Developmental Science

WILEY

Developmental Science

RESEARCH ARTICLE

Canonical Babbling Ratio Development in Infancy: A Systematic Review and Meta-Analysis of Methodological and Ambient Language Influences

Margaret Cychosz[1] | Helen L. Long[2]

[1]Department of Linguistics, Stanford University, Stanford, California, USA | [2]Communication Sciences Program, Department of Psychological Sciences, Case Western Reserve University, Cleveland, Ohio, USA

Correspondence: Margaret Cychosz (mcychosz@stanford.edu)

Received: 14 April 2025 | **Revised:** 5 January 2026 | **Accepted:** 8 January 2026

Keywords: babbling | biomarker | development | infancy | speech

ABSTRACT

Canonical babbling, or the production of adult-like consonant–vowel syllables in infancy, represents a critical milestone in prelinguistic vocal development and predicts later speech and language outcomes. This systematic review and meta-analysis synthesized findings from 42 studies and 1277 infants aged 5–24 months across 16 language environments to examine how methodological and contextual factors influence the most common measure of canonical babbling: the canonical babbling ratio (CBR). Results confirmed a robust, linear increase in CBR with age, reinforcing its role as a consistent developmental marker. Different CBR measures yielded comparable developmental trajectories. Sampling method significantly affected CBR values, with interactive free-play sessions eliciting higher CBRs than naturalistic (LENA) home recordings, particularly in older infants. In contrast, the location of data collection had no effect. Ambient language complexity also shaped CBR: Infants acquiring languages with more complex syllable structures (e.g., English, Dutch) initially exhibited lower CBRs compared to infants acquiring languages with less complex syllable structures (e.g., Mandarin, Spanish). Despite these initial differences, babbling trajectories were predicted to converge by approximately 20 months as infants exposed to languages with more complex syllables demonstrated accelerated CBR growth. Publication bias was detected, with smaller samples more likely to report inflated CBRs. To address this issue, simulation-based analyses are reported to estimate sample size recommendations for improved precision in future research. Together, these results support CBR as a meaningful developmental marker while offering practical guidance for future research directions and continued clinical applications.

1 | Introduction

Speech and language acquisition begins well before the onset of first words, with human infant vocalizations providing building blocks of spoken language. Among prelinguistic vocalizations, canonical babbling (CB) is a well-established developmental milestone that typically emerges in the second half of the first year of life (Oller 2000; Stark 1980; Oller 1978; 1980). CB marks the onset of consistent production of well-formed consonant-vowel (CV) syllables, with the entire syllable typically <250 ms in duration. CB syllables exhibit normal phonation, a clearly articulated consonant-vowel margin, vowel nucleus, with a rapid formant transition (<120 ms) between the margin and nucleus (Oller 2000; Stark 1980; Oller et al. 1999). These criteria have

Margaret Cychosz and Helen L. Long contributed equally to this work.

Summary

- Meta-analysis of 42 studies with 1277 infants shows canonical babbling ratio (CBR) increases linearly from 0.12 at 5 months to 0.65 at 24 months across language environments.
- Ambient language syllable complexity significantly affects early CBR trajectories, with initial differences converging by 20 months.
- Interactive free-play sampling methods where caregivers are instructed to "act as they normally do," consistently elicits higher CBRs than naturalistic, home-based recordings.
- Publication bias revealed smaller studies report inflated CBR values; simulation analyses provide sample size recommendations for reliable measurement.

provided a benchmark for identifying speech-like vocalizations in research for over two decades as the emergent frame for the later onset of words.

CB has been observed to emerge in both independent vocal play and interactive turn-taking (Oller et al. 1999; Vihman 2014; Long et al. 2022), providing infants with important practice for later speech production. CB is considered a universal milestone, observed across languages and cultures in typically developing infants (Cychosz et al. 2021; Hitczenko et al. 2023; Oller and Eilers 1988). Delayed onset, often defined as absence of CB by 10 months, is associated with increased risk for speech and language disorders and has been documented in deaf infants (Oller et al. 1985; Eilers and Oller 1994) and those with neurodevelopmental conditions such as autism, Down syndrome, Fragile X syndrome (Lang et al. 2019; Long et al. 2024; Belardi et al. 2017; Long and Hustad 2023; Lynch et al. 1995).

Despite its longstanding role as a marker of typical and atypical development, methodological inconsistency in defining, eliciting, and analyzing CB has complicated cross-study comparison. Early foundational work emphasized the need for explicit, formal definitions to ensure coding reliability, in particular (Lewedag 1995; Oller 2000). However, many subsequent studies have adopted modified or unspecified definitions—often varying in whether glides or liquids are included, whether syllables in words are counted, or what transition durations qualify—introducing substantial measurement variance (Oller et al. 1994; Stoel-Gammon 1989).

One widely used metric, the canonical babbling ratio (CBR), quantifies the proportion of canonical syllables relative to all speech-like syllables in a sample (Oller and Seibert 1988; Oller 2000). CBR was introduced as an objective, quantitative index of CB emergence, with several studies identifying a CBR of 0.15 as the onset threshold for CB when calculated using CB syllables (e.g. Oller and Eilers (1988); Lewedag (1995)), or 0.20 when calculated using utterances containing CB syllables (also noted in Lewedag, 1995). However, it is important to note that several studies have emphasized parental report, when elicited through open-ended interviews, as a reliable indicator of canonical or reduplicated babbling onset, in some cases more consistent than coder-derived CBR estimates (Oller et al. 1998; Fagan 2009; Koopmans-van Beinum and van der Stelt 1986). Moreover, although CBR provides a numerical index, its calculation still relies on coder judgments about whether each infant sound qualifies as canonical. These considerations highlight that the metric is sensitive to methodological decisions, including the choice of denominator (syllables vs. utterances), inclusion or exclusion of "true" consonant types that exclude glides and glottals (Stoel-Gammon 1989), and sampling context (home vs. lab, daylong vs. structured play). While CB onset has received substantial research attention, the present meta-analysis focuses on CBR trajectories as a function of age, method, and language environment to provide a quantitative synthesis of this widely used measure.

TABLE 1 | The 5 main CBR measures typically employed.

CBR measure	Calculation
CBR^{utt}	Number of canonical syllables/Total utterances
CBR^{syl}	Number of canonical syllables/Total syllables
CBR^{utter}	Number of canonical utterances/Total utterances
$TCBR^{utt}$	Number of true canonical syllables/Total utterances
$TCBR^{syl}$	Number of true canonical syllables/Total syllables

Note: Table adapted from Lang et al., *Clinical Linguistics & Phonetics*, 2021. Reproduced here with permission from Taylor & Francis.
Abbreviations: CBR, canonical babbling ratio; TCBR, true canonical babbling ratio.

1.1 | Canonical Babbling Ratio Measures

CBR has been in use for over three decades as a core index of advanced babbling (as discussed in Lee et al. 2018). Five main variants have emerged (as reviewed in Lang et al. 2021, and summarized in Table 1), differing in numerator, denominator, and "true" consonant inclusion criteria. This lack of standardization produces variability unrelated to developmental status, limiting cross-study comparison. Standardizing CBR computation remains a critical step toward establishing robust developmental norms.

2 | Canonical Babbling Across Settings

CBR values can also differ across recording contexts, particularly *where* (i.e., home, lab) and *how* (i.e., sampling method) infant vocalizations were recorded and analyzed (Lee et al. 2018; Lewedag et al. 1994). Traditional laboratory settings, where caregivers are instructed to engage with infants "as they normally do" in interactive free play often yield higher CBRs than daylong naturalistic home recordings, where caregiver attention varies and interaction is less directive or continuous (Lewedag et al. 1994; Long et al. 2024; Tamis-LeMonda et al. 2017). Such differences may reflect infant alertness and engagement rather than elicitation per se, given that much infant babbling is endogenously produced (Long et al. 2020; Oller et al. 2013). Understanding these contextual effects is essential for interpreting CBR findings across studies.

3 | Canonical Babbling Across Languages

CB is considered a cross-linguistic milestone (Oller et al. 1994; 2016), but language-specific phonotactic constraints and prosody may influence babbling profiles (Morgan and Wren 2018). For example, Mandarin-learning infants have shown slightly higher CBRs than English-learning infants, possibly reflecting syllable structure constraints (Lee et al. 2017). Yet systematic evaluation of cross-linguistic differences has not been conducted, thus babbling developmental trajectories across studies of diverse language environments warrant further synthesis.

4 | Research Aims

Given the definitional variability, methodological inconsistency, and recent studies on cross-linguistic patterns (Hitczenko et al. 2023; Cychosz et al. 2021), a systematic review and meta-analysis of CBR in typically developing infants is needed. This study aims to

1. Examine the typical developmental trajectory of CBR across infancy from the onset of CB (as early as 5 months) through the end of the prelinguistic period (24 months) and determine the pooled mean CBR values at different ages,
2. Assess whether different CBR measures have an effect on the observed trajectory of CBR development,
3. Determine whether ambient language exposure systematically affects CBR values in typically developing infants,
4. Evaluate the effect of methodological contexts (e.g., location and sampling methods) on typical CBR outcomes.

Across these aims, we hypothesize that (1) CBRs will increase with age across studies, reflecting a consistent typical developmental trajectory reported across prior research; (2) different CBR methods will not interact with age, consistent with prior validation studies (Nyman and Lohmander 2018); (3) CBRs will not differ by ambient language, supporting CB as a universal milestone in early human speech and language development (Oller et al. 2016), and (4) studies conducted in laboratory settings will report higher CBRs than those conducted in the home, and studies using interactive free play will yield higher CBRs than daylong naturalistic recordings due to differences in how vocal production is stimulated. As a final addition, we conduct a simulation-based analysis to estimate sample size recommendations for improved precision of CBR estimates for other researchers going forward.

By addressing these aims, this study will provide an empirically grounded synthesis of CBR norms, highlight methodological best practices, and inform both research and clinical monitoring of early vocal development.

5 | Methods

5.1 | Search Strategy

On September 4, 2024, we conducted a comprehensive literature search across three electronic databases as follows: PubMed (NCBI), Scopus (Elsevier), and Web of Science (Clarivate). Key search terms included "canonical babble" OR "canonical babbling" OR "canonical vocalization" OR "speech-like vocalization" OR "babble onset" OR "canonical syllable," using Boolean operators across each database (Appendix A). To sample broadly, there were no initial search limits applied. In addition to the database search, we conducted a gray literature search using Google Scholar and performed a hand search of citations and reference lists from studies identified during full-text screening. All search results were uploaded to the Covidence software platform (Veritas Health Innovation 2025) where duplicates were automatically and manually removed before review. Both authors then manually screened the remaining records using the eligibility criteria defined below. This work did not require an ethics board approval as no human subjects were involved.

5.2 | Eligibility Criteria

Studies were included if they examined CB development in typically developing infants at any time point between 5 and 24 months using CBR. Peer-reviewed papers published in English from the last 40 years (1984–2024) were eligible, including dissertations, research studies, and conference proceedings. Both prospective and retrospective studies were eligible. We included only observational study data; however, baseline data from experimental studies were also included. Studies examining CB in the context of any ambient language were considered eligible. Studies without quantitative CBR data, including data unable to be extracted from aggregated groups or figures, were excluded. In some clinical comparison studies, infants initially identified as having an increased likelihood of a condition due to genetic factors (e.g., sibling studies) were later classified as typically developing. We included these samples only if clear criteria were defined confirming their typical development.

5.3 | Screening

Our initial search yielded 1329 results, with three additional articles identified through other sources such as gray literature or reference lists. After removing 300 duplicates, 1029 records remained for title and abstract screening. Following this initial screening phase, 174 articles met preliminary inclusion criteria and were advanced to full-text review, where eligibility was assessed against all inclusion and exclusion criteria. Of these, 131 articles were excluded. The most common reasons for this were that CBR was not measured in the paper, CBR was measured but values were not reported, the paper was not written in English for the review team to evaluate, or the paper duplicated data that were reported in an already included paper. A final total of 42 studies were selected for inclusion and underwent data extraction and analysis. Both title-abstract and full-text screening were conducted by two members of the research team (the two authors and a trained undergraduate research assistant). Discrepancies were resolved through consensus during regular review meetings. The full screening process is illustrated in Figure 1.

5.4 | Data Extraction

Data were extracted from each study separately by at least two members of the study team (both authors and one research assistant) using the Covidence platform. Reviewers met regularly to resolve discrepancies by consensus discussion.

The extracted data included multiple variables to synthesize findings across studies to address our research questions. These included the following:

1. Participant demographics: Key characteristics of the sample, including total sample size, infant ages studied (range and mean), gender distribution, and ambient language environment
2. Study design: Whether the study used longitudinal, cross-sectional, case-control, or other design types.
3. Study location: Setting where data were collected, such as home, laboratory, hospital, or other contexts.
4. Sampling method: The approach used to engage infants in vocalization, including caregiver instructions to engage "as they normally do" at home with their infants (i.e., interactive free play), daylong naturalistic recordings (e.g., language environmental analysis [LENA] recordings), or other methods.
5. Elicitation method: The interaction partner(s) during vocal sampling, such as caregiver–infant, infant–researcher, or environmental recording.
6. Coding method: Whether CB was hand-coded by trained experts or annotation was crowd-sourced through citizen scientists.
7. CBR measure: The specific CBR calculation method used as reported in Table 1 or other measures as reported.
8. CBR: Mean CBRs were extracted from each study by infant age and language (including sample sizes within each). Standard deviations (SD) were extracted if reported or calculated by the research team. Additional details on missing SD data for some samples are provided in "Meta Analysis Plan."

5.5 | Quality Appraisal

Due to substantial heterogeneity in study design and sample characteristics, we included a quality appraisal to evaluate methodological rigor. Existing appraisal tools developed for health-related clinical trials were not appropriate for the observational studies of typical development included in this review. Given the multidisciplinary nature of the evidence base, we developed a custom 10-item checklist (Appendix B) to assess studies from both clinical and nonclinical disciplines. The checklist was structured to maintain core standards of rigor while remaining flexible across diverse study designs.

All studies were appraised by at least one member of the study team. To minimize bias, no author judged their own prior work; instead, another study team member was assigned to these cases (five total studies that belonged to one of the authors were assigned to another study team member). After completing the checklist, reviewers assigned a subjective judgment of "Good," "Fair," or "Poor" to each study. Specifically, each study's methodological quality was judged qualitatively as good, fair, or poor based on the overall pattern of ratings across individual checklist criteria. In line with recommendations for judgment in appraisal tools developed for clinical studies (Hong et al. 2018), no total scores were calculated; instead, these categorical judgments reflected the reviewers' confidence in the rigor and interpretability of each study' s methods and findings.

5.6 | Statistical Plan

To investigate the factors potentially influencing CBR values, we employed mixed effects modeling, incorporating both fixed and random effects to account for study-level and within-study variability. The fixed effects structure included the key predictor of interest (e.g., age, language environment). Age was centered and scaled at 5 months, so model coefficients can be interpreted as the value at 5 months of age (the earliest age in our sample), or in the case of parameters interacting with age, the developmental change from 5 months. The random effects structure was specified as `(~1|study_id/effect_id)`, where `study_id` accounts for between-study variation, and `effect_id` represents unique combinations of each individual study and age, in months. This nested structure was chosen because studies often reported multiple data points from different ages (e.g., a given study could report data from 5-month-olds, 6-month-olds, etc.), making it necessary to model age-related dependencies within studies. In other words, this random effect modeling approach accounts for the fact that observations of the same age, or from the same study, are more likely to be similar to one another than observations from either a different study and/or a different age.

All models were built step-wise, starting with the baseline model consisting of the random effects structure described above and `CBR_measure` (matching vs. mismatched, explained in detail below) to account for potential differences in CBR values due to measurement differences. Then the variable of interest was added to the model and, if of theoretical interest, its interaction with age was tested. We used a restricted maximum likelihood (REML) estimator to calculate tau ^{2} (between-study variance), as REML has been shown to be more conservative and less biased than other estimators, particularly for meta-analyses with heterogeneous studies. All modeling was conducted in R (RStudioTeam 2024) using the `metafor` package (Viechtbauer 2010). Data visualizations were generated using `ggplot2` (Wickham 2016). See project documentation for software package versions. Some parts of initial analysis code were developed using Claude-Sonnet after which the first author reviewed and edited the code line by line, including adding additional tests, prior to adding any results to the manuscript.

Given variability in reporting formats across studies, we took several data preparation steps before modeling. First, for studies reporting individual infant CBR values, data were aggregated

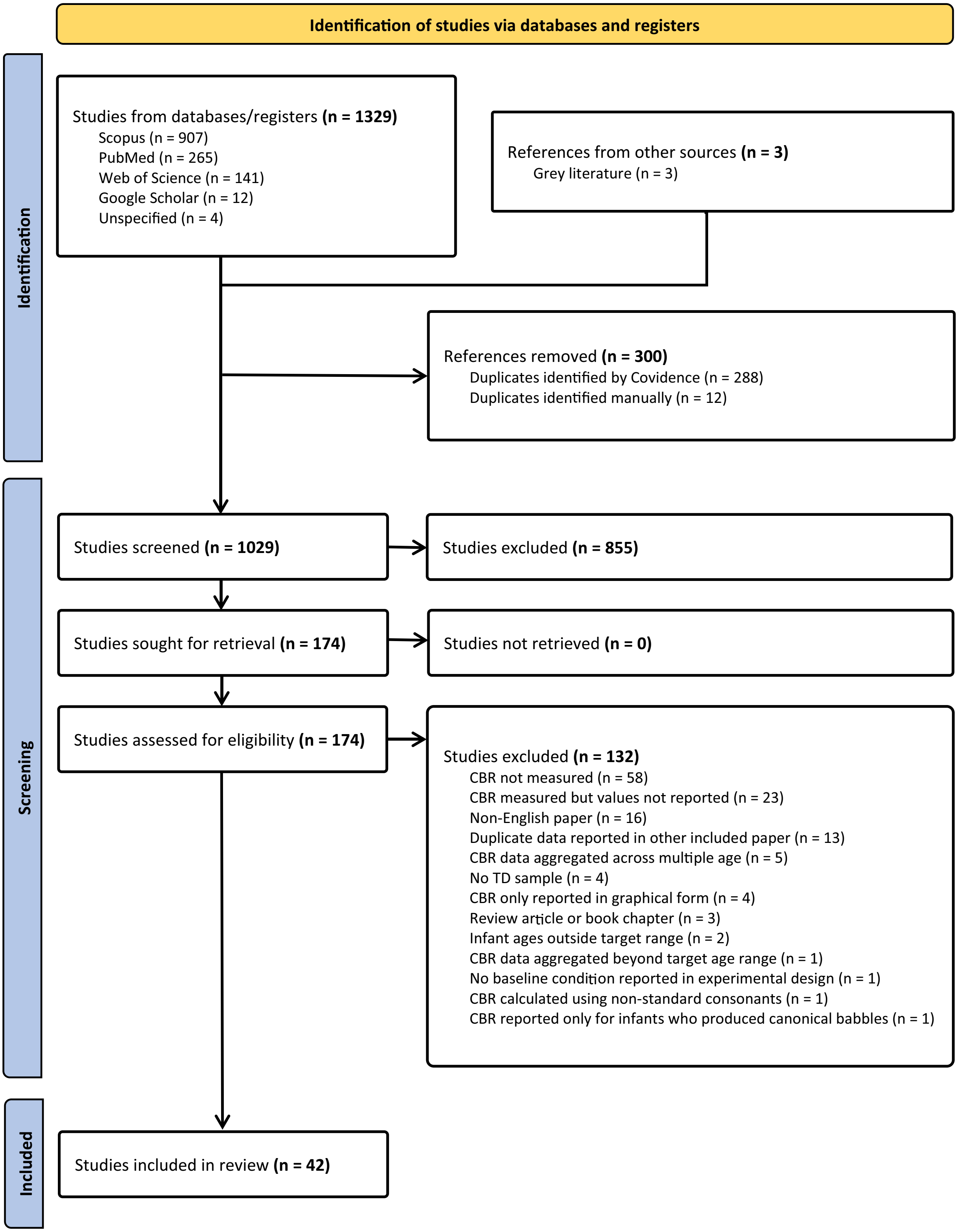

FIGURE 1 | PRISMA flow diagram illustrating the study selection process for the systematic review and meta-analysis of studies examining canonical babbling ratios (CBR) in typically developing infants aged 5–24 months. Of 1329 records identified, 1029 were screened, and 42 studies met inclusion criteria for final analysis.

across infants by study and age (in months) to ensure consistency in effect size estimation. Second, we addressed two types of missing SD data. Missing SDs arose in two scenarios as follows: (1) studies reporting only one infant at a given age, and (2) studies reporting group means without accompanying SDs. To handle these missing SDs, we applied an imputation method based on the coefficient of variation (CV = SD/mean), which we calculated from studies with complete SD data. This approach allowed us to maintain age-appropriate imputation while ensuring all studies could be included in the analysis. To prevent computational issues from zero or near-zero variances, we set a minimum variance threshold of 0.0001 to ensure numerical stability.

Finally, because different studies used different methods to calculate CBR (see Table 4), we classified CBR measures into two broad categories as follows: those with matching numerator and denominator (e.g., syllables/syllables or utterances/utterances) and those with mismatched numerator and denominator (e.g., syllables/utterances). This grouping was informed by prior validation studies (Lang et al. 2021; Nyman and Lohmander 2018), which found high consistency among measures within each category. Grouping measures at this level allowed us to account for systematic differences in CBR calculation while preserving statistical power and avoiding overfitting that would result from modeling sparsely represented individual CBR formulas.

6 | Results

6.1 | Descriptive Characteristics of Included Studies

6.1.1 | Infant Data and Research Designs Across Studies

Our systematic review initially identified 1329 studies meeting the inclusion criteria published between 1989 and 2024. Within these works, some research groups reported CBRs for the same infants across multiple studies, often addressing different research questions. However, to ensure sample independence, our criteria required that *each infant be represented only once per age (in months) within our final dataset*. Consequently in a later stage of screening, 13 studies that otherwise met inclusion criteria were excluded because they duplicated CBR data for the same infants at the same age as another included study. For studies using longitudinal designs, we retained repeated measures of infants who contributed more than one time point. For cross-sectional studies, each infant contributed data at only one time point. Thus, while some infants had multiple observations in longitudinal datasets, no infant was counted more than once for any given age.

After these adjustments, our final sample comprised 42 total studies (full list available in Table 2) reporting on a total of 1277 unique infants. Of these, 25 studies [59.52%] contained at least some longitudinal data, resulting in a total of 2944 infant observations across all time points (hereafter referred to as "observations").[1] The average age at observation was 11.60 months (SD = 4.38). Among the 30 studies (71.43%) that reported infant gender information: The gender distribution was N = 415 male (51.94% of infants among the studies that reported gender) and N = 384 female (48.06%). Table 3 presents the key characteristics of the included studies.

6.1.2 | CBR Measures

We identified five different CBR measures used across studies. Specifically, N = 22 (52.38%) studies used $\mathrm{CBR}^{\mathrm{syl}}$ (canonical syllables/total syllables), N = 4 (9.52%) used $\mathrm{CBR}^{\mathrm{utt}}$ (canonical syllables/total utterances), N = 11 (26.19%) used $\mathrm{CBR}^{\mathrm{utter}}$ (canonical utterances/total utterances), N = 2 (4.76%) used $\mathrm{TCBR}^{\mathrm{syl}}$ (true canonical syllables/total syllables), and N = 3 (7.14%) used $\mathrm{TCBR}^{\mathrm{utt}}$ (true canonical syllables/total utterances). Several studies used other measures, including N = 4 studies (9.52%) that used the Stark Assessment of Early Vocal Development-Revised (SAEVD-R; Nathani et al. 2006), and N = 1 study that used the mean babbling level (MBL; Stoel-Gammon 1989). Both SAEVD-R and MBL calculate canonical and noncanonical vocalizations using utterance-level calculations so we converted these measures to $\mathrm{CBR}^{\mathrm{utter}}$ to address our research questions (the descriptive statistics above reflect our final classification of SAEVD-R and MBL). See Table 4 for methodological details of the included studies.

6.1.3 | Location, Sampling, and Elicitation Methods

The majority of studies collected data in either a home setting (N = 22 [52.38%]) or a laboratory setting (N = 12 [28.57%]). A smaller number of studies were conducted in other environments, including hospitals or medical clinics (N = 2 [4.76%]), a school or daycare center (N = 1), or multiple environments, such as a daycare/home (N = 2), lab/home (N = 1), or clinic/home (N = 1). Location was not reported for one paper. For the meta-analysis, we compared CBR only between home and lab settings, as sample sizes for other locations were too small for meaningful statistical comparisons.

Within these settings, sampling methods varied. The most common sampling method was an interactive free play session (N = 22 [52.38%]) where caregivers were provided a selection of toys and instructed to engage with their infant naturally, followed by daylong audio recordings of the home environment (N = 10 [23.81%]). The remaining studies used semi-structured play protocols with a predefined set of interaction scenarios (N = 5 [11.90%]), retrospective home recordings self-selected and submitted to researchers by caregivers (N = 3 [7.14%]), or an experimental manipulation (N = 1, measures from the baseline condition were used). As above, we analyzed CBR only by interactive free play versus daylong home recordings because of the small number of studies that used other sampling methods.

Elicitation method also varied in terms of the interaction partner, with infants engaging with caregivers, researchers, or both. The most common method used caregiver–infant interactions (N = 17 [40.48%]), followed by daylong home recording of the environment (N = 12 [28.57%]). The remaining studies used caregiver–infant–researcher interactions (N = 8 [19.05%]), infant–researcher interactions (N = 2 [4.76%]), or a combination of these methods (N = 3 [7.14%]). Elicitation method is only reported descriptively here but was not included in the meta-analysis.

TABLE 2 | Included Studies.

Paper	Age Range (months)	Sample Size	Language Environment(s)	Location	Elicitation Method	CBR Measure
Albert et al. 2022	10–15	34	English	School/daycare	Parent-child interaction	CBR^{utt}
Andruski et al. 2013	12–13	1	English, Spanish	Not stated	Parent-child interaction	CBR^{syl}
Athari et al. 2021	6–14	9	English	Laboratory	Parent-child interaction	CBR^{utt}
Bartl-Pokorny et al. 2022	9–11	3	German	Home	Parent-child interaction	CBR^{utter}
Chan et al. 2022	12–24	2	Mandarin	Hospital or home	Parent-child; Parent-child-lab staff; Child-lab staff interaction	CBR^{utter}
Chapman et al. 2003	9–14	15	English	Home	Parent-child interaction	$TCBR^{syl}$
Chapman et al. 2001	9–10	15	English	Home	Parent-child interaction	$TCBR^{syl}$
Chen and Kent 2010	6–17	24	Mandarin	Home/daycare	Parent-child; Naturalistic recording (LENA)	CBR^{syl}
Chen et al. 2011	8–24	1	Mandarin	Home	Parent-child; Naturalistic recording (LENA)	CBR^{syl}
Cychosz et al. 2021	6–24	35	English, Tseltal, Yélî, Tsimané, Quechua	Home	Naturalistic recording (LENA)	CBR^{utter}
Deveney and Kyvelidou 2020	6–12	10	English	Home	Parent-child interaction	CBR^{syl}
D'Odorico et al. 2011	12–18	15	Italian	Laboratory	Parent-child; Parent-child-lab staff interaction	CBR^{utt}
Ertmer and Jung 2012	6–12	11	English	Laboratory	Parent-child interaction	SAEVD-R converted to CBR^{utter}
Ertmer et al. 2013	15–24	11	English	Laboratory	Parent-child interaction	SAEVD-R converted to CBR^{utter}
Gipson et al. 2021	12	41	English, Spanish-English	Home and lab	Naturalistic recording (LENA)	CBR^{syl}
Goldstein and Schwade 2008	9–10	60	English	Laboratory	Parent-child interaction	CBR^{syl}
Ha 2024	6–14	20	Korean	Home	Naturalistic recording (LENA)	CBR^{syl}
Ha et al. 2021	9–21	28	English	Home	Naturalistic recording (LENA)	CBR^{syl}
Hamrick et al. 2023	5–20	36	English	Home	Naturalistic recording (LENA)	CBR^{syl}

(Continues)

Paper	Age Range (months)	Sample Size	Language Environment(s)	Location	Elicitation Method	CBR Measure
Hitczenko et al. 2024	5–24	186	English, French, Tsimané, Yélî, Solomon*, Vanuatu*	Home	Naturalistic recording (LENA)	CBR^{utter}
Jones et al. 2003	16.71	14	English	Home	Parent-child interaction	CBR^{syl}
Lang et al. 2021	9	20	German	Home	Parent-child; Naturalistic recording (LENA)	CBR^{utter}
Lee et al. 2018	6–11	8	Mandarin (small amounts of Southern Min and Cantonese)	Home	Naturalistic recording (LENA)	CBR^{syl}
Lohmander et al. 2011	12	21	Swedish	Hospital/clinic	Parent-child-lab staff interaction	$TCBR^{utt}$
Lohmander et al. 2017	10	30	Swedish	Laboratory	Parent-child-lab staff interaction	$TCBR^{utt}$
Long et al. 2024	5–13	127	English	Home	Naturalistic recording (LENA)	CBR^{syl}
Long et al. 2023	6–19	10	English	Home	Naturalistic recording (LENA)	CBR^{syl}
Lopez et al. 2020	12–15	53	English	Home	Naturalistic recording (LENA)	$TCBR^{utt}$
Molemans et al. 2012	6–24	40	Dutch	Home	Parent-child interaction	CBR^{syl}
Murillo and Capilla 2016	9–15	11	Spanish	Laboratory	Parent-child-lab staff interaction	CBR^{syl}
Myers-Jennings 1994	7–11	25	English	Home/daycare	Parent-child interaction	CBR^{syl}
Oller et al. 1997	6–18	18	Bilingual English-Spanish	Laboratory	Parent-child-lab staff interaction	CBR^{syl}
Patten et al. 2014	9–18	14	Not listed but English assumed	Home	Naturalistic recording (LENA)	CBR^{syl}
Paul et al. 2011	6–12	31	English	Laboratory	Parent-child interaction	$TCBR^{utt}$
Price et al. 2006	15–23	6	Mandarin pre-adoption, English post-adoption	Home	Parent-child-lab staff interaction	CBR^{syl}
Rvachew et al. 2005	9–18	10	English	Hospital/clinic	Parent-child interaction	CBR^{utt}
Smith et al. 2010	9–19	6	English	Home	Child-lab staff interaction	CBR^{syl}
Stoel-Gammon 1989	9–18	32	English	Laboratory	Parent-child; Parent-child-lab staff interaction	CBR^{utter}
Ward et al. 2023	6–15	18	Australian English	Home	Naturalistic recording (LENA)	SAEVD-R converted to CBR^{utter}

(Continues)

TABLE 2 | (Continued)

Paper	Age Range (months)	Sample Size	Language Environment(s)	Location	Elicitation Method	CBR Measure
Warren 2018	8–16	15	English**	Laboratory	Parent-child interaction	SAEVD-R converted to CBR^{utter}
Wass et al. 2022	12	82	Not listed but English assumed	Home	Parent-child interaction	CBR^{syl}
Yankowitz et al. 2022	6–12	73–129	English	Laboratory	Child-lab staff interaction	CBR^{syl}

Note: LENA = Language ENvironment Analysis system, a wearable recording device. *Both Solomon and Vanuatu are language environments where children are exposed to two or more of several different languages; see Methods. **One child was exposed to English/Indian/Vietnamese, and another English/Arabic.

TABLE 3 | Summary of study and participant characteristics.

Characteristic	*N*	%	*M* (SD)	Range
Study information				
Number of studies	42			
Publication years				1989–2024
Participant characteristics				
Total observations	2944			
Study sample size (unique infants)			30.40 (37.47)	1–186
Infant age at observation (months)			11.60 (4.38)	5–24
Gender distribution				
Males (*N* observations)	415	51.94		
Females	384	48.06		
N of studies that did not report gender	12	28.57[a]		
Ambient language information				
Number of distinct lang. environments [b]	16			
English (*N* observations)	1760	59.50		
Dutch	688	23.26		
Bilingual English-Spanish	162	5.48		
Swedish	51	1.72		
Mandarin	50	1.69		
Solomon	42	1.42		
Korean	40	1.35		
Spanish	35	1.18		
German	29	0.98		
Italian	25	0.85		
Tsimane'	17	0.57		
Yélî Dnye	16	0.54		
Vanuatu	12	0.41		
French	10	0.34		
Tseltal Mayan	5	0.17		
Bilingual Spanish+Quechua	2	0.07		

Note: [a] Percentages for gender distribution are calculated based on studies that reported gender information. Studies could report gender information for subjects without de-aggregating by gender for CBR reports.
[b] We refer to these as "language environments" as the Solomon and Vanuatu environments are highly multilingual, and children are exposed to two or more of several languages; see Hitczenko et al. (under review) for detail.

TABLE 4 | Summary of study methods.

Methodological characteristics	*N* studies	%
Study location		
Home	22	52.38
Laboratory	12	28.57
Clinic	2	4.76
Daycare	1	2.38
Multiple settings	4	9.52
Sampling methods		
Interactive free play	22	52.38
Home audio recordings	10	23.81
Structured play	5	11.90
Retrospective home recordings	3	7.14
Experimental manipulation	1	2.38
Elicitation methods		
Parent–infant interactions	17	40.48
Daylong recordings	12	28.57
Parent–infant–researcher interactions	8	19.04
Infant–researcher interactions	2	4.76
Naturalistic parent–infant interactions	3	7.14
CBR measures		
CBR^{syl}	22	52.38
CBR^{utt}	4	9.52
CBR^{utter}	11	26.19
$TCBR^{syl}$	2	4.76
$TCBR^{utt}$	3	7.14

Note: CBR^{syl} = canonical syllables/total syllables; CBR^{utt} = canonical syllables/total utterances; CBR^{utter} = canonical utterances/total utterances; $TCBR^{syl}$ = true canonical syllables/total syllables; $TCBR^{utt}$ = true canonical syllables/total utterances. One study did not report location.

6.1.4 | Coding Method

Nearly all of the studies ($N = 40$ [95.24%]) used human coding by trained research assistants to classify vocalizations as canonical and noncanonical. The remaining studies ($N = 2$ [4.76%]) used crowd-sourced annotation by nonexperts.

6.1.5 | Ambient Language Environment

The studies included children exposed to 16 different language environments. See Table 3 for complete list and distribution across the sample.

6.2 | Quality Appraisal

Of the 42 included studies, 27 were rated as Good quality (64.29%), 11 as Fair quality (26.19%), and four as Poor quality (9.52%) based on our custom 10-item appraisal checklist (see Appendix B). The most common methodological concerns were

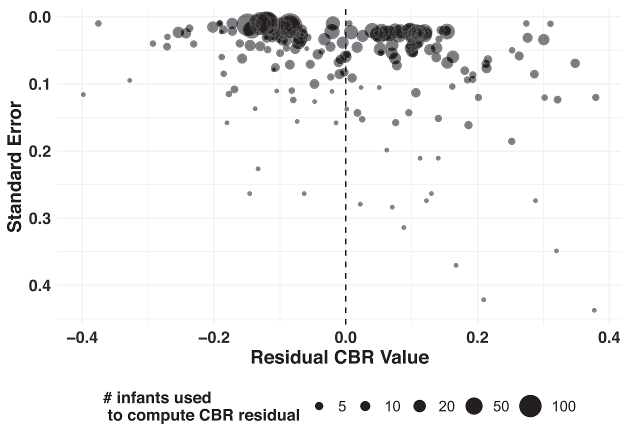

FIGURE 2 | Funnel plot showing the relationship between residual CBR values and standard error. Residuals represent deviations from the meta-analytic model that controls for age and CBR measurement type. Each point represents a combination of study and infant age, in months ($N = 190$ effect sizes). Point size refers to the number of infants used to compute each CBR value (range = 1–129). Points in the legend are discrete, to facilitate interpretation, but point size is continuous. The asymmetrical distribution, with more points in the lower right quadrant than the lower left, suggests publication bias, which was confirmed via Egger's test; see text for detail. *Note:* SEs are reported here, while SDs were used in the primary meta-analysis models.

insufficiently detailed eligibility criteria ($N = 12$, 28.57%) and inadequate acknowledgment of study limitations ($N = 12$, 28.57%). To evaluate the potential influence of study quality on our meta-analysis, we conducted a sensitivity analysis excluding the four studies rated as Poor quality (see Supporting Materials I). Results were consistent with the full dataset, indicating that including these studies did not meaningfully alter our findings. Thus, we retained all studies in our full analyses reported below.

6.3 | Publication Bias

We first report the results of our publication bias analysis. To assess bias, we conducted Egger's test for funnel plot asymmetry (Figure 2). Results indicate evidence of publication bias, as demonstrated by a model intercept significantly different from 0 ($\beta = -0.09$, $p < 0.001$): As standard error (SE) decreases (i.e., precision increases), CBRs tended to be lower than the model/meta-analytic average, indicating that smaller studies may be overestimating CBR relative to larger, more precise studies.[2] Additionally, a positive, significant effect of SE in the model indicates that as SE increases (i.e., precision decreases), CBR also increases ($\beta = 2.02$, $p = 0.002$). This pattern is consistent with small-study effects, where studies with samples that have lower statistical power tend to report inflated effect sizes due to sampling variability or selective reporting.

To mitigate the impact of publication bias in our meta-analysis, we used inverse variance weighting, where each sample's contribution was adjusted based on its estimated precision (variance = SD^2/n). This approach ensured that larger, more precise samples received greater weight in the analysis, reducing the influence of small-sample studies on overall estimates. In addition, to inform field standards on adequate sample size, we conducted

TABLE 5 | Summary of meta-regression model for CBR by infant age and CBR measure type.

Parameter	Estimate	SE	*z*	*p* value	95% CI
Intercept	0.12	0.03	4.79	<0.001***	0.07–0.17
Age (months)	0.03	0.001	19.25	<0.001***	0.02–0.03
CBR measure [mismatched]	0.04	0.06	0.74	0.459	−0.07–0.16

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

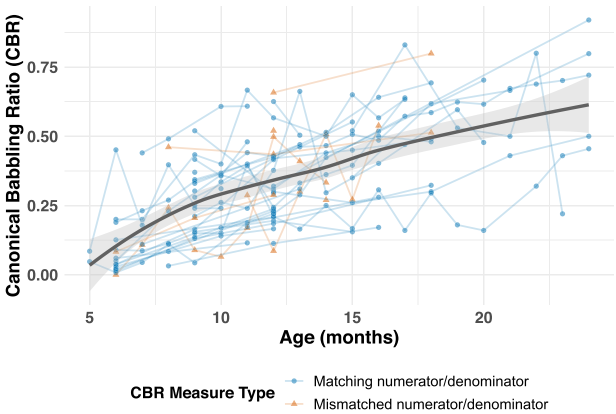

FIGURE 3 | Canonical babbling ratio (CBR) derived from meta-analysis by infant age, in months. Individual data points represent a single combination of study by age (i.e., CBR data from infants of a particular age within a study). Thin lines connect longitudinal observations from individual studies. The dark gray line shows the locally weighted smoothing curve (loess) fit to raw data. Gray ribbon represents standard error.

simulation-based estimations to determine the minimum number of infants required to reliably estimate CBR at each month from 5 to 24 months. Results of these simulations are reported in Section 3.6.

6.4 | Meta-Analysis of CBR by Age

To examine how CBR changes between 5 and 24 months, we added `Infant Age` (continuous variable, in months) as a fixed effect to the baseline model predicting CBR values.[3] `Infant Age` significantly improved model fit (likelihood ratio test: $\chi^2 = 175.76$, $p < 0.001$), and showed a significant, positive effect on CBR ($\beta = 0.03$, SE = 0.001, $p < 0.001$): For each additional month of age, CBR increases by approximately 0.03 units (see Table 5 for model summary and Figure 3 for visualization). We tested an interaction between `Infant Age` and `CBR measure` to assess whether different measurement approaches influence CBR developmental trajectories. This interaction did not improve model fit ($p = 0.80$), suggesting that the relationship between CBR and age remains consistent across measurement approaches. Based on model fit statistics (AIC difference of 2), we retained the simpler main-effects model without the interaction term, but report the interaction model summary in Supporting Materials I.

Finally, we evaluated whether CBR development follows a linear or quadratic trajectory with age by comparing models with linear versus quadratic age terms. The model with the quadratic term did not significantly improve fit over the linear model ($p =$ 0.35), suggesting that CBR development progresses at a relatively consistent rate between 5 and 24 months. In Table 6, we list the model estimates for each month of development and in Appendix C, we list the raw CBR values, and sample size used for their calculation, for each month of development.

TABLE 6 | Model-estimated CBR by infant age, in months, for all $N = 16$ language environments (center) and just English (right), derived from meta-analysis ($N = 42$ studies, $N = 1277$ unique infants, $N = 2944$ observations).

	Predicted CBR [95% CI]	
Age (months)	**All languages**	**English**
5	0.12 [0.07, 0.17]	0.14 [0.09, 0.20]
6	0.15 [0.1, 0.20]	0.17 [0.12, 0.23]
7	0.18 [0.13, 0.22]	0.2 [0.14, 0.25]
8	0.2 [0.16, 0.25]	0.23 [0.17, 0.28]
9	0.23 [0.19, 0.28]	0.25 [0.20, 0.30]
10	0.26 [0.21, 0.31]	0.28 [0.23, 0.33]
11	0.29 [0.24, 0.33]	0.31 [0.25, 0.36]
12	0.31 [0.27, 0.36]	0.33 [0.28, 0.38]
13	0.34 [0.30, 0.39]	0.36 [0.31, 0.41]
14	0.37 [0.32, 0.42]	0.39 [0.34, 0.44]
15	0.40 [0.35, 0.44]	0.41 [0.36, 0.47]
16	0.43 [0.38, 0.47]	0.44 [0.39, 0.49]
17	0.45 [0.40, 0.50]	0.47 [0.42, 0.52]
18	0.48 [0.43, 0.53]	0.50 [0.44, 0.55]
19	0.51 [0.46, 0.56]	0.52 [0.47, 0.58]
20	0.54 [0.48, 0.59]	0.55 [0.49, 0.60]
21	0.56 [0.51, 0.62]	0.58 [0.52, 0.63]
22	0.59 [0.54, 0.64]	0.60 [0.55, 0.66]
23	0.62 [0.56, 0.67]	0.63 [0.57, 0.69]
24	0.65 [0.59, 0.70]	0.66 [0.60, 0.72]

Note: Values represent model-predicted CBRs from meta-regression, controlling for CBR measure.
Abbreviation: CI, confidence interval.

6.5 | Additional Factors Affecting CBR

6.5.1 | Language Environment

To investigate the influence of language environment on CBR, we first compared English versus non-English learning

TABLE 7 | Meta-analytic model results for factors potentially affecting CBR development.

Model	Model parameter	Estimate	SE	z	p value	95% CI
+/− English	Intercept	0.09	0.03	3.03	0.002**	[0.032, 0.151]
	Language: Not English	0.06	0.03	1.69	0.091	[−0.009, 0.120]
	Age	0.03	0.00	10.19	<0.001***	[0.023, 0.033]
	CBR measure: Mismatched	0.06	0.05	1.07	0.283	[−0.049, 0.166]
	Language: Not English*Age	−0.00	0.00	−0.44	0.660	[−0.008, 0.005]
Syllable complexity	Intercept	0.20	0.04	5.21	<0.001***	[0.124, 0.275]
	Syll. complexity: More complex	−0.10	0.04	−2.81	0.005**	[−0.174, -0.031]
	Age	0.02	0.00	8.56	<0.001***	[0.017, 0.027]
	CBR measure: Mismatched	0.05	0.06	0.85	0.393	[−0.064, 0.163]
	Syll. complexity: More complex*Age	0.01	0.00	2.05	0.041*	[0.000, 0.013]
Location	Intercept	0.12	0.03	4.13	<0.001***	[0.063, 0.178]
	Location: Unfamiliar	−0.03	0.05	−0.51	0.608	[−0.130, 0.076]
	Age	0.03	0.00	17.04	<0.001***	[0.023, 0.029]
	CBR measure: Mismatched	0.05	0.06	0.91	0.364	[−0.063, 0.171]
	Location: Unfamiliar*Age	0.00	0.00	1.29	0.197	[−0.002, 0.011]
Multilingual	Intercept	0.12	0.03	4.65	<0.001***	[0.070, 0.171]
	Multilingual status: Multilingual	−0.06	0.05	−1.23	0.217	[−0.155, 0.035]
	Age	0.03	0.00	17.39	<0.001***	[0.025, 0.031]
	CBR measure: Mismatched	0.04	0.06	0.74	0.458	[−0.070, 0.156]
	Multilingual status: Multilingual:Age	0.01	0.00	1.03	0.301	[-0.005, 0.015]
Sampling method	Intercept	0.07	0.04	1.79	0.073†	[−0.006, 0.141]
	Method: Interactive free play	0.09	0.05	1.81	0.071†	[−0.007, 0.181]
	Age	0.02	0.00	10.29	<0.001***	[0.019, 0.028]
	CBR Measure:Mismatched	−0.04	0.05	−0.68	0.498	[−0.144, 0.070]
	Method: Interactive free play*Age	0.01	0.00	2.52	0.012*	[0.002, 0.013]

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

environments by adding a binary "`+/−English`" parameter to the model controlling for age and CBR measure type. Neither a main effect of `+/−English` environments (p = 0.06), nor its interaction with `Infant Age`, improved baseline model fit (p = 0.66).

We then conducted a more nuanced analysis by classifying languages according to `Syllable Structure Complexity` (classified into a binary category of *Less Complex*: Mandarin, Korean, Italian, Spanish, Tseltal, Tsimane, Yélî Dnye and *More Complex*: English, Dutch, German, French, Swedish). For this analysis, we removed infants in multilingual environments where the languages had mismatched syllable complexity (e.g., Mandarin and English) or we could not determine the specific languages of exposure (e.g., Solomon and Vanuatu environments where infants are exposed to two or more of several different languages). Both the main effect of `Syllable Complexity` and its interaction with `Infant Age` improved upon baseline model fit: Infants learning languages with more complex syllable structures demonstrated significantly lower initial CBR values at 5 months ($\beta = -0.10$, $p = 0.005$), but showed faster developmental trajectories with a significant syllable complexity*age interaction ($\beta = 0.01$, $p = 0.040$; see Table 7). Model predictions suggest that trajectories would converge around 20.3 months (Figure 4), indicating that syllable complexity shapes early foundations of babbling, and that early differences in CBR are gradually overcome through accelerated development. Overall, these results demonstrate an effect of ambient language structure, specifically syllable complexity, on CBR development between 5 and 24 months of age, showing that this vocal milestone is sensitive to language structure.

As a final step to evaluate potential effects of the language environment on CBR, we evaluated potential differences in the measure by the infants' multilingual status. We compared monolingual versus multilingual environments by adding a binary "`Multilingual vs. Monolingual`" parameter to the model controlling for age and CBR measure type. Neither a main effect of `Multilingual vs. Monolingual` environments (p = 0.47), nor its interaction with `Infant Age`, improved baseline model fit (p = 0.31), indicating no differences in early CBR trajectories by infants' exposure to one versus two or more languages.

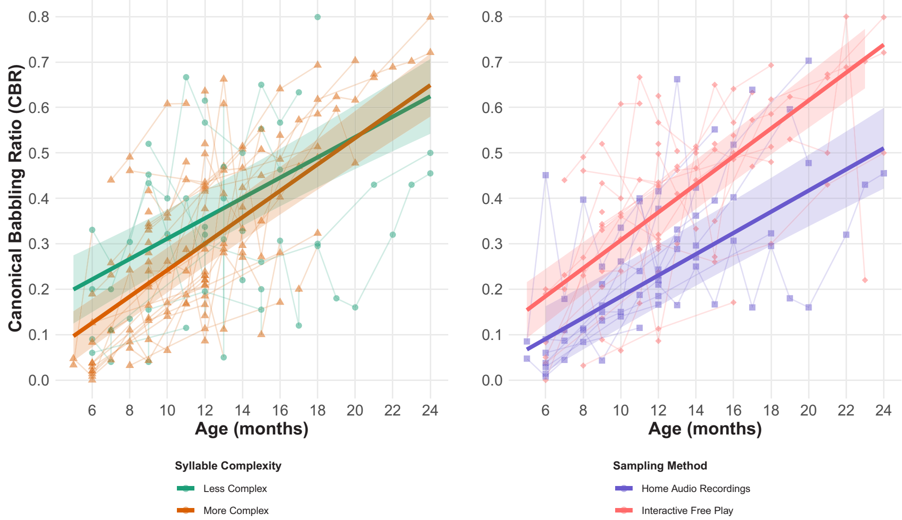

FIGURE 4 | Factors affecting CBR development from 5 to 24 months. In both plots, thick lines represent meta-analytic model-predicted values and ribbons represent 95% confidence intervals around model fit. Thin lines and individual points show individual study trajectories and observations of study by age. Left: Effect of syllable complexity, comparing languages with less (green) versus more (orange) syllable structure complexity. Right: Effect of sampling method, comparing interactive free play sessions, often in the lab, (pink) versus naturalistic home recordings (purple). Both plots: Syllable complexity shows convergence over developmental time, while the effect of sampling method increases with age.

6.5.2 | Location

Neither the main effect of testing location (laboratory vs. home settings; $p = 0.94$), nor its interaction with age ($p = 18$), improved upon the baseline model suggesting that the location of data collection does not impact CBR measurement between 5 and 24 months.

6.5.3 | Sampling Method

We evaluated how sampling method may impact CBR measurement by comparing interactive free play (typically lab-based with researcher present) versus naturalistic home recordings (e.g., LENA recordings). After controlling for age and CBR measurement approach, results showed a significant interaction between sampling method and age ($\beta = 0.01$, $p = 0.012$), with interactive free play eliciting higher CBR values compared to naturalistic home recordings (Table 7). The interaction with age indicates that the disparity between sampling methods grows larger over developmental time: At 5 months, there was no effect of sampling method on CBR estimation ($p = 0.07$), but significant differences emerged by 6 months of age ($p = 0.04$; Figure 4). Unlike the syllable complexity effect where trajectories converge, the difference between sampling methods continued to diverge throughout development, suggesting older infants may be more sensitive to sampling method when producing CB.

6.6 | Simulations to Estimate Appropriate Sample Sizes

We conducted a simulation to determine the minimum necessary sample size of infants to achieve a reliable measurement of CBR for each month of development from 5 to 12 months, and every 3 months from 12 to 24. For the simulation, we first extracted the meta-analytic estimates of predicted CBR values at each month (see Table 6) and observed SD for that age to reflect the empirical between-infant variability observed across studies (see Appendix C).[4] For each combination of sample size (5–200) and age (5–24 months), we conducted 1000 runs randomly generating CBRs based on the model-estimated CBR and raw data SD for that age (e.g., $n = 5$, $n = 6 \ldots 200$ 6-month-olds, 7-month-olds, etc.). Finally, we calculated key precision metrics including relative error (mean error divided by true CBR) and 95% confidence interval coverage probability (Figure 5).

To determine the minimum sample size we recommend at each age, we identified the smallest sample that achieved the following three different precision thresholds: margins of error of 0.07, 0.05, and 0.03 (Table 8). These results can thus help researchers ensure their estimates of CBR are reliable based on infant age and precision required/desired.

7 | Discussion

This systematic review and meta-analysis synthesized findings from 42 studies with a combined sample of 1277 infants to examine the typical trajectory of CB using the CBR, one of the most widely used quantitative metrics of CB development. Our goal was to identify typical patterns in CBR and evaluate how methodological and linguistic factors may shape observed babbling outcomes. Consistent with a large body of literature beginning in the late 1970s (Oller 1978, 1980, 2000; Stark, 1980), we observed a robust, linear increase in CBR from 5 to 24 months across the combined sample of infants. We also found no evidence that the developmental slope of CBR differed by CBR metric, supporting the use of CBR as a robust developmental indicator in relative terms, although the long-standing call for explicit operational definitions remain (Oller 2000; Lewedag 1995). Regarding contextual effects, our findings suggest that sampling method

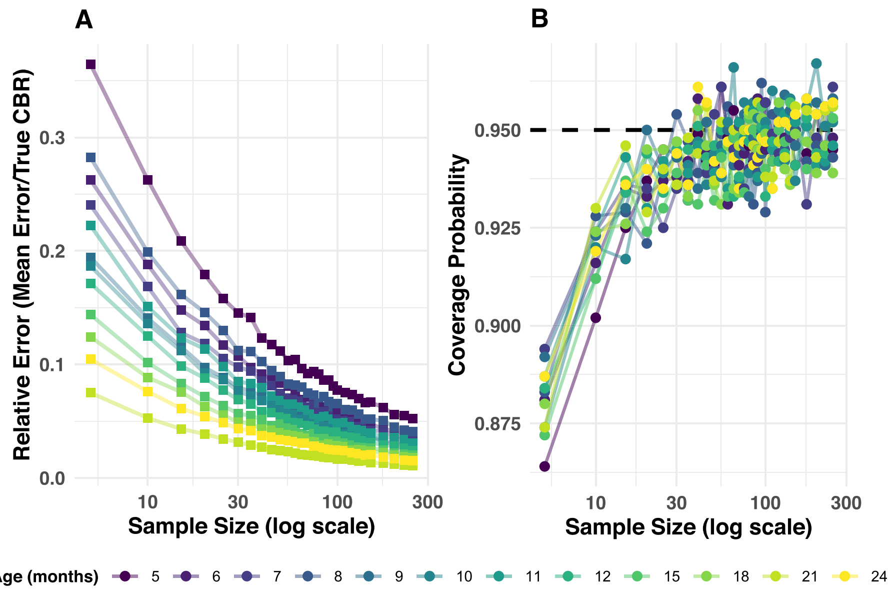

FIGURE 5 | Simulation results for sample size recommendations in CBR measurement based on 1000 simulation runs/sample size. Results are derived from meta-analytic model-estimated means and raw standard deviations from participant data, by age. *A* = Relative error (mean error/CBR) by sample size and infant age (months). Relative error decreases rapidly up to approximately 30 participants, with diminishing returns for larger samples. *B* = 95% confidence interval coverage probability by sample size and infant age (months). The dashed line represents the ideal 95% confidence interval coverage rate.

(naturalistic home audio recordings vs. interactive free play, usually in the lab with a researcher present), rather than location (lab vs. home), has the strongest influence on measured CBR. Our analysis of ambient language effects found that syllable structure complexity was associated with early differences in CBR that diminished by approximately 20 months, suggesting that while CB remains a broadly shared milestone, its early trajectory may be influenced by the phonotactic properties of the input language.

Our systematic review revealed substantial variability in how CB has been measured and reported, echoing concerns long raised by others (Oller and Eilers 1988; Lewedag 1995), that without explicit operational definitions, cross-study comparisons remain constrained. Specifically, the 42 included studies varied in study design types, locations of data collection, sampling contexts for eliciting vocalization, coding methods, CBR calculations, and language environments studied. Over half of the infants were exposed to English, and a slightly greater proportion of males (52%) than females (48%) were represented in the data. A large number of studies ($N = 12$, 28%) did not report infant gender data. Five different CBR measures were identified, alongside two non-CBR measures of vocal development (i.e., SAEVD-R and MBL), with CBR^{syl} being the most common. Potentially, the popularity of CBR^{syl} reflects its ability to capture finer-grained analysis of within-utterance syllable structures, despite the time-consuming nature and potential ambiguity of syllable-level coding, an issue highlighted in prior work that has advocated for an utterance-level measure (CBR^{UTTER}) as a more reliable and efficient alternative (Nyman and Lohmander 2018). These patterns largely align with prior reviews highlighting methodological inconsistencies across studies of babbling in both typically developing and clinical populations (Morgan and Wren 2018; Lang et al. 2019; Yankowitz et al. 2022; Long et al. 2023). Our review adds to this work by quantifying these inconsistencies across a larger set of studies and specifying how they map onto differences in age, setting, and ambient language. We concluded our analyses by reporting the results of a simulation to estimate minimum sample size requirements to reliably estimate infant CBRs at different ages, for different levels of precision. Our simulation showed that the required sample size varied substantially by infant age—as few as $N = 51$ infants could be required to estimate CBR with a 0.3 simulated margin of error at 6 months, but this would increase to $N = 110$ at 8 months and $N = 138$ at 11 months. These sample size estimates should thus assist other researchers, measuring CBRs at different infant ages, establish sample size estimates going forward.

Studies were most often conducted in the home, using interactive free play (where adults "act as they normally do" with infants), and involved caregiver–infant interactions, although the exact context varied across studies. We also note that nearly all studies published prior to 2021 relied on trained human coders, whereas two of the more recent studies applied non-expert citizen scientist coding methods (Cychosz et al. 2021; Hitczenko et al. under review). Increasing advancements in technology, such as employing child-centered daylong audio recorders in home environments (e.g., LENA technology) and the use of citizen science platforms (e.g., MTurk and Zooniverse) to facilitate large-scale data annotation, provide promising, less time- and cost-intensive alternatives to human coding by trained assistants. However, these approaches also raise questions about the reliability and comparability of CBR analysis across studies. For example, more work is needed to evaluate how multiple nonexpert annotators on citizen science platforms compare to a single in-lab expert annotator, particularly when used for developmental screening purposes (Semenzin et al. 2021). As the field continues to pursue scalable and ecologically valid

TABLE 8 | Recommended minimum sample sizes for CBR measurement.

Age (months)	Expected CBR	Simulated margin of error	Relative margin (%)	Min. recommended sample size
5	0.12	0.03	25.0	61
5	0.12	0.05	41.7	22
5	0.12	0.07	58.3	11
6	0.15	0.03	20.0	51
6	0.15	0.05	33.3	19
6	0.15	0.07	46.7	10
7	0.18	0.03	16.7	62
7	0.18	0.05	27.8	22
7	0.18	0.07	38.9	11
8	0.20	0.03	15.0	110
8	0.20	0.05	25.0	39
8	0.20	0.07	35.0	20
9	0.23	0.03	13.0	72
9	0.23	0.05	21.7	26
9	0.23	0.07	30.4	13
10	0.26	0.03	11.5	84
10	0.26	0.05	19.2	30
10	0.26	0.07	26.9	15
11	0.29	0.03	10.3	138
11	0.29	0.05	17.2	50
11	0.29	0.07	24.1	25
12	0.31	0.03	9.7	97
12	0.31	0.05	16.1	34
12	0.31	0.07	22.6	18
15	0.40	0.03	7.5	109
15	0.40	0.05	12.5	40
15	0.40	0.07	17.5	20
18	0.48	0.03	6.2	123
18	0.48	0.05	10.4	45
18	0.48	0.07	14.6	23
21	0.56	0.03	5.4	61
21	0.56	0.05	8.9	22
21	0.56	0.07	12.5	12
24	0.65	0.03	4.6	171
24	0.65	0.05	7.7	61
24	0.65	0.07	10.8	31

methods, trade-offs between efficiency and interpretability will need to be more explicitly considered. These patterns also highlight the multidisciplinary nature of research on early vocal development, where developmental, clinical, and computational approaches have historically relied on different measurement conventions and elicitation contexts. As prior reviews have emphasized, a lack of consensus around measurement can limit the translational utility of these findings for early identification (Morgan and Wren 2018; Lang et al. 2019; Long et al. 2023).

7.1 | Canonical Babbling Ratios Increase With Age

Our meta-analysis confirmed a robust, linear increase in CBR for typically developing infants between 5 and 24 months, reinforcing its value as a stable and generalizable developmental marker. This trajectory was consistent across studies despite methodological differences in calculation, supporting the early assertion that CBR provides a reliable quantitative index of CB development (Oller and Eilers 1988; Oller 2000). Our inclusion of ages beyond the prelinguistic period through 24 months also suggest that, on average, infants continue to increase their production of canonical vocalizations throughout the early lexical period (Hitczenko et al. 2023). Our model-derived benchmarks—CBR of 0.15 by 6 months, 0.23 by 9 months, and 0.31 by 12 months—align with prior work identifying 0.15–0.20 as an approximate onset threshold (Lewedag 1995; Oller and Eilers 1988; Nyman et al. 2021; Yankowitz et al. 2022) and confirm that most infants surpass this range between 6 and 8 months. This trajectory can inform ongoing work aiming to quantify estimates to support the early identification of speech and communication disorders (Yankowitz et al. 2022; Lohmander et al. 2021; Belardi et al. 2017; Long et al. 2024; Lynch et al. 1995). Importantly, we note that these benchmarks were derived from all CBR metrics given our finding of no significant differences across metric type; however, additional work is necessary to confirm this. Also, our analysis of publication bias revealed that studies with smaller sample sizes tended to report larger CBR effect sizes. This suggests that small-study effects may inflate perceived developmental progress in the literature and highlights the need for more rigorous sample size planning across clinical and nonclinical studies. This meta-analysis thus allowed a synthesis of decades of work documenting age-related increases in canonical syllable production, while also highlighting that small-sample studies in the literature tend to report larger effects, ultimately underscoring the need for adequately powered designs in future research.

7.1.1 | Consistent CBR Trajectories Across Measures

Despite substantial variability in CBR calculations, we found no evidence that developmental slopes differed by CBR metric type across matched (e.g., syllables/syllables, utterances/utterances) or mismatched (e.g., syllables/utterances) formulas. This parallels historical validation studies showing strong correlations among alternative measures (Nyman and Lohmander 2018; Lang et al. 2021), and supports the robustness of CBR as a developmental indicator. However, absolute values do shift depending on CBR computation choices. Specifically, the mismatched formula (syllables/utterances), often used in early CB research, may yield inflated CBR values because a single canonical syllable can increase the numerator even if most other syllables in an utterance are noncanonical (Oller et al. 1994; Molemans et al. 2012). This method may obscure variability in the proportion of noncanonical syllables within utterances, particularly in infants at earlier developmental stages who produce a high number of precanonical syllables that may be inconsistently identified. Conversely, matched formulas offer finer granularity but are more labor-intensive to code, making them more susceptible to coder ambiguity when labeling different syllable types. Despite differences in absolute CBR values, our meta-analysis found consistent developmental trajectories across measurement types, suggesting that both approaches capture similar growth patterns over time. Still, these shifts ultimately matter when establishing onset thresholds, comparing across studies, or using calculations for clinical purposes, re-affirming the long-standing recommendation to report the precise numerator and denominator definitions used (Oller 2000).

7.2 | Sampling Method, But Not Location, Affects Canonical Babbling Elicitation

Contrary to some earlier expectations that laboratory settings would inflate vocalization rates (Lewedag et al. 1994; Lee et al. 2018), we found no consistent difference in CBRs between laboratory and home locations. Instead, the structure of the sampling method (i.e., interactive free play vs. daylong naturalistic recordings) was the stronger determinant. Specifically, interactive free play elicited higher CBRs, a pattern consistent with prior reports that socially engaging contexts stimulate canonical syllable production (Lewedag 1995; Long et al. 2022; Tamis-LeMonda et al. 2017). Such differences may reflect infant *capacity* rather than a typical frequency, while naturalistic recordings likely provide more ecologically valid estimates at the cost of lower observed CBRs. This ultimately reinforces the caution that both sampling method and elicitation contexts should be matched to the intended construct of interest.

Also worth noting is that the smaller number of studies using semi-structured play protocols ($N = 5$) and retrospective home recordings ($N = 4$) were not included in our statistical comparisons due to their small samples. While excluded from quantitative analysis, these methods introduce distinct sources of variability that warrant further consideration. In particular, retrospective home recordings that allow caregivers to choose which interactions to record and submit may introduce selection bias. Parents may be more likely to record when their child is highly engaged, vocal, or alert, potentially inflating CBR estimates compared to more continuous or randomly sampled contexts. Semi-structured play protocols may vary in the degree of standardization and adult involvement across studies, further complicating cross-study comparisons. These nuances underscore the need for future work to more explicitly document how methodological sampling choices shape the interpretation of CB trajectories.

7.3 | Reduced Canonical Babbling Levels When Exposed to More Complex Ambient Language Structures

Our analysis of ambient language effects revealed that infants exposed to languages with more complex syllable structures initially produced lower CBRs, converging with their peers by approximately 20 months. Although CB is known to be a universal, biologically driven milestone (Oller et al. 2016), others have argued that CBR milestones were more driven by an infant's motor maturation and thus typical CBR development was not expected to differ cross-linguistically (Peute and Casillas 2022). Our results suggest that its early trajectory may be slightly modulated by phonotactic properties of the input language, an idea consistent with early cross-linguistic work (de Boysson-Bardies et al. 1984). Specifically, our results revealed a reliable effect of syllable complexity on CBR: Infants exposed to languages with more complex syllable structures initially exhibit lower CBRs. These results suggest that while CB is a broadly occurring and biologically grounded milestone, its trajectory may not be entirely universal. Instead, it appears modifiable by the structural properties of the infant' s linguistic environment, particularly in the early stages of development. Although our formal analysis was limited to distinguishing only between languages with "more complex" (e.g., English, Dutch) versus "less complex" (e.g., Spanish, Mandarin) structures, we concluded that these ambient effects appeared to disappear by 20 months. This supports the view that ambient language complexity influences the timing, but not necessarily the ultimate attainment, of CB. Much more work is needed in this area to build upon this early result, however. Only 16 unique language environments were represented in our literature search, for example, and this is woefully inadequate to start answering questions about how ambient language structure drives early milestones in speech development. Current advances in this area will continue to be driven by large-scale collection and annotation of infant speech samples from language environments under-represented in language acquisition research (Hitczenko et al. under review; Tey et al. under review), and this is a critical area of study to continue examining how early language exposure shapes speech development across diverse linguistic environments. Yet, these findings advance a more nuanced account of babbling as a developmental phenomenon shaped by both motoric maturation and exogenous linguistic input.

7.3.1 | No Reliable Effect of Multilingualism on Canonical Babbling Patterns

As part of our study of ambient language effects, we found no consistent differences in CBR trajectories between monolingual and multilingual infants. Again, we encourage caution in the interpretation of this null effect: First, given that we found effects of ambient language complexity upon CBR trajectories, we find it likely that effects of multilingualism may interact in complex ways with the number and structure of the specific languages being examined. For example, one could still predict different CBR trajectories between an infant exposed to two languages with relatively less syllabic structure complexity and an infant exposed to three languages with varying syllabic complexity. This review was not able to formally evaluate differences like these, as the literature does not yet have sufficient reports of CBR trajectories from infants with this fine-grained variability in ambient language exposure. We also encourage caution in the interpretation of the null result here because we had relatively few effect sizes to evaluate—just six studies reported data from some infants from multilingual backgrounds and they were almost always Spanish-English. Thus, we see a ripe area for future research: Millions of infants are exposed to and subsequently acquire spoken competence in two or more languages, yet these infants are extremely under-represented in language acquisition research, especially in the CBR literature. We thus especially encourage future research in this area.

8 | Limitations

Our review and meta-analysis had several key limitations that should be considered when interpreting the results. First, our ambient language analysis was limited by treating language using binary variables (English vs. non-English, more versus less complex syllable structures), which prevented us from examining specific language effects or identifying particular phonotactic features beyond syllabic complexity that might drive the observed CBR differences. This coarse phonological classification likely obscures more nuanced phonological features that may impact CB across languages. Future research should incorporate larger, more balanced samples across languages and developmental stages to examine more specify features of complexity to offer a more precise understanding of how specific phonological characteristics shape the emergence of canonical syllables. For example, moving toward treating language as a continuous spectrum of phonotactic complexity would allow for more nuanced investigation of cross-linguistic differences.

Second, our review was restricted to English-language publications, potentially limiting the diversity of language environments that were incorporated into our review. Including non-English literature could have enhanced our ability to make more nuanced comparisons of CBR patterns across language environments. The exclusion of non-English papers likely biased our sample toward studies conducted in English-speaking countries.

Third, the variability in sampling methodologies constrained our analysis. While we confirmed that naturalistic all-day recordings produce lower rates of CB compared to interactive free play sessions, we could not thoroughly compare all variability in location, sampling method, and elicitation contexts. This raises important questions about ecological validity in CB research—interactive free play may produce artificially elevated rates that do not reflect everyday vocal production levels, though they might better capture infants' maximum capabilities. Future studies should explore optimal methodologies for eliciting CB in ways that balance naturalistic settings with sufficient vocal production.

Additionally, we acknowledge that our syllable complexity categorization may inadvertently capture broader socio-cultural differences, as the complex syllable structure group predominantly comprises Western European languages while the simple structure group includes more diverse, non-Western language communities. This fact makes it difficult to disentangle purely linguistic from potential cultural influences on early vocal development.

Finally, we encountered considerable inconsistency in how canonical consonants were defined across studies. Some explicitly included or excluded glides and liquids, while others calculated a TCBR by excluding these sounds altogether. However, we do not know the extent to which some studies did or did not code for "true" canonical syllables if the authors were not fully transparent in their coding scheme. In addition, no study in our review differentiated between canonical syllables occurring in identifiable words versus nonword contexts, which is a particularly difficult distinction to make in early development when much speech production is ambiguous in its intent. This lack of specification means that CBRs likely included syllables within words across all ages studied, including older ages (up to 24 months) when canonical syllables can occur in real-word productions. Because authors were not always transparent about these coding details, we could not reliably classify studies beyond grouping metrics into two broad categories (matching vs. mismatched numerators/denominators). Future research should adopt standardized definitions, explicitly report coding criteria for consonant inclusion and word status, and compare the relative utility of different CBR measures under these conditions.

9 | Conclusion

We report a systematic review and meta-analysis of CB measured using the CBR across 42 studies and 1277 unique infants. Results confirm a robust, linear developmental trajectory from 5 to 24 months, with CBR increasing from 0.12 to 0.65 during this period. Two additional key factors significantly influence CBR measurements: syllable complexity of the ambient language and sampling method. Infants learning more complex syllable languages (e.g., English, Dutch) initially show lower CBRs but accelerated development, converging with peers acquiring languages with less syllable structure complexity (e.g., Mandarin) by 20 months. Semi-structured interactive, free play contexts consistently elicit higher CBRs than naturalistic home recordings—a difference that increases with age. We uncovered some publication bias, where studies with smaller samples may be over-represented in the literature, so we provide simulation-based sample size recommendations as practical guidance for future research. These findings establish valuable developmental benchmarks while highlighting the need for methodological consistency when measuring and interpreting this universal milestone of speech development in infancy.

Acknowledgements

Thank you to Kriti Beswal, Noel Guzman, and Kat Castro for assistance with screening and data extraction. Additional thanks to Hyunjoo Yoo and Megan Burkhardt-Reed for assistance in study interpretation. The authors also gratefully acknowledge the following individuals who shared their data and/or assisted in its interpretation: Kasia Hitczenko, Anette Lohmander, Steven Gillis, Lisa Hamrick, and Rachel Albert.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in OSF at https://osf.io/tks5h/overview.

Endnotes

[1] It was not possible to determine the unique number of infants at each time point because longitudinal studies often only reported the overall sample size and not the specific number of infants included at each time point (some time points had missing data or infants had been loss to attrition).

[2] Because our publication bias analysis focused on the developmental trajectory of CBR rather than a traditional binary comparison seen elsewhere in meta-analytic research (e.g., treatment vs. control, high vs. low socioeconomic status), some study*age combinations resulted in $N = 1$ sample sizes, even within studies that included many infants. To evaluate the impact of small sample bias, we conducted a sensitivity analysis, rerunning Egger's test after excluding study*age samples of <3 and <5 infants. The presence of publication bias remained significant, confirming that smaller sample sizes were associated with larger-than-expected effect sizes.

[3] Effect sizes for the age models were derived from the combination of infant age and study for a total of $N = 190$ effect sizes. The only exception was for Yankowitz et al. (2022), where there were two reports of 6-month-olds and two reports of 12-month-olds, one for high-risk negative controls and one for low-risk controls.

[4] For 5-month-olds, we used the SD from 6- and 7-month-olds (SD = 0.12) due to limited available data at this age ($N = 6$).

References

Albert, R. R., M. Ernst, and C. D. Vallotton. 2022. "Infant Vocalizations Elicit Simplified Speech in Childcare." *Infancy* 28, no. 2: 322–338.

Andruski, J. E., E. Casielles, and G. Nathan. 2013. "Is Bilingual Babbling Language-specific? Some Evidence From a Case Study of Spanish–English Dual Acquisition." *Bilingualism: Language and cognition* 17, no. 3: 660–672. https://doi.org/10.1017/S1366728913000655.

Andruski, J. E., E. Casielles, and G. Nathan. 2014. "Is Bilingual Babbling Language-specific? Some Evidence From a Case Study of Spanish–English Dual Acquisition." *Bilingualism: Language and Cognition* 17, no. 3: 660–672. https://doi.org/10.1017/S1366728913000655.

Athari, P., R. Dey, and S. Rvachew. 2021. "Vocal Imitation Between Mothers and Infants." *Infant Behavior and Development* 63: 101531. https://doi.org/10.1016/j.infbeh.2021.101531.

Bartl-Pokorny, K. D., F. B. Pokorny, D. Garrido, B. W. Schuller, D. Zhang, and P. B. Marschik. 2022. "Vocalisation Repertoire at the End of the First Year of Life: an Exploratory Comparison of Rett Syndrome and Typical Development." *Journal of Developmental and Physical Disabilities* 34, no. 6: 1053–1069.

Belardi, K., L. R. Watson, R. A. Faldowski, et al. 2017. "A Retrospective Video Analysis of Canonical Babbling and Volubility in Infants With Fragile X Syndrome at 9–12 Months of Age." *Journal of Autism and Developmental Disorders* 47, no. 4: 1193–1206.

Chan, C. J., L. M. Chen, and L. W. Chen. 2022, November. "Early Speech Production in Infants and Toddlers Later Diagnosed With Cerebral Palsy: a Retrospective Study." In Proceedings of the 34th Conference on Computational Linguistics and Speech Processing (ROCLING 2022) (214–220).

Chapman, K. L., M. Hardin-Jones, and K. A. Halter. 2003. "The Relationship Between Early Speech and Later Speech and Language Performance for Children With Cleft Lip and Palate." *Clinical Linguistics & Phonetics* 17, no. 3: 173–197. https://doi.org/10.1080/0269920021000047864.

Chapman, K. L., M. Hardin-Jones, J. Schulte, and K. A. Halter. 2001. "Vocal Development of 9-Month-Old Babies with Cleft Palate." *Journal of Speech, Language, and Hearing Research* 44, no. 6: 1268–1283. https://doi.org/10.1044/1092-4388(2001/099).

Chen, L. M., and R. D. Kent. 2010. "Segmental Production in Mandarin-learning Infants." *Journal of child language* 37, no. 2: 341–371.

Chen, L. M., C. C. Lee, and T. W. Kuo. 2011, December. "Measures of Early Phonetic Development: a Longitudinal Analysis." In 2011 IEEE International Symposium on Multimedia (524–529). IEEE.

Cychosz, M., A. Cristia, E. Bergelson, et al. 2021. "Vocal Development in a Large-scale Crosslinguistic Corpus." *Developmental Science* 24, no. 5: e13090. https://doi.org/10.1111/desc.13090.

de Boysson-Bardies, B., L. Sagart, and C. Durand. 1984. "Discernible Differences in the Babbling of Infants According to Target Language." *Journal of Child Language* 11, no. 1: 1–15.

DeVeney, S., and A. Kyvelidou. 2020. 11 Early Diagnostic Signs of Autism: Preliminary Findings for Infant Vocalizations. On Under-Reported Monolingual Child Phonology 19.

D'Odorico, L., M. Majorano, M. Fasolo, N. Salerni, and C. Suttora. 2011. "Characteristics of Phonological Development as a Risk Factor for Language Development in Italian-speaking Pre-term Children: a Longitudinal Study." *Clinical Linguistics & Phonetics* 25, no. 1: 53–65.

Eilers, R. E., and D. K. Oller. 1994. "Infant Vocalizations and the Early Diagnosis of Severe Hearing Impairment." *Journal of Pediatrics* 124, no. 2: 199–203.

Ertmer, D. J., and J. Jung. 2012. "Prelinguistic Vocal Development in Young Cochlear Implant Recipients and Typically Developing Infants: Year 1 of Robust Hearing Experience." *Journal of Deaf Studies and Deaf Education* 17, no. 1: 116–132.

Ertmer, D. J., J. Jung, and D. T. Kloiber. 2013. "Beginning to Talk Like an Adult: Increases in Speech-Like Utterances in Young Cochlear Implant Recipients and Typically Developing Children." *American Journal of Speech-Language Pathology* 22, no. 4: 591–603.

Fagan, M. K. 2009. "Mean Length of Utterance Before Words and Grammar: Longitudinal Trends and Developmental Implications of Infant Vocalizations." *Journal of Child Language* 36, no. 3: 495–527.

Gipson, T. T., G. Ramsay, E. E. Ellison, E. R. Bene, H. L. Long, and D. K. Oller. 2021. "Early Vocal Development in Tuberous Sclerosis Complex." *Pediatric Neurology* 125: 48–52. https://doi.org/10.1016/j.pediatrneurol.2021.08.009.

Goldstein, M. H., and J. A. Schwade. 2008. "Social Feedback to Infants' Babbling Facilitates Rapid Phonological Learning." *Psychological Science* 19, no. 5: 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x.

Hitczenko, K., E. Bergelson, M. Casillas, H. Colleran, M. Cychosz, and A. Cristia. 2023. "The Development of Canonical Proportion Continues Past Toddlerhood." In *Proceedings of the International Congress of the Phonetic Sciences*.

Hitczenko, K., E. Bergelson, M. Casillas, et al. Under review. "A Cross-Linguistic Study of the Effect of Early Experience on Vocal Development.".

Hitczenko, K., L. Peurey, W. Havard, et al. Under review. "Speech Maturity Dataset: A Cross-Cultural Corpus of Naturalistic Child and Adult Vocalizations.".

Hong, Q. N., S.à bregues, G. Bartlett, et al. 2018. "The Mixed Methods Appraisal Tool (MMAT) Version 2018 for Information Professionals and Researchers." *Education for Information* 34, no. 4: 285–291.

Koopmans-van Beinum, F. J., and J. M. van der Stelt. 1986. "Early Stages in the Development of Speech Movements." In *Precursors of Early Speech: Proceedings of an International Symposium Held at the Wenner-Gren Center*, 37–50. Palgrave Macmillan UK.

Lang, S., K. D. Bartl-Pokorny, F. B. Pokorny, et al. 2019. "Canonical Babbling: A Marker for Earlier Identification of Late Detected Developmental Disorders?" *Current Developmental Disorders Reports* 6, no. 3: 111–118.

Lang, S., K. Willmes, P. B. Marschik, D. Zhang, and A. Fox-Boyer. 2021. "Prelexical Phonetic and Early Lexical Development in German-Acquiring Infants: Canonical Babbling and First Spoken Words." *Clinical Linguistics & Phonetics* 35, no. 2: 185–200.

Lee, C.-C., Y. Jhang, L.-m. Chen, G. Relyea, and D. K. Oller. 2017. "Subtlety of Ambient-Language Effects in Babbling: A Study of English- and Chinese-Learning Infants at 8, 10, and 12 Months." *Language Learning and Development* 13, no. 1: 100–126.

Lee, C.-C., Y. Jhang, G. Relyea, L.-m. Chen, and D. K. Oller. 2018. "Babbling Development as Seen in Canonical Babbling Ratios: A Naturalistic Evaluation of All-Day Recordings." *Infant Behavior and Development* 50: 140–153.

Lewedag, V. L. 1995. "Patterns of Onset of Canonical Babbling Among Typically Developing Infants." Ph.D. thesis, University of Miami.

Lewedag, V. L., D. K. Oller, and M. P. Lynch. 1994. "Infants' Vocalization Patterns Across Home and Laboratory Environments." *First Language* 14, no. 42-43: 49–65.

Lohmander, A., L. R. Westberg, S. Olsson, B. I. Tengroth, and T. Flynn. 2021. "Canonical Babbling and Early Consonant Development Related to Hearing in Children With Otitis Media With Effusion With or Without Cleft Palate." *Cleft Palate Craniofacial Journal* 58, no. 7: 894–905.

Long, H. L., D. Bowman, H. Yoo, M. Burkhardt-Reed, E. Bene, and D. K. Oller. 2020. "Social and Endogenous Infant Vocalizations." *PLoS ONE* 15, no. 8: e0224956.

Long, H. L., L. Christensen, S. Hayes, and K. C. Hustad. 2023. "Vocal Characteristics of Infants at Risk for Speech Motor Involvement: A Scoping Review." *Journal of Speech, Language, and Hearing Research: JSLHR* 66, no. 11: 4432–4460.

Long, H. L., and K. C. Hustad. 2023. "Marginal and Canonical Babbling in 10 Infants at Risk for Cerebral Palsy." *American Journal of Speech-Language Pathology* 32, no. S4: 1835–1849.

Long, H. L., G. Ramsay, E. R. Bene, et al. 2024. "Canonical Babbling Trajectories Across the First Year of Life in Autism and Typical Development." *Autism* 28: 3078–3091.

Long, H. L., G. Ramsay, U. Griebel, et al. 2022. "Perspectives on the Origin of Language: Infants Vocalize Most During Independent Vocal Play but Produce Their Most Speech-Like Vocalizations During Turn Taking." *PLOS ONE* 17, no. 12: e0279395.

Lynch, M. P., D. K. Oller, M. L. Steffens, S. L. Levine, D. L. Basinger, and V. Umbel. 1995. "Onset of Speech-Like Vocalizations in Infants With Down Syndrome." *American Journal of Mental Retardation* 100, no. 1: 68–86.

Molemans, I., R. Van Den Berg, L. Van Severen, and S. Gillis. 2012. "How to Measure the Onset of Babbling Reliably?" *Journal of Child Language* 39, no. 3: 523–552.

Morgan, L., and Y. E. Wren. 2018. "A Systematic Review of the Literature on Early Vocalizations and Babbling Patterns in Young Children." *Communication Disorders Quarterly* 40, no. 1: 3–14.

Nathani, S., D. J. Ertmer, and R. E. Stark. 2006. "Assessing Vocal Development in Infants and Toddlers." *Clinical Linguistics & Phonetics* 20, no. 5: 351–369.

Nyman, A., and A. Lohmander. 2018. "Babbling in Children With Neurodevelopmental Disability and Validity of a Simplified Way of Measuring Canonical Babbling Ratio." *Clinical Linguistics & Phonetics* 32, no. 2: 114–127.

Nyman, A., S. Strömbergsson, and A. Lohmander. 2021. "Canonical Babbling Ratio – Concurrent and Predictive Evaluation of the 0.15 Criterion." *Journal of Communication Disorders* 94: 106164.

Oller, D. K. 1978. "Infant Vocalization and the Development of Speech." *Allied Health and Behavioral Sciences* 1: 523–549.

Oller, D. K. 1980. "The Emergence of the Sounds of Speech in Infancy." In *Child Phonology, Volume 1: Production*, edited by G. Yeni-Komshian, J. Kavanagh, and C. Ferguson, 93–112. Academic Press.

Oller, D. K. 2000. *The Emergence of the Speech Capacity*. Lawrence Erlbaum Associates.

Oller, D. K., E. H. Buder, H. L. Ramsdell, A. S. Warlaumont, L. Chorna, and R. Bakeman. 2013. "Functional Flexibility of Infant Vocalization and the Emergence of Language." *Proceedings of the National Academy of Sciences* 110: 6318–6632.

Oller, D. K., R. Eilers, A. Neal-Beevers, and A. Cobo-Lewis. 1998. "Late Onset Canonical Babbling: A Possible Early Marker of Abnormal Development." *American Journal of Mental Retardation: AJMR* 103: 249–263.

Oller, D. K., and R. E. Eilers. 1988. "The Role of Audition in Infant Babbling." *Child Development* 59, no. 2: 441–449.

Oller, D. K., R. E. Eilers, D. H. Bull, and A. E. Carney. 1985. "Pre-Speech Vocalizations of a Deaf Infant: A Comparison With Normal

Metaphonological Development." *Journal of Speech and Hearing Research* 28: 47–63.

Oller, D. K., R. E. Eilers, A. R. Neal, and H. K. Schwartz. 1999. "Precursors to Speech in Infancy: The Prediction of Speech and Language Disorders." *Journal of Communication Disorders* 32: 223–245.

Oller, D. K., R. E. Eilers, M. L. Steffens, M. P. Lynch, and R. Urbano. 1994. "Speech-Like Vocalizations in Infancy: An Evaluation of Potential Risk Factors." *Journal of Child Language* 21, no. 1: 33–58.

Oller, D. K., U. Griebel, and A. S. Warlaumont. 2016. "Vocal Development as a Guide to Modeling the Evolution of Language." *Topics in Cognitive Science* 8, no. 2: 382–392.

Oller, D. K., and J. M. Seibert. 1988. "Babbling in Prelinguistic Mentally Retarded Children." *American Journal on Mental Retardation* 92: 369–375.

Peute, B., and M. Casillas. 2022. "(Non-)Effects of Linguistic Environment on Early Stable Consonant Production: A Cross Cultural Case Study." *Glossa: A Journal of General Linguistics* 7, no. 1: 1–28.

RStudioTeam. 2024. *RStudio: Integrated Development for R*. RStudio Inc.

Semenzin, C., L. Hamrick, A. Seidl, B. L. Kelleher, and A. Cristia. 2021. "Describing Vocalizations in Young Children: A Big Data Approach Through Citizen Science Annotation." *Journal of Speech, Language, and Hearing Research* 64, no. 7: 2401–2416.

Stark, R. E. 1980. "Stages of Speech Development in the First Year of Life." In *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson, 73–92. Academic Press.

Stoel-Gammon, C. 1989. "Prespeech and Early Speech Development of Two Late Talkers." *First Language* 9, no. 6: 207–223.

Tamis-LeMonda, C. S., Y. Kuchirko, R. Luo, K. Escobar, and M. H. Bornstein. 2017. "Power in Methods: Language to Infants in Structured and Naturalistic Contexts." *Developmental Science* 20, no. 6: e12456.

Tey, K., S. Walker, A. Seidl, et al. Under review. "The Development of Canonical Proportion as a Function of Community, Multilingualism, and Target Language's Syllable Complexity.".

Veritas Health Innovation. 2025. "Covidence Systematic Review Software." https://www.covidence.org/.

Viechtbauer, W. 2010. "Conducting Meta-Analyses in *R* With the **metafor** Package." *Journal of Statistical Software* 36, no. 3: 1–48.

Vihman, M. M. 2014. *Phonological Development: The First Two Years*. 2nd ed. Wiley-Blackwell.

Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Yankowitz, L. D., V. Petrulla, S. Plate, et al. 2022. "Infants Later Diagnosed With Autism Have Lower Canonical Babbling Ratios in the First Year of Life." *Molecular Autism* 13, no. 1: 28.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1: desc70139-sup-0001-SuppMat.pdf.

Appendix A: Search Strategy

PubMed search terms (9/5/24): 265 items "canonical babble" OR "canonical babbling" OR "canonical vocalization" OR "speech-like vocalization" OR "babble onset" OR "canonical syllable")

Web of Science search terms (9/5/24): 141 items ALL = "canonical babble" OR "canonical babbling" OR "canonical vocalization" OR "speech-like vocalization" OR "babble onset" OR "canonical syllable")

Scopus search terms (9/5/24): 907 items All"canonical babble" OR "canonical babbling" OR "canonical vocalization" OR "speech-like vocalization" OR "babble onset" OR "canonical syllable")

Appendix B: Quality Appraisal Questions

1. Is the research aim or hypothesis clearly stated and relevant to the field of study?
2. Are the characteristics of the study sample or participants described in sufficient detail?
3. Are the inclusion and exclusion criteria for participants clearly stated?
4. Is the methodology clearly described and appropriate for addressing the research question?
5. Are the measurement tools or data collection methods valid, reliable, and appropriate for the study design?
6. Were appropriate statistical or qualitative analyses conducted, and were they suitable for the type of data collected?
7. Does the study report on the reliability and validity of the data or findings?
8. Are the study' s results clearly presented and supported by the data?
9. Are limitations of the study acknowledged and discussed, including potential sources of bias or confounding factors?
10. Are funding sources and potential conflicts of interest disclosed?

Final quality appraisal: Good, Fair, or Poor

Appendix C: Raw Canonical Babbling Ratio Measures

TABLE C.1 | Observed canonical babbling ratio by age and language group, mean (SD).

	All languages		English	
Age (months)	**All languages (*N*)**	**All languages CBR**	**English (*N*)**	**English CBR**
5	6	0.07 (0.03)	5	0.04 (0.01)
6	436	0.09 (0.11)	351	0.04 (0.04)
7	77	0.18 (0.12)	14	0.19 (0.16)
8	217	0.24 (0.16)	155	0.21 (0.19)
9	258	0.26 (0.13)	157	0.22 (0.1)
10	269	0.29 (0.14)	175	0.23 (0.18)
11	87	0.36 (0.18)	25	0.3 (0.18)
12	721	0.33 (0.15)	532	0.28 (0.15)
13	106	0.39 (0.14)	38	0.39 (0.18)
14	134	0.4 (0.1)	71	0.38 (0.1)
15	118	0.39 (0.16)	38	0.29 (0.15)
16	109	0.44 (0.15)	39	0.41 (0.19)
17	67	0.55 (0.22)	15	0.34 (0.19)
18	113	0.51 (0.17)	43	0.53 (0.14)
19	34	0.48 (0.21)	1	0.6 (NA)
20	39	0.49 (0.24)	5	0.59 (0.16)
21	42	0.57 (0.12)	9	0.67 (NA)
24	44	0.68 (0.2)	9	0.8 (NA)

Note: Values represent means and standard deviations of observed canonical babbling ratios across studies. N = total sample size across all studies for each age group. NA for SD indicates that the CBR estimate was derived from only one study for that age group.