}-CDI 1 gestures or N_{YOUth}-CDI 2 production. This is in line with Rowland et al. (2022) who found that the reverse SES effect for infants was far less prevalent in the gesture scale across ten cross-linguistic CDI datasets. Gestures may be more easily observable and do not require as much interpretation, making them less susceptible to reporting biases. Unlike gestures, word production still requires a small amount of interpretation because caregivers are instructed to also check SPEAKS for vocabulary items when their child produces synonyms or production errors. In addition, caregivers may over-report their child's vocabulary if they think larger vocabularies are desirable. This social stigma is less prominent for children's gesture repertoires which makes the gesture scale less susceptible to caregiver reporting biases. Lastly, we found a positive effect of maternal education on the lab-administered PPVT-III-NL during Wave 2. This result is in line with previous studies finding that a higher SES, often measured through maternal education, correlates with larger vocabularies (Hoff, 2003; Huttenlocher et al., 2010). This could suggest that an advantage of maternal education only emerges later in children's development, although an effect on infants' vocabularies could be obscured by caregiver reporting biases or floor effects.

An alternative explanation to consider is that the reverse SES effect found for caregiver-reported comprehension and production during infancy is real. That would imply that infants of lower SES families start out with larger vocabularies compared to infants of higher SES families. However, we believe that this explanation is unlikely. Previous studies assessing a range of language-related abilities in children, including language processing and early use of gestures, show that children of higher SES families tend to outperform children of lower SES families (e.g., Fernald et al., 2013; Rowe & Goldin-Meadow, 2009). Although the SES difference could be explained by several different reasons, including differences in genetics or the environment, we believe that it is unlikely that the reverse-SES effect can be attributed to a real effect. Nevertheless, this recurrent finding across studies should not be dismissed without further thought, and future studies should examine SES differences in other language-related measures of infants.

Girls have an advantage over boys

The results show that girls have an advantage over boys on N_{YOUth}-CDI 1 gestures and N_{YOUth}-CDI 2 production. Previous studies have also frequently reported an advantage for girls using CDIs (Eriksson et al., 2012; Feldman et al., 2005; Frank et al., 2021; Reese &

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

Read, 2000). The results of our study suggest that the gender difference could start with a difference in children's gesture repertoires during infancy. Infants' gestures are known to influence children's later vocabularies (see Brooks & Meltzoff, 2008; Colonnesi et al., 2010; Rowe & Goldin-Meadow, 2009). Recently, Germain et al. (2022) also showed that 14-month-old girls produce more gesture types than boys using caregiver reports. Our results add to this finding by showing that a difference in gestures between boys and girls is already present before their first birthday. The gesture scale could be the only scale that shows enough variability across infants, resulting in sufficient variation to detect the gender effect early on. Our findings are also in line with the hypothesis that gender differences are more prevalent in vocabulary production than vocabulary comprehension (see Bornstein et al., 1998; Feldman et al., 2005; Frank et al., 2021; Qi et al., 2003). This could explain the absence of a significant gender effect on the PPVT-III-NL. Another possible explanation for this is that the gender effect on N_{YOUth}-CDIs is the result of a reporting bias. Caregivers could expect that girls are more verbal than boys, influencing how they fill out the vocabulary checklist. However, we suspect that this is unlikely because we also found a significant gender effect on word production during Wave 2. Caregiver reports on word production (rather than comprehension) and toddlers (rather than infants) are less susceptible to reporting biases. Frank et al. (2021) also showed that cross-linguistically, the advantage for girls is more prominent in caregiver reports of word production than word comprehension. This suggests that girls truly have an advantage over boys — at least in their expressive vocabularies.

No effects of gestational duration and birth weight

We did not find any effects of gestational duration or birth weight on children's vocabularies in this non-clinical sample. This does not support earlier findings that preterm infants are at risk of having smaller vocabularies later in life than full-term infants (e.g., Foster-Cohen et al., 2007; Guarini et al., 2009; Sansavini et al., 2011). Nevertheless, some studies suggest that only extremely preterm children (under 28 weeks) and/or children of very low birth weight (under 1500 g) have language delays (Barre et al., 2011; Kern & Gayraud, 2007). None of the children included in our sample fall under those criteria. Therefore, it is possible that we did not find any differences because gestational duration and birth weight predominantly affect the more extreme cases. Future studies should examine the effects of gestational duration and birth weight in longitudinal samples that include very to extremely preterm children and/or children of very low birth weights – but the results of our study do not provide evidence for the generalisability of these predictors across large, healthy samples.

Multilinguals know fewer words than monolinguals

We lastly examined the effects of children in the Netherlands growing up with more than one language. The results show that monolingual toddlers have larger receptive vocabularies measured with the PPVT-III-NL, but not larger productive vocabularies measured with the N_{YOUth}-CDIs. Given the fact that multilingual toddlers are not exposed to as much Dutch language input as their monolingual peers, and vocabulary development is heavily influenced by the quantity and quality of exposure (Hoff, 2003), we expected

https://doi.org/10.1017/S0305000924000199 Published online by Cambridge University Press

multilingual toddlers to have smaller vocabularies when measuring only one of their languages (in line with Blom et al., 2020; De Houwer et al., 2014; Hoff et al., 2012). In the N_{YOUth}-CDIs, caregivers were instructed to also check SPEAKS on vocabulary items when their child produces a synonym. Arguably, these instructions yielded large variability in how multilingual caregivers filled out the checklists. It is plausible that some multilingual caregivers also accepted translations for vocabulary items which could explain the absence of a negative effect of multilingualism on the N_{YOUth}-CDIs. Our sample could also have been too homogeneous because all caregivers who participated in the YOUth cohort study were required to be able to fill out Dutch questionnaires to participate. This resulted in a small number of multilingual children in our sample that may not have been sufficient to detect an effect of multilingualism on caregiver reports, especially given the potential variability in how multilingual caregivers filled out the reports. Lastly, we found no effect of multilingual input on gestures, which is in line with a recent study that did not find an effect of multilingualism on 14-month-old infants' gestures measured with CDIs (Germain et al., 2022). Even though infants' gesture repertoires are an early indicator of their later vocabulary size, they are likely independent of specific language exposure and therefore not affected by multilingual language input. Importantly, whether multilingualism affects early vocabulary in one language seems dependent on the type of vocabulary measure (i.e., a lab measurement vs. caregiver report).

Limitations

Although we found some effects of maternal education in the expected directions based on previous studies using socio-demographically diverse samples (Feldman et al., 2000), our sample is rather homogeneous and overrepresents highly educated mothers. A lack of diversity makes SES differences less apparent. The results of our study suggest that caregiver reports of infants' vocabulary comprehension and vocabulary production, but not gestures, are negatively affected by maternal education. According to previous studies, infants of lower SES may be using fewer gestures during caregiver-child interactions (Rowe & Goldin-Meadow, 2009). Although we did not examine gesture rates, we did not find an effect of maternal education on infants' gesture repertories. Future longitudinal cohort studies with more diverse samples should re-evaluate whether gesture repertoires show any SES differences and how these may affect the predictive value of gestures in diverse samples.

We also want to draw attention to the large age range of children included in Wave 2. This was a decision made by the YOUth cohort study for reasons orthogonal to the present study. While we controlled for children's ages in the statistical models, the large age range could have impacted the results. Some demographic effects may explain more variation in the first few years of life, but not at later ages. By grouping all children aged 2 to 5 together, we may have underestimated some of the demographic effects on vocabulary that become weaker predictors across development. In addition, many children in Wave 2 were too old for the N-CDI 2. This could have caused a ceiling effect on production measured by caregiver reports for toddlers. In order to address this possibility, we also fitted all models excluding those children (see OSF), which did not change the results. This tentatively suggests that we can use N-CDIs while sampling a large age range of young children, which is beneficial to longitudinal cohort studies with repeated measurements.

Conclusions

The results of our longitudinal study including over four hundred Dutch children suggest that the effects of widely discussed demographic predictors on children's vocabularies are dependent on children's ages and the type of vocabulary task being used. Except for age, none of the predictors remained constant across development or the different measurement tasks. We found a disadvantage for males in infants' gestures and toddlers' word production. We found a negative effect of maternal education on infants' caregiver-reported vocabulary, but a positive effect on the lab-administered receptive vocabulary task. Lastly, we found a negative effect of multilingualism – but only for the lab-administered receptive vocabulary task. The results imply that research findings can be influenced by children's age or the vocabulary task being used in a specific study. This is important to consider for child language researchers in future studies who aim to explain variation in vocabulary development. One advantage of cohort studies with repeated measurements is to gain better insights into which predictors have temporary or weak effects on development. Given our results, we would recommend researchers to sample diverse groups of children – including a broad age range – and use more than one vocabulary outcome when examining predictors of individual variation to gain a more comprehensive understanding of the duration, directionality, and magnitude of the effects on variation in children's vocabulary across development. Predictors can differentially affect children's gesture development during infancy and their expressive and comprehensive vocabularies across development. We also found that effects can shift over time, at least from infancy to toddlerhood. This corroborates that we should examine large, longitudinal samples cross-linguistically to determine the generalisability and robustness of key predictors of children's language development.

Acknowledgements. We are grateful to all families who participate in the YOUth study. YOUth is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). A complete listing of the study investigators and study management can be found at https://www.uu.nl/en/research/youth-cohort-study/about-us/who-isinvolved. YOUth investigators and management designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the YOUth study investigators or YOUth management.

Data statement. YOUth is a longitudinal study that aims to produce and safely store FAIR and high-quality data. The data can be accessed for both use and verification purposes upon request (see https://www.uu.nl/en/research/youth-cohort-study/data-access). The *R* script and other materials can be found online: https://osf.io/vj72c/.

Competing interest. The authors declare none.

References

Barre, N., **Morgan, A.**, **Doyle, L. W.**, & **Anderson, P. J.** (2011). Language abilities in children who were very preterm and/or very low birth weight: A meta-analysis. *The Journal of Pediatrics*, **158**(5), 766–774. https://doi.org/10.1016/j.jpeds.2010.10.032

Bavin, E. L., **Prior, M.**, **Reilly, S.**, **Bretherton, L.**, **Williams, J.**, **Eadie, P.**, **Barrett, Y.**, & **Ukoumunne, O. C.** (2008). The Early Language in Victoria Study: Predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language*, **35**(3), 687–701. https://doi.org/10.1017/S0305000908008726

Bergelson, E., **Soderstrom, M.**, **Schwarz, I.-C.**, **Rowland, C. F.**, **Ramírez-Esparza, N.**, **Hamrick, L. R.**, **Marklund, E.**, **Kalashnikova, M.**, **Guez, A.**, **Casillas, M.**, **Benetti, L.**, **van Alphen, P.**, & **Cristia, A.** (2023).

Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences*, **120**(52), e2300671120. https://doi.org/10.1073/pnas.2300671120

Blom, E., **Boerma, T.**, **Bosma, E.**, **Cornips, L.**, **van den Heuij, K.**, & **Timmermeister, M.** (2020). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language*, **40**(2), 151–171. https://doi.org/10.1177/0142723719892794

Bornstein, M. H., **Haynes, M. O.**, & **Painter, K. M.** (1998). Sources of child vocabulary competence: A multivariate model. *Journal of Child Language*, **25**(2), 367–393. https://doi.org/10.1017/s0305000998003456

Brooks, R., & **Meltzoff, A. N.** (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, **35**(1), 207–220. https://doi.org/10.1017/s030500090700829x

Cattani, A., **Abbot-Smith, K.**, **Farag, R.**, **Krott, A.**, **Arreckx, F.**, **Dennis, I.**, & **Floccia, C.** (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders*, **49**(6), 649–671. https://doi.org/10.1111/1460-6984.12082

Colonnesi, C., **Stams, G. J. J. M.**, **Koster, I.**, & **Noom, M. J.** (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, **30**(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

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

Eriksson, M., **Marschik, P. B.**, **Tulviste, T.**, **Almgren, M.**, **Pérez Pereira, M.**, **Wehberg, S.**, **Marjanovič-Umek, L.**, **Gayraud, F.**, **Kovacevic, M.**, & **Gallego, C.** (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, **30**(2), 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x

Feldman, H. M., **Dale, P. S.**, **Campbell, T. F.**, **Colborn, D. K.**, **Kurs-Lasky, M.**, **Rockette, H. E.**, & **Paradise, J. L.** (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*, **76**(4), 856–868. https://doi.org/10.1111/j.1467-8624.2005.00882.x

Feldman, H. M., **Dollaghan, C. A.**, **Campbell, T. F.**, **Kurs-Lasky, M.**, **Janosky, J. E.**, & **Paradise, J. L.** (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*, **71**(2), 310–322. https://doi.org/10.1111/1467-8624.00146

Fenson, L., **Dale, P. S.**, **Reznick, J. S.**, **Bates, E.**, **Thal, D. J.**, **Pethick, S. J.**, **Tomasello, M.**, **Mervis, C. B.**, & **Stiles, J.** (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59**(5), i–185. https://doi.org/10.2307/1166093

Fenson, L., **Marchman, V. A.**, **Thal, D. J.**, **Dale, P. S.**, **Reznick, J. S.**, & **Bates, E.** (2007). *The MacArthur Communicative Development Inventories: User's guide and technical manual* (Second edition). Paul H. Brookes Publishing Co., Inc.

Fernald, A., **Marchman, V. A.**, & **Weisleder, A.** (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, **16**(2), 234–248. https://doi.org/10.1111/desc.12019

Foster-Cohen, S., **Edgin, J. O.**, **Champion, P. R.**, & **Woodward, L. J.** (2007). Early delayed language development in very preterm infants: Evidence from the MacArthur-Bates CDI. *Journal of Child Language*, **34**(3), 655–675. https://doi.org/10.1017/S0305000907008070

Frank, M. C., **Braginsky, M.**, **Yurovsky, D.**, & **Marchman, V. A.** (2021). *Variability and consistency in early language learning: The Wordbank project*. MIT Press. https://langcog.github.io/wordbank-book/

Germain, N., **Gonzalez-Barrero, A. M.**, & **Byers-Heinlein, K.** (2022). Gesture development in infancy: Effects of gender but not bilingualism. *Infancy*, **27**(4), 663–681. https://doi.org/10.1111/infa.12469

Guarini, A., **Sansavini, A.**, **Fabbri, C.**, **Alessandroni, R.**, **Faldella, G.**, & **Karmiloff-Smith, A.** (2009). Reconsidering the impact of preterm birth on language outcome. *Early Human Development*, **85**(10), 639–645. https://doi.org/10.1016/j.earlhumdev.2009.08.061

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, **74**(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

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

Houston-Price, C., **Mather, E.**, & **Sakkalou, E.** (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, **34**(4), 701–724. https://doi.org/10.1017/S0305000907008124

Huttenlocher, J., **Waterfall, H.**, **Vasilyeva, M.**, **Vevea, J.**, & **Hedges, L. V.** (2010). Sources of variability in children's language growth. *Cognitive Psychology*, **61**(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Kern, S., & **Gayraud, F.** (2007). Influence of preterm birth on early lexical and grammatical acquisition. *First Language*, **27**(2), 159–173. https://doi.org/10.1177/0142723706075790

Kidd, E., & **Donnelly, S.** (2020). Individual differences in first language acquisition. *Annual Review of Linguistics*, **6**(1), 319–340. https://doi.org/10.1146/annurev-linguistics-011619-030326

Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, **22**(6), 665. https://doi.org/10.5351/CSAM.2015.22.6.665

Kuvač-Kraljević, J., **Blaži, A.**, **Schults, A.**, **Tulviste, T.**, & **Stolt, S.** (2021). Influence of internal and external factors on early language skills: A cross-linguistic study. *Infant Behavior and Development*, **63**, 101552. https://doi.org/10.1016/j.infbeh.2021.101552

Maechler, M., **Rousseeuw, P.**, **Croux, C.**, **Todorov, V.**, **Ruckstuhl, A.**, **Salibian-Barrera, M.**, **Verbeke, T.**, **Koller, M.**, **Conceicao, E. L. T.**, & **Anna di Palma, M.** (2022). *robustbase: Basic robust statistics.* http://robustbase.r-forge.r-project.org/

Ogneva, A., & **Pérez-Pereira, M.** (2023). Communicative and linguistic factors influencing language development at 30 months of age in preterm and full-term children: A longitudinal study using the CDI. *Frontiers in Psychology*, **14**. https://doi.org/10.3389/fpsyg.2023.1177161

Onland-Moret, N. C., **Buizer-Voskamp, J. E.**, **Albers, M. E. W. A.**, **Brouwer, R. M.**, **Buimer, E. E. L.**, **Hessels, R. S.**, **de Heus, R.**, **Huijding, J.**, **Junge, C. M. M.**, **Mandl, R. C. W.**, **Pas, P.**, **Vink, M.**, **van der Wal, J. J. M.**, **Hulshoff Pol, H. E.**, & **Kemner, C.** (2020). The YOUth study: Rationale, design, and study procedures. *Developmental Cognitive Neuroscience*, **46**, 100868. https://doi.org/10.1016/j.dcn.2020.100868

Pan, B. A., **Rowe, M. L.**, **Spier, E.**, & **Tamis-Lemonda, C.** (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language*, **31**(3), 587–608. https://doi.org/10.1017/S0305000904006270

Pérez-Pereira, M., & **Cruz, R.** (2018). A longitudinal study of vocabulary size and composition in low risk preterm children. *First Language*, **38**(1), 72–94. https://doi.org/10.1177/0142723717730484

Qi, C. H., **Kaiser, A. P.**, **Milan, S. E.**, **Yzquierdo, Z.**, & **Hancock, T. B.** (2003). The performance of low-income, African American children on the Preschool Language Scale-3. *Journal of Speech, Language, and Hearing Research*, **46**(3), 576–590. https://doi.org/10.1044/1092-4388(2003/046)

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

Reese, E., & **Read, S.** (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. *Journal of Child Language*, **27**(2), 255–266. https://doi.org/10.1017/S0305000900004098

Reilly, S., **Bavin, E. L.**, **Bretherton, L.**, **Conway, L.**, **Eadie, P.**, **Cini, E.**, **Prior, M.**, **Ukoumunne, O. C.**, & **Wake, M.** (2009). The Early Language in Victoria Study (ELVS): A prospective, longitudinal study of communication skills and expressive vocabulary development at 8, 12 and 24 months. *International Journal of Speech-Language Pathology*, **11**(5), 344–357. https://doi.org/10.1080/17549500903147560

Rowe, M. L., & **Goldin-Meadow, S.** (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, **323**(5916), 951–953. https://doi.org/10.1126/science.1167025

Rowland, C., **Krajewski, G.**, **Meints, K.**, **Łuniewska, M.**, **Kochańska, M. K.**, & **Alcock, K.** (2022). *CDI Demographics.* https://osf.io/hwg4c/

Sansavini, A., **Guarini, A.**, **Savini, S.**, **Broccoli, S.**, **Justice, L.**, **Alessandroni, R.**, & **Faldella, G.** (2011). Longitudinal trajectories of gestural and linguistic abilities in very preterm infants in the second year of life. *Neuropsychologia*, **49**(13), 3677–3688. https://doi.org/10.1016/j.neuropsychologia.2011.09.023

Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Hartcourt Assessment BV.

Simonsen, H. G., **Kristoffersen, K. E.**, **Bleses, D.**, **Wehberg, S.**, & **Jørgensen, R. N.** (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, **34**(1), 3–23. https://doi.org/10.1177/0142723713510997

Tomasello, M., & **Mervis, C. B.** (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, **59**, 174–179. https://doi.org/10.1111/j.1540-5834.1994.tb00186.x

van Baar, A. L., **Ultee, K.**, **Gunning, W. B.**, & **Soepatmi, S.** (2006). Developmental course of very preterm children in relation to school outcome. *Journal of Developmental and Physical Disabilities*, **18**(3), 273–293. https://doi.org/10.1007/s10882-006-9016-6

Washington, J. A., & **Craig, H. K.** (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in Schools*, **30**(1), 75–82. https://doi.org/10.1044/0161-1461.3001.75

Zink, I., & **Lejaegere, M.** (2002). *N-CDI's: Lijsten voor communicatieve ontwikkeling. Aanpassing en hernormering van de MacArthur CDI's van Fenson et al.* Acco.

Zink, I., & **Lejaegere, M.** (2003). *N-CDI's: Korte vormen, aanpassing en hernormering van de MacArthur Short Form Vocabulary Checklist van Fenson et al.* Acco.

Cite this article: van der Klis, A., Junge, C., Adriaans, F., & Kager, R. (2025). Examining Dutch children's vocabularies across infancy and toddlerhood: Demographic effects are age-specific and task-specific. *Journal of Child Language* **52,** 709–728, https://doi.org/10.1017/S0305000924000199