LANGUAGE DEVELOPMENT RESEARCH

An Open Science Journal

Volume 5 | Issue 3 | December 2025
ISSN 2771-7976

About the Journal

Language Development Research: An Open-Science Journal was established in 2020 to meet the field's need for a peer-reviewed journal that is committed to fully open science: LDR charges no fees for readers or authors, and mandates full sharing of materials, data and analysis code. The intended audience is all researchers and professionals with an interest in language development and related fields: first language acquisition; typical and atypical language development; the development of spoken, signed or written languages; second language learning; bi- and multilingualism; artificial language learning; adult psycholinguistics; computational modeling; communication in nonhuman animals etc. The journal is managed by its editorial board and is not owned or published by any public or private company, registered charity or nonprofit organization.

TalkBank System

Language Development Research is the official journal of the **TalkBank system**, comprising the CHILDES, PhonBank, HomeBank, FluencyBank, Multilingualism and Clinical Banks, the CLAN software (used by hundreds of researchers worldwide to analyze children's spontaneous speech data), and the Info-CHILDES mailing list, the de-facto mailing list for the field of child language development with over 1,600 subscribers.

Diamond Open Access

Language Development Research is published using the Diamond Open Access model (also known as “Platinum” or “Universal” OA). The journal does not charge users for access (e.g., subscription or download fees) or authors for publication (e.g., article processing charges).

Hosting

The **Carnegie Mellon University Library Publishing Service** (LPS) hosts the journal on a Janeway Publishing Platform with its manuscript management system (MMS) used for author submissions.

License

Peer Review and Submissions

All submissions are reviewed by a minimum of two peer reviewers, and one of our Action Editors, all well- established senior researchers, chosen to represent a wide range of theoretical and methodological expertise. Action Editors select peer reviewers based on their expertise and experience in publishing papers in the relevant topic area.

Submissions and Publication Cycle

We invite submissions that meet our criteria for rigour, without regard to the perceived novelty or importance of the findings. We publish general and special-topic articles (“Special Collections”) on a rolling basis to ensure rapid, cost-free publication for authors.

Language Development Research is published once a year, in December, with each issue containing the articles produced over the previous 12 months. Individual articles are published online as soon as they are produced. For citation purposes, articles are identified by the year of first publication and digital object identifier (DOI).

Table of Contents

Volume 5, Issue 3, December 2025

Phonological representations at the onset of reading acquisition: steady use of phonological detail from preschool to 2nd grade

Anne Bauch
Claudia K. Friedrich
Ulrike Schild
University of Tuebingen, Germany

Abstract: We tracked the developmental path of aspects of spoken word recognition in the beginning years of reading acquisition in German L1 speaking children. Speech processing of phonological feature variation in voicing was tested in preschool, 1st and 2nd grade. During the word onset priming test, spoken words (targets; “Kino”, Engl. cinema) followed spoken syllables (primes) that were either identical to target word onsets (“Ki“), deviated in the onset speech sound in voicing (“Gi”) or were unrelated (“Ba”). Event-related potentials (ERP) and lexical decision latencies were recorded. Results showed a comparable pattern from preschool to 2nd grade. ERP effects emerged around 100 – 300 ms, replicating previous findings for voicing variations. Children’s faster lexical decisions with increasing age were not paralleled in ERP timing differences between age groups. Thus, from a developmental perspective, emerging and increasing reading skills might not relate to increasing sensitivity for phonological feature variation in the tested aspects of spoken word recognition.

Keywords: spoken word processing; lexical access; event-related potentials, literacy acquisition; preschoolers.

Corresponding author(s): Anne Bauch, Department of Psychology, University of Tuebingen, Schleichstrasse 4, 72076 Tuebingen, Germany. Email: anne.bauch@uni-tuebingen.de.

ORCID ID(s): https://orcid.org/0000-0003-4661-2523 (AB); https://orcid.org/0000-0002-2877-349X (CKF); https://orcid.org/0000-0001-9683-8446 (US)

Citation: Bauch, A., Friedrich, C.K., & Schild, U. (2025). Phonological representations at the onset of reading acquisition: steady use of phonological detail from preschool to 2nd grade. *Language Development Research, 5*(3), 1–32. http://doi.org/10.34842/ldr2025

Introduction

Speech unfolds over time and so word beginnings diverge when more identifying information becomes available for that word. Adult listeners habitually predict words before all the information is available in the speech stream. They appear to face this sequential nature of speech by parallel processing of multiple words. They consider several word candidates that match the input at a given point in time. This parallel processing is a basic feature of psycholinguistic models of spoken word recognition (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris & McQueen, 2008; for review see Weber & Scharenborg, 2012). The strength of activation reflects overlap as well as mismatches between the input and stored words (e.g., Allopenna et al., 1998; Soto-Faraco et al., 2001). The mismatch is detected at the level of phonological features (as an example for a phonological feature, namely voicing, consider the different onset of the English words "bin" and "pin"). A single feature mismatch is enough to delay spoken word recognition (e.g., Friedrich et al., 2009), while more feature mismatches add further delay (e.g., Connine et al., 1993; Slowiaczek et al., 1987). In the present study, we aim to track the developmental trajectory of sensitivity to such phonological feature variation in voicing variations in children during the onset of reading acquisition in middle childhood. This developmental period has been associated with the plasticity of phonological processing (e.g., Goswami, 2000).

Initially, enhanced phonological processing in readers was demonstrated by performance in tasks that measure explicit understanding and processing of speech units (*phonological awareness* in general, or *phonemic awareness* for single speech sounds in particular; Anthony & Lonigan, 2004; Goswami, 2000; see also Caravolas & Bruck, 1993; Hulme et al., 2005). For example, reading children outperformed prereaders on phoneme segmentation(Ehri & Wilce, 1980; Treiman & Cassar, 1997; Tunmer & Nesdale, 1985), and literate adults outperformed illiterate adults on phoneme addition and deletion (Morais et al., 1979). Whether reading experience only affects metalinguistic, post-lexical levels (skills associated with phonemic awareness; see Cutler & Davis, 2012; Mitterer & Reinisch, 2015) or also earlier stages of speech processing (like pre-lexical and lexical stages before word access, henceforth automatic stages) itself is disputed. Some authors have argued that reading experience updates aspects of spoken word recognition (e.g., Dehaene et al., 2015; Harm & Seidenberg, 2004; Muneaux & Ziegler, 2004; Pattamadilok et al., 2010; Taft, 2006; Ziegler & Ferrand, 1998). They assume that phonological representations along the speech recognition pathway are restructured by reading and its precursor functions such as phonological and phonemic awareness and grapheme knowledge (Dehaene et al., 2015; Harm & Seidenberg, 2004; Taft, 2006). So called "orthographic consistency effects" seem to back up the assumption that orthographic information is involved in auditory word activation in metalinguistic tasks. For example, listeners tended to identify rhyming word pairs faster, when the pairs shared similar phonology and orthography (e.g., "house – mouse") as when the spelling differed between the words (e.g., "flow –

though"; Pattamadilok et al., 2007, 2014; Perre et al., 201; Ventura et al., 2004; Ziegler & Ferrand, 1998). Furthermore, brain imaging revealed that when listening to speech, adult readers showed higher activity in brain regions which are associated with processing phonological information (compared to illiterate adults; Chang et al., 2010; Dehaene et al., 2010; Mesgarani et al., 2014; see Monzalvo & Dehaene-Lambertz, 2013 for a replication with 6-year-old reading vs. non-reading children). However, other studies have questioned the existence of an intimate relationship between online speech processing and literacy. For instance, by comparing brain activity to speech perception tasks in adults with varying degrees of literacy skills (from illiterate to proficient readers), Hervais-Adelman et al. (2021) did not find evidence that direct brain responses to speech differed between groups with different literacy levels. Instead, the authors suggested that literacy instead might rather promote connectivity between different brain regions that are involved in speech processing, like graphomotor areas and the left posterior superior temporal gyrus, which is associated with the categorical representation of speech sounds.

The analysis of event-related potentials (ERPs) provides detailed insight in temporal processes of word recognition and is suited to investigate word activation processes during an unfolding speech signal. In relation to potential orthographic effects that might occur while listening to spoken language, ERPs have, for example, been used to investigate word-level based auditory orthographic consistency effects through various tasks, such as lexical decision tasks (Perre & Ziegler, 2008), rhyme judgement tasks (Pattamadilok et al., 2011) or non-linguistic Go-NoGo tasks (Perre et al., 2011). Those studies found considerable evidence of an activation of orthographic cues as early as 100 – 300ms after word onset.

Using the ERP analysis, we previously tried to take a closer look at which factors of reading acquisition might potentially modulate word activation at automatic and post-lexical stages of speech processing and sensitivity to phonological feature variation (Bauch et al., 2021). For 10 weeks, pre-literate 6-year-old German native speaking preschoolers participated daily in short games that were either targeting skills in phonemic awareness (solely or in combination with grapheme knowledge, e.g., onset phoneme identification tasks among others) or took part in a control intervention that trained arithmetic skills (for detailed information about the control training, see Schild et al., 2020). In the phonemic trainings, children were especially sensitized for a set of phonemes that differed in the German language only in one phonological feature, namely voicing feature (/g/ and /k/). After the training had taken place, we were interested in whether the specific training in precursor functions of reading might modulate the processing of subtle phonemic mismatch in comparison to the control group. All children participated in a word onset fragment priming paradigm with a lexical decision task. We tested sensitivity to phonemic mismatch by means of ERP amplitude differences and reaction times. Children listened to target words that were preceded by prime syllables that either matched in their initial phoneme (Identity condition, e.g., “Ki – Kino”, Engl. cinema), differed in the voicing feature (Variation condition, e.g., “Gi – Kino”) or were unrelated (Control condition, e.g., “Ba – Kino”). In previous studies using the same paradigm, adults (e.g., Friedrich et al., 2009) as well as reading preschoolers and 2nd grade children but not preliterate preschoolers (Schild et al., 2011), depicted a graded activation pattern of response times and ERPs that depended on the goodness-of-fit between the onset phoneme of syllable prime and target word. In those studies, prime-target pairs varied in place of articulation (e.g., “Non – Monster”) and revealed differentiating ERP amplitudes emerging between 300 – 400ms (referred to as P350 effect) after stimulus onset, indicating that participants used phonological feature variation for multiple word activation during lexical access. While the effect manifested bilaterally in preschool children, 2nd graders showed a left-lateralization that was comparable to the pattern found in adults (e.g., Friedrich et al., 2009), indicating developmental plasticity in hemispheric lateralization that might be independent of literacy acquisition. From these results, we concluded that multiple activation of phonological matching word forms in pre-reading preschoolers appeared to be more tolerant to variation in place of articulation than in reading children and adults. Reading children potentially might use more phonological detail than pre-readers for activating word candidates that match the input (as reflected in graded P350 effects in their ERPs).

Results from the training study (Bauch et al., 2021) partially backed up those previous findings: While reaction time latencies indicated similar post-lexical processing of the phonemic mismatch in all training groups and an adult control group, we found evidence for enhanced phonological processing at early stages of phonological perception (around 100 – 300 ms after stimulus onset) bilaterally over anterior regions for children in the phonemic trainings only. ERP amplitudes to matching prime-target pairs and partially mismatching prime-target pairs revealed that preliterate preschoolers who participated in a phonemic awareness training processed phonemic variation with more sensitivity than preliterate preschoolers who had received the control training. Specifically, mismatching word onset feature started to impact speech processing around 100 ms after word onset, which is a time window that was previously associated with the N100/T-complex and enhanced early auditory and phonological analysis of speech input, as well as auditory attention mechanisms (Connolly, 1993; Diesch & Luce, 2000; Näätänen & Picton, 1987; O'Rourke & Holcomb, 2002; Poeppel et al., 1997; Sanders, Newport, & Neville, 2002; Wolpaw & Penry, 1975). Altogether, we concluded from our previous studies (Bauch et al., 2021; Schild et al., 2011) that like adults (Friedrich et al., 2009; Schild et al. 2012), children use phonological detail for processing spoken language and that processing of phonological detail might be enhanced in children who either have explicit reading expertise (Schild et al., 2011) or have been trained in precursor functions of reading (Bauch et al., 2021), compared to same aged children without reading-related training. The results also indicated that different phonological features might be processed at different processing stages, although direct comparisons between the studies could not be drawn due to methodological differences and different levels of literacy experience between the participants.

To further investigate the role of reading expertise on phonological processing of spoken language especially in a longitudinal approach, in the current study we followed up the children of our training groups in a longitudinal approach. We aimed to gain more insight in the development of phonological representations beyond a cross-sectional comparison as done in Schild et al. (2011). Furthermore, we were interested in how phonological processing of voicing alterations might develop during the first years of formal reading instruction. Specifically, we sought to investigate whether the found enhanced phonological processing of voicing alterations in the trained children might be a short-term product of the training or might undergo further development after the children started learning to read. Children from our training study (Bauch et al., 2021) were re-invited to our laboratory at the end of their first and their second grade. The subjects participated in the same word onset priming paradigm with identical stimulus material as during preschool assessments (see Bauch et al., 2021). If enhanced sensitivity to voicing mismatch during phonological encoding was a temporal by-product of our explicit phonemic awareness training tailored to this phonological feature, this effect in the children might have vanished after a year of attending 1st grade. However, if increasing levels of phonological awareness and reading skills through formal teaching directly relate to increasing sensitivity for phonological feature variation in early automatic speech processing, we expected a stronger priming effect for ERPs as well as for reaction times in 2nd graders, compared to 1st graders, compared to preschoolers.

Methods

Procedure

All participants were part of a training study that was conducted during their final year of kindergarten in their respective kindergarten institutions (approximately 6 months prior to entering elementary school, a transition that is accomplished in Germany within the 6th year of life). The children received a training of precursor functions of reading (phonemic awareness training only or in combination with letter knowledge) or an arithmetic control training (details see Schild et al., 2020). In total, $N = 102$ monolingual children participated in the training study. Ten additional bilingual children also attended the training to maintain the integrity of the pre-school groups, but they did not participate in the study and data collection. After the ten-week-training period, the preschoolers took part in an individual testing session lasting about 30-40 minutes, in which we obtained explicit measurements of language and general cognitive abilities. Furthermore, the children attended one session at our laboratory. Here, they conducted a reaction time experiment with EEG recording that took about 30-40 min to complete.

All children and parents were invited to participate in two follow-up sessions at the end of 1st and 2nd grade. We obtained measures of reading skill as well as explicit measurements of language (about 60 min) for 1st and 2nd graders. We repeated the reaction time experiment with EEG recording with the pupils in each grade in an additional session at our laboratory (about 30-40 min, more details below).

Participants

The trainings were carried out at local kindergartens in the city of Tuebingen, Germany. Before the training started, parents and children received written information about the project and gave their written consent to participate in the whole study (including all three measurements). The ethical committee of the German Psychological Association (Ethikkommission der Deutschen Gesellschaft für Psychologie, 08.2014) advised us regarding the procedures we adopted in this study. There were no ethical concerns raised by the committee.

Originally, we invited all children from the former training study (Bauch et al., 2021) to again participate in the longitudinal study. Because the ERP analysis results differed for preschoolers from the phonemic training groups and the control training group, we did not plan a collapse of all data for the present study. Instead, we aimed for separate analysis for the different training groups. Unfortunately, only 9 children from the original control training group (N = 21) contributed complete data for all three points of measurements (preschool, 1st and 2nd grade), which led to inconclusive analysis results for this particular group of participants. Hence, for the present study we only considered data from children who had previously received one of the two phonemic interventions[1]. This sample included N = 46 children (n = 24 from the phonemic awareness only group, n = 22 from the combined phonemic awareness and letter knowledge group).

A detailed description of training results is presented in Bauch et al. (2021). As pre-post-intervention comparisons in the intervention study revealed, both groups with a phonemic training showed an increased performance on phonological awareness test compared to the control group. Furthermore, ERP analysis revealed sensitivity for phonological mismatch in both groups with a phonemic awareness training, but tolerance for phonological mismatch in the control group. In neither the phonological

[1] In the original analysis reported in Bauch et al. (2021) we found that children from both phonemic trainings showed similar ERP processing in the time window of interest (100-300ms), but not the control group. For the present longitudinal analysis, we therefore planned separate analyses for the phonemic groups and the control group. As only 9 children of the control group contributed complete data sets over the two follow up years, we decided to drop the control group from further analyses. We decided to refrain from merging the 9 valid data sets from the control group with the 28 data sets from the phonemic training groups because of the previously found differences in the baseline neurophysiological processing of the phonemic groups and the control group.

awareness test nor in ERPs, did we find group differences between children that received a pure phonemic awareness training and children with a combined phonemic-orthographic training. Additionally, we compared preschooler's letter knowledge before and after the training. However, there was no significant difference between the growth of letter knowledge between these children and children of a group that exercised on precursor functions of mathematical abilities. This indicated that there was no advantage of letter knowledge in any of the training group that exceeded maturation effects. Consequently, we decided to collapse data sets from both phonemic awareness training groups for the present analysis.

From the 46 preschoolers who had received phonemic awareness training, 28 fulfilled the following inclusion criteria: (1) Parents of the child were native speakers of German and German was the only language spoken at home. (2) The child was not identified as an early reader in preschool via the "Ein Leseverständnistest für Erst- bis Sechstklässler" reading test (ELFE 1-6, Lenhard & Schneider, 2006). A child was considered to be an early reader in this test when they were able to read aloud single unknown words at the subtest "Word Comprehension". (3) The child was not able to read words (except for their own name) in preschool. (4) The child participated and completed all standardized tests in all three points of measurement (preschool, 1^{st} grade, 2^{nd} grade). (5) We were able to obtain EEG recordings from the child at each point of measurement that provided enough segments for analysis (i.e., EEG recordings contained only a minimal amount of noise and provided a minimum of 15 segments per condition (40% of segments per condition) for ERP analysis). (6) The child's error rate in the lexical decision task was below the cut-off rate (for missing words > 20%; for incorrect responses to pseudo-words > 80%). (7) In 1^{st} as well as in 2^{nd} grade, the child's scores in the reading test "Würzburger Leise Leseprobe - Revision" (WLLP-R, Schneider et al., 2011) and in all subtests of the phonological awareness test "Test zur Erfassung der phonologischen Bewusstheit und Benennungsgeschwindigkeit" (TEPHOBE, Mayer, 2011) were at least at average or above average. The final sample size of $N = 28$ aligned with sample sizes of previous studies with the same paradigm and analysis design in children and in adults, which yielded robust findings of the effects in question (e.g., for children: $n = 21$-24 in Bauch et al., 2021; $n = 13$-19 in Schild et al., 2011; for adults: $n = 20$-25 in Friedrich et al., 2009; Schild et al., 2012). Thus, we considered the sample size sufficient for analysis. Out of the 28 datasets we considered for the present analysis, $n = 13$ had received a combined phonemic-orthographic training in preschool.

Table 1 summarizes demographic information and sample characteristics. Preschool children did not have advanced reading skills, but rudimentary knowledge of letters (e.g., knowledge of the letters in their given names). At preschool, all children scored at least average in the phonological awareness test TEPHOBE. For none of the children, parents reported neurological or hearing problems. All children had normal or corrected to normal eyesight. Handedness for all participants was obtained via the "Edinburgh Handedness Inventory" (EHI, Oldfield, 1971).

Training of phonemic awareness

Children received a daily training of phonemic awareness over a period of ten consecutive weeks. Each session ran for approximately 10 to 15 minutes and was conducted by instructed collegiate and doctoral members of the Department of Psychology, Eberhard-Karls-University Tuebingen, Germany. Each session consisted of two to three short games that focused on the training of phoneme onset detection (e.g., identifying the first sound in a given object) and on the training of phoneme synthesis and analysis (e.g., segmenting words to their single phonemes and vice versa, e.g., segmenting the word "gold" in its respective phonemes). The training program was adapted from Küspert and Schneider (2008) and Plume and Schneider (2004). For more details on the training study materials, see Bauch et al., (2021).

Table 1. ***Demographic data and mean results of the standardized tests***

Variable	Preschool	1st Grade	2nd Grade
Sex (male/female)	16/12	16/12	16/12
Mean age (*SD*)[a]	73.78 (4.71)	85.92 (5.03)	96.60 (5.13)
Mean LQ (*SD*)	54.86 (57.31)	-	-
Mean TEPHOBE (*SD*) Total Score	21.64 (3.90)	24.25 (3.70)	26.28 (1.18)
Mean Letter Knowledge (*SD*); capital/small letters	11.00/7.64 (4.09/3.92)	-	-
Mean WLLP-R (*SD*)	-	43.28 (13.68)	73.32 (19.55)

Note. Presented results include standardized tests on handedness (LQ; Oldfield, 1971), phonological awareness (max = 28, TEPHOBE Total Score; Mayer, 2011), letter knowledge (max = 15 for capital and small letters) and reading speed (max = 140, WLLP-R; Schneider et al., 2011). By the end of the 1st and 2nd grade children knew all capital and small letters from the letter knowledge test. Reading speed was only assessed at 1st and 2nd grade, handedness LQ was once measured at preschool. [a] In months, at post-test. Laterality index (LQ) between -100 to -29 indicates left handedness, LQ between -28 to 48 indicates no preference in handedness, LQ between 49 to 100 indicates right handedness.

Test materials

Phonological awareness was tested during preschool, 1st grade and 2nd grade. Reading skills were obtained in the two follow-up sessions in school. Handedness was tested once in preschool.

Phonological awareness. Phonological awareness was measured with the TEPHOBE (Mayer, 2011). This test is available in a version for preschoolers and 1st graders combined and a version for 2nd graders. The TEPHOBE version for preschool children and 1st graders contains the four subtests *Synthesis of Onset and Rhyme, Phoneme Synthesis, Rhyming,* and *Categorization of Initial Sounds*. Due to ceiling effects in the preschoolers, we decided to test 1st graders with the TEPHOBE version for 2nd graders. This version contains five subtests, *Rhyming, Categorization of Onset Phonemes, Categorization of Offset Phonemes, Phoneme Elision* and *Phoneme Reversal*. The latter was excluded as it did not assess a skill relevant for our research question.

Letter knowledge and reading skills. The children were asked to name 15 capital and their corresponding small letters to measure their rudimentary letter knowledge in preschool (G, K, B, P, A, E, I, U, O, D, T, S, W, H, R). We tested 1st and 2nd graders once with one version of the WLLP-R (Schneider et al., 2011). This reading test assesses reading speed in elementary school children. The WLLP-R is available in two versions, which contain the same items but in changed sequence. 1st graders were tested with the A version, 2nd graders with the B version. In preschool, we used the ELFE 1-6 (Lenhard & Schneider, 2006) to identify early readers who were later excluded from the study. The ELFE 1-6 measures reading comprehension of 1st to 6th graders in three subtests (*Comprehension of Words, Comprehension of Sentences, Comprehension of Texts*). Children were excluded when they were able to read and understand words that were given in the subtest *Comprehension of Words,* ergo when they were able to read single given words.

Experimental stimuli and procedure

The experimental material was identical to the material we used in Bauch et al. (2021). We used 74 monomorphemic disyllabic German nouns as targets (see Table A1 in the appendix). All of the nouns were stressed on the first syllable. Half of the nouns started with the phonemes /g/ or /k/, the other half started with /b/ or /p/. The latter phonemes had not been a set of sounds that we trained in the interventions and served as a set of control phonemes that also differed in the voicing feature, in order to track potential generalization effects across different set of phonemes. 74 pseudo-words were added as distractors for the lexical decision task. We generated them by extracting the second syllable of each target word and substituting them with the second syllable of another target word. For example, the second syllable of "Ki*no*" (Engl. cinema) was inserted as the second syllable in "Buers*te*" (Engl. brush) and vice versa, resulting in the two pseudo-words "Ki*te*" and "Buer*no*" (both words do not exist in the German language). Primes were created from the first syllable of each target word.

The prime-target combination varied across three conditions. In the Identity condition, prime and target completely matched (e.g., “Ki - Kino”). In the Variation condition, the prime varied from its assigned target in the voicing of its initial sound (e.g., “Gi - Kino”). In the Control condition, the prime and the target were unrelated insofar as their first syllables contained different phonemes and, additionally, the first phoneme differed in place of articulation as well as in voicing to maximize differences between prime and target (e.g., “Ba - Kino”). Furthermore, prime-target pairs in the control condition never matched in the respective vowels following the initial consonants. A pseudo-word appeared instead of a target in 33% of the trials. Primes and pseudo-words were combined according to the different conditions in the same way as the primes and targets. Targets appeared once in each condition, pseudo-words only once in total.

A male and a female native German speaking actor and actress produced the spoken material. The primes were taken from words spoken by the male speaker while the targets and pseudo-words were taken from the female speaker to prevent mere acoustical priming effects. None of the speakers was aware of the purpose of the study.

Children completed a unimodal auditory word fragment priming experiment with EEG recording. In total, 296 trials (222 targets and 74 pseudo-words) were presented, which appeared in twelve blocks. In eight blocks, the children listened to 25 trials and in four blocks to 24 trials. Targets were not repeated within a block. Trials were randomized within each block. The sequence of the blocks was balanced across participants. We introduced the experiment as a “Word-Catching-Game”. Children were instructed to press the space bar as fast and as correctly as possible whenever they heard a real word and refrain from responding whenever they heard a pseudo-word. Each trial started with the presentation of a fixation picture (1x1 cm, a smiley) in the middle of the screen. After 500 ms the auditory prime was presented. The auditory target or a pseudo-word followed 200 ms after offset of the prime to create a comparable and adequate baseline period for the ERPs. Visual feedback (3x7 cm) was provided for about two seconds in every case the child responded correctly to a target (a smiley flying into a basket) or incorrectly pressed the space bar for a pseudo-word (a little ghost appeared in the middle of the screen). The next trial started 1.5 seconds after feedback offset. No feedback was given whenever the child missed a target. In this case, the next trial started 3.5 seconds after the onset of the target. After each block, a short break was provided. Half of the children used the index finger of their right hand, while the other children used the index finger of their left hand to press the space bar.

Electrophysiological recording

We used 46 active Ag/AgCl electrodes (Brain Products) attached into an elastic cap (Electro Cap International, Inc.) for the continuous EEG recording according to the international 10-20 system (bandpass filter 0.01-100 Hz, BrainAmp Standard, Brain Products, Gilching, Germany). The reference and the ground electrodes were placed on the tip of the nose and in the electrode cap at position AF3, respectively. Two additional electrodes were placed below each eye. Two eye-calibration blocks were presented before and after the experiment. EEG data was processed with the Brain Electrical Source Analysis Software (BESA, MEGIS Software GmbH, Version 5.3). We applied the surrogate Multiple Source Eye Correction (Berg & Scherg, 1994) implemented in BESA for eye-movement artifact correction. For offline analysis, the signal was re-referenced to an average reference. All artifact rejection was computed manually and by visual inspection. Individual noisy channels were linearly interpolated for all trials (M = 3.40, SD = 1.72, *Range* = 0-9). Results reported in the main text were based on recordings filtered offline with a 0.3 Hz high-pass filter. As pointed out by a reviewer, strong high-pass filter might carry the risk of EEG distortion (e.g., Tanner et al., 2015). Therefore, and as suggested by the reviewer, we also considered a re-analysis of the ERP recordings filtered at 0.1 Hz. Those results are reported in Table A2 in the appendix. The re-analysis with 0.1 Hz filter obtained the same significant interaction effects as we found with our original 0.3 Hz filter analysis, therefore we opted for reporting our original analysis with 0.3 Hz filter in the following results section. ERPs were computed only for targets with correct responses, starting from the beginning of the speech signal until 700 ms post-stimulus onset, with a 200 ms pre-stimulus baseline.

Data analysis

Explicit tests. We applied a repeated measures ANOVA with the within-factor *Age* (Preschool vs. 1st grade vs. 2nd grade).

Reaction times and errors. Reaction times (RT) shorter than 200 ms and longer than 2000 ms were removed from analysis. A repeated measures ANOVA with the within-factors *Condition* (Identity vs. Variation vs. Control) and the within-factor *Age* (Preschool vs. 1st grade vs. 2nd grade) was applied. The same procedure was used for the analysis of errors in word trials (omissions).

Event-related potentials. In order to analyze N100 as well as P350 effects, and to keep the analysis closer to the analysis we carried out in our previous studies, four lateral regions of interest (ROI, anterior-left: F9, F7, F3, FT9, FT7, FC5, FC1, T7, C5; posterior-left: C3, TP9, TP7, CP5, CP1, O9, P3, PO9, O1; anterior-right: F10, F8, F4, FT10, FT8, FC6, FC2, T8, C6; posterior-right: C4, TP10, TP8, CP6, CP2, P8, P4, PO10, O2) were identified prior to analyses. Averaged ERPs across each participant and each condition entered analysis. ERP amplitudes were computed with the same ANOVA as the reaction times, with the additional factors *Region* (anterior vs. posterior) and *Hemisphere* (left vs. right). To make the present analysis comparable to the results of Schild et al. (2011) and Bauch et al. (2021), we adapted the same time windows in the present study. This resulted in a first-time window ranging from 100 to 300 ms and a second time window from 300 to 400 ms. Both time windows preceded the behavioral responses. The following result section will only report the highest-ranking significant interactions of *Condition* with significant post hoc comparisons. In case of significant interactions, further follow-up ANOVAs and *t*-tests were computed. All *t*-test results reported below were subject to a Holm-Bonferroni correction.

Results

Explicit tests

In the test for phonological awareness, the ANOVA revealed a main effect of *Age* ($F(2, 54) = 17.80$, $p < .001$, $p^2 = .40$). Children scored best on this when they were at the end of 2nd grade, medium when they were at the end of 1st grade and lowest when they were in preschool. All time points differed significantly from each other, all $t(27) \geq 2.65$, $p \leq .01$, $d \geq .36$. Also in the speed reading test, the ANOVA revealed a main effect of *Age* ($F(1, 27) = 140.83$, $p < .001$, $p^2 = .84$). Children scored higher in the reading test in the 2nd grade, compared to the 1st grade, $t(27) = 11.86$, $p \leq .001$, $d = 15.52$.

Reaction time and error analysis

The ANOVA for reaction times revealed a main effect of *Condition* ($F(2, 54) = 87.49$, $p < .001$, $p^2 = .76$). Response times differed significantly from each other in each condition, all $t(27) \geq 5.69$, $p \leq .0001$, $d \geq .1.01$. Across all points of measurement, children responded fastest to the Identity condition ($M = 956.78$ ms, $SD = 99.78$ ms), followed by medium response times in the Variation condition ($M = 988.79$ ms, $SD = 98.65$ ms), and slowest response times in the Control condition ($M = 1061.30$ ms, $SD = 109.57$ ms). Furthermore, a main effect of *Age* ($F(2, 54) = 10.03$, $p < .001$, $p^2 = .27$) revealed that across all trials, children responded faster as pupils (1st grade: $M = 987.79$ ms, $SD = 114.52$ ms; 2nd grade: $M = 954.70$ ms, $SD = 135.78$ ms) than they responded as preschoolers ($M = 1064.38$ ms, $SD = 126.40$ ms), both $t(27) \geq 3.30$, $p \leq .002$, $d \geq .59$. There was no difference between the overall response times obtained at the end of the 1st and 2nd year of schooling, $t(27) = 1.24$, n.s. We did not find an interaction effect of *Age* x *Condition* ($F(4, 108) = 1.71$, n.s.). Figure 1 illustrates the mean response times as a function of age and condition.

The overall error rate across children and conditions was 2.84% ($SD = 1.65$, Range 1.05% - 7.50%). The ANOVA revealed a main effect of *Age* ($F(2, 54) = 21.75$, $p < .001$, $p^2 = .45$). At the end of their 1st and 2nd grade, children made less mistakes than they made in preschool, both $t(27) \geq 4.90$, $p \leq .0001$, $d \geq .70$. While in preschool, children missed on average 4.76% ($SD = 3.23\%$) “yes” responses to words, the error rate dropped to 1.88% ($SD = 1.45\%$) and 1.44% ($SD = 1.48\%$) at the end of 1st and 2nd grade, respectively. Overall error rates at the end of 1st and 2nd grade, did not differ significantly, $t(27) = 1.31$, n.s. There was no significant main effect of Condition ($F(2, 54) = 1.95$, n.s.), and no interaction between *Age* x *Condition*, $F(4, 108) = 0.85$, n.s.

Event-related potentials

Figure 2 presents the ERP effects for anterior and posterior regions for each age group. Figure 3 presents averaged priming effects between the three age groups, suggesting that there were no timing differences of ERP deflections paralleling the reaction time differences obtained at the three ages.

100 – 300 ms, N100. The ANOVA revealed significant interactions of *Condition* x *Region* ($F(2, 54) = 26.51$, $p < .001$, $p^2 = .50$) and *Condition* x *Hemisphere* ($F(2, 54) = 7.84$, $p = .001$, $p^2 = .23$). A graded ERP priming pattern emerged when anterior and posterior regions were considered separately (as guided by the significant interaction of the factors *Condition* and *Region*). Amplitudes from the Identity condition and from the Variation condition were both more negative over anterior and more positive over posterior regions than amplitudes from the Control condition, all $t(83) \geq 2.93$, $p \leq .004$, $d \geq .34$. Crucially, amplitudes from the Identity condition were also more negative (resp. positive) than amplitudes from the Variation condition, $t(83) \geq 2.13$, $p \leq .03$, $d \geq .23$.

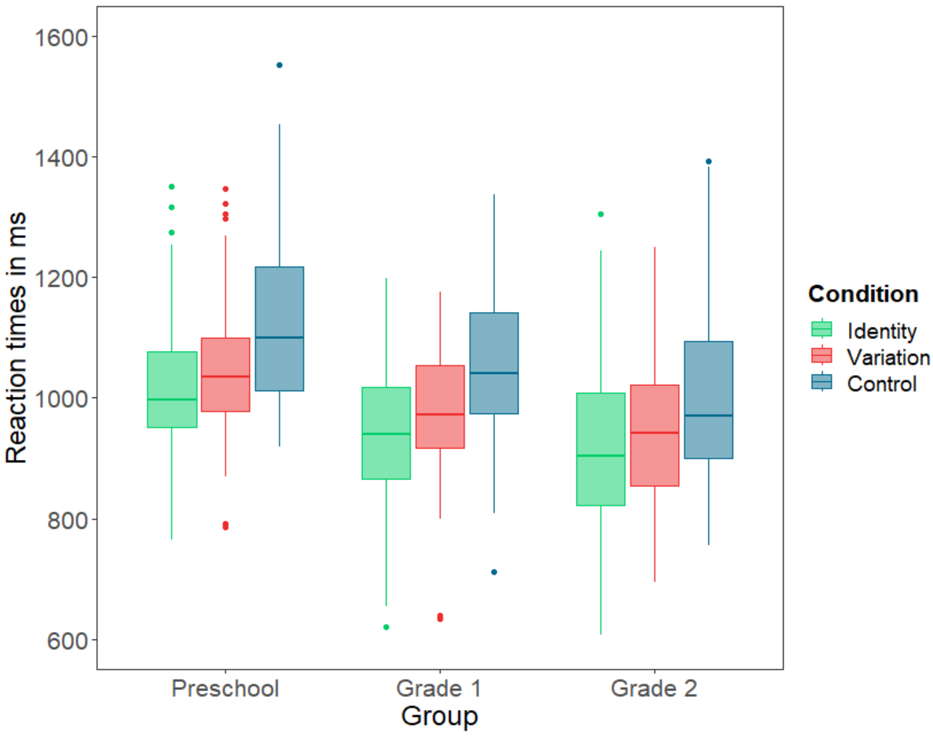

Figure 1. ***Mean reaction times and quantiles of the three conditions (Identity,***

Variation and Control) for each age group (Preschool, Grade 1, Grade 2).

At the same time, there were indices of rough priming, not differentiating the Identity and the Variation condition, for the left hemisphere (the significant interaction of the factors *Condition* and *Hemisphere* guided separate consideration of both hemispheres). Over the left hemisphere, amplitudes in the Identity and Variation condition were both more negative than in the Control condition, both $t(83) \geq 4.01$, $p \leq .0001$, $d \geq .46$. There was no difference between amplitudes from the Identity and Variation condition, $t(83) = 0.07$, n.s. Over the right hemisphere, amplitudes in the Variation condition were more negative than amplitudes from the Control condition, $t(83) = 3.22$, $p = .001$, $d = 0.34$. There were no significant differences between the Identity and Variation condition, nor between the Identity and Control condition, all $t(83) \leq 1.63$, n.s.

Additionally, and as suggested by a reviewer as post-hoc analysis, we included a latency analysis of peaks to consider for timing differences in neurophysiological processing between the three age groups. There were no significant differences in the latencies between the age groups, $F(2, 81) = 0.46$, $p = .631$, $p^2 = .01$. Figure 3 presents averaged priming

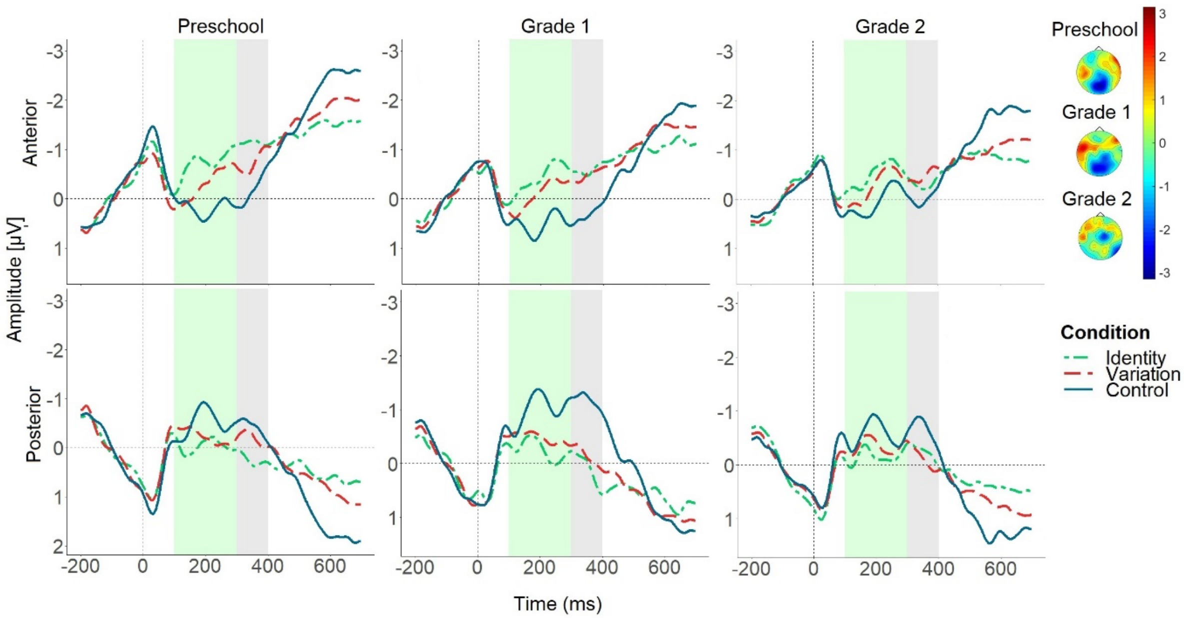

Figure 2. ***Mean ERP effects over anterior and posterior regions for all groups. Identity condition (green, short-dashed line), Variation condition (red, long-dashed line) and Control condition (blue, solid line). The light green bar marks the analysis area for the time window ranging from 100 to 300 ms. The grey bar marks the analysis area for the time window ranging from 300 to 400 ms. Topographical voltage maps***

indicate difference waves for the Variation condition minus the Identity condition for each age group. Topographical voltage maps represent averaged amplitude differences for all age groups for the time window ranging from 100 to 300 ms, during which significant differences between the Identity condition and the Variation condition occurred.

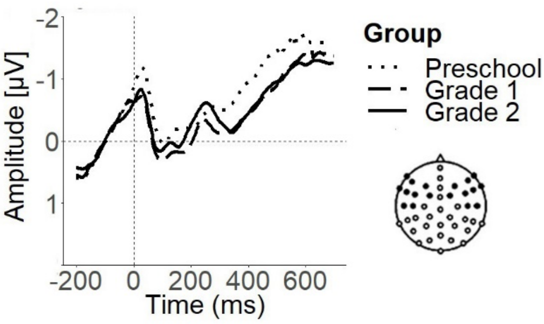

Figure 3. ***Averaged amplitudes for all conditions for each age group. Preschoolers (dotted line), 1st graders (dashed line), 2nd graders (solid line). Electrode position map with marked electrodes for anterior regions.***

effects between the three age groups, suggesting that there were no timing differences of ERP deflections paralleling the reaction time differences obtained at the three ages.

300 – 400 ms, P350. Again, the ANOVA revealed two significant interactions, one of the factors *Condition* x *Region* (F(2, 54) = 23.74, p < .001, p^2 = .47), and another one of the factors *Condition* x *Hemisphere* (F(2, 54) = 6.18, p = .004, p^2 = .19).

Both interactions pointed to rough priming, not differentiating the Identity and the Variation condition in the second time window. Guided by the significant interaction of the factors *Condition* and *Region*, we analyzed anterior and posterior regions separately. There were no differences between amplitudes from the Identity and Variation condition (both t(83) ≤ 1.26, n.s.), but amplitudes from both conditions were more negative over anterior and more positive over posterior regions than amplitudes rising from the Control condition, all t(83) ≥ 4.02, p ≤ .0001, d ≥ .44. Guided by the significant interaction of the factors *Condition* and *Hemisphere*, we also analyzed left and right regions separately. Over the left and right hemisphere, only amplitudes from the Variation condition were more negative than amplitudes from the Control condition, t(83) ≥ 2.72, p ≤ .007, d ≥ .28. We found no differences in the comparisons between the amplitudes from both Identity and Variation condition and Identity and Control condition, all t(83) ≤ 1.88, n.s.

To sum up, we found graded response times and early graded ERP priming patterns differentiating all three conditions across all three tested ages. There were no timing differences in the ERPs paralleling speeded lexical decisions when children were able to read.

Discussion

In a longitudinal study, we tested how children process spoken words after a training of phonemic awareness in preschool, and as a function of formal reading and writing instruction in elementary school. The training comprised ten weeks with daily ten-minute training sessions (see Bauch et al., 2021). For the second and third measurement, we tested the children who had participated in the training at the end of their first and second year of elementary school. We were interested in the degree of speech detail that children considered for different aspects of spoken word recognition, and in the timing of those aspects as a function of reading acquisition. To this end, we recorded lexical decision latencies and ERPs to targets presented in spoken word onset priming at all three measurements. We considered different responses to targets overlapping with preceding primes (e.g., "Ki - Kino") compared to partially mismatching targets (e.g., "Gi - Kino") as informative regarding the amount of speech detail that children exploit. ERPs indicated that children were able to exploit phonological feature variation at all points of measurements. In the ERPs, matching and partially mismatching targets elicited differences substantiating 100 to 300 ms after target word onset across all ages, replicating previous results for phonemic variations in voicing for preschoolers sensitized to voicing mismatches (Bauch et al., 2011). Reduced amplitudes within this time window have been interpreted to be related to facilitated auditory processing and phonological encoding (Friedrich et al., 2009; Lange & Röder, 2006; O'Rourke & Holcomb, 2002; Praamstra & Stegeman, 1993; Sanders & Astheimer, 2008; Sanders & Neville, 2003; Schild et al., 2014; Schild et al., 2012).

Steady priming effects in the ERPs across the three age groups indicated that the processing of phonological detail did not change with emerging reading experience (1st grade) and prolonged reading experience (2nd grade). After preschool, children's phonemic awareness training continued during this longitudinal study when they entered elementary school, where they received formal instruction on the phonological principle. Indeed, children constantly improved in phonological awareness measures in offline explicit phonological awareness tasks. Children scored lowest on these measures of phonological awareness when they were in kindergarten and highest when they were in 1st and 2nd grade. This is in line with various studies showing that specific aspects of phonological awareness profit from reading acquisition (Ehri & Wilce, 1980; Morais et al., 1979; Treiman & Cassar, 1997; Tunmer & Nesdale, 1985). Additionally, other factors such as general maturation processes and/or linguistic exposure might underlie development of phonological awareness related skills in children aged 6 to 8 years (Bentin et al., 1991; Cunningham & Carroll, 2011). Yet, this growth in metalinguistic speech processing was not reflected in the amount of detail children used during spoken word processing at pre-lexical level. These results might imply that sensitivity for phonemic mismatch was sufficiently adapted at the the end of the preschool training. Thus, and contrary to our expectation, children do not appear to specify their phonological representations as a function of developing phonological awareness, phonemic awareness or more broadly emerging reading skills. Rather, they appear to reach a threshold level of which might be sufficient to facilitate phonetically mediated access and strategic mechanisms across middle childhood.

Still, we interpret the current results in favor of the assumption that, in middle childhood, automatic stages of speech processing are modulated by facilitated phonological processing via precursors of literacy such as phonological and phonemic awareness (Dehaene et al., 2015; Harm & Seidenberg, 2004; Pattamadilok et al., 2010; Taft, 2006). Former research pointed to the mutual relationship between learning to read and phonemic awareness (Deacon et al., 2013; Perfetti et al., 1987). At an initial stage of reading, decoding letters to corresponding sounds is crucial for understanding the alphabetic script (Anthony & Francis, 2005). Thus, learning to read emphasizes and triggers the refinement of explicit phonological representations which feeds back on their implicit counterparts stored in the mental lexicon. Taking into consideration that adult participants in the training study showed no specific sensitivity to voicing mismatches in the ERPs (Bauch et al., 2021), one might speculate that phonemic information regarding voicing is heightened when readers initially become sensitive to small differences between phonemes. Later on, at least for voicing, such differences might become less important for pre-lexical processing.

Again, in the present study, ERP differences obtained for matching and partially mismatching targets for voicing feature emerged somewhat earlier than the formerly obtained ERP differences for place variation in adults (e.g., Friedrich et al., 2009; Schild et al., 2012) and reading children (Schild et al., 2011). In the present study, ERP differences were evident between 100 and 300 ms after target word onset across all age groups. As the timing of ERP differences in the present study was consistent for all age groups (and hence was not restricted to the training), these results further suggest that timing differences across the various studies using word fragment priming relates to the different features varied in the studies. While initial place varied in partially mismatching prime-target pairs in the former studies (e.g. Friedrich et al., 2009 for adults; Schild et al., 2011 for reading children), initial voicing varied in the present study. We speculate that due to the used stimuli, voicing information in this study might be earlier available in the signal (vibration of voiced speech sounds starts with their onset) than place information (formant information indicating place develops across the speech sound). While voiced plosives in German lack pre-voicing and are therefore available relatively late in the signal (Geiss et al., 2022), our stimuli also consisted of unvoiced plosives with longer VOT times. This might relate to an on average earlier timing of respective ERP differences elicited by voicing feature compared to the place feature. Especially the timing of word form activation might depend on the availability of phonological information and might not always need 300 ms after target word onset. However, as the current study does not allow for direct timing comparisons between the features, future studies will be needed to investigate this question in further detail.

Similarly to the ERP results, children showed delayed responses for partially mismatching targets compared to matching targets in their lexical decisions across all three measurements. While ERPs exclusively reflect prime-target overlap in phonological information, additional factors are associated with lexical decision latencies. For example, participants made delayed responses (compared to an unrelated condition) for target words that partially matched the preceding primes (e.g., "Ana - Anorak") when a better matching completion of the prime existed (e.g., "Ananas", Engl. pineapple; Friedrich et al., 2013). This contrasted with reduced ERP amplitudes for partially overlapping targets compared to an unrelated condition (e.g., "Idi - Anorak"). It was concluded that overlapping words receive bottom-up activation from the primes (as reflected in ERP amplitude reduction), but that better matching words either hinder selection of partially overlapping words or interfere with the lexical decision response (as reflected in delayed lexical decisions). Speeded speech processing is another aspect that reaction times, but not ERPs, capture. Congenitally blind adults made faster lexical decisions in unimodal auditory word onset priming, but their ERPs did not reflect timing differences compared to hearing controls (Schild & Friedrich, 2018). This result suggested that the adult system realizes speeded speech processing via facilitated post-lexical, strategic aspects of processing rather than via facilitated phonological encoding and lexical mapping (which appear optimally adjusted to the input in both hearing and congenitally blind adults).

In the present study, age-related differences only emerged for mean response latencies. Overall, children responded fastest as 2nd graders, with medium speed as 1st graders and slowest as preschoolers. Age-related speeding of responses and the overall decreasing error rates with increasing age might have several triggers, including general and motoric maturation that – among other aspects like repeated testing – might affect speeded motor reactions or enhanced attention and concentration spans. Yet, it is important to note that these factors do not contribute to respective speeding of ERP deflections. This dissociation finds a parallel in a former word onset priming study with congenitally blind and sighted adults (Schild & Friedrich, 2018). In combination with the present results, both studies imply modifications of lexical decision responses between groups that might reflect different adjustments and proficiency in processing auditory input. While lexical decision responses might be sensitive to late strategic mechanisms that interfere with the yes-responses to the targets, ERPs might more closely relate to rapid phonologically mediated lexical access and phonological processes that are less prone to strategic, post-lexical modulations. We might conclude that, comparable to adults, children realize speeded speech processing via relatively late aspects of processing rather than via facilitated phonological encoding and lexical mapping. That is, already in childhood, the timing of input-related implicit aspects of phonologically mediated lexical access appear optimally adjusted to the input.

If one speculates that (besides maturation effects) literacy experience could also affect the performance on a metalinguistic task such as the lexical decision, faster lexical decision latencies with increasing age seem to be in accord with the assumption that reading experience fosters prediction in language processing (Huettig & Pickering, 2019). Eye tracking studies already suggest that proficient readers predict spoken language faster than less proficient readers and illiterate adults (Mishra et al., 2012), and children who are good readers are more efficient in predicting than those who are less-proficient readers (Mani & Huettig, 2014). The present data suggest that enhanced reading proficiency from preschool to 2nd grade might foster priming of lexical decision responses in spoken word recognition. Predictions within a priming paradigm can aid responses for related prime-target pairs. As only the timing of lexical decisions, but not the timing of ERPs, varied with increasing reading proficiency, we might conclude that predictions modulate selection of word candidates in the speech recognition process rather than bottom-up activation of potential word candidates.

Conclusion

With this study, we took a developmental approach on how phonological sensitivity of different aspects of spoken word processing evolves during the very beginning of learning to read. The importance of this work lies in the longitudinal approach in the investigation of neuronal plasticity of phonological representations in middle childhood. The findings suggest a complex relationship between phonemic awareness, reading acquisition, and spoken word processing. Preschool children trained in phonemic awareness showed detailed implicit and explicit spoken word processing. The findings stress the importance of phonological awareness for phonological word processing in early stages of literacy. While meta-linguistic processing continued to develop after children started learning to read, processing of voicing variations did not become more detailed once children gained reading experience and stable knowledge of letters after 1st and 2nd grade. In that, the findings implicate that while the development of metalinguistic phonological skills and underlying neuronal phonological processing might be closely related (e.g., Bauch et al., 2021; Dehaene et al., 2015; Harm & Seidenberg, 2004), diverging pathways of both aspects of phonological processing might emerge. However, unlike adults (Bauch et al., 2021), primary school children still appear to use voicing variation for gradually modulating access to stored phonological representations during spoken word recognition. Thus, beginning readers' pre-lexical processing of phonological feature variation might still profit from training of conscious understanding of the language's structure during formal schooling. It remains to be determined when sensitivity to this feature begins decreasing in implicit spoken word recognition. Furthermore, the findings implicate diverging neuronal processes for different phonological features. We acknowledge the need for more research to understand when sensitivity to certain phonological features decreases in implicit spoken word recognition, emphasizing the ongoing evolution of children's language processing abilities.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439. https://doi.org/10.1006/jmla.1997.2558

Anthony, J. L., & Francis, D. J. (2005). Development of Phonological Awareness. *Current Directions in Psychological Science, 14*(5), 255-259. https://doi.org/10.1111/j.0963-7214.2005.00376.x

Anthony, J. L., & Lonigan, C. J. (2004). The Nature of Phonological Awareness: Converging Evidence From Four Studies of Preschool and Early Grade School Children. *Journal of Educational Psychology, 96*(1), 43–55. https://doi.org/10.1037/0022-0663.96.1.43

Bauch, A., Friedrich, C. K., & Schild, U. (2021). Phonemic training modulates early speech processing in pre-reading children. *Frontiers in Psychology, 12*(1959). https://doi.org/10.3389/fpsyg.2021.643147

Becker, A., Schild, U., & Friedrich, C. K. (2018). Tracking independence and merging of prosodic and phonemic processing across infancy. *Developmental Science, 21*(2). https://doi.org/10.1111/desc.12525

Bentin, S., Hammer, R., & Cahan, S. (1991). The Effects of Aging and First Grade Schooling on the Development of Phonological Awareness. *Psychological Science, 2*(4), 271-275. https://doi.org/10.1111/j.1467-9280.1991.tb00148.x

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology, 90*(3), 229-241. https://doi.org/10.1016/0013-4694(94)90094-9

Caravolas, M., & Bruck, M. (1993). The effect of oral and written language input on children's phonological awareness: A cross-linguistic study. *Journal of Experimental Child Psychology, 55*(1), 1–30. https://doi.org/10.1006/jecp.1993.1001

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R.T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13*(11), 1428-1432. https://doi.org/10.1038/nn.2641

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the Beginnings of Spoken Words Have a Special Status in Auditory Word Recognition? *Journal of Memory and Language, 32*(2), 193-210. https://doi.org/10.1006/jmla.1993.1011161

Connolly, J. F. (1993). The influence of stimulus intensity, contralateral masking and handedness on the temporal N1 and the T complex components of the auditory N1 wave. *Electroencephalography and Clinical Neurophysiology, 86*(1), 58-68. https://doi.org/10.1016/0013-4694(93)90067-6

Cunningham, A. J., & Carroll, J. M. (2011). Reading-related skills in earlier- and later-schooled children. *Scientific Studies of Reading, 15*(3), 244–266. https://doi.org/10.1080/10888431003706309

Cutler, A., & Davis, C. (2012). An Orthographic Effect in Phoneme Processing, and Its Limitations. *Frontiers in Psychology, 3*(18). https://doi.org/10.3389/fpsyg.2012.00018

Deacon, H., Benere, J., & Pasquarella, A. (2013). Reciprocal Relationship: Children's Morphological Awareness and Their Reading Accuracy Across Grades 2 to 3. *Developmental Psychology, 49*(6), 1113-1126. https://doi.org/10.1037/a0029474

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience, 16*(4), 234-244. https://doi.org/10.1038/nrn3924

Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science, 330*(6009), 1359-1364. https://doi.org/ 10.1126/science.1194140

Diesch, E., & Luce, T. (2000). Topographic and Temporal Indices of Vowel Spectral Envelope Extraction in the Human Auditory Cortex. *Journal of Cognitive Neuroscience, 12*(5), 878-893. https://doi.org/10.1162/089892900562480

Ehri, L. C., & Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Applied Psycholinguistics, 1*(4), 371-385. https://doi.org/10.1017/S0142716400009802

Friedrich, C. K. (2005). Neurophysiological correlates of mismatch in lexical access. *BMC Neuroscience, 6*(64), 64. https://doi.org/10.1186/1471-2202-6-64

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. (2004). ERPs Reflect Lexical Identification in Word Fragment Priming. *Journal of Cognitive Neuroscience, 16,* 541-552. https://doi.org/10.1162/089892904323057281

Friedrich, C. K., Felder, V., Lahiri, A., & Eulitz, C. (2013). Activation of words with phonological overlap. *Frontiers in Psychology, 4,* 556. https://doi.org/10.3389/fpsyg.2013.00556

Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological Evidence for Underspecified Lexical Representations: Asymmetries With Word Initial Variations. *Journal of Experimental Psychology Human Perception & Performance, 34,* 1545-1559. https://doi.org/10.1037/a0012481

Friedrich, C. K., Schild, U., & Röder, B. (2009). Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. *Biological Psychology, 80*(1), 105-113. https://doi.org/10.1016/j.biopsycho.2008.04.012

Geiss, M., Gumbsheimer, S., Lloyd-Smith, A., Schmid, S., & Kupisch, T. (2022). VOICE ONSET TIME IN MULTILINGUAL SPEAKERS: ITALIAN HERITAGE SPEAKERS IN GERMANY WITH L3 ENGLISH. *Studies in Second Language Acquisition, 44*(2), 435-459. doi:10.1017/S0272263121000280

Goswami, U. (2000). Phonological representations, reading development and dyslexia: towards a cross-linguistic theoretical framework. *Dyslexia, 6*(2), 133-165151. https://doi.org/10.1002/(SICI)1099-0909(200004/06)

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662-720. https://doi.org/10.1037/0033-295x.111.3.662

Hervais-Adelman, A., Kumar, U., Mishra, R. K., Tripathi, V. N., Guleria, A., Singh, J. P., & Huettig, F. (2021). How does literacy affect speech processing? Not by enhancing cortical responses to speech, but by promoting connectivity of acoustic-phonetic and graphomotor cortices. *Journal of Neuroscience, 42* (47), 8826-8841. 10.1523/JNEUROSCI.1125-21.2022

Huettig, F., & Pickering, M. J. (2019). Literacy Advantages Beyond Reading: Prediction of Spoken Language. *Trends in Cognitive Sciences, 23*(6), 464-475. https://doi.org/10.1016/j.tics.2019.03.008

Hulme, C., Caravolas, M., Málková, G., & Brigstocke, S. (2005). Phoneme isolation ability is not simply a consequence of letter-sound knowledge. *Cognition, 97*(1), B1–B11. https://doi.org/10.1016/j.cognition.2005.01.002

Küspert, P., & Schneider, W. (2008). *Hören, lauschen, lernen - Sprachspiele für Kinder im Vorschulalter* (6. Aufl.). Vandenhoeck & Ruprecht.

Lange, K., & Röder, B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. *Journal of Cognitive Neuroscience, 18*(5), 715-729. https://doi.org/10.1162/jocn.2006.18.5.715

Lenhard, W., & Schneider, W. (2006). *ELFE 1-6 - Leseverständnistest für Erst- bis Sechstklässler*. Hogrefe.

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: a study of children developing proficiency in reading. *Journal of Experimental Child Psychology, 126*, 264-279. https://doi.org/10.1016/j.jecp.2014.05.004

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102. https://doi.org/10.1016/0010-0277(87)90005-9

Mayer, A. (2011). *TEPHOBE. Test zur Überprüfung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit*. Reinhardt Verlag.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86. https://doi.org/10.1016/0010-0285(86)90015-0

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science, 343*(6174), 1006-1010. https://doi.org/10.1126/science.1245994

Mishra, R., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eyemovements are modulated by reading ability - Evidence from Indian low and high literates. *Journal of Eye Movement Research, 5*. https://doi.org/10.16910/jemr.5.1.3

Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language, 85*, 116-134. https://doi.org/10.1016/j.jml.2015.08.005

Monzalvo, K., & Dehaene-Lambertz, G. (2013). How reading acquisition changes children's spoken language network. *Brain and Language, 127*(3), 356-365. https://doi.org/10.1016/j.bandl.2013.10.009

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*(4), 323-331. https://doi.org/10.1016/0010-0277(79)90020-9

Muneaux, M., & Ziegler, J. C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes, 19*(5), 641-660. https://doi.org/10.1080/01690960444000052

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395. https://doi.org/10.1037/0033-295X.115.2.357

O'Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology, 60*(2-3), 121-150. https://doi.org/10.1016/S0301-0511(02)00045-5

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Pattamadilok, C., Knierim, I. N., Kawabata Duncan, K. J., & Devlin, J. T. (2010). How does learning to read affect speech perception? *Journal of Neuroscience, 30*(25), 8435-8444. https://doi.org/10.1523/JNEUROSCI.5791-09.2010

Pattamadilok, C., Morais, J., Colin, C., & Kolinsky, R. (2014). Unattentive speech processing is influenced by orthographic knowledge: evidence from mismatch negativity. *Brain and Language, 137*, 103-111. doi:10.1016/j.bandl.2014.08.005

Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language and Cognitive Processes, 22*(5), 700-726. doi:10.1080/01690960601049628

Pattamadilok, C., Perre, L., & Ziegler, J. C. (2011). Beyond rhyme or reason: ERPs reveal task-specific activation of orthography on spoken language. *Brain and Language, 116*(3), 116-124. https://doi.org/10.1016/j.bandl.2010.12.002

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic Knowledge and Learning to Read are Reciprocal: A Longitudinal Study of First Grade Children. *Merrill-Palmer Quarterly, 33*(3), 283-319. Retrieved from http://www.jstor.org/stable/23086537

Perre, L., Bertrand, D., & Ziegler, J. C. (2011). Literacy affects spoken language in a non- linguistic task: An ERP study. *Frontiers in Psychology, 2,* 1-8. https://doi.org/10.3389/fpsyg.2011.00274

Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain research, 1188,* 132–138. https://doi.org/10.1016/j.brainres.2007.10.084

Plume, E., & Schneider, W. (2004). *Hören, lauschen, lernen 2. Spiele mit Buchstaben und Lauten für Kinder im Vorschulalter*. Vandenhoeck & Ruprecht.

Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P. L., & Marantz, A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters, 221*(2), 145-148. https://doi.org/10.1016/S0304-3940(97)13325-0

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research, 1*(2), 73-86. https://doi.org/10.1016/0926-6410(93)90013-U

Sanders, L. D., & Astheimer, L. B. (2008). Temporally selective attention modulates early perceptual processing: event-related potential evidence. *Perception & Psychophysics, 70*(4), 732-742. https://doi.org/10.3758/PP.70.4.732

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research, 15*(3), 228-240. https://doi.org/10.1016/S0926-6410(02)00195-7

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience, 5*(7), 700–703. https://doi.org/10.1038/nn873

Schild, U., Bauch, A., & Nuerk, H.-C. (2020). A Finger-based Numerical Training Failed to Improve Arithmetic Skills in Kindergarten Children beyond Effects of an Active Non-Numerical Control Training. *Frontiers in Psychology, 11*(529). https://doi.org/10.3389/fpsyg.2020.00529

Schild, U., Becker, A., & Friedrich, C. K. (2014). Processing of syllable stress is functionally different from phoneme processing and does not profit from literacy acquisition. *Frontiers in Psychology, 5*(530). https://doi.org/10.3389/fpsyg.2014.00530

Schild, U., & Friedrich, C. K. (2018). What determines the speed of speech recognition? Evidence from congenitally blind adults. *Neuropsychologia, 112,* 116-124. https://doi.org/10.1016/j.neuropsychologia.2018.03.002

Schild, U., Röder, B., & Friedrich, C. K. (2011). Learning to read shapes the activation of neural lexical representations in the speech recognition pathway. *Developmental Cognitive Neuroscience, 1*(2), 163-174. https://doi.org/10.1016/j.dcn.2010.11.002

Schild, U., Röder, B., & Friedrich, C. K. (2012). Neuronal spoken word recognition: The time course of processing variation in the speech signal. *Language and Cognitive Processes, 27,* 159-183. https://doi.org/10.1080/01690965.2010.503532

Schneider, W., Blanke, I., Faust, V., & Küspert, P. (2011). *Würzburger Leise Leseprobe - Revision. Ein Gruppentest für die Grundschule*. Hogrefe.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of experimental psychology. Learning, memory, and cognition, 13*(1), 64–75. https://doi.org/10.1037//0278-7393.13.1.64

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and Suprasegmental Mismatch in Lexical Access. *Journal of Memory and Language, 45*(3), 412-432. https://doi.org/10.1006/jmla.2000.2783

Taft, M. (2006). Orthographically Influenced Abstract Phonological Representation: Evidence from Non-rhotic Speakers. *Journal of Psycholinguistic Research, 35*(1), 67-78. https://doi.org/10.1007/s10936-005-9004-5

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology, 52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Treiman, R., & Cassar, M. (1997). Can children and adults focus on sound as opposed to spelling in a phoneme counting task? *Developmental Psychology, 33*(5), 771-780.

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology, 77*(4), 417-427. https://doi.org/10.1037/0022-0663.77.4.417

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes, 19*(1), 57-95. doi:10.1080/01690960344000134

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *WIREs Cognitive Science, 3*(3), 387-401. https://doi.org/10.1002/wcs.1178

Wolpaw, J. R., & Penry, J. K. (1975). A temporal component of the auditory evoked response. *Electroencephalography and Clinical Neurophysiology, 39*(6), 609-620. https://doi.org/10.1016/0013-4694(75)90073-5

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683-689. https://doi.org/10.3758/BF03208845

Data, code and materials availability statement

The Editor, Ben Ambridge, granted an exemption (13th September 2023) to materials sharing for (a) the SPY GAME, which includes copyright pictures (though the protocol for the test is shared), (b) the Edinburgh Handedness Inventory (used in German translation), whose copyright status is unclear, and the following tests which are subject to copyright: (c) TEPHOBE (Mayer, A. (2011). TEPHOBE. Test zur Überprüfung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit. Reinhardt Verlag; (d) WLLP-R (Schneider, W., Blanke, I., Faust, V., & Küspert, P. (2011). Würzburger Leise Leseprobe - Revision. Ein Gruppentest für die Grundschule. Hogrefe/); (e) ELFE 1-6 (Lenhard, W., & Schneider, W. (2006). ELFE 1-6 - Leseverständnistest für Erst- bis Sechstklässler. Hogrefe.). All other data, questionnaires, test sheets and codes are accessible at OSF storage and provide anonymous access under the following link: https://osf.io/sr4wv/?view_only=9fc86cb0875e4b81bd8d3ec1a9b23b50

Ethics statement

The study involving human participants was reviewed and approved by Ethikkommission der Deutschen Gesellschaft für Psychologie. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Authorship and Contributorship Statement

US conceived the study. AB took primary responsibility for drafting the manuscript and conducted the study and analyzed the data. AB, CF, and US contributed to design of the training and the tasks. CF and US commented on drafts. All authors read and approved the submitted version.

Declaration of conflict of interests

The authors declare that there are no conflicts of interests for this report.

Acknowledgements

We are especially grateful to Dr. Anne Rau for scheduling trainings and measurements in close and kind interaction with kindergartens and parents, as well as for carefully reading a former version of this manuscript. We would further like to thank Janina Althaus, Nina Bee, Sabine Braasch, Clemens Bromann, Jana Denkinger, Lena Flik, Eva Franz, Stella Gellner, Birte Herter, Isolde Kolb, Ruth Keßler, Janina Krempel, Johanna Lechner, Larissa Lorch, Monika Peterová Jarošová, Angelika Schmitt, Anna-Lena Schultz, Daniel-Niels Tönsing, Gülsan Usta, Charlotte Veil, Marlene Weller and Miriam Wild for assistance in the training and in collecting the data. We also warmly thank all kindergartens, parents and especially children who participated in this study. The study was part of the dissertation written by Anne Bauch.

Appendix A

Table A1. ***List of stimuli.***

Target words	Pseudo-words	Target words	Pseudo-words
Gei-er (vulture)	Geine	Pap-pe (cardboard)	Papke
Gei-ge (violin)	Geise	Pul-ver (powder)	Pulbel
Git-ter (grid)	Gitsche	Pum-pe (pump)	Pumle
Gloc-ke (bell)	Glocpe	Pud-ding (pudding)	Pudhe
Gra-ben (trench)	Grany	Pup-pe (doll)	Pupte
Gren-ze (border)	Grenhe	Pic-kel (pimple)	Picsche
Gru-be (pit)	Gruza	Po-ny (pony)	Poben
Grup-pe (group)	Grupzle	Piz-za (pizza)	Pizbe
Guer-tel (belt)	Guerbe	Pan-ne (breakdown)	Panze
Gum-mi (rubber)	Gumse	Peit-sche (whip)	Peitter
Gur-ke (cucumber)	Gurbon	Pom-mes (fries)	Pombel
Kaff-ee (coffee)	Kaffnen	Bri-lle (glasses)	Brissen
Ka-ter (male cat)	Kaffel	Bie-ne (bee)	Bieer
Kat-ze (cat)	Katne	Brun-nen (fountain)	Brunnee
Ker-ze (candle)	Kertel	Bon-bon (candy)	Bonke
Ket-te (chain)	Ketzel	Bru-der (brother)	Bruchen
Keu-le (mace)	Keusen	Bam-bus (bamboo)	Bamsche
Ki-no (cinema)	Kite	Ba-by (baby)	Bave
Kir-che (church)	Kirber	But-ter (butter)	Butche
Kir-sche (cherry)	Kirbus	Bue-gel (stirrup)	Buede
Kis-sen (pillow)	Kisle	Bruec-ke (bridge)	Bruecfer
Kis-te (box)	Kiskel	Brem-se (break)	Bremken
Kno-chen (bone)	Knoder	Buer-ste (brush)	Buerno
Kno-ten (knot)	Knore	Brau-se (shower)	Braunig
Koe-nig (king)	Koese	Bom-be (bomb)	Bomtel
Kof-fer (trunk)	Kofke	Bir-ne (pear)	Birgel
Krae-he (crow)	Kraeding	Bue-hne (stage)	Buehmel
Kraeu-ter (herbage)	Kraeude	Blu-me (flower)	Bluchen
Kral-le (claw)	Kralpe	Blue-te (blossom)	Bluene
Krei-de (chalk)	Kreigel	Buef-fel (buffalo)	Buefter
Kroe-te (toad)	Kroepe	Be-sen (broom)	Bele
Kro-ne (crown)	Krote	Blu-se (blouse)	Bluge
Krue-mel (crumbs)	Kruehne	Bla-se (bubble)	Blami
Kue-che (kitchen)	Kueter	Bi-ber (beaver)	Biche
Kue-ken (chicken)	Kuekse	Bee-re (berry)	Beerten
Kur-ve (curve)	Kurby	Beu-tel (bag)	Beuze
Kut-sche (carriage)	Kutkel	Bre-zel (pretzel)	Brete

Table A2. ***EEG analysis ANOVA effects for both time windows of interest (100-300ms and 300-400ms) with EEG signal being preprocessed with 0.1hz high pass filter.***

Effect	df	*F*	*p*	p^2
		100 - 300 ms		
Condition	2	1.57	.174	.05
Region	1	3.48	.073	.11
Hemisphere	1	0.00	.973	<.01
Group	2	0.67	.514	.02
Condition x Region	2	31.48	<.001**	.54
Condition x Hemisphere	2	7.07	.002**	.21
Condition x Group	4	0.53	.716	.02
Region x Hemisphere	1	0.13	.716	<0.1
Region x Group	2	1.08	.347	.04
Hemisphere x Group	2	0.16	.813	.01
Condition x Region x Hemisphere	2	0.37	.695	.01
Condition x Region x Group	4	0.80	.494	.03
Condition x Hemisphere x Group	4	1.07	.373	.04
Region x Hemisphere x Group	2	0.06	.899	<.01

Table A2 (continued)

Effect	df	*F*	*p*	p^2
		300 - 400 ms		
Condition	2	1.52	.228	.05
Region	1	2.57	.120	.09
Hemisphere	1	0.18	.678	.01
Group	2	0.59	.559	.02
Condition x Region	2	23.61	<.001***	.47
Condition x Hemisphere	2	8.45	.001**	.24
Condition x Group	4	1.94	.108	.07
Region x Hemisphere	1	2.98	.096	.10
Region x Group	2	1.26	.292	.04
Hemisphere x Group	2	0.85	.410	.03
Condition x Region x Hemisphere	2	0.12	.886	<.01
Condition x Region x Group	4	1.50	.219	.05
Condition x Hemisphere x Group	4	1.62	.174	.06
Region x Hemisphere x Group	2	0.93	.399	.03
Condition x Region x Hemisphere x Group	4	1.10	.361	.04

License

Comparing language input in homes of young blind and sighted children: Insights from daylong recordings

Erin Campbell
Department of Psychology & Neuroscience, Duke University, Durham, NC, USA
Department of Education & Human Development Boston University, Boston, MA, USA

Lillianna Righter
Eugenia Lukin
Elika Bergelson
Department of Psychology & Neuroscience, Duke University, Durham, NC, USA
Department of Psychology, Harvard University, Cambridge, MA, USA

Abstract: We compared everyday language input to young congenitally-blind children with no additional disabilities (N=15, 6–30 mo., M:16 mo.) and demographically-matched sighted peers (N=15, 6–31 mo., M:16 mo.). By studying whether the language input of blind children differs from their sighted peers, we aimed to determine whether, in principle, the language acquisition patterns observed in blind and sighted children could be explained by aspects of the speech they hear. Children wore LENA recorders to capture the auditory language environment in their homes. Speech in these recordings was then analyzed with a mix of automated and manually-transcribed measures across various subsets and dimensions of language input. These included measures of quantity (adult words), interaction (conversational turns and child-directed speech), linguistic properties (lexical diversity and mean length of utterance), and conceptual features (talk centered around the here-and-now; talk focused on visual referents that would be inaccessible to the blind but not sighted children). Overall, we found broad similarity across groups in speech quantitative, interactive, and linguistic properties. The only exception was that blind children's language environments contained slightly but significantly more talk about past/future/hypothetical events than sighted children's input; both groups received equivalent quantities of "visual" speech input. The findings challenge the notion that blind children's language input diverges substantially from sighted children's; while the input is highly variable across children, it is not systematically so across groups, across nearly all measures. The findings suggest instead that blind children and sighted children alike receive input that readily supports their language development, with open questions remaining regarding how this input may be differentially leveraged by language learners in early childhood.

Keywords: blindness; language input; child-directed speech; daylong audio recordings; vision

Corresponding author(s): Erin Campbell, Wheelock Deaf Center, Department of Education and Human Development, Boston University, Boston, MA, 02215, USA. Email: eecamp@bu.edu.

ORCID ID(s): Erin Campbell: https://orcid.org/0000-0002-8770-2317; Lillianna Righter: https://orcid.org/0009-0008-8732-4441; Eugenia Lukin: https://orcid.org/0009-0001-0387-4112; Elika Bergelson: https://orcid.org/0000-0003-2742-4797

Citation: Campbell, E. E., Righter, L., Lukin, E., Bergelson, E. (2025). Comparing language input in homes of young blind and sighted children: Insights from daylong recordings. *Language Development Research*, *5*(3), 33–71. http://doi.org/10.34842/ldr2025-846

Introduction

The early language skills of blind children are highly variable. Some children demonstrate age-appropriate vocabulary and grammar from the earliest stages of language learning, while others experience substantial language delays (Bigelow, 1987; E. E. Campbell et al., 2024; Landau & Gleitman, 1985). By adulthood, however, blind individuals are fluent language-users, even demonstrating faster lexical processing skills than sighted adults (Loiotile et al., 2020; Röder et al., 2003; Röder et al., 2000; though cf., Sak-Wernicka, 2017 for discussion of possible pragmatic differences). The causes of early variability and the potential ability (or need) to "catch up" remain poorly understood: what could make the language learning problem different or initially more difficult for the blind child? Here, we compare the language environments of blind children to that of their sighted peers. In doing so, we begin to untangle the role that perceptual input plays in shaping children's language environment and better understand the interlocking factors that may contribute to variability in blind children's early language abilities.

Why Would Input Matter?

Among both typically-developing children and children with developmental differences, language input has been found to predict variability in language outcomes (Anderson et al., 2021; Gilkerson et al., 2018; Huttenlocher et al., 1991, 2010; Rowe, 2008, 2012). At a coarse level, children who are exposed to more speech (or sign, Watkins, Pittman, & Walden, 1998) tend to have stronger language outcomes and produce more speech themselves (Anderson et al., 2021; Bergelson et al., 2023; Gilkerson et al., 2018; Huttenlocher et al., 1991; Rowe, 2008).

Previous research suggests that the structure and content of the language input (often referred to as input "quality")[1] is even more influential than the amount of speech alone (Hirsh-Pasek et al., 2015; Rowe, 2012). Rowe and Snow (2020) categorized the makeup of the input along three dimensions: interactive features (e.g., parent responsiveness, speech directed *to* child vs. overheard, conversational turn-taking), linguistic features (e.g., lexical diversity, grammatical complexity), and conceptual features (i.e., the extent to which input focuses on the *here-and-now*).

In examining interactive features, previous studies have indicated that back-and-forth communicative exchanges (also known as conversational turns) between caregivers and children are predictive of better language outcomes across infancy (Donnellan et al., 2020; Goldstein & Schwade, 2008) and toddlerhood (Hirsh-Pasek et al.,

[1] We avoid the term "quality" here as it carries potential biases regarding linguistic norms (MacLeod & Demers, 2023).

2015; Romeo et al., 2018). Another way to quantify caregiver and infant interaction is by looking at how much speech is directed *to* the child (e.g. as opposed to an overheard conversation between adults). The amount of child-directed speech in children's input (at least in Western contexts, Casillas et al., 2020) has been linked to children's vocabulary size and lexical processing (Rowe, 2008; Shneidman et al., 2013; Weisleder & Fernald, 2013).

Under the linguistic umbrella, we can measure the kinds of words used (often measured as lexical diversity, type-token ratio), and the ways they are combined (syntactic complexity, often measured by mean length of utterance). Both parameters have been found to correlate with children's language growth: sighted toddlers who are exposed to a greater diversity of words in their language input are reported to have larger vocabulary scores (N. J. Anderson et al., 2021; Hsu et al., 2017; Huttenlocher et al., 2010; Rowe, 2012; Weizman & Snow, 2001). Likewise, the diversity and complexity of syntactic constructions in parental language input has been associated with both children's vocabulary growth and structural diversity in their own productions (de Villiers, 1985; Hadley et al., 2017; Hoff, 2003; Huttenlocher et al., 2002, 2010; Naigles & Hoff-Ginsberg, 1998).

Finally, the conceptual dimension of language input aims to capture the extent to which the language signal maps onto present objects and ongoing events in children's environments (Rowe & Snow, 2020). As children develop, their ability to represent abstract referents improves (Bergelson & Swingley, 2013; Kramer et al., 1975; Luchkina et al., 2020). Decontextualized language input—that is, talking about past, future, or hypothetical events, or people and items that are not currently present in the environment—may be one contributing factor (Rowe, 2013). Greater prevalence of decontextualized language in input to toddlers has been found to predict aspects of children's own language in kindergarten and beyond (Demir et al. 2015; Rowe, 2012; Uccelli et al., 2019).

From this (necessarily abridged) review, it appears that many factors in the language input alone link to how sighted children learn about the world and language, but that children also learn from sensory, conceptual, and social knowledge. Many cues for word learning are visual: for example, empirical work finds that sighted children can leverage visual information like parental gaze, shared visual attention (Tomasello & Farrar, 1986), pointing (Lucca & Wilbourn, 2018), and the presence of salient objects in the visual field (Yu & Smith, 2012). Because these visual cues are inaccessible to blind children, language input may take on a larger role in the discovery of word meaning (E. E. Campbell & Bergelson, 2022). Syntactic structure, in particular, provides critical cues to word meaning, such as the relationship between two entities that aren't within reach, or are intrinsically unobservable or ambiguous (Gleitman, 1990). But in order to evaluate whether language input plays a larger role for blind versus sighted children's learning, it is worth first establishing whether blind and sighted children's language input differs. That is, children with different sensory access could differentially make use of the same kind of language input, or they could apply the same learning mechanisms to input with different properties– a debate carried over from work with typically-sighted children (Newport et al., 1977). Either way, characterizing the input across potentially relevant dimensions is a helpful first step.

Why would the input differ between blind and sighted children?

Speakers regularly tailor their speech to communicate efficiently with the listener (Grice, 1975). Across many contexts, research finds that parents are sensitive to their child's developmental level and tune language input accordingly (Newport et al., 1977; Snow, 1972; Vygotsky, 1978). One example is child-directed speech, wherein parents speak to young children with exaggerated prosody and slower speech (Bernstein Ratner, 1984; Fernald, 1989; Moser et al., 2022; Newport et al., 1977), which are in some cases helpful to the young language learner (Thiessen et al., 2005). For instance, parents tend to repeat words more often when interacting with infants than with older children or adults (Fernald & Morikawa, 1993; Snow, 1972). Communicative tailoring is also common in language input to children with disabilities, who have been found to receive simplified, more directive language input, and less interactive input compared to typically-developing children (Dirks et al., 2020; Yoshinaga-Itano et al., 2020). In other contexts, language input to children with disabilities has been shown to be more multimodal, such that parents more frequently combine communicative cues (e.g., speech and touch, Abu-Zhaya et al., 2019) when interacting with deaf children, compared to their typically-hearing peers.

In addition to tailoring communication to children's developmental level, speakers also adjust their conversation in accordance to their conversational partner's sensory access (Gergle et al., 2004 for adults; and Grigoroglou et al., 2016 for adults and 4–6-year-old children). For example, in a noisy environment, adults will often adapt the acoustic-phonetic features of their speech to make it easier for their interlocutor to understand them (Hazan & Baker, 2011), demonstrating sensitivity to even temporary sensory conditions. When describing scenes, adult speakers tend to provide the information their listeners lack but seem to avoid redundant visual description (Grice, 1975; Ostarek et al., 2019). During in-lab tasks with sighted participants, participants in several studies verbally provide visually-absent cues when an object is occluded to their partner (Hawkins et al., 2021; Jara-Ettinger & Rubio-Fernandez, 2021; Rubio-Fernandez, 2019). These results suggest that adults (Gergle et al., 2004; Hazan & Baker, 2011), children (e.g., Grigoroglou et al., 2016), and even infants (Chiesa et al., 2015; Ganea et al., 2018; Senju et al., 2013) can flexibly adapt communication to the visual and auditory abilities of their partner.

Taking these results into consideration, and given the strong verbal abilities of blind adults (Loiotile et al., 2020; Röder et al., 2000, 2003), we might expect parents of blind children to verbally describe visual information in the child's environment or otherwise structure interactions to align with their child's strengths and abilities. But prior research doesn't yield a clear answer on whether sighted parents modify language input to blind children. Several studies suggest differences in the conceptual features: caregivers of blind children restrict conversation to things that the blind child is currently engaged with, rather than attempt to redirect their attention to other stimuli (Andersen et al., 1993; J. Campbell, 2003; Kekelis & Andersen, 1984; though cf., Moore & McConachie, 1994). Studies of naturalistic input to blind children report that parents use *fewer* declaratives and *more* imperatives than parents of sighted children, suggesting that blind children might be receiving less description than sighted children (Kekelis & Andersen, 1984; Landau & Gleitman, 1985; though cf., Lukin et al., 2023; Pérez-Pereira & Conti-Ramsden, 2001). Other studies report that parents adapt their interactions to their children's visual abilities, albeit in specific contexts. Tadić, Pring, and Dale (2013) find that in a structured book-reading task, parents of blind children provide more descriptive utterances than parents of sighted children. Further, parents of blind children have been found to provide more tactile cues to initiate interactions or establish joint attention (Preisler, 1991; Urwin, 1983, 1984), which may serve the same social role as shared gaze in sighted children and take advantage of children's access to other senses (e.g., touch). These mixed results suggest that parents of blind children might alter language input in some domains but not others. The apparent conflict in results may be exacerbated by the difficulty of recruiting specialized populations to participate in research: the small (in most cases, single-digit) sample sizes of prior work limit our ability to generalize about any differences in the input to blind vs. sighted infants.

The Present Study

Children can and do learn language in a variety of input scenarios (Gleitman & Newport, 1995), but if language input differs systematically between blind and sighted infants and toddlers, capturing this variation may reveal a more nuanced picture of how infants use the input to learn language. In the present study, we examine daylong recordings of the naturalistic language environments of blind and sighted children in order to characterize the input to each group. Using both automated measures and manual transcription of these recordings, we analyze several characteristics that have been previously suggested to be information-rich learning cues, including overall amount of environmental language (adult word count), interaction (conversational turn count, proportion of child-directed speech), conceptual features (temporal displacement, sensory modality), and linguistic complexity (type-token ratio and mean length of utterance). Though the present study is largely exploratory, we took the directionality of previously reported results as our (admittedly limited) starting point. Thus, based on prior research, we made the tentative predictions that blind vs. sighted children would have input featuring less interactivity (fewer conversational turns and less child-directed speech; Rowland, 1984; Grumi et al., 2021), less linguistic complexity (Bulk et al., 2020; lower type-token ratio and shorter utterances, Chernyak, n.d.; Dirks et al., 2020; FamilyConnect, n.d.; Lorang et al., 2020), and conceptual content focused more on the child's locus of attention (more here-and-now speech and fewer visual words, Andersen et al., 1993; J. Campbell, 2003; Kekelis & Andersen, 1984); we have no a priori hypotheses regarding adult word count.

Method

Participants

This study included 15 congenitally-blind infants and their families[2]. To be eligible, participants had to be 6–30 months old, have severe to profound visual impairment (i.e. at most light perception), no additional disabilities (developmental delays, intellectual disabilities, or hearing loss), and be exposed to ≥ 75% English at home. Blind participants were recruited through ophthalmologist referral, preschools, early intervention programs, social media, and word of mouth. Blindness in our sample was caused by a range of conditions, including cataracts (n=3), Leber's Congenital Amaurosis (n=1), Microphthalmia (n=2), Ocular albinism (n=2), Optic Nerve Hypoplasia (n=2), Retinal Detachments (n=1), and Retinopathy of Prematurity (n=1). Etiology was unknown in 2 participants, and 2 participants had multiple contributing conditions. Caregivers were also asked to complete a demographics survey and the MacArthur-Bates Communicative Development Inventory (CDI, Fenson et al., 1994) within one week of the home language recording.

To control for the wide age range of the study, each blind participant was matched to a sighted participant, based on age (± 6 weeks), sex, maternal education (± one education level), and number of siblings (± 1 sibling). Sighted matches were drawn from multiple existing corpora: two children from VanDam et al. (2015) and VanDam et al. (2016); five children from Bergelson (2015) and Bergelson et al. (2019); one child from Ramírez-Esparza et al. (2014); two children from Warlaumont et al. (2016); two from Wang et al. (2022); and two from Rowland et al. (2018)[3]. There was no recording available that matched two blind participants' demographic characteristics; we therefore collected recordings from two sighted children *de novo*. See Table 1 for sample demographic characteristics.

[2] One family contributed two recordings for the same blind child. In the present study, we used only the first recording from that participant.

[3] These two sighted children are from the UK, the rest from North America. While recognizing this potential limitation, we have no a priori reason to predict that North American and UK English learners should differ meaningfully in our language measures, especially given our broader demographic matching procedure.

Table 1. ***Demographic characteristics of the blind and sighted samples. For continuous variables, range and mean are provided. For categorical variables, percentages by level are provided.***

Variable	Blind (N = 15)	Sighted (N=15)
Age (months)	6–30, 15.8 (8.2)	6–32, 16.1 (8.1)
Sex	Female: 44% Male: 56%	Female: 44% Male: 56%
Number of Older Siblings	0–2, 0.5 (0.8)	0–3, 1.1 (1)
Maternal Education	Some college: 19% Associate's degree: 6% Bachelor's degree: 31% Graduate degree: 44%	Some college: 6% Associate's degree: 12% Bachelor's degree: 56% Graduate degree: 6%
Race	American Indian or Alaska Native: 6% Black or African American: 6% Multiracial: 19% White: 69% Unknown: 0%	American Indian or Alaska Native: 0% Black or African American: 6% Multiracial: 6% White: 56% Unknown: 31%
Ethnicity	Hispanic or Latino: 19% Not Hispanic or Latino: 81% Unknown: 0%	Hispanic or Latino: 0% Not Hispanic or Latino: 62% Unknown: 38%

Recording Procedure:

For the recording portion of the study, caregivers of participating infants received a LENA wearable audio recorder and vest (Ganek & Eriks-Brophy, 2016; Gilkerson & Richards, 2008). They were instructed to place the recorder in the vest on the day of their scheduled recording and put the vest on their child from the time they woke up until the recorder automatically shut off after 16 hours (setting the vest nearby during baths, naps, and car rides). Actual recording length ranged from 8 hours 17 minutes to 15 hours 59 minutes (Mean: 15 hours 6 minutes).

Processing:

The audio recordings were first processed by the LENA proprietary software (Xu et al., 2009), creating algorithmic measures such as conversational turn count and adult word count. Each recording was then run through an in-house automated sampler that selected 15- non-overlapping 5-minute segments, randomly distributed across the duration of the recording. Each segment consists of 2 core minutes of annotated time, with 2 minutes of listenable context preceding the annotation clip and 1 minute of additional context following. Because these segments were sampled randomly, across participants roughly 27% of the random 2-minute coding segments contained no speech at all. For questions of *how much does a phenomenon occur*, random sampling schemes can help avoid overestimating speech in the input, but for questions of input *content*, randomly selected samples may be too sparse (Pisani et al., 2021).

Therefore, we chose to annotate 5 additional (non-overlapping) 2-minute segments specifically for their high density of speech. To select these segments of dense talk, we first conducted an automated analysis of the audio file using the voice type classifier for child-centered daylong recordings (Lavechin et al., 2021) which identified segments likely containing human speech. The entire recording was divided into 2-minute chunks, each ranked highest to lowest by the total duration of the speech segments contained within the chunk. We annotated the 5 highest-ranked segments of each recording. These high-volubility segments allow us to more closely compare our findings to studies classifying the input during structured play sessions, which paint a denser and differently-proportioned makeup of the language input (Bergelson et al., 2019). In sum, 30 minutes of randomly-sampled input and 10 minutes of high-volubility input (40 minutes total) were annotated per child.

Annotation:

Manual annotation of the selected segments was conducted using the ELAN software (Brugman & Russel, 2009). Trained annotators listened through each 2-minute segment plus its surrounding context and coded it using the ACLEW annotation scheme (Soderstrom et al., 2021). For more information about this scheme, see the ACLEW homepage. Speech by people other than the target child was transcribed using an adapted version of the CHAT transcription style (MacWhinney, 2019; Soderstrom et al., 2021). Because the majority of target children in the project are pre-lexical, utterances (e.g. babble) produced by the target child are not yet transcribed. Speech was then further classified by the addressee of each utterance: child, adult, both an adult and a child, pets or other animals, unclear addressee, or a recipient that doesn't fit into another category (e.g., voice control of Siri or Alexa, prayer to a metaphysical entity).

Manual Annotation Training and Reliability. All annotators are tested on the ACLEW scheme prior to beginning corpus annotation, until they reach 95% agreement or better with a "gold standard" coder for segmentation and utterance classification. Training often takes upwards of 20 hours of annotation practice. Following the first pass by annotators, all files were reviewed by a highly-trained "superchecker" to ensure consistency between coders and check for errors. Over a span of three years, 15 trained annotators contributed to this dataset. Ten percent of clips were re-transcribed to assess reliability; further reliability data are provided in corresponding sections below.

Extracting Measures of Language Input:

To go from our dimensions of interest (word count, interactiveness, linguistic, conceptual), to quantifiable properties, we used a combination of automated measures (generated by the proprietary LENA algorithm, Xu et al., 2009) and manual measures (generated from the transcriptions and classifications made by our trained annotators). Altogether, this corpus presently includes approximately 453 hours of audio, 15994 utterances, and 63665 words. LENA measures were calculated over the whole day, and then normalized by recording length. Transcription-based word count and interactiveness analyses were conducted on the random samples only, to capture a more representative estimate. Linguistic and conceptual analyses were conducted on all available annotations to maximize the amount of speech over which we could calculate them. These measures are described below and summarized in Table 2.

Quantity.

Automated Word Count. To derive this count, the LENA algorithm segments the recording into clips which are then classified by speaker's perceived gender (male/female), age (child/adult), and distance (near/far), as well as several non-human speaker categories (e.g., silence, electronic noise). Only segments that are classified as nearby male or female adult speech are then used by the algorithm for its subsequent Adult Word Count (AWC) estimation (Xu et al., 2009). Validation work suggests that this automated count correlates strongly with word counts derived from manual annotations (Cristia et al., 2020; r = .71 – .92, Lehet et al., 2021), and meta-analytic work finds that AWC is associated with children's language outcomes across developmental contexts (e.g., autism, hearing loss, Wang et al., 2020). Because the recordings varied in length (8 hours 17 minutes to 15 hours 59 minutes), we normalized AWC by dividing by recording length[4].

Manual Word Count. We also calculated a manual count of speech in the children's environment. Manual Word Count (MWC) is simply the number of intelligible words in our transcriptions of each child's recording. Speech that was too far or muffled to be intelligible, as well as speech from the target child and electronic speech (TV, radio, toys) are excluded from this count. Unlike LENA's AWC, MWC contains speech from other child speakers in the environment (e.g., siblings), not just from adults.

By using automated *and* manual word count, we hope to capture complementary

[4] To make these measures more comparable, we present both the Automated Word Count and the Manual Word Count in terms of words per hour.

estimates of the amount of speech children are exposed to. While AWC is considered less accurate than manual annotation, it is commonly used due to its ability to readily provide an estimate of the adult speech across the whole day. MWC, because it comes from human annotations, is the gold-standard for accurate speech estimates, but due to feasibility, is only derived from 30 minutes of the recording (sampled in 2-minute clips, at random, as described above).

Interaction.

Conversational Turn Count. One common metric of communicative interaction (e.g., Ganek & Eriks-Brophy, 2018; Magimairaj et al., 2022) is conversational turn count (or CTC), an automated measure generated by LENA (Xu et al., 2009). Like AWC, a recent meta-analysis finds that CTC is associated with children's language outcomes (Wang et al., 2020). After tagging vocalizations for speaker identity, the LENA algorithm looks for alternations between adult and target child speech in close temporal proximity (within 5 seconds). This can erroneously include non-contingent interactions (e.g., mom talking to dad while the infant babbles to herself nearby), and therefore inflate the count especially for younger ages and in houses with multiple children (Ferjan Ramírez et al., 2021). Still, this measure correlates moderately well with manually-coded conversational turns (*r*s=0.28-0.75, Busch et al., 2018; Ferjan Ramírez et al., 2021; Ganek & Eriks-Brophy, 2018), and because participants in our sample are matched on both age and number of siblings, CTC overestimation should not be biased towards either group.

Proportion of Child-Directed Speech. Our other measure of interaction is the proportion of utterances that are child-directed, derived from the manual annotations. Each proportion was calculated as the number of utterances (produced by someone other than the target child) tagged with a child as the addressee, out of the total number of utterances. Annotator agreement for addressee was 93%, with a kappa of 0.90 [CI: 0.89–0.91].

Linguistic Features.

Type Token Ratio. As in previous work (e.g., Montag et al., 2018; Pancsofar & Vernon-Feagans, 2006; Templin, 1957), we calculated the lexical diversity of the input by dividing the number of unique words by the total number of words (i.e., the type-token ratio). Because the type-token ratio changes as a function of the number of words in a sample (Montag et al., 2018; Richards, 1987), we first standardized the size of the sample by cutting the manual annotations in each recording into 100-word bins. We then calculated the type-token ratio within each of these bins by dividing the number of unique words in each bin by the number of total words (~100) and then averaged the type-token ratio across bins for each child[5]. This provided a measure of lexical diversity: per 100 words, how many unique words are children exposed to?

MLU. We also analyzed the syntactic complexity of children's language input, approximated as mean utterance length in morphemes. Each utterance in a child's input was tokenized into morphemes using the 'morphemepiece' R package (Bratt & Harmon, 2022). We then calculated the mean length of utterance (number of morphemes) in each audio recording. We manually checked utterance length in a random subset of 10% of the utterances (n = 2826 utterances), which yielded an intra-class correlation coefficient of 0.94 agreement with the morphemepiece approach (CI: 0.94–0.95, $p < .001$), indicating high consistency.

Conceptual Features. Our analysis of the conceptual features aims to measure the extent to which language input centers around the *"here and now"*: things that are currently present or occurring that a child may attend to in real time. We approximate *here-and-now*ness using lexical and morphosyntactic properties of the input.

Proportion of temporally displaced verbs. We examined the displacement of *events* (focusing on the "now" aspect of here-and-now) discussed in children's linguistic environment, via properties of the verbs in their input. We are attempting to highlight semantic features of the language environment with a morphosyntactic proxy. We do so here by categorizing utterances based on the syntactic and morphological features of verbs, since these contain some time information in their surface forms. We assigned each utterance a temporality value: utterances tagged "displaced" describe events that take place in the past, future, or irrealis space, while utterances tagged "present" describe current, ongoing events. This coding scheme roughly aligns with both the temporal displacement and future hypothetical categories in Grimminger et al. (2020; see also: Hudson, 2002; Lucariello & Nelson, 1987). That is, for this event temporality-based measure, rather than focusing on whether any of the noun referents in an utterance are present or attended to by the child, we focus on whether the events concerning them are presently occurring and salient.

To do this, we used the udpipe package (Wijffels, 2023) to tag the transcriptions with parts of speech and other lexical features, such as tense, number agreement, or case inflection. To be marked as present, a verb either had to be marked with both present tense and indicative mood or appear in the gerund form with no marked tense (e.g. 'you talking to Papa?'). Features that could mark an utterance as displaced included past tense, presence of a modal, presence of 'if', or presence of 'gonna'/'going to',

[5] Computing TTR over the entire sample instead of averaging over 100-word bins rendered the same pattern of results.

'have to', 'wanna'/'want to', or 'gotta'/'got to'[6], since these typically indicate future events, belief states and desires, rather than real-time events. In the case of utterances with multiple verbs, we selected the features from the first verb or auxiliary, as a proxy for hierarchical dominance. Utterances without verbs were excluded. A small number of verb-containing utterances in our corpus were left "ambiguous" (n = 1440/8930), either because they were fragments or because the automated parser failed to tag any of the relevant features. We manually checked verb temporality in a random subset of 10% of the utterances (n = 825). Notably, we did not simply verify whether the tagger accurately identified tense and aspect. Rather, human coders holistically tagged the utterance as decontextualized or not, factoring in meaning, context, and syntax, providing a stronger test of reliability against the tagger's verb-tense based assessment. Human judgments of event temporality aligned with the automated tense tagger 76% of the time (kappa = 0.56, CI: 0.56-0.62, p = .050), indicating substantial agreement, with the majority of discrepancies occurring on utterances the tagger categorized as ambiguous.

Proportion of highly visual words. In addition to this general measure of decontextualized language, we include one measure that is uniquely decontextualized for blind children: the proportion of words in the input with referents that are highly and exclusively visual. We first filter the input to only content words (excluding, for example: *the, at, of*). We then categorize the perceptual modalities of words' referents using the Lancaster Sensorimotor Norms, which are ratings from sighted adults noting the extent to which a word evokes a sensory experience in a given modality (Lynott et al., 2020). Each of the approximately 40,000 words in the Lancaster Sensorimotor Norms gets a score for each of 6 sensory modalities (auditory, haptic, gustatory, interoceptive, olfactory, visual). In this rating system, words with higher ratings in a given modality are more strongly associated with perceptual experience in that modality, and a word's dominant perceptual modality is the modality which received the highest mean rating. We tweak this categorization in two ways: we categorized content words that received relatively low ratings across all modalities (<3.5./5) as predominantly *amodal*, and content words whose ratings were distributed across modalities were categorized as *multimodal*[7]. Using this system, each of the content words in children's input were categorized into their primary perceptual modality;

[6] Only the "-to" forms of these verbs are pulled specifically into the "displaced" category, because they specifically select phrasal complements. Sentences like "I want that ball" are treated as having a separate verb than "wanna;" in this case the utterance would be tagged as present tense and put into the "present" category since it is grounded in present objects and events.
[7] Words with perceptual exclusivity scores < 0.5 (calculated as a word's range of ratings across modalities divided by the sum of ratings across modalities, Lynott et al., 2020) were re-categorized as multimodal. The cut-offs for classifying amodal and multimodal words were chosen based on authors' intuitions regarding what thresholds seemed to classify the words well into amodal, multimodal, and visual phenomena. That said, results are robust across a range of thresholds, and all data are provided to interested readers should they be interested in considering other values.

76% of the words in our corpus had a corresponding word in the Lancaster ratings and could be categorized in this way. For each child, we extracted the proportion of exclusively *visual* words in their home speech sample. Examples of visual words include: “blueprint”, “see”, “color”, “sky”, “pictures”, “lighting”, “moon”, “glowing”.

Results

Comparing Properties of Language Input

Our study assesses whether language input to blind children is different from the language input to sighted children, along the dimensions of word count, interaction, linguistic, and conceptual properties. We test for group differences using paired t-tests or non-parametric Wilcoxon signed rank tests, when a Shapiro-Wilks test indicates that the variable is not normally distributed (summarized in Table 2). Because this analysis involves multiple tests against the null hypothesis (*that there is no difference in the language input to blind vs. sighted kids*), we use the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) to control false discovery rate (Q = .05) for each set of analyses (word count, interaction, linguistic, conceptual). Because each dimension’s analysis consists of two statistical tests, our Benjamini-Hochberg critical values were $p < 0.025$ for the smaller p value and $p < 0.05$ for the larger p value. The results are summarized in Table 2: how each measure was calculated; what portion of the recording the measure was calculated over; whether a parametric or non-parametric test was used; the mean, median, and range for blind and sighted children, and the raw (uncorrected) p-value of the test comparing groups. Only the proportion of displaced verbs reached significance at our corrected $p < .025$ threshold for significance.

Table 2. ***Summary of language input variables.***

Variable	Description	Portion of Recording	Test	Blind Mean, Median, Range	Sighted Mean, Median, Range	*p* value
Adult Word Count	Estimated number of words in recording categorized as nearby adult speech by LENA algorithm.	Whole day	t-test	2124, 1808, 779–3968 words/hour	2117, 2047, 951–3216 words/hour	.984
Manual Word Count	Number of word tokens from speakers other than target child.	Random	t-test	3994, 3504, 1208–7288 words/hour	4598, 4296, 780–8668 words/hour	.307

Conversational Turn Count	Count of temporally close switches between adult and target-child vocalizations, divided by recording length.	Whole day	Wilcoxon test	66, 49, 26–180 turns/hour	71, 65, 18-–69 turns/hour	.811
Prop. Child-Directed Speech	Number of utterances tagged with child addressee out of total number of utterances, from speakers other than target child.	Random	t-test	0.55, 0.60, 0.19–1	0.57, 0.57, 0.09–0.95	.978
Type-Token Ratio	Average of the type-token ratios (number of unique words divided by number of total words) for each of the 100-word bins in their sample.	Random + High-Vol	t-test	0.64, 0.65, 0.58–0.67	0.63, 0.63, 0.54–0.69	.353
Mean Length of Utterance	Average number of morphemes per utterance	Random + High-Vol	t-test	5.53, 5.28, 4.13–7.71 morphemes	4.97, 5.11, 4.09–5.87 morphemes	.063
Prop. Displaced Verbs	Proportion of verbs that refer to past, future, or hypothetical events	Random + High-Vol	t-test	0.34, 0.33, 0.24–0.43	0.29, 0.3, 0.13–0.39	.018*
Prop. Visual	Proportion of words in the input with high visual association ratings and low ratings for other perceptual modalities	Random + High-Vol	Wilcoxon test	0.1, 0.08, 0.04–0.21	0.11, 0.1, 0.06–0.22	.421

Overall Quantity. We first compared the language input to blind and sighted children using two measures of the number of words in their environment: LENA's automated Adult Word Count and our transcription-derived Manual Word Count. Despite wide variability in the number of words children hear (Range from Manual Word Count: 604–3644 $words_{blind}$, 390–4334 $words_{sighted}$ per hour), along both word count measures, blind and sighted children did not differ (Adult Word Count: $t(14) = -0.02$, $p = .984$; Manual Word Count: $t(14) = 1.06$, $p = .307$); see Figure 1.

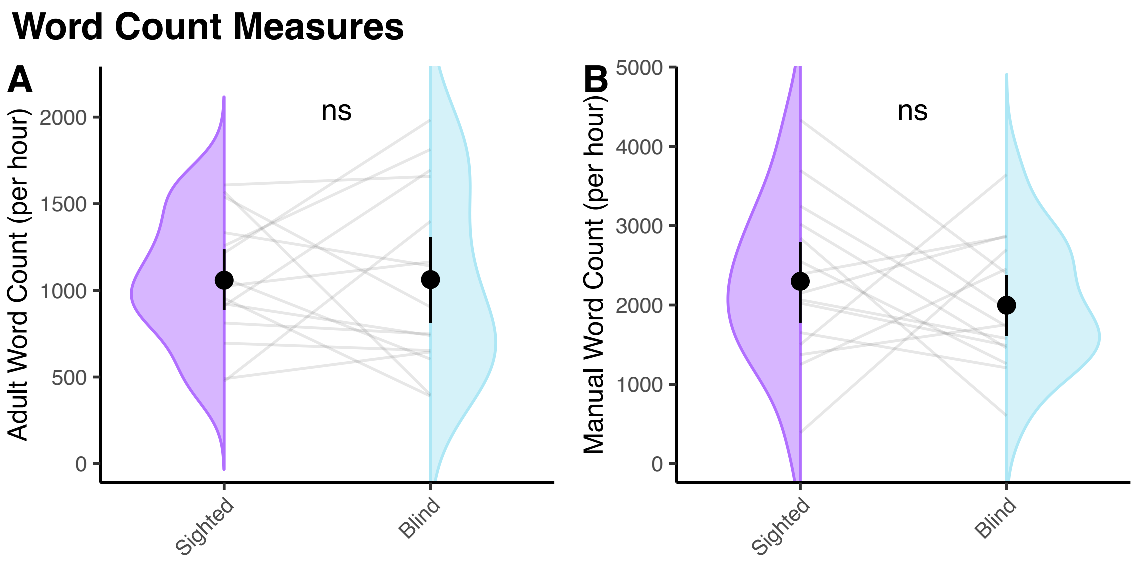

Figure 1. ***Comparing LENA-generated adult word counts (left) and transcription-based word counts in the input of blind and sighted children. Violin density represents the distribution of word counts for each group. Grey lines connect values from matched participants. Black dot and whiskers show standard error around the mean. Neither measure differed between groups.***

Interaction. Our corpus also revealed no significant difference in the amount of interaction with the child, measured as the proportion of child-directed speech ($t(14) = 0.24$, $p = .811$) or in conversational turn counts to blind children versus to sighted children ($W = 61$, $p = .978$). Across both groups, child-directed speech constituted approximately 56% of the input, and children were involved in an estimated 34 conversational turns per hour (based on the LENA automated metric); see Figure 2.

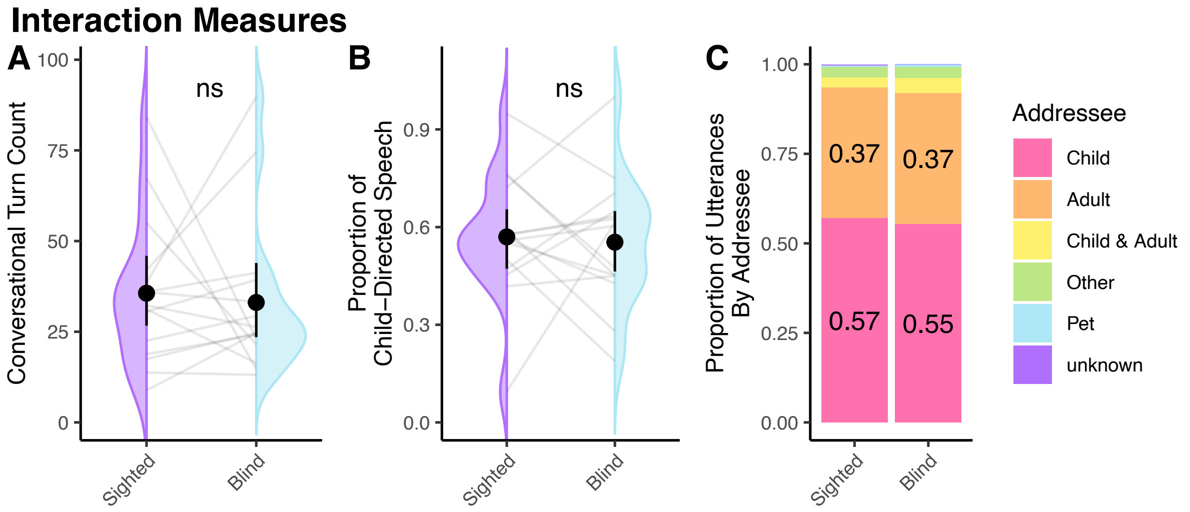

Figure 2. ***Comparing LENA-generated conversational turn counts (left) and proportion of utterances in child-directed speech (center). Violin density represents the distribution of values for each group. Grey lines connect values from matched participants. Black dot and whiskers show standard error around the mean. The full breakdown by addressee is shown in the rightmost panel. Neither conversational turn count nor proportion of child-directed speech differed between groups.***

Linguistic Features. Similarly, neither linguistic variable differed across groups: blind and sighted children's input had comparable type-token ratios ($t(14) = -0.96$, $p = .353$) and utterance lengths ($t(14) = -2.02$, $p = .063$). Children in our samples heard on average 64 unique words per hundred words and 5.20 morphemes per utterance; see Figure 3.

Conceptual Features. Lastly, we compared two measures of the conceptual features of language input: the proportion of temporally displaced verbs and the proportion of highly visual words; see Figure 4. We found that blind children heard a higher proportion of displaced verbs than sighted children ($t(14) = -2.68$, $p = .018$), which on average equates to 22 more utterances about past, future, or hypothetical events per hour. We found no significant difference across groups in the proportion of highly visual words[8] (W = 75, $p = .421$), which constituted roughly 10% of the input for both groups.

[8] And similarly, there were no significant group differences in the proportions of auditory words, tactile words, or non-visual-but-still-perceptual words.

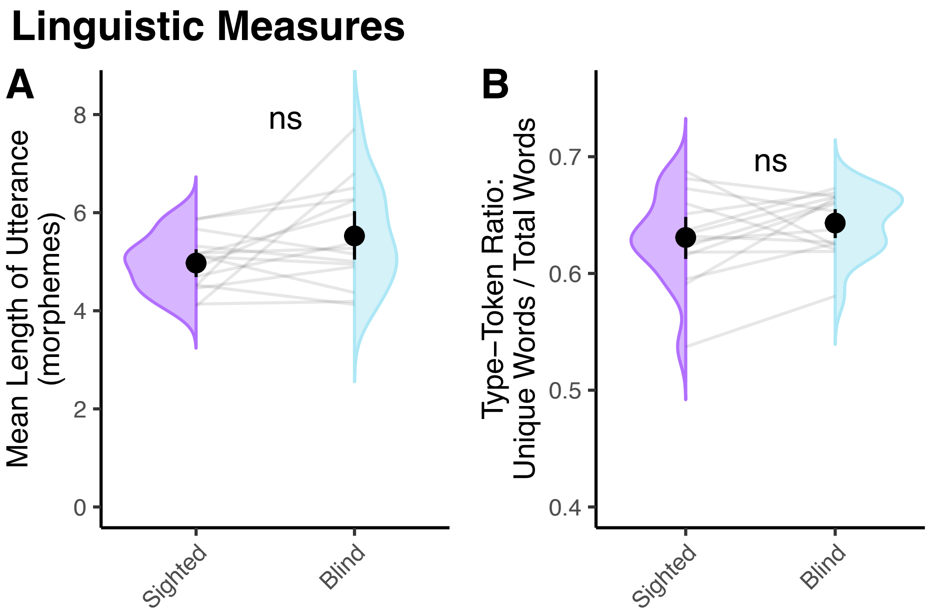

Figure 3. ***Comparing linguistic features: Mean length of utterance (left) and type-token ratio (right). Violin density represents the distribution of values for each group. Grey lines connect values from matched participants. Black dot and whiskers show standard error around the mean. Utterances in blind children's input were significantly longer, and type- token ratio was significantly higher. Note that the y-axis on the type-token ratio plot has been truncated.***

Evidence of Absence? To explore the extent to which any observed lack of difference could be interpreted as equivalence – *that blind and sighted children's input did not differ*–, we also conducted equivalence tests for variables that did not differ significantly across groups. Thus, for adult word count, manual word count, conversational turn count, proportion of child-directed speech, type-token ratio, MLU, and proportion of visual words, we conducted two one-sided equivalence tests (Lakens, 2017) against a small, moderate, and large effect sizes (Cohen's $|d| < 0.3$, $|d| < 0.5$, $d < 0.7$, respectively). Given our relatively small sample, for all but the largest effect sizes tested, results were inconclusive, i.e. it remains possible there are small to moderate differences in the input across the blind and sighted groups. For adult word count, conversational turn count, proportion of child-directed speech, and proportion of highly visual words, we found evidence for equivalence (i.e. a significant equivalence test) when the Cohen's D threshold is set at $|0.7|$. Full equivalence test results are available in the Supplementals.

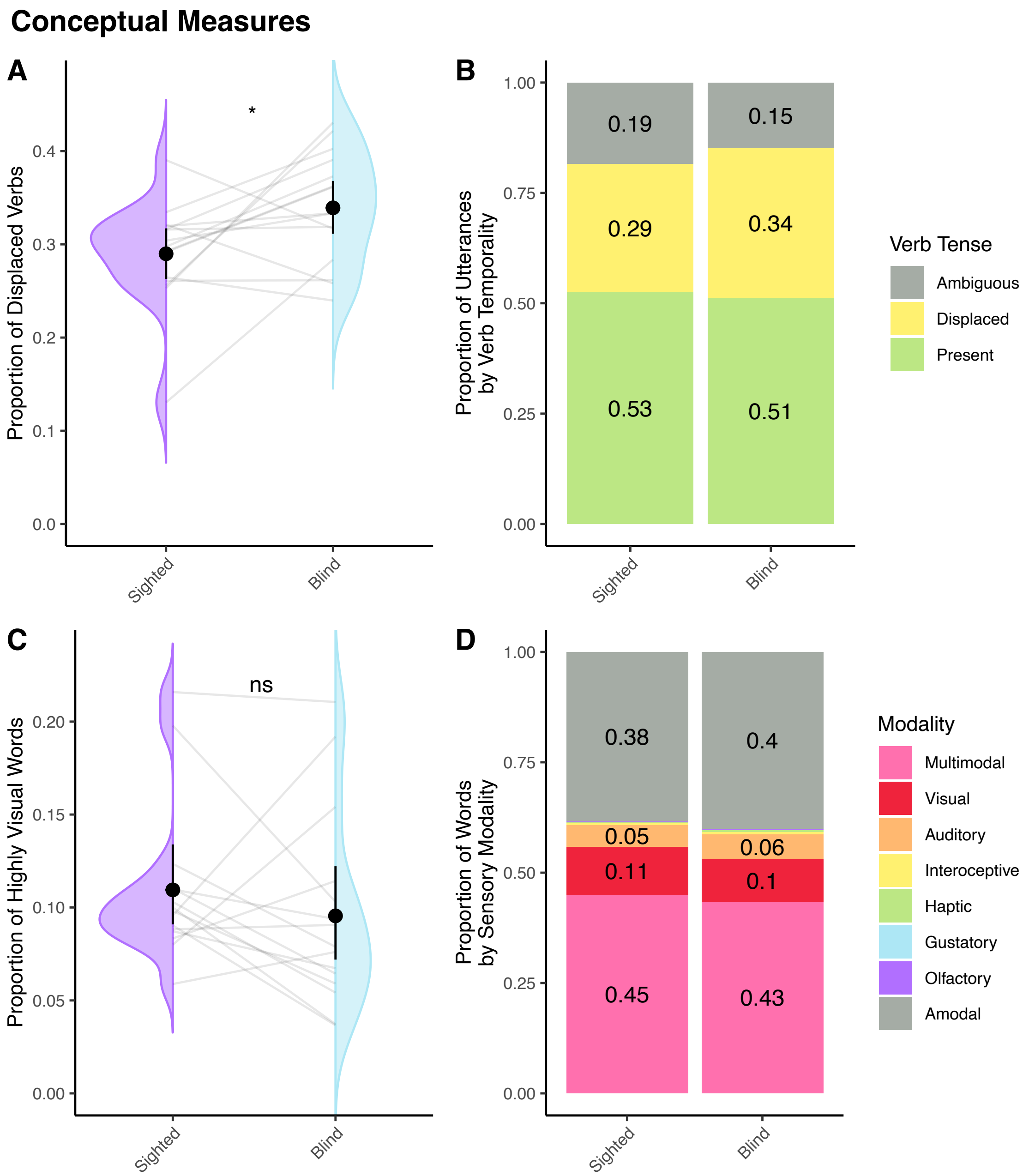

Figure 4. ***Left col: Comparing proportion of temporally displaced verbs (top) and proportion of highly visual words (bottom). Violin density represents the distribution of values for each group. Grey lines connect values from matched participants. Black dot and whiskers show standard error around the mean. Right col: Full distribution of verb types (top) and sensory modality (bottom) by group, collapsing across participants. Blind children's input contained significantly more temporally displaced verbs. Notably, the groups did not differ in the proportion of highly visual words.***

Age Differences. Lastly, we used a series of linear models (each predicting one of our input variables, based on an interaction between age and group) to explore whether input characteristics differed for younger vs. older children. We note that these analyses are extremely exploratory but are offered in the spirit of transparency to comment on the developmental trends in a limited sample. For the number of words in the input, the proportion of child-directed speech, MLU, and the proportion of temporally displaced verbs, we did not find that the input differed across age for either of our groups (*p*s > .05 for all interaction terms). We found that the number of conversational turns increased across age, such that for each month older, children took part in 1.78 more conversational turns per hour ($p = .004$), and this effect did not differ across groups. The proportion of visual words in children's input increased across developmental time for sighted children (by ~0.72% per month, $p < .001$) but not for blind children. An opposite pattern arose for amodal words: across developmental time, sighted children had fewer amodal words in their input (-0.39% fewer per month, $p < .001$) whereas blind children had marginally more (by ~0.26% per month, $p = .082$). This interaction with age was not observed for any of the other sensory modalities. Tables and figures for these exploratory models are available in Supplementals.

Discussion

In this study, we analyzed the everyday language input to 15 young congenitally-blind children alongside a carefully peer-matched sighted sample using LENA audio recorders. While still relatively modest in absolute terms, this is a larger and more naturalistic sample than has previously been leveraged by prior work with this low-incidence population. We found that along the word count, interaction, and linguistic dimensions, caregivers talked similarly to blind and sighted children, with small but potentially notable differences in conceptual content of the input. We discuss each of these results further below.

Word Count

Across two measures of input word count, one estimated from the full sixteen-hour recording (Adult Word Count) and one precisely measured from a 30-minute samples from the day (Manual Word Count), blind and sighted children were exposed to similar amounts of speech in the home. Word count was highly variable within groups, but we found no evidence for between group differences, though it remains a possibility that there are smaller effects that we were unable to detect. This lack of difference runs counter to two folk accounts of language input to blind children: 1) that sighted parents of blind children might talk less because they don't share visual common ground with their children; 2) that parents of blind children might talk more to compensate for their children's lack of visual input. Instead, we find a similar amount of speech across groups.

Interaction

We quantified interaction in two ways: through the LENA-estimated conversational turn count and through the proportion of child-directed speech in our manual annotations. Again, we found no differences across groups in the amount of parent-child interaction. This finding contrasts with previous research; other studies tend to report less interaction in dyads where the child is blind (Nagayoshi et al., 2017; Rogers & Puchalski, 1984; as measured by responsiveness, Tröster & Brambring, 1992; initiations of interactions, Andersen et al., 1993; Dote-Kwan, 1995; Kekelis & Andersen, 1984; Moore & McConachie, 1994; Tröster & Brambring, 1992; caregiver dominance of the conversation, Kekelis & Andersen, 1984; or "weak and inconsistent" responses to blind infants' vocalizations, Rowland, 1984). Our use of daylong audio recordings might explain this apparent discrepancy in results. For one thing, many prior studies (e.g., Kekelis & Andersen, 1984; Moore & McConachie, 1994; Pérez-Pereira & Conti-Ramsden, 2001; Preisler, 1991) involve videorecordings in the child's home, with the researcher present. Like other young children, blind children distinguish between familiar individuals and strangers and react with trepidation to the presence of a stranger; for blind children, this reaction may involve "quieting", wherein children cease speaking or vocalizing when they hear a new voice in the home (Fraiberg, 1975; McRae, 2002). By having a researcher present during the recordings[9], prior research may have artificially suppressed blind children's initiation of interactions. Even naturalistic, observer-free videorecordings appear to inflate aspects of parental speech, relative to daylong audio recordings (Bergelson et al., 2019).

Additionally, a common focus in earlier interaction literature is to measure visual cues of interaction, such as shared gaze or attentiveness to facial expressions (Baird, Mayfield, & Baker, 1997; Nagayoshi et al., 2017; Preisler, 1991; Rogers & Puchalski, 1984). For example, Nagayoshi et al. (2017) write: "Infants with visual impairment were characterized by high likelihood of developmental delays and problematic behaviors; they tended not to turn their face or eyes toward their mothers." We can't help but wonder: are visual markers of social interaction the right yardstick to measure blind children against? In line with MacLeod and Demers (2023), perhaps the field should move away from sighted indicators of interaction "quality", and instead situate blind children's interactions within their own developmental niche, one that may be better captured with auditory- or tactile-focused measures. While daylong audio recordings excel at capturing extended, naturalistic spoken language use, they miss non-verbal information, like proximity, touch, or physical properties of the referent. In contrast, video recordings could provide rich information about these multimodal features. Future work should consider integrating these approaches to provide a more comprehensive view of blind children's interactions.

[9] Fraiberg (1975) writes "these fear and avoidance behaviors appear even though the observer, a twice-monthly visitor, is not, strictly speaking, a stranger." (pg. 323).

Linguistic Features

Along the linguistic dimension, we measured type-token ratio and mean length of utterance. Parents of children with disabilities (including parents of blind children, e.g., Chernyak, n.d.; FamilyConnect, n.d.) are often advised to use shorter, simpler sentences with their children; correspondingly, previous work finds that parents of children with disabilities tend to find that parents do use shorter, simpler utterances (e.g., Down syndrome, Lorang et al., 2020; hearing loss, Dirks et al., 2020). While language input patterns among these populations may not necessarily generalize to blind children, the societal infantilization of disabled people broadly, including blind individuals (Bulk et al., 2020; Hernandez Padilla & Arias Valencia, 2024), might lead to differences in how caregivers structure their input. We therefore hypothesized that caregivers might provide shorter utterances and less lexically diverse input to blind children compared to their sighted peers. Instead, we found that blind children heard indistinguishable input by these metrics, with, if anything, a (marginally significant) trend towards longer sentences in their input. Contrary to the advice often given to parents, evidence suggests that, longer, more complex utterances are associated with better child language outcomes in both typically-developing children (Hoff & Naigles, 2002) and children with cognitive differences (Sandbank & Yoder, 2016). And similarly, higher lexical diversity is associated with larger vocabulary (Anderson et al., 2021; Hsu et al., 2017; Huttenlocher et al., 2010; Rowe, 2012; Weizman & Snow, 2001). Regardless, the present analysis did not reveal robust statistical evidence that, at least on the group level, caregivers systematically provide utterances with different length or lexical diversity as a function of whether their child could see.

Conceptual Features

Although there are many potential ways to measure the conceptual features of language, we chose to capture *here-and-now*-ness by measuring the proportion of temporally displaced verbs (i.e., targeting non-present events) and the proportion of highly visual words. We found that blind children heard roughly 5% more temporally displaced verbs than sighted peers. This measure is imperfect: in using tense as a proxy for conceptual features, it fails to adjudicate, for example, the decontextualized nature of a “make-believe” utterance in the present tense, or the salience of a past-tense utterance describing an event that happened seconds before[10]. Nonetheless, we believe this captures similar or higher amounts of signal relative to more costly manually-annotated measures. Moreover, though blind and sighted participants were

[10] One concern about this metric is its treatment of multi-clause utterances. For example, in “I went to the grocery store and now I’m watching TV”, “went” is not syntactically higher than “watching” but our classification system would rely on the tense of “went” alone. In practice, only 1.7% of utterances in our dataset contain verbs both before and after a conjunction, while 11.05% contain syntactic subordination, where the tense of the highest verb is most appropriate to assess.

exposed to a similar proportion of highly visual words, the referents of these words are by definition inaccessible to the blind participants. Our conceptual results suggest that blind children's input could be *less* focused on the *here-and-now*.

The extent to which blind children's language input is centered on the *here-and-now* has been contested in the literature (Andersen et al., 1993; J. Campbell, 2003; Kekelis & Andersen, 1984; Moore & McConachie, 1994; Urwin, 1984). This aspect of language input is of particular interest because early reports suggest that blind children's own use of decontextualized language develops later than sighted children's (Bigelow, 1990; Urwin, 1984). Could such a difference be attributed to an absence of decontextualized language in the input? Our results suggest this is unlikely: we found that blind children's input contained *more* decontextualized language (as indicated by verb temporality) rather than less. Speculatively, this may be because visually-oriented, sighted caregivers find a perceptual common ground for discussion, instead of replacing visually-grounded conversation with sensory modalities that the child can access. For example, while riding on a train, parents of sighted children may discuss the changing scenery outside the window, which is present, perceptually accessed by both parent and child, and salient as a topic of conversation. Present, perceptually available features of the environment for the blind child, such as the rumble of the train and velvety feel of the seats, may be less salient to the sighted parent as a topic of discussion, which may lead the caregiver to choose to talk about events that happened earlier in the day or their plans upon arriving home. Past and future events are experienced via mental representation rather than perceptually for caregiver and child alike. This is a potential avenue for broadening the concept of joint attention as a fundamental feature of conversation and language acquisition beyond shared visual reference.

Our findings indicate that sighted caregivers used a comparable amount of 'highly visual' words when speaking to their blind children and their sighted peers, as measured using sensorimotor norms derived from sighted adults (Lynott et al., 2020). While these norms offer a valuable framework for analyzing input from sighted caregivers, it is important to consider the semantic implications for blind children themselves. Kerr and Johnson (1991) reported that blind adults rated traditionally visual words, like 'sky,' as evoking more varied and multimodal mental imagery, including tactile and spatial experiences. Future work developing sensory norms specifically tailored to blind individuals would provide valuable insights into these children's perceptual-semantic mappings. In the meantime, our findings suggest that while caregivers do not reduce their use of visual words when interacting with blind children, these words could potentially take on unique semantic dimensions within the linguistic and sensory environments of blind learners. Without further information about the social and perceptual context, it is difficult to determine the motivation of any differences we find in the input's conceptual features (e.g. in decontextualized speech). As more dense annotation becomes available, we look forward to further work exploring the social and environmental contexts of conceptual information as it unfolds across discourse.

Without further information about the social and perceptual context, it is difficult to determine the motivation of any differences we find in the input's conceptual features (e.g. in decontextualized speech). As more dense annotation becomes available, we look forward to further work exploring the social and environmental contexts of conceptual information as it unfolds across discourse.

It is worth underscoring again how much variability there is within groups and how much consistency there is between groups. One could imagine a world in which the language environments of blind and sighted children are radically different from each other. Our data do not support that hypothesis. Rather, we find similarity in word count, interaction, and linguistic properties, alongside modest differences in conceptual properties. That is, in line with recent work highlighting immense within-group variability across many different socio-cultural and linguistic contexts (Bergelson et al., 2023), our blind and sighted groups here have large within-group variability but very few between-group differences. Despite strikingly different visual experiences, young blind and sighted learners have at best modest differences in their speech environments.

Connecting to Language Outcomes

Our results uncover no systematic group differences in word count, amount of language interaction, or linguistic complexity parents provide to blind vs. sighted children, at least as measured here. When we do see differences, language input to blind children looks more conceptually complex or perceptually unavailable. In other populations, complexity of this sort is linked with *more* sophisticated child language outcomes (Demir et al., 2015; Rowe, 2012; Uccelli et al., 2019), so it is not the case that blind children's language input is "impoverished" in this sense.

In our modestly-sized, predominantly pre-lexical sample, linking language input to children's language outcomes directly is not yet feasible, but prior literature allows us to speculate on two possibilities. First, if input effects pattern similarly for blind and sighted children, we would expect blind and sighted children alike to benefit from more input (Anderson et al., 2021; Gilkerson et al., 2018; Huttenlocher et al., 1991; Rowe, 2008), more interactive input (Donnellan et al., 2020; Goldstein & Schwade, 2008; Hirsh-Pasek et al., 2015; Romeo et al., 2018; Rowe, 2008; Shneidman et al., 2013; Weisleder & Fernald, 2013), more linguistically complex input (Anderson et al., 2021; de Villiers, 1985; Hadley et al., 2017; Hoff, 2003; Hsu et al., 2017; Huttenlocher et al., 2002, 2010; Naigles & Hoff-Ginsberg, 1998; Rowe, 2012; Weizman & Snow, 2001), and more conceptually complex input (Demir et al., 2015; Rowe, 2012; Uccelli et al., 2019).

At the same time, however, recent results show that blind children have a roughly half-year delay in their productive vocabulary, relative to sighted peers (E. E. Campbell et al., 2024). If properties of the language input play a role in this delay, this raises the second possibility: that language input affects acquisition *differently* for blind children than it does for sighted children. Under this possibility, blind children would benefit from *less* complex language input, and the equivalencies in word count, linguistic complexity and interactivity alongside the increased conceptual complexity we find here would, in theory, contribute to early vocabulary delays.

To show our cards, we are inclined towards option one: that blind children benefit from language input in the same ways as sighted peers (Landau & Gleitman, 1985), and that this additionally extends to the benefits of receiving more conceptually complex language input. Language regularly supports learning in the absence of direct sensory perception (e.g., reading a book about mythical creatures). Given the language skills of blind adults (Loiotile et al., 2020; Röder et al., 2003; Röder et al., 2000), it is undeniable that language is a rich source of meaning for blind individuals as well (E. E. Campbell & Bergelson, 2022; Lewis et al., 2019; van Paridon et al., 2021). Testing each of these predictions– as well as whether links between language input and language outcomes change across developmental time– awaits further research.

In either case, if properties of language input do influence blind children's language outcomes, attempting to train parents to talk differently may be unfruitful. While some input-focused interventions show promise (Huber et al., 2023; Roberts et al., 2019), such interventions often fail to change parental speech patterns on more extended timescales (e.g., McGillion et al., 2017; Suskind et al., 2016).

Conclusion

In summary, our study compared language input in homes of 15 blind and 15 sighted infants. We found that both groups received language input with similar quantities of speech, interactivity, and linguistic complexity. Additionally, blind children were exposed to input that had somewhat more conceptual complexity, with more decontextualized talk and words for less perceptually-available (visual) referents. This suggests that young blind children are being exposed to a rich linguistic environment that differs only modestly from the language input of sighted children. Our study does not imply that parents should change their communication styles, but rather highlights the language experiences of blind children. Future research linking input measures to language development and cognitive abilities of blind and sighted children alike would be a fruitful and welcome next step.

References

Abu-Zhaya, R., Kondaurova, M. V., Houston, D., & Seidl, A. (2019). Vocal and tactile input to children who are Deaf or Hard of Hearing. *Journal of Speech, Language, and Hearing Research, 62*(7), 2372–2385. https://doi.org/10.1044/2019_JSLHR-L-18-0185

Andersen, E., Dunlea, A., & Kekelis, L. (1993). The impact of input: Language acquisition in the visually impaired. *First Language, 13* (37), 23–49. https://doi.org/10.1177/014272379301303703

Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., & Madigan, S. (2021). Linking quality and quantity of parental linguistic input to child language skills: A meta-analysis. *Child Development, 92*(2), 484–501. https://doi.org/10.1111/cdev.13508

Baird, S. M., Mayfield, P., & Baker, P. (1997). Mothers' interpretations of the behavior of their infants with visual and other impairments during interactions. *Journal of Visual Impairment & Blindness, 91*(5), 467–483. https://doi.org/10.1177/0145482X9709100507

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289–300. Retrieved from https://www.jstor.org/stable/2346101.

Bergelson, E. (2015). *Bergelson Seedlings HomeBank Corpus*. TalkBank. https://doi.org/10.21415/T5PK6D

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science, 22* (1), e12715. https://doi.org/10.1111/desc.12715

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C., Ramírez-Esparza, N., R. Hamrick, L., . . . 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

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition, 127*(3), 391–397. https://doi.org/10.1016/j.cognition.2013.02.011

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother–child speech. *Journal of Child Language, 11,* 557–578.

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

Bigelow, A. (1990). Relationship between the development of language and thought in young blind children. *Journal of Visual Impairment & Blindness, 84* (8), 414–419. https://doi.org/10.1177/0145482X9008400805

Bratt, J. & Harmon, J., (2022). *Morphemepiece: Morpheme tokenization.*

Brugman, H., & Russel, A. (2009). Annotating multimedia / multi-modal resources with ELAN. *Proceedings of the Fourth International Conference on Language Resources and Evaluation.*

Bulk, L. Y., Smith, A., Nimmon, L., & Jarus, T. (2020). A closer look at opportunities for blind adults: Impacts of stigmatization and ocularcentrism. *British Journal of Visual Impairment, 38*(3), 270–283. https://doi.org/10.1177/0264619620911424

Busch, T., Sangen, A., Vanpoucke, F., & van Wieringen, A. (2018). Correlation and agreement between Language ENvironment Analysis (LENA™) and manual transcription for Dutch natural language recordings. *Behavior Research Methods, 50*(5), 1921–1932. https://doi.org/10.3758/s13428-017-0960-0

Campbell, E. E., & Bergelson, E. (2022). Making sense of sensory language: Acquisition of sensory knowledge by individuals with congenital sensory impairments. *Neuropsychologia, 174,* 108320. https://doi.org/10.1016/j.neuropsychologia.2022.108320

Campbell, E. E., Casillas, R., & Bergelson, E. (2024). The role of vision in the acquisition of words: Vocabulary development in blind toddlers. *Developmental Science, n/a*(n/a), e13475. https://doi.org/10.1111/desc.13475

Campbell, J. (2003). Maternal directives to young children who are blind. *Journal of Visual Impairment & Blindness, 97*(6), 355–365. https://doi.org/10.1177/0145482X0309700604

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development, 91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Chernyak, P. (n.d.). 3 Ways to Teach Your Blind or Visually Impaired Child to Talk. https://www.wikihow.life/Teach-Your-Blind-or-Visually-Impaired-Child-to-Talk.

Chiesa, S., Galati, D., & Schmidt, S. (2015). Communicative interactions between visually impaired mothers and their sighted children: Analysis of gaze, facial expressions, voice and physical contacts. *Child: Care, Health and Development, 41* (6), 1040–1046. https://doi.org/10.1111/cch.12274

Cristia, A., Bulgarelli, F., & Bergelson, E. (2020). Accuracy of the Language Environment Analysis system segmentation and metrics: A systematic review. *Journal of Speech, Language, and Hearing Research*, *63*(4), 1093–1105. https://doi.org/10.1044/2020_JSLHR-19-00017

de Villiers, J. (1985). Learning how to use verbs: Lexical coding and the influence of the input. *Journal of Child Language, 12*(3), 587–595. https://doi.org/10.1017/S0305000900006668

Demir, Ö. E., Rowe, M. L., Heller, G., Goldin-Meadow, S., & Levine, S. C. (2015). Vocabulary, syntax, and narrative development in typically developing children and children with early unilateral brain injury: Early parental talk about the "there-and-then" matters. *Developmental Psychology, 51*(2), 161–175. https://doi.org/10.1037/a0038476

Dirks, E., Stevens, A., Kok, S., Frijns, J., & Rieffe, C. (2020). Talk with me! Parental linguistic input to toddlers with moderate hearing loss. *Journal of Child Language, 47*(1), 186–204. https://doi.org/10.1017/S0305000919000667

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science, 23*(1), e12843. https://doi.org/10.1111/desc.12843

Dote-Kwan, J. (1995). Impact of mothers' interactions on the development of their young visually impaired children. *Journal of Visual Impairment & Blindness, 89* (1), 46–58. https://doi.org/10.1177/0145482X9508900109

FamilyConnect. (n.d.). Understanding the Stages of Language Development for Babies Who Are Blind.

Fenson, L., Dale, P., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), 1-185. https://doi.org/10.2307/1166093

Ferjan Ramírez, N., Hippe, D. S., & Kuhl, P. K. (2021). Comparing automatic and manual measures of parent–infant conversational turns: A word of caution. *Child Development*, *92*(2), 672–681. https://doi.org/10.1111/cdev.13495

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*(6), 1497–1510. https://doi.org/10.2307/1130938

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637–656. https://doi.org/10.1111/j.1467-8624.1993.tb02933.x

Fraiberg, S. (1975). The development of human attachments in infants blind from birth. *Merrill-Palmer Quarterly*, *21*, 315–334. Retrieved from https://www.jstor.org/stable/23083880

Ganea, N., Hudry, K., Vernetti, A., Tucker, L., Charman, T., Johnson, M. H., & Senju, A. (2018). Development of adaptive communication skills in infants of blind parents. *Developmental Psychology*, *54*(12), 2265–2273. https://doi.org/10.1037/dev0000564

Ganek, H., & Eriks-Brophy, A. (2016). The Language ENvironment Analysis (LENA) system: A literature review. *Proceedings of the Joint Workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, 24–32. Umeå, Sweden: LiU Electronic Press.

Ganek, H., & Eriks-Brophy, A. (2018). Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders*, *72*, 77–85. https://doi.org/10.1016/j.jcomdis.2017.12.005

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, *23*(4), 491–517. https://doi.org/10.1177/0261927X04269589

Gilkerson, J., & Richards, J. (2008). *The LENA Natural Language Study*. Boulder, CO: LENA Foundation.

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics, 142*(4), e20174276. https://doi.org/10.1542/peds.2017-4276

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*(1), 3–55. Retrieved from https://www.jstor.org/stable/20011341

Gleitman, L. R., & Newport, E. L. (1995). The invention of language by children: Environmental and biological influences on the acquisition of language. In *An Invitation to Cognitive Science. Language: An invitation to cognitive science, Vol. 1, 2nd ed* (pp. 1–24). Cambridge, MA, US: The MIT Press.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19* (5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Grice, H. P. (1975). Logic and conversation. In *Syntax and Semantics 3: Speech Acts* (pp. 41-58). Brill.

Grigoroglou, M., Edu, U., & Papafragou, A. (2016). Are children flexible speakers? Effects of typicality and listener needs in children's event descriptions. *Proceedings of the Annual Meeting of the Cognitive Science Society, 38,* 782-787. Retrieved from https://escholarship.org/uc/item/3s13v7sg

Grimminger, A., Rohlfing, K. J., Lüke, C., Liszkowski, U., & Ritterfeld, U. (2020). Decontextualized talk in caregivers' input to 12-month-old children during structured interaction. *Journal of Child Language, 47*(2), 418–434. https://doi.org/10.1017/S0305000919000710

Grumi, S., Cappagli, G., Aprile, G., Mascherpa, E., Gori, M., Provenzi, L., & Signorini, S. (2021). Togetherness, beyond the eyes: A systematic review on the interaction between visually impaired children and their parents. *Infant Behavior and Development, 64,* 101590. https://doi.org/10.1016/j.infbeh.2021.101590

Hadley, P. A., Rispoli, M., Holt, J. K., Papastratakos, T., Hsu, N., Kubalanza, M., & McKenna, M. M. (2017). Input subject diversity enhances early grammatical growth: Evidence from a parent-implemented intervention. *Language Learning and Development, 13*(1), 54–79. https://doi.org/10.1080/15475441.2016.1193020

Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science, 45*(3), e12926. https://doi.org/10.1111/cogs.12926

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America, 130* (4), 2139–2152. https://doi.org/10.1121/1.3623753

Hernandez Padilla, M. L., & Arias Valencia, S. A. (2024). “The way they treat us reflects how they see us”: Aspects contributing to social inequities in people with visual disability, a qualitative study. *British Journal of Visual Impairment*, 02646196241265616. https://doi.org/10.1177/02646196241265616

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., . . . Suma, K. (2015). The contribution of early communication quality to low-income children’s language success. *Psychological Science, 26* (7), 1071–1083. https://doi.org/10.1177/0956797615581493

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., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*(2), 418–433. https://doi.org/10.1111/1467-8624.00415

Hsu, N., Hadley, P. A., & Rispoli, M. (2017). Diversity matters: Parent input predicts toddler verb production. *Journal of Child Language, 44*(1), 63–86. https://doi.org/10.1017/S0305000915000690

Huber, E., Ferjan Ramírez, N., Corrigan, N. M., & Kuhl, P. K. (2023). Parent coaching from 6 to 18 months improves child language outcomes through 30 months of age. *Developmental Science, 26*(6), e13391. https://doi.org/10.1111/desc.13391

Hudson, J. A. (2002). "Do you know what we’re going to do this summer?": Mothers’ talk to preschool children about future events. *Journal of Cognition and Development, 3*(1), 49–71. https://doi.org/10.1207/S15327647JCD0301_4

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*(3), 337–374. https://doi.org/10.1016/s0010-0285(02)00500-5

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

Jara-Ettinger, J., & Rubio-Fernandez, P. (2021). The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychological Review*, *129*(6), 1394–1413. https://doi.org/10.1037/rev0000345

Kekelis, L. S., & Andersen, E. S. (1984). Family communication styles and language development. *Journal of Visual Impairment & Blindness*, *78*(2), 54–65. https://doi.org/10.1177/0145482X8407800202

Kerr, N. H., & Johnson, T. H. (1991). Word norms for blind and sighted subjects: Familiarity, concreteness, meaningfulness, imageability, imagery modality, and word associations. *Behavior Research Methods, Instruments, & Computers*, *23*(4), 461–485. https://doi.org/10.3758/BF03209988

Kramer, J. A., Hill, K. T., & Cohen, L. B. (1975). Infants' development of object permanence: A refined methodology and new evidence for Piaget's hypothesized ordinality. *Child Development*, *46*(1), 149–155. https://doi.org/10.2307/1128843

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child* (pp. xi, 250). Harvard University Press.

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2021). *An open-source voice type classifier for child-centered daylong recordings*. arXiv. https://doi.org/10.48550/arXiv.2005.12656

Lehet, M., Arjmandi, M. K., Houston, D., & Dilley, L. (2021). Circumspection in using automated measures: Talker gender and addressee affect error rates for adult speech detection in the Language ENvironment Analysis (LENA) system. *Behavior Research Methods, 53*(1), 113–138. https://doi.org/10.3758/s13428-020-01419-y

Lewis, M., Zettersten, M., & Lupyan, G. (2019). Distributional semantics as a source of visual knowledge. *Proceedings of the National Academy of Sciences, 116*(39), 19237–19238. https://doi.org/10.1073/pnas.1910148116

Loiotile, R., Lane, C., Omaki, A., & Bedny, M. (2020). Enhanced performance on a sentence comprehension task in congenitally blind adults. *Language, Cognition and Neuroscience, 35*(8), 1010–1023. https://doi.org/10.1080/23273798.2019.1706753

Lorang, E., Venker, C. E., & Sterling, A. (2020). An investigation into maternal use of telegraphic input to children with Down syndrome. *Journal of Child Language, 47* (1), 225–249. https://doi.org/10.1017/S0305000919000503

Lucariello, J., & Nelson, K. (1987). Remembering and planning talk between mothers and children. *Discourse Processes*, *10*(3), 219–235. https://doi.org/10.1080/01638538709544673

Lucca, K., & Wilbourn, M. P. (2018). Communicating to learn: Infants' pointing gestures result in optimal learning. *Child Development, 89*(3), 941–960. https://doi.org/10.1111/cdev.12707

Luchkina, E., Xu, F., Sobel, D., & Morgan, J. (2020). Sixteen-month-olds comprehend unanchored absent reference [Preprint]. *Open Science Framework*. https://doi.org/10.31219/osf.io/5tc6d

Lukin, E., Campbell, E. E., Righter, L., & Bergelson, E. (2023). Comparing utterance composition and conversational content in everyday language input to blind and sighted toddlers. Presented at MPAL 2023 in Nijmegen, Netherlands, and BUCLD 48 in Boston, MA.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods, 52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

MacLeod, A., & Demers, C. (2023). Transmitting white monolingual Anglo-American norms: A concept analysis of “quality of language” in parent-child interactions. *Applied Psycholinguistics, 44*(4), 431–559. https://doi.org/10.1017/S014271642300005X

MacWhinney, B. (2019). *CHAT Manual*. https://doi.org/10.21415/3MHN-0Z89

Magimairaj, B., Nagaraj, N., Caballero, A., Munoz, K., & White, K. (2022). A systematic review of the effects of LENA-based feedback on parent-child language interactions in families with young children. *Journal of Early Hearing Detection and Intervention, 7*(3), 47–60. https://doi.org/10.26077/6c72-973b

McGillion, M., Pine, J. M., Herbert, J. S., & Matthews, D. (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *Journal of Child Psychology and Psychiatry, 58*(10), 1122–1131. https://doi.org/10.1111/jcpp.12725

McRae, K. A. (2002). *Attachment in blind infants: A systematic investigation using Ainsworth’s Strange Situation.* (PhD thesis). University of Toronto, Toronto, Canada.

Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early word learning environments. *Cognitive Science, 42 Suppl 2* (Suppl 2), 375–412. https://doi.org/10.1111/cogs.12592

Moore, V., & McConachie, H. (1994). Communication between blind and severely visually impaired children and their parents. *British Journal of Developmental Psychology, 12*(4), 491–502. https://doi.org/10.1111/j.2044-835X.1994.tb00650.x

Moser, C., Lee-Rubin, H., Bainbridge, C., Atwood, S., Simson, J., Knox, D., . . . Mehr, S. (2022). Acoustic regularities in infant-directed vocalizations across cultures. *Nature Human Behavior, 6*(11), 1545–1556. https://doi.org/10.1038/s41562-022-01410-x

Nagayoshi, M., Hirose, T., Toju, K., Suzuki, S., Okamitsu, M., Teramoto, T., . . . Takeo, N. (2017). Related visual impairment to mother-infant interaction and development in infants with bilateral retinoblastoma. *European Journal of Oncology Nursing: The Official Journal of European Oncology Nursing Society, 28*, 28–34. https://doi.org/10.1016/j.ejon.2017.02.002

Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children’s early verb use. *Journal of Child Language, 25*(1), 95–120. https://doi.org/10.1017/S0305000997003358

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to Children: Language Input and Acquisition* (pp. 109–149). Cambridge University Press.

Ostarek, M., van Paridon, J., & Montero-Melis, G. (2019). Sighted people's language is not helpful for blind individuals' acquisition of typical animal colors. *Proceedings of the National Academy of Sciences, 116*(44), 21972–21973. https://doi.org/10.1073/pnas.1912302116

Pancsofar, N., & Vernon-Feagans, L. (2006). Mother and father language input to young children: Contributions to later language development. *Journal of Applied Developmental Psychology, 27*(6), 571–587. https://doi.org/10.1016/j.appdev.2006.08.003

Pérez-Pereira, M., & Conti-Ramsden, G. (2001). The use of directives in verbal interactions between blind children and their mothers. *Journal of Visual Impairment & Blindness, 95*(3), 133–149. https://doi.org/10.1177/0145482x0109500302

Pisani, S., Gautheron, L., & Cristia, A. (2021). *Long-form recordings: From A to Z.*

Preisler, G. M. (1991). Early patterns of interaction between blind infants and their sighted mothers. *Child: Care, Health and Development, 17* (2), 65–90. https://doi.org/10.1111/j.1365-2214.1991.tb00680.x

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17*(6), 880–891. https://doi.org/10.1111/desc.12172

Richards, B. (1987). Type/Token Ratios: What do they really tell us? *Journal of Child Language, 14*(2), 201–209. https://doi.org/doi:10.1017/s0305000900012885

Roberts, M. Y., Curtis, P. R., Sone, B. J., & Hampton, L. H. (2019). Association of parent training with child language development: A systematic review and meta-analysis. *JAMA Pediatrics, 173*(7), 671–680. https://doi.org/10.1001/jamapediatrics.2019.1197

Röder, B., Demuth, L., Streb, J., & Rösler, F. (2003). Semantic and morpho-syntactic priming in auditory word recognition in congenitally blind adults. *Language and Cognitive Processes, 18*(1), 1–20. https://doi.org/10.1080/01690960143000407

Röder, B., Rösler, F., & Neville, H. J. (2000). Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia, 38*(11), 1482–1502. https://doi.org/10.1016/S0028-3932(00)00057-9

Rogers, S. J., & Puchalski, C. B. (1984). Social characteristics of visually impaired infants' play. *Topics in Early Childhood Special Education, 3*(4), 52–56. https://doi.org/10.1177/027112148400300409

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science, 29*(5), 700–710. https://doi.org/10.1177/0956797617742725

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*(1), 185–205. https://doi.org/10.1017/S0305000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Rowe, M. L. (2013). Decontextualized language input and preschoolers' vocabulary development. *Seminars in Speech and Language, 34*(4), 260–266. https://doi.org/10.1055/s-0033-1353444

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. *Journal of Child Language, 47*(1), 5–21. https://doi.org/10.1017/S0305000919000655

Rowland, Charity. (1984). Preverbal communication of blind infants and their mothers. *Journal of Visual Impairment & Blindness, 78*(7), 297–302. https://doi.org/10.1177/0145482X8407800701

Rowland, Caroline, Durrant, S., Peter, M., Bidgood, A., Pine, J., & Jago, L. S. (2018). *The Language 0-5 Project.* https://doi.org/10.17605/OSF.IO/KAU5F

Rubio-Fernandez, P. (2019). Overinformative speakers are cooperative: Revisiting the Gricean maxim of quantity. *Cognitive Science, 43*(11), e12797. https://doi.org/10.1111/cogs.12797

Sak-Wernicka, J. (2017). *Blind people's pragmatic abilities*. Cambridge Scholars Publishing.

Sandbank, M., & Yoder, P. (2016). The association between parental Mean Length of Utterance and language outcomes in children with disabilities: A correlational meta-analysis. *American Journal of Speech-Language Pathology, 25*(2), 240–251. https://doi.org/10.1044/2015_AJSLP-15-0003

Senju, A., Tucker, L., Pasco, G., Hudry, K., Elsabbagh, M., Charman, T., & Johnson, M. H. (2013). The importance of the eyes: Communication skills in infants of blind parents. *Proceedings. Biological Sciences*, *280*(1760), 20130436. https://doi.org/10.1098/rspb.2013.0436

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language, 40*(3), 672–686. https://doi.org/10.1017/S0305000912000141

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*, 549–565. https://doi.org/10.2307/1127555

Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Alam, F., Warlaumont, A. S., & Bunce, J. (2021). Developing a cross-cultural annotation system and meta-corpus for studying infants' real world language experience. *Collabra: Psychology, 7*(1), 23445. https://doi.org/10.1525/collabra.23445

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., . . . Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of Child Language, 43*(2), 366–406. https://doi.org/10.1017/S0305000915000033

Tadić, V., Pring, L., & Dale, N. (2013). Story discourse and use of mental state language between mothers and school-aged children with and without visual impairment. *International Journal of Language & Communication Disorders, 48*(6), 679–688. https://doi.org/10.1111/1460-6984.12040

Templin, M. C. (1957). *Certain language skills in children; their development and interrelationships* (pp. xviii, 183). University of Minnesota Press.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454–1463. https://doi.org/10.2307/1130423

Tröster, H., & Brambring, M. (1992). Early social-emotional development in blind infants. *Child: Care, Health and Development, 18*(4), 207–227. https://doi.org/10.1111/j.1365-2214.1992.tb00355.x

Uccelli, P., Demir-Lira, Ö. E., Rowe, M. L., Levine, S., & Goldin-Meadow, S. (2019). Children's early decontextualized talk predicts academic language proficiency in midadolescence. *Child Development, 90*(5), 1650–1663. https://doi.org/10.1111/cdev.13034

Urwin, C. (1983). Dialogue and cognitive functioning in the early language development of three blind children (142–161). In Mills, A. E. (ed.), Language acquisition in the blind child: normal and deficient. London: Croom Helm.

Urwin, C. (1984). Language for absent things: Learning from visually handicapped children. *Topics in Language Disorders, 4*(4), 24–37.

van Paridon, J., Liu, Q., & Lupyan, G. (2021). How do blind people know that blue is cold? Distributional semantics encode color-adjective associations. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43,* 2671–2677. Retrieved from https://escholarship.org/uc/item/6sq7h506

VanDam, M., Oller, D. K., Ambrose, S. E., Gray, S., Richards, J. A., Xu, D., . . . Moeller, M. P. (2015). Automated vocal analysis of children with hearing loss and their typical and atypical peers. *Ear and Hearing, 36*(4), e146–e152. https://doi.org/10.1097/AUD.0000000000000138

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., Palma, P. D., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language, 37*(02), 128–142. https://doi.org/10.1055/s-0036-1580745

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes.* Harvard University Press.

Wang, Y., Cooke, M., Reed, J., Dilley, L., & Houston, D. M. (2022). Home auditory environments of children with cochlear implants and children with normal hearing. *Ear and Hearing, 43*(2), 592–604. https://doi.org/10.1097/AUD.0000000000001124

Wang, Y., Williams, R., Dilley, L., & Houston, D. M. (2020). A meta-analysis of the predictability of LENA™ automated measures for child language development. *Developmental Review, 57*, 100921. https://doi.org/10.1016/j.dr.2020.100921

Warlaumont, A. S., Pretzer, G. M., Mendoza, S., & Walle, E. A. (2016). *Warlaumont HomeBank Corpus*. TalkBank. https://doi.org/10.21415/T54S3C

Watkins, S., Pittman, P., & Walden, B. (1998). The Deaf Mentor Experimental Project for young children who are deaf and their families. *American Annals of the Deaf, 143*(1), 29–34. https://doi.org/10.1353/aad.2012.0098

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*(2), 265–279. https://doi.org/10.1037/0012-1649.37.2.265

Wijffels, J. (2023). *UDPipe*.

Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the LENA Language Environment Analysis System in Young Children's Natural Home Environment* (pp. 1–16). The LENA Foundation.

Yoshinaga-Itano, C., Sedey, A. L., Mason, C. A., Wiggin, M., & Chung, W. (2020). Early intervention, parent talk, and pragmatic language in children with hearing loss. *Pediatrics, 146* (Supplement 3): S270–S277. https://doi.org/10.1542/peds.2020-0242F

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Data, Code, and Materials Availability Statement

LENA data, anonymized transcripts, and code for all analyses presented in this article are available at https://osf.io/dcnq6/.

Ethics Statement

This study received approval from the Duke University Institutional Review Board. All families consented to take part in this research.

Authorship and Contributorship Statement

Erin Campbell conceived of the study; collected the LENA recordings from blind participants and curated the recordings from sighted participants; transcribed recordings; performed data analysis and visualization; wrote the first draft of the manuscript and revised it many times over. **Lillianna Righter** oversaw the massive transcription effort for this corpus, including hundreds of hours of transcribing, training a small army of research assistants, and ensuring that annotation was reliable; she also contributed to the first draft of the paper, data analysis (most especially on our linguistic analyses), and revisions. **Eugenia Lukin** transcribed recordings, helped to validate our transcriptions, and helped revise the manuscript. **Elika Bergelson** conceived of the study; acquired funding; provided the LENA recorders; supervised data collection, transcription, analysis, and visualization; and revised the manuscript. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the manuscript.

Acknowledgements

We wish to thank the following individuals who generously contributed recordings for the sighted matches: Anne Warlaumont, Derek Houston, Nairán Ramírez-Esparza, Mark VanDam, Caroline Rowland. We thank Zhenya Kalenkovich and Alex Emmert for careful code review.

License

Phonological deficits in developmental dyslexia in a psycholinguistic framework: unguided phonological encoding

Aki Tapionkaski
Independent Practitioner, Clinical Neuropsychologist, Tampere, Finland

Sanna Tapionkaski
Independent Practitioner, Speech and Language Therapist, Tampere, Finland

Abstract: Developmental dyslexia (DD) is associated with phonological deficits. In this article, we will discuss the phonological deficits occurring in DD by applying psycholinguistic models of speech production (Levelt et al., 1999; Dell, 1986). According to these models, output phonology is created through a process called phonological encoding, which operates in a close relationship with internal speech monitoring. We will argue that in decoding (reading by using letter-sound correspondences instead of whole word recognition) phonological word-forms are assembled within the speech output system by utilising phonological encoding and, more specifically, through a process that we call *unguided phonological encoding*. In this process, encoding is done in a piece-by-piece fashion without access to an active word-form, which means that internal speech monitoring cannot function in a normal manner. This makes the process less regulated and more prone to difficulties. We argue that DD is related to difficulties in unguided phonological encoding. We will consider the implications of our theoretical hypotheses, first, from a clinical perspective by providing a qualitative description of typical blending difficulties among children with DD at our clinic. Second, we will discuss earlier research literature on DD by focusing on how difficulties in unguided phonological encoding could explain the widely researched features of DD. Finally, we will outline a few possibilities for how our theoretical hypotheses could be tested. We suggest that the focus on the interplay between phonological encoding and internal speech monitoring provides a framework with which we can ask new questions about the phonological difficulties associated with DD.

Keywords: developmental dyslexia; phonological representations; phonological encoding; speech monitoring

Corresponding author: Aki Tapionkaski, Psykologikeskus Koite, Puutarhakatu 16 B, 5.krs, 33210, Tampere. Email: aki.tapionkaski@koite.fi

ORCID IDs: Aki Tapionkaski: https://orcid.org/0000-0002-5537-2472, Sanna Tapionkaski: https://orcid.org/0000-0003-3180-7678

Citation: Tapionkaski, A., & Tapionkaski, S. (2025). Phonological deficits in developmental dyslexia in a psycholinguistic framework: unguided phonological encoding. *Language Development Research, 5*(3), 72–103. http://doi.org/10.34842/ldr2025-852

Introduction

Developmental dyslexia (DD) is a condition characterised by slow and inaccurate reading. In particular, it is related with poor decoding skills (reading by using letter-sound correspondences instead of whole word recognition). DD has been associated with deficits in the spoken language system (Snowling, 2000), particularly with phonological deficits in three areas: phonological awareness, short term memory and rapid naming (Wagner & Torgesen, 1987; Bus & Van IJzendoorn, 1999; Snowling, 1998). However, it remains under dispute whether phonological deficits cause some, all or any of the observed reading difficulties and if they do, what the mechanisms involved are.

We will discuss the phonological deficits occurring in DD in a psycholinguistic framework by applying well known and extensively tested model(s) of speech production (Levelt et al., 1999; Dell, 1986). In contrast to some other psycholinguistic models used in the context of dyslexia, these models of speech production, significantly, discriminate between *lexical representations* and *output representations*. Lexical representations are permanent and contain only basic information of word-forms (phonemes and possibly stress) whereas output representations are fully formed phonological entities, words or utterances, complete with all the phonological features that speech contains. These output representations are not stored for long periods of time and, consequently, they are formed anew each time when needed. In these psycholinguistic models, the output representations are formed through a process called phonological encoding, which operates in a close relationship with internal speech monitoring. By applying this framework, we will discuss how the mechanisms related to phonological encoding and internal speech monitoring may cause the observed difficulties in dyslexics' reading and phonological tasks.

It has been hypothesised that problems in phonological representations cause the reading difficulties in dyslexia (Shankweiler et al., 1979; de Gelder & Vroomen, 1991; Fowler, 1991; Elbro, 1998; Mody et al., 1997; Manis et al., 1997; Swan & Goswami; 1997). Some hypotheses locate the deficit specifically in output representations (Hulme & Snowling, 1991; Ramus, 2001) but these hypotheses have not been confirmed in experimental research. Instead, several experiments have found no clear signs of deficits in dyslexics' representations by investigating output in word repetition, non-word repetition and sentence production tasks (e.g. Ramus, 2008; Szenkovits et al. 2016). However, we suggest that designing experiments that also consider the role of internal monitoring processes in the formation of output representations might lead to different outcomes.

Internal speech monitoring serves the purpose of regulation and error-correction during the phonological encoding process (Gauvin & Hartsuiker, 2020; Nozari, 2018).

In short, internal monitoring uses phonological information in the phonological lexicon to supervise and aid the phonological encoding process (the forming of output representations). Monitoring is targeted not only to the end-products of phonological encoding but also to the process itself (for an overview: Nozari & Novic, 2017). In the experiments targeted to chart deficits in dyslexics' output representations (Ramus, 2008; Szenkovits et al., 2016), tasks (e.g. word and non-word repetition, sentence production) have allowed for the normal functioning of internal speech monitoring, that is, the full information about the end-product has been available during the process either in the lexicon (words) or in short-term memory (non-words). In contrast, we will suggest that in decoding situations phonological encoding operates pre-lexically. In such situations, as the end-product (e.g. word) only emerges in a piece-by-piece fashion, internal monitoring would have no access to phonological information of the complete end-product in the lexicon or short-term memory and phonological encoding would have to operate with reduced monitoring possibilities. This may cause the formation of output representations to be more prone to difficulties in decoding situations in comparison to (normal) speech production.

We argue that reading and speech production share the fundamental mechanism of building phonological forms: they both rely on phonological encoding. To be more precise, we suggest that during the decoding process output representations are formed before lexical involvement by using smaller phonological components as an input (e.g. individual phonemes, syllables). This is in line with the literature in reading research that indicates that at the very early stages of reading, before lexical involvement, the reading process is sequential, sound based and follows the features of speech (for reviews: Pollatsek, 2015; Leinenger, 2014). However, we suggest that phonological encoding functions differently during speech production and word decoding. During speech production, internal speech monitoring is able to regulate the encoding process by comparing the output to the existing phonological lexicon. In contrast, internal speech monitoring cannot function in a normal manner during decoding as the end-product is not known, which makes the decoding process less regulated. This means that whenever we read a word that we do not recognise (and consequently use decoding), we build larger phonological entities from smaller pieces without the knowledge of the end-product during the process. In this situation, we argue, phonological encoding operates under reduced monitoring possibilities.

The involvement of the speech output system in decoding (and reading) is supported by many research findings. There is evidence that the same serial phonological encoding mechanism is used in naming objects and reading their names (Roelofs, 2004). Dyslexics' rapid naming is more significantly facilitated by phonological cues in comparison to controls, which suggests that slowness in naming is affected by problems in phonological encoding (Truman & Hennessey, 2006). There is also evidence that the masked onset priming effect, which is closely related to the serial nature of the reading process, actually takes place within the speech output system (Kinoshita, 2000; Kinoshita & Woollams, 2002). Additionally, neuroimaging evidence suggests that brain areas associated with speech production are involved in reading much earlier than would be expected if decoding happened outside speech system (Uno et al., 2025; Cornelissen et al., 2009).

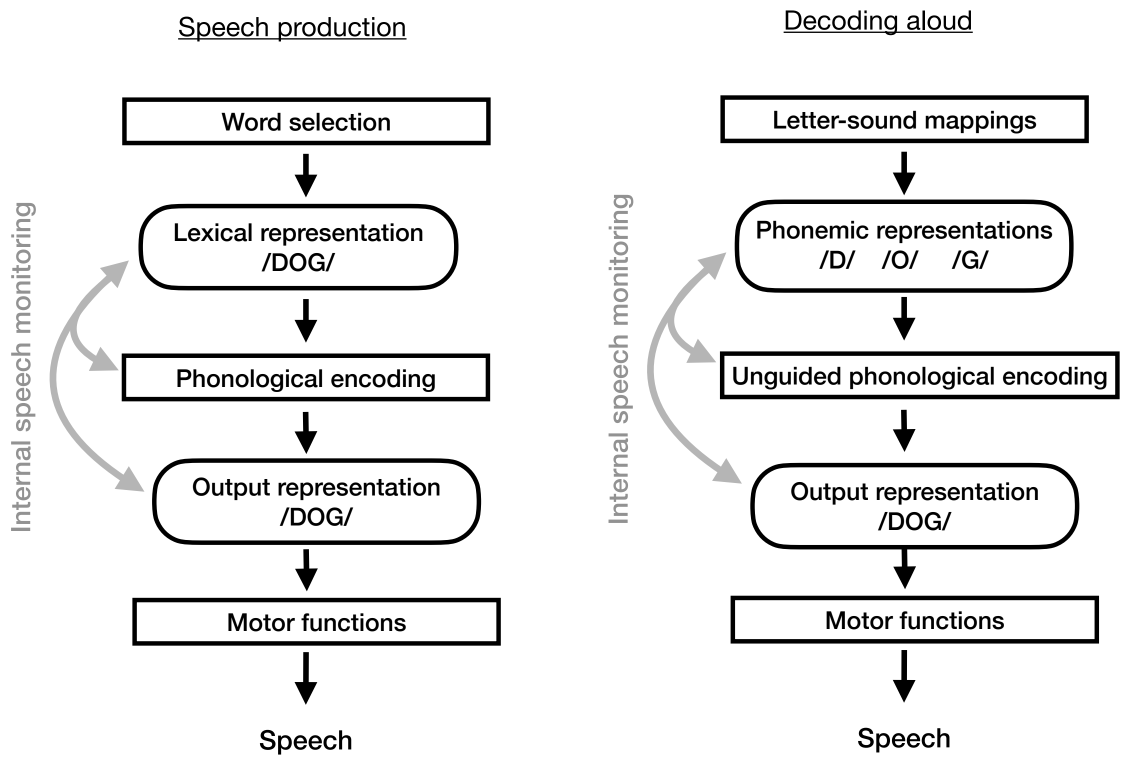

Figure 1. ***Simplified models of speech production and decoding aloud. During decoding, internal speech monitoring has no access to full word-form during phonological encoding – a situation here referred to as unguided phonological encoding. Note that the model of decoding aloud is not meant to represent a complete model of word recognition but only to illustrate the interplay between phonological encoding and internal speech monitoring during decoding.***

Thus, to describe phonological encoding in decoding situations, we will extend the notion of phonological encoding into situations where it is done without access to an active word-form either in the lexicon or short-term memory – a process that we call *unguided phonological encoding*. In speech production, this process is also active, for example, when generating impromptu, continuous nonsense-speech by randomly combining phonemes and/or syllables without preparation. In such a situation, there is no model of the end-product available and the formation of output representations (by phonological encoding) operates in a piece-by-piece fashion. Within the presented framework, our hypothesis can be divided to two parts:

1. Building phonological word-forms during decoding relies on pre-lexical phonological encoding. That is, decoding utilises *unguided* phonological encoding (where there is no access to an active word-form either in the lexicon or short-term memory).

2. DD is related to difficulties in *unguided* phonological encoding.

It is assumed that beginner-level readers learn to decode through phonological blending (e.g. Rose, 2006). Blending is the process by which smaller phonological components (e.g. phonemes) are built into larger entities (e.g. words) in volitional tasks. We argue that blending and decoding both rely on the same mechanism in building phonological forms: that is, they rely on phonological encoding. The current consensus maintains that dyslexics do not have any specific difficulties in blending but that they demonstrate more general phonological problems (for a review: Melby-Lervåg et al., 2012). Indeed, while dyslexics have considerable decoding difficulties, they seldom fail on generally used blending tasks. We suggest that this discrepancy could be explained by the manner in which the blending process has been defined and measured in earlier research. Typically, blending data in research (and in clinical settings) is collected using blending tasks that contain only simple phonological units in simple phonological environments and where fluency is not measured (for example: combine sounds /c/, /a/, /t/). In our clinical practice, we have developed tasks in which blending needs to be completed in changing and complex phonological environments in a fluent manner, that is, in tasks that resemble reading situations. In these types of tasks, we have observed distinct blending difficulties among children with DD.

Thus, our key argument is that decoding and blending rely on phonological encoding. As phonological encoding is the process where output phonology is assembled (e.g. Keating, 2000), placing phonological blending inside that process could be considered self-evident. However, in research on phonological awareness it is not typically specified where blending takes place or whether it shares any processes with speech production. As for decoding, in many theories of word recognition (e.g. Coltheart et al., 2001; Seidenberg & McClelland, 1989), decoding is assumed to take place outside the speech system. To illustrate our argument, we will offer a brief discussion of the phonological information flow in decoding and in our oral blending tasks.

As mentioned above, dyslexia is related to poor decoding skills (Lyon et al., 2003). In transparent orthographies in particular, this is typically reflected as disfluent decoding (Wagner et al., 2022). This phenomenon fits poorly to the assumption that decoding would take place outside the speech system. If this was the case, the finished word forms would be fed into the speech production chain only when completed and thus one would expect that decoding aloud would consist of periods of silence followed by fluent speech. The silent periods would occur when phonological words were being build outside the speech system, and once the feed (phonological word form) would be sent to the speech system, normal, fluent speech would follow. This is not, however, what one usually observes when decoding is disfluent. Typically, the utterances are slow and sometimes discontinuous; it appears that disfluent decoders are working on the task at the same time as they are producing the answer. To us, this suggests that the feed to the speech production chain does not seem to be whole word-forms, but smaller phonological components that are then combined inside the speech production chain. This argument also applies to the blending tasks that we will discuss. That is, we hypothesise that during blending phonological word forms are build inside the speech output system.

In this article, we have started by outlining the hypothesis that both decoding and blending utilise phonological encoding and that DD is related to difficulties to *unguided* phonological encoding in specific. We will next proceed to describe our observations of blending difficulties among children with DD. These observations are made in clinical settings, and our aim is not to provide rigorous data but to illustrate our theoretical argument. We will suggest that, first, in contrast with current consensus, there appear to be severe problems in the blending skills of children with DD in a clinical context. Second, we will introduce a set of new types of blending tasks that can be used to measure blending in settings that resemble reading situations. We maintain that the study of blending difficulties among children with DD might offer important knowledge not only about dyslexia but about the functioning of phonological encoding. In the section that follows, we will provide a more detailed discussion of the psycholinguistic model of speech production and unguided phonological encoding. We will then move on to examine a body of earlier research literature on the speech and language difficulties associated with reading problems and discuss how the psycholinguistic model of phonological encoding and speech monitoring could explain the key features of DD. Finally, we will briefly suggest how our hypotheses could be tested. We argue that the focus on the interplay between phonological encoding and internal speech monitoring provides a framework with which we can ask new questions about the phonological difficulties associated with DD.

Blending Difficulties Among Dyslexics

Measuring Blending

In previous research as well as in clinical practice, the principal method for measuring phonological blending abilities has been straightforward: participants are given phonological items, such as phonemes, syllables, or words, and are asked to blend those together. This method is used in many phonological awareness assessment tools such as CTOPP-2 (Wagner et al., 2013) or PAT-2 (Robertson & Salter, 2007), as well as in numerous in-lab assessment packs. While children with DD have shown some difficulties in these types of blending tasks, these difficulties have not stood out from difficulties in other areas of phonological awareness (Melby-Lervåg et al. 2012). This method of measuring blending skills treats blending as an isolated phenomenon, which is most probably intended, but it misses certain features that are essential when blending is used for reading purposes. In the context of reading, blending would be done in constantly changing phonological environments, by combining varying linguistic structures at a very swift pace. In contrast, in typical blending tasks the items are fairly uncomplicated, and the answers are not timed, which means that any difficulties with increasing phonological complexity and/or fluency are not measured.

Interestingly, when children with DD are exposed to more complex blending tasks, rather clear patterns of blending difficulties begin to emerge. This phenomenon is what we have observed in our clinical practice, where we have developed certain types of blending tasks to support the reading skills of Finnish children with DD. Our tasks target a very definite set of skills, namely building a larger phonological structure from smaller items without access to lexical representations and with a limited possibility to use orthographic strategies (including full-word recognition). To ensure that the tasks are completed by relying on phonological blending rather than orthographic or guessing strategies, we use non-words or meaningless syllables as target structures and/or provide the blended items entirely or partly orally. Moreover, by manipulating the pace of the tasks, we make it harder for the children to apply error-correction procedures before giving answers, which makes both blending errors as well as disfluency more readily observable. Finally, in our tasks we use changing phonological environments and vary between tasks where the children are given the task items orally, in writing, or partly orally and partly in writing.

By applying these types of blending tasks, we have observed major difficulties in blending among children with DD – in our case, Finnish children with DD that have fluent grapheme-phoneme associations to all letters of the alphabet and no apparent speech-sound problems. Among the children with DD at our clinic, not only are the blending difficulties common (rather than residual) but they also reflect certain patterns. We have divided the observed difficulties in blending tasks into two groups: *errors* and *laboriousness*. Although these groups of difficulties have a resemblance to the types of reading problems associated with DD - that is, accuracy- and rate-related problems (e.g. Lovett, 1984; Wolf & Bowers, 1999) - we have deliberately chosen to use different terminology as the observed phenomenon is not reading but blending.

Blending Errors

Significantly, the blending *errors* that children with DD make fall into categories that greatly resemble phonological error patterns that occur in the speech of children with phonological disorder (PD) (see e.g. Dodd, 2014), including substitution errors, assimilation, omissions, additions, and metatheses, usually occurring at the level of single phonemes. Speech errors in PD are not necessarily higher in phonotactic probability in comparison to error-free utterances, and this is also the case with blending errors – they are not necessarily higher in phonotactic probability in comparison to error-free blending. The observed error types occur both with vowels and consonants in various phonological contexts. While we have not tried to map out all types of potential errors based on our clinical observations, certain phonological environments, such as consonant clusters and diphthongs, seem to be more prone to errors. As with the errors in PD, it is conceivable that there are certain patterns depending on the phonotactics of the language in question. Our aim here is not to describe a set of typical patterns in Finnish but to present a few observations to illustrate some of the error types that children with DD at our clinic typically demonstrate. Moreover, what we see at the clinic are idiosyncratic patterns of errors that may vary from a general difficulty related to several phonemes to very narrow difficulties with one or two phonemes in specific syllable structures and combinations. The following examples are collected from different children with DD between ages 9-12.

To start with the case of consonant clusters, the errors that emerge often look like this: when a child is asked to blend a written syllable with another syllable given orally, for instance *sin* and /so/, the produced answer involves consonant errors (e.g. [sinto] instead of *sinso*). In our tasks, the errors do not necessarily emerge if only two items are blended (e.g. into one syllable/word) but become apparent when the same blending process is repeated in a series of items – which is closer to the process of reading out loud or producing continuous speech. We usually see errors emerging in a task in which the child is asked to continuously blend a list of written syllables with an orally given syllable, for instance, adding /so/ to the end of five different syllables *sin, ket, lok, mat, vel,* which would correctly result in *sinso, ketso, lokso, matso, velso* but may, in fact, come out as [sinto kesso kolso matso veltso]. That is, the blending difficulties emerge in the form of several different patterns, including substituting /s/

with /t/ in [sinto]/*sinso,* an assimilation in [kesso]/*ketso,* a metathesis in [kolso]/*lokso* and an addition of a /t/ in [veltso]/*velso*. It is important to note that the children do not have any trouble reading the first syllables without the orally added second syllable, yet in the blending process errors may also emerge in the orthographically given item. What is more, the same children do not have any difficulties in fluently *repeating* the blended targets if the full model is given, that is, they do not have any trouble producing the problematic sound combinations in speech (here blending a CVC syllable with a CV syllable beginning with /s/).

Despite the fact that errors are more likely to occur in more complex contexts, it is not uncommon that children with DD also struggle with blending two phonemes into a syllable if they need to do this repeatedly, and this is something that we often see if the child has difficulties with vowels. For example, in a task in which a child is asked to add a certain vowel after a set of written letters, the vowel begins to vary (e.g. written consonants *k, l, s, m, p* + sound /a/ should result in *ka la sa ma pa* but becomes [ka, la, si, ma, pe]). Vowel errors also emerge in a task where children need to blend two CV syllables (e.g. written syllables *pe so ri ma tu* + /ka/ should result in *peka, soka, rika, maka, tuka* but becomes [pek ka soki riki maki ka tuk ka]). Often there are detectable patterns, such as the strong tendency to substitute /i/ for /a/ in the previous example. Some children have a rather general difficulty here, whereas others are more prone to making vowel errors after certain consonants, and yet others have a greater tendency to make errors with certain vowels regardless of the imminent phonological context. In the previous example, apart from difficulties with vowels, the child also has difficulties with the length of the consonant (in Finnish the length of phonemes is a distinctive feature).

In the examples above, the difficulties have occurred while producing a series of single syllables or short non-words. However, with practice, children with DD often learn to blend items and to read fluently at the level of syllables or short words - that is, in simple phonological contexts. Yet their difficulties remain in more complex or alternating phonological environments. Creating blending tasks involving complex phonological environments that rely solely on orally given items is challenging due to restrictions of working memory. Thus, we often chart blending difficulties emerging in more complex phonological contexts in tasks where children need to blend an increasing number of syllables, most or all of which are given orthographically. Typically, a child may be fluent in reading syllables and blending syllables into short words (e.g. syllables *kor* and *lik* into [korlik]) but begins to struggle with longer words, for instance, in blending syllables *kor* + *lik* + *rap* which should result in *korlikrap* but becomes a series of erroneous attempts: [korliptak ... korlitpak]. The child in question has a specific difficulty with voiceless plosives /k/, /p/, /t/, as well as combinations of /l/ and /r/ sounds. Again, we should note that there are no difficulties in repeating the words after a full model. What we also see here are the immediate self-corrections, as the child monitors their attempts, spots the errors and tries to correct them.

Laboriousness of Blending

Among the children with DD at our clinic, an entirely different type of difficulty is the laboriousness of blending. This concerns children for whom the whole process of joining phonological items together is very laborious and slow. Although laboriousness can be observed in various blending contexts, it is often markedly present in complicated and/or rare phonological structures. Yet, as in the previous examples, it may also emerge in very specific contexts, related to specific phonemes, such as adding to /t/ sound to a list of syllables (e.g. *pen, kor, ril* + /t/ should result in *pent, kort, rilt* but becomes [pent ... kor ... ril ... t], where the process is not only very arduous but partly erroneous). Again, the child is fully able to read the given syllables fluently without the added sound, and if the child is asked to repeat the target structures, they have no difficulties with the sound combinations. That is, the difficulty is only present in the situation when blending occurs without an activated word model of the end product. As with the blending errors, the laboriousness seems to be a phenomenon that involves variation and idiosyncratic patterns: for some children, it may manifest in all kinds of blending tasks as a general difficulty while for others it only emerges in specific phonological contexts in which blending becomes more arduous.

To sum up, in our clinical practice we have observed major blending difficulties among children with DD in a context where they need to assemble larger phonological units from smaller ones without the possibility to rely (completely) on orthographic strategies or lexical representations. Each child that we have treated in our clinic has had blending difficulties, although the type and severity of these difficulties vary greatly. Some of the children with DD at our clinic mainly demonstrate error-proneness while others mainly struggle with laboriousness. For some children, difficulties only arise in one or two problematic phonological contexts. In contrast to children with very narrow difficulties, children with severe dyslexia typically demonstrate difficulties in a range of phonological contexts and their overall blending processes are characterised by laboriousness. We argue that these difficulties are not residual or random and thus raise significant questions about the role of blending skills in dyslexia.

Our clinical observations suggest a connection between blending skills and reading skills. First, in our tasks, blending difficulties are significantly more pronounced in the performance of children with DD, whereas normal readers learn these tasks quickly and are able to perform them fluently and with no or very few errors. Moreover, the gravity of blending difficulties corresponds closely with the gravity of reading difficulties, and the structures that cause difficulty in blending tasks also cause difficulty in reading. Finally, our clinical experience suggests that improvement in blending skills in our tasks also results in improvement in decoding. It should be noted that blending difficulties occur across different task types, that is, in tasks presented orally, in writing or partly orally and partly in writing. Thus, any mechanisms that cause these difficulties seem to be in operation in each case.

We suggest that these difficulties can be explained by applying psycholinguistic models of speech production (Levelt et al., 1999; Dell, 1986) where blending difficulties would emerge at the level of phonological encoding. This may seem obvious, given that the difficulties that we have observed occur at the level of phonology usually in terms of phonemic errors. Nevertheless, one might ask for further reasoning for locating blending processes in a psycholinguistic *speech* production model, rather than looking for explanations in other cognitive functions. As regards the types of blending tasks that we use, we argue that the difficulties in blending we have observed are not related to problems in visual processes, speech perception, phonological working memory, or attention control. First, blending difficulties emerge whether or not part of the items is given orthographically; thus, it seems that visual processes are not the root of the problem. Second, difficulties emerge whether or not speech perception was required in the task. Third, when support is provided to working memory and executive function, the problems with blending persist. Moreover, if there were difficulties in short term memory and/or executive function, we would expect random errors in tasks. Instead, what we have observed are regular and consistent difficulties in terms of errors and laboriousness. Thus, we propose that the source of blending difficulties is in the operation of phonological encoding. However, to be able to explain how blending is done without meaningful words as a starting point, the model of speech production also needs to describe *speech that is produced without reference to active word-forms*. To achieve this, we introduce a component that we call *unguided phonological encoding*.

Phonological Deficits in a Psycholinguistic Framework

Phonological Encoding and Unguided Phonological Encoding

Detailed descriptions of phonological encoding in the speech production chain can be found in Levelt et al. (1999), Roelofs & Meyer (1998) and Dell (1986). Briefly, in the speech production chain, the process of phonological encoding follows stages where a word form is accessed in the mental lexicon and goes first through morphological encoding. In phonological encoding, this word form is then retrieved piece by piece through two parallel and at least partly independent processes: the retrieval of segmental information (phonemes) and the retrieval of metrical information (number of syllables, stress, etc.). These processes are called segmental spell-out and metrical spell-out. Next, using segment-to-frame association the retrieved segments are combined in a metrical frame. This recombination is made incrementally, from left to right, and it can begin to operate without completed information of the word form, suspending or resuming the recombination process according to the availability of segmental or metrical information. There is some uncertainty about how much metrical information word forms in the mental lexicon contain. According to the most parsimonious view (Levelt et al, 1999), almost all metrical information is created during the encoding process. Whether or not this is true, at least part of metrical information is necessarily created during the phonological encoding process because the syllabification of a phonological word - that is, the end result of the phonological encoding process affected by morphological processes and the phonological environment of the word - does not always follow the syllabification of a lexical word. Additionally, there are processes related to prosody and stress. Once completed, the phonological word is ready for further processes in the speech production chain, including phonetic encoding and articulation. It is important to note that in the phonological encoding process the word form is not retrieved as a whole but it is combined from pieces. It is assumed that this piece-by-piece combination of word forms each time that they are used (spoken) serves the purpose of generating connected speech (Levelt, 1992).

There are several models (e.g. Hanley et al., 2004) that describe how speech can be produced without the involvement of the mental lexicon, which takes place, for example, in auditory repetition of non-words. To our knowledge, however, there has been no discussion of how phonological encoding operates when producing speech without a reference to any activated word form(s). This sort of situation occurs, for example, when generating impromptu utterances of nonsense-speech. The fact that this sort of speech production is possible indicates that the phonological encoding process can operate in a piece-by-piece fashion without a model of the end-product. Moreover, this means that there is a notable amount of flexibility in the processes that take place in phonological encoding.

Blending and decoding, we argue, are other instances where phonological encoding operates without access to a complete word form. Instead, the input consists of small phonological components (e.g. phonemes or syllables) and the knowledge of their order. How would this change the encoding process? The input would be taken into the segmental and metrical spell-out processes. As in the typical encoding process (using the word form from the mental lexicon), also here the phonemes would not necessarily be retrieved at once, but the process would begin with some of the components, and more would be retrieved during the process. Similarly to the typical phonological encoding process, the segment to frame association would be done incrementally, from left to right, and it would be operated without all information being available, suspending or resuming the process according to the availability of segmental or metrical information. In contrast with the typical process, during the segment to frame association there would be less knowledge or no knowledge at all about the metrical frame. The missing metrical information would have to be created during phonological encoding, which might or might not change the process. After combining the segments into a metrical frame, the rest of the processes (creating intonation, stress etc.) would be dependent on the available information. In sum, during an oral blending task or during decoding the phonological encoding process would differ from the typical situation in that there would be 1) less knowledge on the metrical frame and 2) less knowledge about stress. We call this type of process *unguided phonological encoding* to differentiate it from the way in which phonological encoding operates during typical speech production. We argue that in the unguided process - in which information about the lexical word form is missing - the speech monitoring and error correction processes cannot function in the same manner as they do in typical speech production.

Speech Monitoring

In psycholinguistic models of speech production, phonological encoding functions in close connection with monitoring and error correction processes that take place during speech production (for reviews: Gauvin & Hartsuiker, 2020; Postma, 2000). These processes concern both the form and the content of speech. During speech production, the monitoring processes take place at two different stages: the external loop monitors the output (uttered speech), and the internal loop monitors the production of phonological word forms before initiating the motoric planning of speech production. In the internal loop, monitoring is targeted at the phase where phonological forms are first assembled (e.g. Levelt et al., 1999) and it enables regulation and fast corrections before the motoric planning of speech is carried out (Gauvin & Hartsuiker, 2020; Nozari, 2018). The phonological error correction is found to be among the earliest forms of speech related self-correction (Clark, 1982), observable even before the age of two (Clark, 1982; Forrester, 2008; Laakso, 2006). At the age of three, children already use internal and external speech monitoring flexibly in different situations (Manfra et al, 2016).

It is widely agreed that the external monitoring utilises the speech comprehension system. However, there is some disagreement on the mechanisms of internal monitoring. Some argue that internal monitoring also relies on the speech comprehension system (Levelt, 1993; Roelofs, 2020). According to this view, the phonological word assembled in phonological encoding is sent to a separate monitoring unit (a monitor) that compares this data with the corresponding representations. However, this view conflicts with neuropsychological evidence that suggests a dissociation between the internal monitoring and the monitoring of the speech of others, as well as a double dissociation between comprehension skills and error detection in one's own speech (see Nozari et al., 2011 for a review of the evidence). According to another view, internal monitoring uses process-related information (Nozari et al., 2011; Nozari, 2020; Gauvin & Hartsuiker; 2020). In this view, internal monitoring focuses on the amount of conflict in phoneme selection. As regards the monitoring of form, this model assumes a layer of lexical nodes and another layer of phoneme nodes with reciprocal connections. A conflict may be detected at the lexical node based on the amount of feedback from the phoneme nodes. A small amount of feedback would signal that the correct phonemes are not activated. This account assumes that internal monitoring is deeply interconnected with the mechanisms of phonological encoding. Indeed, Nozari (2018) has suggested that error detection and correction are only a small part of what internal monitoring does. She proposes that internal monitoring primarily assesses a need for control over the various stages of speech production. If conflict-related activity is increased at any stage, more control is allocated to resolve the possible problem before any errors emerge. Different accounts of internal monitoring are not in conflict with each other. On the contrary, they may complement each other. Recently, Nozari (2024) has proposed a multi-process view of monitoring with several different monitoring mechanisms operating simultaneously.

Importantly, all the discussed mechanisms of internal monitoring are dependent on lexical-level information, that is, on knowledge of the word-form that is being encoded. In unguided phonological encoding – as in oral blending tasks or in impromptu non-sense speech – information about the complete word form is missing, which means that internal monitoring cannot operate in the same manner as in typical speech production. We suggest that this makes the process of unguided phonological encoding more prone to difficulties in comparison with phonological encoding in typical speech production.

Many difficulties in blending and decoding in dyslexia could be explained by errors in phoneme selection and difficulties in finding the correct metrical frame during phonological encoding. In each case, the regulatory support provided by internal monitoring during typical speech production may compensate for the difficulties and problems that become apparent only during unguided encoding. Additionally, it seems that the very process of combining the retrieved segments (e.g. phonemes) may be prone to difficulties (slowness, arduousness) without regulatory support from internal monitoring. This is reflected in the observed slowness in combining even single phonemes (/k/, /a/ -> /ka/) or syllables (/ka/, /to/ -> /kato/) in blending and decoding tasks. Importantly, examination of the differences between phonological encoding and unguided phonological encoding provides a new tool for investigating the mechanisms of creating output representations.

Phonological Encoding and Speech Monitoring in Developmental Dyslexia

There exists a wide range of theoretical proposals about the causes of DD. The phonological deficit theory, discussed in the introduction, remains as the most influential one. Other proposals include problems in rapid temporal processing (Tallal, Miller, & Fitch, 1993), magnocellular abnormalities (Stein & Walsh, 1997), sluggish attentional shifting (Hari & Renvall, 2001), anchoring difficulties (Ahissar, 2007), procedural learning problems (Nicolson & Fawcett, 2007), and a phonological access deficit (Ramus & Szenkovits, 2008). None of these theories or proposals has been able to provide a sufficient account of the phenomena related to DD and currently many scientists have turned their attention to multifactorial explanations of dyslexia (Pennington, 2006; McGrath et al. 2020). Here we will take an alternative route and argue that reading and speech production share certain core mechanisms that can be located in the processes of phonological encoding, and that reading difficulties in DD are a consequence of deficits in these processes.

We hypothesise that DD (in absence of other language difficulties, including PD and developmental language disorder, DLD) results from systematic problems in phonological encoding (related to error-proneness and laboriousness) but with intact internal speech monitoring. During language development speech monitoring has compensated for the encoding problems and, as a result, no major speech difficulties have emerged. For these children, encoding difficulties become apparent only when encoding needs to function without the availability of end-product information (complete word-forms) and thus with reduced monitoring possibilities. That is, when reading practice begins and unguided phonological encoding becomes essential. We suggest that decoding difficulties result from deficits in unguided phonological encoding. We will next discuss how this hypothesis fits with language-related features of dyslexia.

Our theory would predict difficulties among dyslexics in tasks that require a creation of output representations in difficult conditions, that is, that either utilise unguided phonological encoding or utilise typical phonological encoding in such a way that monitoring is not able to compensate for the encoding problems (speeded conditions or complex and novel words/phrases). In previous research, dyslexics have shown difficulties in tasks measuring certain speech abilities, rapid naming, short-term memory, and phonological awareness. We note that the used tasks have required the forming of output representations. To be more specific, they have involved either phonological encoding in difficult conditions or unguided phonological encoding. We argue that the findings in these areas can be explained by our hypothesis.

To start with speech abilities, a consequence of our model is that dyslexics should have speech difficulties in situations where speech monitoring cannot compensate for deficits in encoding. These difficulties could emerge in two forms: error-proneness and slowness (slowness caused by the laboriousness of blending). Indeed, there is a notable amount of evidence of such difficulties: dyslexics make speech production errors and misarticulations in phonologically complex words (Snowling, 1981; Brady et al., 1989). They cannot produce simple or complex phrases as quickly as normal readers, and in complex phrases they make more mistakes (Catts, 1989). They are slow in syllable repetition (Wolff et al., 1990). They have difficulties in sentence repetition (Moll et al., 2015) and in non-word repetition (Kamhi et al., 1988), as well as many sorts of small anomalies in speech (McArthur et al., 2000; Vellutino, 1979). In his discussion of dyslexics' slowness and error-proneness in speech repetition tasks, Catts (1989) concluded that “dyslexics may have difficulties in the planning stage of speech production”. However, to our knowledge, this idea has not been developed further.

Second, a slower performance in rapid naming tasks could also be explained by problems in phonological encoding, laboriousness in particular. There are at least three possible factors causing slow performance. First, laboriousness in phonological encoding can potentially slow down expression of words (as in Catts, 1989). Second, labouriousness in unguided phonological encoding might cause difficulties in producing words very close to each other. Hypothetically, in the case of naming numbers, for instance, a dyslexic might be inclined to keep the phonological entities apart (e.g. /tu:/ /faɪv/ /naɪn/) whereas a non-dyslexic is able to blend them together (e.g. /tu:faɪvnaɪn/), thus speeding up the performance. Third, encoding problems demand an increased use of cognitive resources for speech regulation in speech production, which could slow down the progression in the task. We should note that difficulties resulting from these three factors would be more evident in speeded, serial tasks (in contrast with individual naming tasks) because in these tasks the phonological environment would be more challenging and variating and slowness in speech production would be easier to observe. This, indeed, is a pattern of performance observed in dyslexics (Araújo & Faísca, 2019). It has also been shown that dyslexics' difficulties in rapid automatised naming (RAN) tasks are more pronounced in conventional naming tasks that require the articulation of specific names in comparison to RAN-like categorisation tasks (cancellation or RAN yes/no tasks) (Georgiou et al., 2013). This is in line with our hypothesis.

Third, phonological encoding problems would cause difficulties in many tasks that measure short-term or working memory. As pointed out by Elliot & Grigorenko (2024), to succeed in such tasks one must be able to protect memory representations from interference or decay. As these tasks typically involve spoken responses, any difficulties in phonological encoding would introduce interference compromising the overall performance. For example, in non-word repetition tasks, dyslexics are typically able to repeat short non-words correctly and difficulties only emerge when repeating long non-words (e.g. Marshall & van der Lely, 2009). The interpretation has been that the poor performance results from a narrower short-term memory rather than production problems. However, the interference due to encoding difficulties would provide an alternative explanation of the situation. As we have demonstrated in our examples, those difficulties are often evident in more demanding phonological environments such as long non-words. Apart from interference, phonological encoding problems could also affect performance in memory tasks in another way. In a serial task involving verbal items (e.g. two, five, eight, two, six, eight), intact phonological encoding enables the combination of task items into larger phonological entities in a flexible manner to support memory performance (e.g. /tu:faɪv/, /eɪttu:/, /sɪikseɪt/ or /tu:faɪveɪt/, /tu:sɪkseɪt/). This strategy is weakened if one has unguided phonological encoding problems. Additionally, as with rapid naming tasks, the increased demand of cognitive resources for speech regulation may limit performance in memory tasks.

Finally, there is plenty of evidence that DD is related to difficulties in phonological awareness, even if this does not concern all dyslexics (Saksida et al. 2016; Mundy & Hannant, 2020; Dębska et al. 2022). It has been suggested that all phonological operations that are present in phonological awareness tasks reflect one underlying skill, as the evidence suggests that all phonological tasks applied to material of similar complexity (same linguistic level) are highly interrelated (Wagner et al. 1997; Stahl & Murray, 1994; Stanovich et al, 1984; Schatschneider et al. 1999; Anthony & Lonigan, 2004). However, internal monitoring processes that take place during phonological awareness tasks have not been considered when interpreting the results. We suggest that the constant monitoring and error correction procedures are in use during phonological awareness tasks. This means that while the task itself may require one phonological operation, such as segmenting, internal monitoring enables the use of other operations, such as blending for checking the answer. Further, if any error correction procedures are needed, these may, again, involve further phonological operations, such as phoneme substitution. Thus, there may exist a number of various phonological operations as well as monitoring and correction procedures that are carried out before the answer is given. This might explain the observed interconnectedness of different phonological operations. In any case, the process of unguided phonological encoding (blending) is directly involved in several phonological operations and thus it is very likely that difficulties in phonological encoding would have an effect on performance in phonological awareness tasks.

It is also important to consider the role of phonological encoding in phonological awareness tasks in general. It is probable that all phonological tasks associated with phonological awareness utilise phonological encoding. Let us consider, for example, segmenting: school children, including dyslexics, are taught to segment words by producing the word aloud or silently in one's mind, making the rhythm of the word more pronounced in the utterance while consciously attending to this process. In terms of the psycholinguistic speech production framework this means that children are instructed to produce an output representation of the word (the process of phonological encoding) while attending to and manipulating this process in a certain way. Assuming a difficulty in the process of phonological encoding would, indeed, predict difficulties also in segmenting (although not as severe, as in segmenting speech monitoring can operate normally as the target word is known). The same argument can be applied to all phonological operations that are related to phonological awareness. The most apparent strategy, and perhaps even the only possible strategy, for performing them is to produce an output representation while attending to and manipulating the process. In sum, we argue that our theory would predict difficulties in all phonological awareness tasks among dyslexics. However, our framework also offers an explanation for why difficulties in phonological awareness have not been found among all dyslexics: when measuring phonological awareness, the fluency (speed) of the process has not been taken into account. If one assumes that the underlying deficit exists in permanent representations, fluency would play no role. However, if the difficulty is in the *process* of creating output representations, as we argue, measuring fluency would be necessary to chart all aspects of the difficulty.

While difficulties in all the four language-related areas discussed above – certain speech abilities, rapid naming, short-term memory, and phonological awareness – are strongly associated with DD, it is important to note that this is not true of many other phonology-related areas. Dyslexics speak relatively normally, and they show normal or near-normal performance on categorical perception (Hazan et al. 2009), lexical quality recognition (Marshall et al., 2010), prosody (Marshall et al., 2009) and context sensitivity of speech perception (Blomert et al., 2004). The pattern of normal performance in these areas is not sufficiently explained by the phonological theories of DD (for a detailed discussion see Ramus & Ahissar, 2012). However, these varying patterns of performance – difficulties in certain areas and normal performance in others – could be explained by a model where the interplay between phonological encoding and internal monitoring is taken into account.

Testing Our Hypotheses

The first hypothesis in our work is that phonological word-forms during decoding (and blending) are built within the speech output system utilising phonological encoding pre-lexically. We will next discuss the possibilities to test this hypothesis and the evidence from earlier research that there already exists concerning this issue. In our discussion, we also aim to distinguish our hypotheses from other phonological theories of dyslexia.

The role and temporal dynamics of phonological encoding in speech production have been investigated with various methods and there is plenty of knowledge on the subject (for an overview: Kerr et al., 2023). Among the applied methods, we would find dual task set-ups and interference paradigms particularly useful to study the operation of phonological encoding during decoding. For example, we would assume interference in both dual task setups and interference paradigms with silent decoding and speech production, since phonological encoding could not operate on decoding and on speech production at the same time. To our knowledge, this kind of experiments have not been carried out in earlier research. However, there are several studies that report the involvement of phonological encoding in tasks that are closely related to reading. Roelofs (2004) demonstrated in three experiments that the same serial phonological encoding mechanism is used in naming objects and reading their names. Truman and Hennessey (2006) found that dyslexics' rapid naming was more greatly facilitated by phonological cues in comparison to controls, suggesting that slowness in naming is affected by problems in phonological encoding. There is also evidence that masked onset priming effect, which is closely related to the serial nature of the reading process, actually takes place within the speech output system (Kinoshita, 2000; Kinoshita & Woollams, 2002).

Evidence to our claim that decoding utilises phonological encoding pre-lexically could also be provided by a line of research that is called phonological coding. In short, this refers to the situation where during the reading process there is an experience of an "inner voice" or "hearing the words in our heads". This phenomenon is also studied under the concepts of subvocalisation, subarticulation, inner speech, covert speech, speech recoding and phonological recoding. A vast amount of literature in these areas of research indicates that at the very early stages of reading, before lexical involvement, the reading process is sequential, sound based and follows the features of speech (for reviews: Pollatsek, 2015; Leinenger, 2014). However, the interpretations about how phonological coding relates to the speech production chain have been rather cautious – yet, it seems unlikely that there would be no relation at all. This is demonstrated by a statement by Pollatsek (2015): "However, it is not clear that anyone so far has successfully been able to clearly ... demonstrate that phonological coding occurs without any involvement—either overt or covert—from the speech system". We note that nearly all the evidence in these research areas supports the hypothesis that reading (and decoding) utilises the speech output system.

There may even be a direct way to investigate if phonological encoding is utilised during decoding. Research on covert oral behaviour shows that muscles related to speech movements are activated during many language-related situations including silent reading, verbal thinking and verbal meditation (For a review: McGuigan, 1970). On the other hand, nonverbal thinking, music listening, word listening and story listening do not cause similar covert oral behaviour (McGuigan & Bailey, 1969; McGuigan, 1972). There are many interpretations of these findings, but to our knowledge no one has suggested that this covert oral behavior could be related to forming of output representations. In the speech production chain, the forming of output representations (phonological encoding) is followed by motor planning of speech movements. Thus, it is possible that phonological encoding causes a bleeding effect to further down the speech production chain, which could be measured by muscle activation. This hypothesis would be straightforward to test. If this hypothesis would turn out to be true, it would support our hypothesis that phonological encoding is utilised in decoding. It would also enable studying directly whether phonological encoding is utilised during blending or other phonological operations and, perhaps, even to assess the temporal dynamics of many phonology-related processes.

The second part of our hypothesis – that DD is related to difficulties in unguided phonological encoding – could be examined by comparing unguided phonological encoding abilities between dyslexics and typical readers. Studying dyslexics' performance in both typical phonological encoding tasks (e.g. word repetition, non-word repetition) and unguided phonological encoding tasks (e.g. blending tasks with sufficient complexity in which both fluency and correctness are measured) would also allow for distinguishing our theory from other phonology-related hypotheses of dyslexia. Our hypothesis predicts that dyslexics would demonstrate mild difficulties in typical encoding and severe difficulties in unguided encoding. Theories that assume deficit in permanent representations, speeded access to representations or short-term memory functions would not share these predictions.

There are also several predictions concerning blending that are more closely related to clinical work and could be used to test particular aspects of our hypotheses. These include:

1. There is a high correlation between a progress in blending skills and progress in decoding skills.

2. Progress in blending skills would lead to progress in rapid automatised naming and in many tasks that are used to measure verbal short-term memory.

3. Progress in phonological operations that include blending (e.g. blending, phoneme manipulation, syllable manipulation) would produce a more significant progress in decoding skills in comparison to progress in such phonological skills that do not include blending (e.g. segmenting, phoneme detection, syllable detection).

Implications and Future Directions

In this article, we have discussed the phonological deficits occurring in DD in a context of well-known and extensively tested model(s) of speech production (Levelt et al., 1999; Dell, 1986) by focusing on the interplay between the forming of output representations (phonological encoding) and internal speech monitoring. We have provided a conceptually feasible approach of how deficits in phonological encoding may cause the observed difficulties in reading. Our theory posits that speech production and reading share fundamental mechanisms in building up phonological forms. This is in line with the literature in reading research, indicating that at the very early stages of reading, before lexical involvement, the reading process is sequential, sound based and follows the features of speech (for reviews: Pollatsek, 2015; Leinenger, 2014). We will next discuss a few implications of our model for future research.

First, we have extended the definition of phonological encoding by introducing the concept of unguided phonological encoding. It is obvious that speech can be produced (and output representations can be formed) without access to complete word-forms. However, current psycholinguistic models of speech production have not described this phenomenon. Scrutinising the process of unguided phonological encoding would enable novel ways to examine the interplay between phonological encoding and speech monitoring. Moreover, research on unguided phonological encoding could provide a new perspective to study reading. For example, when producing long utterances, phonological encoding is typically carried over simultaneously with the articulation of previous items (Wheeldon & Levelt, 1995; Levelt & Meyer, 2000). Adding reading to this line of research (for example by using dual-task setups with reading and speech production) could offer intriguing possibilities for investigating how phonological encoding operates during continuous speech (or possibly, continuous reading).

Second, our hypotheses provide a new perspective on the process of learning to read. We suggest that when first learning to read, all learners will need to master the forming of output representations with reduced regulatory possibilities by internal monitoring (that is, to master unguided phonological encoding). We do not claim that this is the only process that children work on when they are learning to read. They also need practice to gain fluent letter-sound correspondences, to make direct mappings of letter sequences to sound patterns (including whole word recognition regarding to the most common words), and on how to deal with irregularities of the writing system. However, we argue that achieving fluency in unguided phonological encoding is essential to fluent reading and the failure to achieve it will cause the reading difficulties associated with dyslexia.

Our first hypothesis has consequences for the theories of visual word recognition. Current theories of reading aloud (for overviews: Norris 2013; Perfetti & Helder, 2022) share the assumption that phonological word-forms during reading and decoding are built outside the speech system. This assumption is challenged by behavioural (Roelofs, 2004; Truman & Hennessey, 2006; Kinoshita, 2000; Kinoshita & Woollams, 2002) and neuroimaging (Uno et al., 2025; Cornelissen et al., 2009) evidence indicating that speech output system is active earlier than would be expected based on the current theories of reading aloud. Although we have not introduced an alternative model of word recognition, our first hypothesis offers possibilities for constructing new types of models and, potentially, for adding to our understanding of typical reading process as well.

We also note that the psycholinguistic framework that focuses on the interplay between phonological encoding and internal speech monitoring opens up new possibilities for targeted interventions. According to our theory, interventions in DD should emphasise the practice of unguided phonological encoding skills, that is, blending skills. As there seems to be individual variation in blending difficulties, the practice should be planned individually and targeted to those linguistic structures and phonological environments that are problematic. Our clinical experience favors the described intervention approach, but more rigorous methods are necessary to assess its effectiveness.

Lastly, we will briefly address two other disorders, developmental language disorder (DLD) and phonological disorder (PD). These two disorders have a notable diagnostic and genetic overlap with DD (for a review: Pennington & Bishop, 2009) and all these three disorders are associated with phonological deficits. However, the role of phonological deficits remains under dispute also in DLD and PD. We believe that examining the interplay between internal speech monitoring and phonological encoding could also benefit the research of these two disorders.

We speculate that both DLD and PD may be associated with delayed internal speech monitoring development. Without support from internal speech monitoring any difficulties in phonological encoding – such as the error-proneness and labouriousness – would be observable in speech production. We consider it possible that the combination of difficulties in phonological encoding and delayed internal monitoring is the cause of early speech production difficulties in PD and DLD. Yet, there is very little research on the internal speech monitoring among children with DLD or PD. In general, the role of internal speech monitoring in language development is a neglected research area. In their studies, Navarro-Ruiz and Rallo-Fabra (2001; 2015) have found that in comparison with typically developing children, children with DLD show less metalinguistic, morphological, or syntactic self-repair and almost no phonological self-repair at all in their speech. This pattern of results could be explained by delayed internal speech monitoring development. We are not aware of any research on the role of internal monitoring in PD. However, it is known that the speech perception difficulties in PD are not as severe as production difficulties (for a review: Hearnshaw et al., 2019), suggesting that output representations are more erroneous than lexical representations. This may indicate problems in internal monitoring.

We realise that our article raises more questions than it answers. Our qualitative observations of blending difficulties among children with DD are made in clinical settings and, consequently, only offer a starting point for examining the relationship between phonological encoding and reading. Also, there is little existing research on the two key processes that we have discussed, that is, phonological encoding and internal speech monitoring in relation to language development and language-related disorders. Nevertheless, we believe that it is exactly the focus on these processes that could provide a more comprehensive understanding of the phonological deficits in DD, and perhaps also in PD and DLD, as well as offer new insights for targeted interventions.

References

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences, 11*(11), 458–465. https://doi.org/10.1016/j.tics.2007.08.015

Anthony, J.L., & Lonigan, C.J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96*(1), 43–55. https://doi.org/10.1037/0022-0663.96.1.43

Araújo, S., & Faísca, L. (2019). A meta-analytic review of naming-speed deficits in developmental dyslexia. *Scientific Studies of Reading, 23*(5), 349–368. https://doi.org/10.1080/10888438.2019.1572758

Blomert, L., Mitterer, H., & Paffen, C. (2004). In search of the auditory, phonetic, and/or phonological problems in Dyslexia. *Journal of Speech, Language and Hearing Research*, 47(5), 1030–1047. https://doi.org/10.1044/1092-4388(2004/077)

Brady, S., Poggie, E., & Rapala, M.M. (1989). Speech repetition abilities in children who differ in reading skill. *Language and Speech, 32*(2),109–122. https://doi.org/10.1177/002383098903200202

Bus, A.G., & Van IJzendoorn, M.H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational*

Psychology, 91(3), 403–414. https://doi.org/10.1037/0022-0663.91.3.403

Catts, H.W. (1989). Speech production deficits in developmental dyslexia. *Journal of Speech and Hearing Disorders, 54*(3), 422–428. https://doi.org/10.1044/jshd.5403.422

Clark, E.V. (1982). Language change during language acquisition. In M.E. Lamb, & A.L. Brown (eds) *Advances in Developmental Neuropsychology (*pp. 171-195). Hillsdale, New York: Lawrence Erlbaum.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cornelissen, P. L., Kringelbach, M. L., Ellis, A. W., Whitney, C., Holliday, I. E., & Hansen, P. C. (2009). Activation of the left inferior frontal gyrus in the first 200 ms of reading: evidence from magnetoencephalography (MEG). *PloS One*, *4*(4), e5359. https://doi.org/10.1371/journal.pone.0005359

Dębska, A., Łuniewska, M., Zubek, J., Chyl, K., Dynak, A., Dzięgiel-Fivet, G., Plewko, J., Jednoróg, K., & Grabowska, A. (2022). The cognitive basis of dyslexia in school-aged children: A multiple case study in a transparent orthography. *Developmental Science*, *25*, e13173. https://doi.org/10.1111/desc.13173

De Gelder, B., & Vroomen, J. (1991). Phonological deficits: Beneath the surface of reading-acquisition problems. *Psychological Research, 53*(1), 88–97. https://doi.org/10.1007/bf00867336

Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295x.93.3.283

Dodd, B. (2014). Differential diagnosis of Pediatric speech Sound Disorder. *Current Developmental Disorders Reports, 1*(3), 189–196. https://doi.org/10.1007/s40474-014-0017-3

Elbro, C. (1998). When readingis 'readn' or somthn. Distinctness of phonological representations of lexical items in normal and disabled readers. *Scandinavian Journal of Psychology, 39*(3), 149–153. https://doi.org/10.1111/1467-9450.393070

Elliott, J. G., & Grigorenko, E. L. (2024). *The Dyslexia Debate Revisited*. Cambridge University Press.

Forrester, M.A. (2008). The Emergence of Self-Repair: a case study of one child during the early preschool years. *Research on Language and Social Interaction, 41*(1), 99–128. https://doi.org/10.1080/08351810701691206

Gauvin, H.S., & Hartsuiker, R.J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition, 3*(1). https://doi.org/10.5334/joc.81

Georgiou, G. K., Parrila, R., Cui, Y., & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading? *Journal of Experimental Child Psychology, 115*(1), 218-225. https://doi.org/10.1016/j.jecp.2012.10.015

Hanley, J.R., Dell, G. S., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology, 21*(2–4), 147–158. https://doi.org/10.1080/02643290342000339

Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends in Cognitive Sciences, 5*(12), 525–532. http://dx.doi.org/10.1016/S1364-6613(00)01801-5

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *Journal of Speech Language and Hearing Research, 52*(6), 1510–1529. https://doi.org/10.1044/1092-4388(2009/08-0220)

Hearnshaw, S., Baker, E., & Munro, N. (2019). Speech Perception Skills of Children with Speech Sound Disorders: A Systematic Review and Meta-Analysis. *Journal of Speech Language and Hearing Research, 62*(10), 3771–3789. https://doi.org/10.1044/2019_jslhr-s-18-0519

Hulme, C., & Snowling, M. (1992). Deficits in output phonology: An explanation of reading failure?. *Cognitive Neuropsychology, 9*(1), 47-72. https://doi.org/10.1080/02643299208252052

Kamhi, A.G., Catts, H.W., Mauer, D., Apel, K., & Gentry, B.F. (1988). Phonological and spatial Processing abilities in Language- and Reading-Impaired children. *Journal of Speech and Hearing Disorders, 53*(3), 316–327. https://doi.org/10.1044/jshd.5303.316

Keating, P. (2000). A phonetician's view of phonological encoding, *LabPhon7*. University of Nijmegen, 29 June. Written version retrieved from https://linguistics.ucla.edu/people/keating/LP7-talk.pdf

Kerr, E., Ivanova, B., & Strijkers, K. (2023). Lexical access in speech production: Psycho-and neurolinguistic perspectives on the spatiotemporal dynamics. In Hartsuiker, J. & Strijkers, K. (Eds), *Language Production* (pp. 32-65). Routledge.

Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. *Psychonomic Bulletin & Review, 7,* 133-141. https://doi.org/10.3758/BF03210732

Kinoshita, S., & Woollams, A. (2002). The masked onset priming effect in naming: Computation of phonology or speech planning?. *Memory & Cognition, 30*(2), 237-245. https://doi.org/10.3758/BF03195284

Laakso, M. (2006). Kaksivuotiaiden lasten oman puheen korjaukset keskustelussa [Self-repair in conversation by 2-year-olds]. *Puhe Ja Kieli [Speech and Language,. 26,* 123-135.

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin, 140*(6), 1534–1555. https://doi.org/10.1037/a0037830

Levelt, W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*(1–3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-j

Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(01), 1-38. https://doi.org/10.1017/s0140525x99001776

Levelt, W. J. M. (1993). *Speaking: From intention to articulation*. MIT press.

Levelt, W. J. M., & Meyer, A. S. (2000) Word for word: Multiple lexical access in speech production, *European Journal of Cognitive Psychology, 12*(4), 433–452. https://doi.org/10.1080/095414400750050178

Lovett, M.W. (1984). A developmental perspective on reading dysfunction: Accuracy and rate criteria in the subtyping of dyslexic children. *Brain and Language, 22*(1), 67-91. https://doi.org/10.1016/0093-934X(84)90080-4

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of dyslexia, 53,* 1-14. https://doi.org/10.1007/s11881-003-0001-9

Manfra, L., Tyler, S.L., & Winsler, A. (2016). Speech monitoring and repairs in preschool children's social and private speech. *Early Childhood Research Quarterly,* *37*, 94–105. https://doi.org/10.1016/j.ecresq.2016.04.004

Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Child Psychology*, *66*(2), 211–235. https://doi.org/10.1006/jecp.1997.2383

Marshall, C.R., Harcourt-Brown, S., Ramus, F., & van der Lely, H.K.J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language & Communication Disorders, 44*(4), 466–488. https://doi.org/10.1080/13682820802591643

Marshall, C.R., Ramus, F., & van der Lely, H.J.K. (2010). Do children with dyslexia and/or specific language impairment compensate for place assimilation? Insight into phonological grammar and representations. *Cognitive Neuropsychology, 27*(7), 563–586. https://doi.org/10.1080/02643294.2011.588693

Marshall, C.R., & van der Lely, H.K.J. (2009). Effects of word position and stress on onset cluster production: Evidence from typical development, specific language impairment, and dyslexia. *Language, 85*(1), 39–57. https://doi.org/10.1353/lan.0.0081

McArthur, G.M., Hogben, J.H., Edwards, V.T., Heath, S.M., & Mengler, E.D. (2000). On the "specifics" of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry, 41*(7), 869–874. https://doi.org/10.1111/1469-7610.00674

McGuigan, F. J. (1970). Covert oral behavior during the silent performance of language tasks. *Psychological Bulletin*, *74*(5), 309. https://doi.org/10.1037/h0030082

McGuigan, F. J. (1972). *Covert Response Patterns in Processing Language Stimuli.* Final Report. Institute of Education Sciences. https://eric.ed.gov/?id=ED066980

McGuigan, F. J., & Bailey, S. C. (1969). Covernt response patterns during the procession of language stimuli. *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, *3*(4). https://doi.org/10.30849/rip/ijp.v3i4.569

McGrath, L.M., Peterson, R.L., & Pennington, B.F. (2020). The multiple deficit model: Progress, problems, and prospects. *Scientific Studies of Reading, 24*(1), 7–13. https://doi.org/10.1080/10888438.2019.1706180

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*(2), 322–

352. https://doi.org/10.1037/a0026744

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, *64*(2), 199–231. https://doi.org/10.1006/jecp.1996.2343

Moll, K., Hulme, C., Nag, S., & Snowling, M.J. (2015). Sentence repetition as a marker of language skills in children with dyslexia. *Applied Psycholinguistics, 36*(2), 203–221. https://doi.org/10.1017/s0142716413000209

Mundy, I. R., & Hannant, P. (2020). Exploring the phonological profiles of children with reading difficulties: A multiple case study. *Dyslexia*, 26(4), 411-426. https://doi.org/10.1002/dys.1667

Navarro-Ruiz, M.I., & Rallo-Fabra, L. (2001). Characteristics of mazes produced by SLI children. *Clinical Linguistics & Phonetics, 15*(1–2), 63–66. https://doi.org/10.3109/02699200109167632

Navarro-Ruiz, M. I., & Rallo-Fabra, L. (2015). A developmental study of self-repairs in Spanish normal-speaking children and comparison with a case study of specific language impairment. *Proceedings of the International Symposium on Monolingual and Bilingual Speech*, 239.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? Trends in Neurosciences, *30*(4), 135–141. https://doi.org/10.1016/j.tins.2007.02.003

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences, 17*(10), 517-524. https://doi.org/10.1016/j.tics.2013.08.003

Nozari, N. (2018). How special is language production? Perspectives from monitoring and control. In K.D. Federmeier & D.G. Watson (Eds.), *Psychology of Learning and Motivation,* vol. 68, 179–213. Academic Press. https://doi.org/10.1016/bs.plm.2018.08.006

Nozari, N. (2020). A comprehension- or a production-based monitor? Response to Roelofs (2020). *Journal of Cognition, 3*(1). https://doi.org/10.5334/joc.102

Nozari, N. (2024). An update on monitoring, control, and repair in language production. *PsyArXiv*. https://doi.org/10.31234/osf.io/rmptw

Nozari, N., Dell, G.S., & Schwartz, M.F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63*(1), 1–33. https://doi.org/10.1016/j.cogpsych.2011.05.001

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*(5), 403-410. https://doi.org/10.1177/0963721417702419

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*(2), 385-413. https://doi.org/10.1016/j.cognition.2006.04.008

Pennington, B.F., & Bishop, D.V.M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology, 60*(1), 283–306. https://doi.org/10.1146/annurev.psych.60.110707.163548

Perfetti, C., & Helder, A. (2022). Progress in reading science: Word identification, comprehension, and universal perspectives. In M. J. Snowling, C. Hulme, & K. Nation (Eds.), *The science of reading: A handbook* (2nd ed., pp. 5–35). Wiley. https://doi.org/10.1002/9781119705116.ch1

Pollatsek, A. (2015). The Role of Sound in Silent Reading. *The Oxford Handbook of Reading*, 185–201. https://doi.org/10.1093/oxfordhb/9780199324576.013.12

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition, 77*(2), 97–132. https://doi.org/10.1016/s0010-0277(00)00090-1

Ramus, F. (2001). Outstanding questions about phonological processing in dyslexia. *Dyslexia, 7*(4), 197-216. https://doi.org/10.1002/dys.205

Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology, 29*(1–2), 104–122. https://doi.org/10.1080/02643294.2012.677420

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology, 61*(1), 129–141. https://doi.org/10.1080/17470210701508822

Robertson, C., & Salter, W. (2007). *The phonological awareness test 2 (PAT 2)*. PRO-ED.

Roelofs, A. (2020). Self-Monitoring in Speaking: In defense of a comprehension-based account. *Journal of Cognition, 3*(1). https://doi.org/10.5334/joc.61

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition, 32*, 212-222. https://doi.org/10.3758/BF03196853

Roelofs, A., & Meyer, A.S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*(4), 922–939. https://doi.org/10.1037/0278-7393.24.4.922

Rose, J. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties* (DCSF-00659-2009). DCSF Publications. http://webarchive.nationalarchives.gov.uk/publications/

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J.-F., Bricout, L., Billard, C., Nguyen-Morel, M.-A., Le Heuzey, M.-F., Soares-Boucaud, I., George, F., Ziegler, J. C., & Ramus, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology, 52*(10), 1503–1516. https://doi.org/10.1037/dev0000184

Schatschneider, C., Francis, D.J., Foorman, B.R., Fletcher, J.M. & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91*(3), 439–449. https://doi.org/10.1037/0022-0663.91.3.439

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Shankweiler, D., Liberman, I.Y., Mark, L.S., Fowler, C.A., & Fischer, F.W. (1979). The speech code and learning to read. *Journal of Experimental Psychology, 5*(6), 531–545. https://doi.org/10.1037/0278-7393.5.6.531

Snowling, M.J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research, 43*(2), 219–234. https://doi.org/10.1007/bf00309831

Snowling, M.J. (1998). Dyslexia as a Phonological deficit: Evidence and implications. *Child Psychology and Psychiatry Review, 3*(1), 4–11. https://doi.org/10.1017/s1360641797001366

Snowling, M. J. (2000). *Dyslexia* (2nd ed.). Oxford, UK: Blackwell.

Stahl, S.A., & Murray, B.A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86*(2), 221–234.

https://doi.org/10.1037/0022-0663.86.2.221

Stanovich, K.E., Cunningham, A.E., & Cramer, B.B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38*(2), 175–190. https://doi.org/10.1016/0022-0965(84)90120-6

Stanovich, K.E., & Siegel, L.S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*(1), 24–53. https://doi.org/10.1037/0022-0663.86.1.24

Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences*, 20, 147–152. https://doi.org/10.1016/S0166-2236(96)01005-3

Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, *66*(1), 18-41. https://doi.org/10.1006/jecp.1997.2375

Szenkovits, G., Darma, Q., Darcy, I., & Ramus, F. (2016). Exploring dyslexics' phonological deficit II: Phonological grammar. *First Language, 36*(3), 316-337. https://doi.org/10.1177/0142723716648841

Tallal, P., Miller, S., & Fitch, R.H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. *Annals-New York Academy of Sciences, 682*, 27-47. https://doi.org/10.1111/j.1749-6632.1993.tb22957.x

Truman, A., & Hennessey, N. W. (2006). The locus of naming difficulties in children with dyslexia: Evidence of inefficient phonological encoding. *Language and Cognitive Processes, 21*(4), 361-393. https://doi.org/10.1080/01690960400011746

Uno, T., Takano, K., & Nakamura, K. (2025). Dissecting the causal role of early inferior frontal activation in reading. *Journal of Neuroscience, 45*(2), e0194242024. https://doi.org/10.1523/JNEUROSCI.0194-24.2024

Vellutino, F. (1979). *Dyslexia: Theory and research*. Massachusetts Institute of Technology Press.

Wagner, R.K., Torgesen, J.K., Rashotte, C.A., & Pearson, N.A. (2013). *Comprehensive Test of Phonological Processing, 2nd ed. (CTOPP-2)*. Pro-Ed.

Wagner, R.K., Torgesen, J.K., Rashotte, C.A., Hecht, S.A., Barker, T.A., Burgess, S.R., Donahue, J. & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*(3), 468–479. https://doi.org/10.1037/0012-1649.33.3.468

Wagner, R.K. & Torgesen, J.K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*(2), 192–212. https://doi.org/10.1037/0033-2909.101.2.192

Wagner, R. K., Zirps, F. A., & Wood, S. G. (2022). Developmental dyslexia. In M. J. Snowling, C. Hulme & K. Nation (Eds), *The science of reading: A Handbook* (2nd Ed., pp. 416–438). Wiley.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time-course of phonological encoding. *Journal of Memory and Language, 34*, 311–334. https://doi.org/10.1006/jmla.1995.1014

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*(3), 415–438. https://doi.org/10.1037/0022-0663.91.3.415

Wolff, P.H., Michel, G.F., & Ovrut, M. (1990). The timing of syllable repetitions in developmental dyslexia. *Journal of Speech Language and Hearing Research, 33*(2), 281–289. https://doi.org/10.1044/jshr.3302.281

Author Contribution Statement

Aki Tapionkaski outlined the theoretical framework and wrote the first draft of the manuscript. **Sanna Tapionkaski** completed the descriptive qualitative analysis of blending difficulties. Both authors contributed to the finalisation of the manuscript and approved the final version.

How is language knowledge related to verbal working memory among preschool children? Evidence from bilinguals and monolinguals

Farzaneh Anjomshoae
University of Alberta, Canada

Elena Nicoladis
University of British Columbia, Kelowna, Canada

Anahita Anahita
University of Alberta, Canada

Abstract: The typical language development in bilingual children, which shows some similarities with atypical language development in monolinguals, displays unique characteristics when compared to monolingual counterparts. In order for clinicians to correctly diagnose bilingual children with atypical language development, it is important to have access to diagnostic tools that distinguish typical from atypical development. Measures of children's verbal working memory, such as digit span and nonword repetition, may be just such diagnostic tools. However, some previous studies have suggested that measures of verbal working memory could be related to language-specific knowledge, such as vocabulary. The purpose of the present study was to test whether bilingual children's performance on two verbal working memory tasks were related to their within- language vocabulary scores. Forty French-English bilingual preschoolers and 40 age- matched English monolingual children were administered a standardized vocabulary test in English, a nonword repetition task (NWR), and digit span tasks (both forward and backward). The results showed that the bilingual children scored significantly lower than the monolingual children on the vocabulary test, but not on the nonword repetition task nor on the digit span tasks. Moreover, vocabulary scores were not correlated with the verbal working memory tasks for monolingual children. For bilingual children the NWR was not correlated with vocabulary. NWR seems to be relatively free of language-specific knowledge, at least within this age group. We discuss the clinical implications of these results.

Keywords: phonological working memory; nonword repetition; bilingual advantage; preschool language development

Corresponding author(s): Farzaneh Anjomshoae, Department of Psychology, University of Alberta, Edmonton, T6G 2E9, Canada. Email: anjomsho@ualberta.ca.

ORCID ID(s): Farzaneh Anjomshoae: https://orcid.org/0000-0002-1556-1957; Elena Nicoladis: https://orcid.org/0000-0002-5065-9159; Anahita Shokrkon: https://orcid.org/0000-0001-9982-6467

Citation: Anjomshoae, F., Nicoladis, E., & Shokrkon, A. (2025). How is language knowledge related to verbal working memory among preschool children? Evidence from bilinguals and monolinguals. *Language Development Research, 5*(3), 104–130. http://doi.org/10.34842/ldr2025-634

Introduction

The number of children who learn a second language has been increasing across the world, including in both Canada and the United States (Statistics Canada, 2016; U.S. Census Bureau, 2013). In Canada, for instance, the proportion of English-French bilinguals increased by 3% in 2016 and 2021. As of 2021, 4.6 million Canadians speak a language other than English or French at home (Statistics Canada, 2016). Likewise, in the United States, more than 18% of the population consists of individuals who are over the age of 5 and speak a language other than English in the home (U.S. Census Bureau, 2000). Bilingual children may appear to lag behind in language development, particularly in within-language vocabulary, when compared to same-aged monolinguals in that language (Gross et al., 2014; Meir et al., 2017; Oller et al., 2007; Smithson et al., 2014; Sullivan et al., 2018). However, this perspective is rooted in a monolingual-centered view of language development. Bilinguals' reduced exposure to each language compared to monolinguals does not necessarily lead to developmental delays but rather reflects a different trajectory of language acquisition that may involve weaker connections between lexical items in each language (Bialystok et al., 2010; Thordardottir et al., 2006).

Typical bilingual development can often resemble atypical language development seen in monolinguals, making the assessment and diagnosis of language impairment among bi-multilingual children challenging (Gollan & Ferreira, 2009; Oetting, 2018; Williams & McLeod, 2012). There is a risk of both over- and under-identification of language impairment in bilinguals and multilinguals (Antonijevic-Elliott et al., 2020; Genessee et al., 2004, Paradis et al., 2013), which can be due to the lack of standardized assessment tools in non-English languages and culturally appropriate resources.

One key area where this challenge becomes evident is in vocabulary size evaluation. De Houwer et al. (2014) demonstrated that when bilingual children are evaluated in both their languages, they can exhibit comparable vocabulary sizes to monolingual children in both comprehension and production. This is in contrast to a common misunderstanding that bilingual children may have smaller vocabulary. This finding underscores the importance of considering a bilingual child's abilities in all their languages when making assessments. Therefore, it is critical for speech-language pathologists to utilize different assessment tools and protocols to more accurately identify language development and disorders in bilingual children. Some studies have suggested that bi-multilingual children should be assessed in all of their languages (American Speech-Language-Hearing Association, 2004; Armon-Lotem et al., 2015). However, this is not always feasible due to various reasons including time restrictions, lack of linguistically and culturally appropriate resources, and a dearth of skilled interpreters and bilingual speech and language therapists (Boerma & Blom, 2017).

One measure that could help clinicians differentiate typically developing bilinguals from language- disordered bilinguals is Verbal Working Memory (VWM) (Campbell et al. 1997; Ellis Weismer & Evans, 2002; Ellis Weismer et al., 2000; Engel et al., 2008). VWM refers to the ability of maintaining speech stimuli, such as words and numbers, for a short period of time (van Dun & Mariën, 2016). One example of VWM would be keeping a phone number in mind for a short time until it is dialed. VWM has been shown to be important in language learning (Martin & He, 2004; Shallice & Papagno, 2019). There are different models of working memory (e.g., Engle et al., 1992; Gray et al., 2017). One prominent model of WM has been suggested by Baddeley and Hitch (1974). This model consists of three components: central executive, phonological loop (for verbal and auditory information) and visuospatial sketchpad (for visuospatial information). Baddeley (2000) added another component to this model: the episodic buffer which connects verbal and visual memory. This model draws a distinction between short-term (STM) and working memory (WM). STM refers to memory stored as the stimuli were experienced (as in the phone number example) while WM refers to memory that is manipulated and processed, such as recalling stimuli in the reverse order. WM critically involves the central executive, while STM does not. This model critically assumes a distinction between VWM and STM. However, Buchsbaum and D'Esposito (2019) have proposed that the VWM refers to both keeping and processing information, as the transformation of the stimulus is always in terms of goal-directed behavior. In other words, they do not draw a sharp distinction between VWM and STM. In the present study, we have followed their lead and include measures of both STM and WM under the umbrella of VWM.

These memory stores are important predictors of children's learning new words in both first and second languages (e.g., Gathercole, 1995; Gathercole & Baddeley, 1990; Juffs & Harrington, 2011; Kohonen, 1995; Service, 1992) and in foreign languages (Masoura & Gathercole, 2005). This relationship holds both contemporaneously and longitudinally (Gathercole et al., 1992; 1997; 1999; 2008).

There is also some evidence supporting a causal association between impairments of VWM and learning difficulties (Alloway et al., 2005; Swanson & Siegel, 2001). For these reasons, a valid and reliable measure of VWM could aid clinicians in the assessment of children with communication difficulties. Processing measures, such as VWM tasks, may be less biased tasks for assessing language development since they involve language-general cognitive processes, compared to a more language-specific task such as vocabulary, which are influenced by a child's exposure to a specific language (Campbell et al., 1997; Ellis Weismer & Evans, 2002; Ellis Weismer et al., 2000). It is particularly important when considering bilingual children, who may exhibit uneven language proficiency across their languages. Therefore, VWM tasks enable a more equitable evaluation of a child's potential to learn a foreign language, without relying on the size of vocabulary specific to a particular language. However, one challenge in using VWM measures in the assessment of bilingual children is that it is not entirely clear how previous experience of a particular language plays a role in VWM ability.

The purpose of the present study is to test whether language knowledge is related to VWM performance in bilingual and monolingual children. From a clinical point of view, it is critical to determine how these tasks can be used as an appropriate assessment for bilingual children. In this study, we expected bilingual children to perform worse on an English vocabulary test than monolinguals, as has often been found in prior research (Gross et al., 2014; Meir et al., 2017; Oller et al., 2007; Smithson et al., 2014; Sullivan et al., 2018).

However, bilingual lexical development suggests it is inappropriate to have monolingual-centered expectations for bilingual children (Fennell & Lew-Williams, 2017). If language knowledge is associated with VWM performance, then the bilingual children should perform worse than same-aged monolingual children. In both groups of children, VWM scores should be correlated with vocabulary scores.

To the extent that forward DS and NWR are both measures of STM, performance on the two tasks should be highly correlated. Indeed, Gathercole et al. (1999) investigated the association between verbal memory measuring by DS and NWR, and vocabulary knowledge in monolingual children 4 to 13 years of age. The results of this study indicated a significant correlation between DS and NWR. As noted earlier, some models of VWM assume a distinction between VWM and STM (Schwering & MacDonald, 2020). However, Alloway and Alloway (2010) found that forward and backward DS are highly correlated. This result is consistent with models assuming that VWM and STM are related (Davidson et al., 2006; Jensen et al., 2007; Miller et al., 1960).

Since the DS involves existing words and the NWR does not, one might expect that language knowledge would be more strongly related to DS performance than NWR performance. Alternatively, since the non-words in a NWR task reflect the phonotactics of a particular language, performance on the NWR could be strongly related to language knowledge. Indeed, as predicted, some studies have shown that NWR performance is less dependent on language knowledge than DS tasks among children aged 4 to 9 (Boerma et al., 2015; Chiat & Polišenská, 2016; Windsor et al., 2010). Similarly, some studies have found that DS performance is related to vocabulary for both monolingual (Gathercole et al., 1999) and bilingual (Haman et al., 2017) children. And some studies have shown no relationship between language abilities and NWR performance (Botting & Conti-Ramsden, 2001; Pigdon et al., 2019).

However, a number of studies have shown that performance on NWR tasks is related to language abilities for typically developing monolingual children (Gathercole, 2006; Gathercole et al., 1999). Similarly, some studies show that children with language difficulty often performed this task less accurately, especially for longer nonwords (Adams & Gathercole, 1995; Estes et al., 2007; Munson et al., 2005), With respect to bilinguals, studies have shown that performance on NWR tasks is related to language knowledge in which the task was performed (Ebert et al., 2014; Lee & Gorman, 2012; Parra et al., 2011; Peña et al., 2002; Summers et al., 2010; Thorn & Gathercole, 1999). Other studies have shown that NWR is one of the tasks that best discriminates between bilingual children with and without language difficulties (e.g., Schwob et al., 2021). In sum, the weight of the evidence suggests that there is a relationship between language knowledge and VWM performance for both bilingual and monolingual children as measured by both DS and by NWR. Another way of testing for possible links between linguistic knowledge and VWM is to compare bilingual and monolingual children. If VWM performance involves language-specific knowledge, one prediction is that monolingual children should outperform bilingual children. Language-specific knowledge refers to the grasp and comprehension of linguistic features unique to a particular language, which could lead to differing task performance between monolinguals and bilinguals due to their exposure to different language systems. Indeed, some studies have shown a monolingual advantage on DS performance, both forward (Blom & Boerma, 2017; Fernandes et al. 2007; Wodniecka et al., 2010) and backward (Bialystok, 2010; Bialystok et al., 2008; Blom & Boerma, 2017). However, other studies have shown equivalent performance on DS tasks in monolinguals and bilinguals (Gutiérrez-Clellen et al., 2004; Luo et al., 2010; Martin-Rhee & Bialystok, 2008; Shokrkon & Nicoladis, 2021).

Similarly, with NWR tasks, some studies have shown that monolinguals outperform bilinguals (Engel de Abreu 2011; Kohnert et al., 2006; Windsor et al. 2010). Whereas other studies have found no difference between monolinguals and bilinguals (Lee et al., 2013; Thorn & Gathercole, 1999). Taken together, these studies have shown mixed results about VWM performance between bilinguals and monolinguals. Further studies are needed to understand the contribution of language knowledge to VWM tasks. The purpose of the current study was to test whether language knowledge underlies performance on two VWM tasks: NWR and DS. Since both tasks are thought to measure VWM, we hypothesized that performance would be highly correlated on the NWR and the DS. We predicted that the bilinguals would score lower than monolinguals on an English vocabulary task and therefore lower on VWM tasks. We also tested whether the English vocabulary score is correlated with DS and NWR in both groups.

Method

Participants

The sample of this study included 40 (*M* age = 62.02 months, *SD* = 8.26), English-speaking monolinguals and 40 French-English (*M* age = 61.62 months, *SD* = 8.19) simultaneous bilingual children who were exposed to both languages from birth. The characterization of the bilingual children as simultaneous bilinguals was based on parental reports indicating that the children had been exposed to both French and English from birth. The monolingual children were chosen from a database of 79 children as the closest age matches the bilingual children. We did not have measures of the socioeconomic status (SES) of the children's families, although our recruitment approaches likely targeted high SES families (e.g., we recruited in university area daycares, known to target the children of graduate students and faculty members). To evaluate language dominance among the bilingual children, their parents were asked to best describe their child's French and English language proficiencies. The options were as follows: (a) My child speaks French far better than English, (b) My child speaks French a bit better than English, (c) My child speaks both languages about equally well, (d) My child speaks English a bit better than French, and (e) My child speaks English far better than French. There was no difference between the two groups on age U= 771.00, $n1$ = 40, $n2$ = 40, p = .78. All the children lived in the same western Canadian city.

Analyses in G-Power (Faul et al., 2009) revealed that the total sample size of 29 participants gave us power (1– β = .80) enough to find at least medium effect size (d = .50). We included 40 participants in each group, which not only exceeds the minimum required by the power analysis but also helps account for potential dropouts and provides a more comprehensive analysis of the research questions.

Materials

Receptive vocabulary

The Peabody Picture Vocabulary Test III (PPVT; Dunn & Dunn, 1997), version A, was used to assess the children's receptive vocabulary in English. A high reliability of .92 was reported for this test (Community-University Partnership for the Study of Children, Youth, and Families, 2011). To assess the vocabulary of the children in French, the Échelle de vocabulaire en images Peabody (EVIP; Dunn et al., 1993) was used. This is a standardized instrument tailored for Canadian French that evaluates receptive vocabulary.

Verbal Working Memory

We included two measures of the children's VWM: Digit Span (DS; Richardson, 2007) and Nonword Repetition (NWR; Gathercole & Baddeley, 1996). Under the Baddeley (2000) model, forward DS is thought to tap STM. Backward DS is thought to tap VWM, since it has greater executive function involvement than forward DS (Gerton et al., 2004). In the NWR task, which is thought to tap STM, children repeat some nonsense words with the phonotactics of the target language (Coady & Aslin 2004; Munson et al., 2005, Rispens & Parriger, 2010; Zamuner et al., 2004). These nonwords are often different in length between 2 to 5 syllables (e.g., *ballop, bannifer, blonterstaping* and *altupatory*; Gathercole & Baddeley, 1996). According to Roy and Chiat (2004), NWR tasks are independent of culture and intelligence quotient (IQ).

Digit Span Task

To measure digit span (DS), we used the Wechsler Intelligence Scale for Children (WISC-R; Weschler, 2003), digit span subscale. In this task, children are given a series of digits, and they were asked to recall and repeat the string of digits in the same way (i.e. the participants should repeat 2,4,6 as they heard 2,4,6) and reverse order (backward digit span, they need to repeat 4,7,5, as they heard 5,7,4). For the first trial, participants were given a series of two digits. For every subsequent correct trial, the experimenter added one digit (to a maximum of nine digits). They were given one trial at each string-length until they made an error and if they responded correctly, they were given a longer string. So, scores represent the highest level of error-free performance. A good psychometric property has been reported for forward and backward DS tasks, .89 and .86, respectively (Alloway & Passolunghi, 2011).

Nonword Repetition Task

To measure nonword repetition skills in English, the children's test of nonword repetition (NWR) was used (Gathercole & Baddeley, 1996). This test consists of 40 pseudowords of two to five syllables and each syllable category involves 10 nonwords. Participants were asked to listen to pre-recorded nonwords, spoken by a female native speaker of Canadian English and repeat afterwards. The performance of children taped for scoring and the raw scores were used for analysis. The scoring was based on the procedure of Dollaghan and Campbell (1998), in which, the responses were considered as correct if they made all phonetics correctly, and if they did any insertion, substitution or omission, the response was considered as incorrect. Each correct response scored 1, with the maximum score of 40. For internal consistency, we calculated Cronbach's alpha, which was .81 and showed good reliability.

Procedure

Research ethics approval was obtained from the institutional research ethics board for this study. Informed consent forms were signed by the parents, allowing us to conduct the test on their children. The data for this study come from a larger study that included a battery of language and cognitive tasks with both bilingual and monolingual children (Nicoladis & Mimovic, 2022). The analyses in the current study, however, are distinct and have not been published elsewhere.

In this paper, we report only the results connected to our research questions. The bilingual children participated in two separate language sessions. The English sessions were conducted by a native English speaker and the French sessions by a native French speaker. The order of the two language sessions were counter balanced and the two sessions were separated by about a week.

All children were asked for their verbal consent before any task was undertaken. Within a language session, the order of the tasks was at the discretion of the experimenter, based on the child's engagement. The sessions usually started with the more passive tasks, like the PPVT, which only requires children to point to a picture. Tasks requiring children to speak, like the NWR and the DS were usually administered later in the session. Two bilingual and monolingual children did not speak when asked to repeat non-words on the NWR repetition task. Five bilingual children (three FDS and two BDS) and one monolingual child did not perform the DS tasks. Three bilingual children and one monolingual child did not complete the PPVT. Consistent with guidelines suggested by Allison (2001), we have responsibly managed the incomplete data from the three bilingual children and one monolingual child who did not complete the PPVT, incorporating their results only in analyses where the absence of PPVT data would not compromise the reliability of the findings.

Results

We devised our results section based on best practices in the field and considered the specifics of our dataset. Our use of non-parametric statistical techniques, like the Mann–Whitney U test and Spearman nonparametric correlations, was necessitated by significant deviations from normality in our data, as indicated by the Shapiro-Wilk test results. We looked to relevant studies, such as (Blom et al., 2014; Czapka et al., 2019), for methodological guidance. Although their study used a combination of t-tests and mixed effects models rather than non-parametric tests, their approach to analyzing data separately for each language group was similar to ours and provided us with useful insights into the potential relationships between variables in our own data. The descriptive statistics associated with PPVT, DS and NWR scores across monolingual and bilingual groups, are summarized in Table 1. We tested the normality of all variables, prior to the analyses, by conducting the Shapiro-Wilk test. The results indicated significant deviations from normality for FDS ($W = .86$, $p < .001$) and BDS ($W = .82$, $p < .001$) tests. Therefore, we conducted Mann–Whitney U as a nonparametric test to examine the differences between groups. As expected, the monolingual children had significantly higher PPVT scores than the bilingual children, $U = 443.50$, $n1 = 39$, $n2 = 39$, $p = .004$. However, the results indicated no significant difference for any of the VWM tasks: FDS: $U = 612.50$, $n1 = 39$, $n2 = 38$, $p = .16$; BDS: $U = 738.00$, $n1 = 39$, $n2 = 38$, $p = .97$; and NWR: $U = 721.50$, $n1 = 36$, $n2 = 40$, $p = .70$.

Table 1. ***Descriptive statistics for monolinguals and the bilinguals.***

Measure	Group	N	Mean	SD	Minimum	Maximum
Age	Monolingual	40	62.02	8.26	47	83
	Bilingual	40	61.62	8.19	48	82
PPVT	Monolingual	39	86.30	24.83	39	137
	Bilingual	37	68.37	20.33	20	101
FDS	Monolingual	39	4.32	.72	3	6
	Bilingual	37	4.51	.93	3	7
BDS	Monolingual	39	1.61	1.61	0	6
	Bilingual	38	1.40	1.23	0	3
NWR	Monolingual	38	31.02	4.76	21	40
	Bilingual	38	29.78	5.98	14	40

PPVT = Peabody Picture Vocabulary Test; FDS = forward digit span; BDS = backward digit span; NWR = non-word repetition

Relationship between language ability, digit span and nonword repetition

To examine relations between the verbal working memory, nonword repetition and vocabulary (PPVT scores), Spearman nonparametric correlations were conducted. We present these correlations in the Appendix. As can be seen in the Table A1, age was highly correlated with VWM measures for both monolingual and bilingual children. We therefore partialled out age to test for the relationship between VWM and vocabulary. For the monolinguals, there was no significant correlation between vocabulary scores and FDS, $r_{partial}(33) = .286$, $p = .10$, BDS, $r_{partial}(33) = .076$, $p = .67$, or NWR, $r_{partial}(33) = .225$, $p = .20$. For the bilinguals, there was no significant correlation between vocabulary scores and FDS, $r_{partial}(32) = .288$, $p = .09$, BDS, $r_{partial}(32) = .243$, $p = .12$, or NWR, $r_{partial}(32) = .071$, $p = .69$.

Given how little difference there was between the bilinguals and the monolinguals on the VWM measures, to test for significant partial correlations (controlling for age) between measures of VWM, we combined the groups. The results indicated that forward and backward DS were significantly correlated, $r_{partial}(76) = .253$, $p = .02$, and NWR was correlated with FDS, $r_{partial}(75) = .232$, $p = .04$ and with BDS, $r_{partial}(75) = .381$, $p < .001$. For partial correlations between VWM tasks for each group separately, see Table A2 in the Appendix.

Discussion

The purpose of this study was to test whether VWM, which is assessed by DS and NWR tasks, is linked to English language ability in monolingual and bilingual preschoolers. As expected, in line with many previous studies (Bialystok et al., 2010; Gollan et al.

2007; Namazi & Thordardottir, 2010), the bilinguals scored lower than the monolinguals on English receptive vocabulary. If VWM performance were related to language experience and knowledge, bilinguals would perform worse on VWM tasks than monolinguals. However, our results did not support a bilingual disadvantage on VWM tests: there was no significant difference between monolinguals and bilinguals on either a DS task or a NWR task. Moreover, vocabulary scores were not significantly correlated with performance on either the DS or the NWR tasks, for either bilingual or monolingual children, after we controlled for age. These results are surprising given that some previous research has shown a bilingual disadvantage in VWM measures relative to monolinguals (Bialystok et al. 2010; Thordardottir et al. 2006, Liu & Liu, 2021). However, it is important to keep in mind that other previous studies have also shown no difference between bilinguals and monolinguals on VWM measures (Cockcroft, 2016; Engel de Abreu, 2011; Engel de Abreu et al., 2012).

The inconsistent results across studies examining children's VWM performance could be due to a multitude of factors, including potential differences in methodological tools, such as variations in tasks used to measure VWM, participant characteristics, and operational definitions of bilingualism. Considering the developmental nature of VWM, age and language proficiency of participants may also influence the performance, underscoring the need for consistent and reliable measures when studying VWM (e.g., Bouffier et al., 2020). Moreover, some studies have shown a potential bilingual advantage in executive function, which may extend to WM and VWM (Adi-Japha et al., 2010; Bialystok, 2010; Carlson & Meltzoff, 2008; Prior & MacWhinney, 2010; Yoshida et al., 2011). Previous studies provided evidence that bilingual's higher executive functioning might extend to WM in general and VWM in particular (e.g., Blom et al., 2014; Delcenserie & Genesee, 2016; Kaushanskaya, 2012; Kroll et al., 2002). This advantage might be contingent on the amount of language exposure (Pierce et al., 2017). The bilingual children in this study, despite having lower vocabulary scores than monolinguals, had strong English proficiency due to their extensive exposure to English from an early age. Future longitudinal studies may shed more light on the relationship between the degree of language proficiency and VWM performance. Furthermore, it's noteworthy that the implications of bilingualism may extend to cognitive resilience in later life, potentially delaying cognitive decline and the onset of dementia (van den Noort et al., 2019), underscoring the lifelong impact of bilingualism on cognitive functions. The point that our bilingual children were possibly developing equivalent VWM skills to the monolinguals deserves careful consideration, especially given the influence of age on VWM. As noted, our study engaged children who were around five years old, unlike some other studies that involved older children (for instance, Lee and Gorman, 2012, worked with seven-year-olds). We acknowledge that the two-year difference is significant in early childhood development, potentially influencing not just language acquisition but also other skills like reading.

It's crucial to recognize that this study did not intend to make direct comparisons with studies involving different age groups. Instead, our goal was to provide a snapshot of VWM and language performance among bilingual and monolingual children at this particular stage of development. The statement about bilingual children potentially developing equivalent VWM skills to the monolinguals was speculative and meant to hint at potential trajectories of development, rather than provide definitive conclusions.

Future studies could focus on longitudinal designs, tracing the development of VWM and language proficiency over time. Such work could provide more definitive insights into the rate and pattern of VWM development among bilingual and monolingual children. Furthermore, it might help in better understanding the interaction between language proficiency and VWM performance.

With respect to the NWR task, our results add more evidence to the debate about the extent to which NWR is related to language knowledge or as a language-free task (Alloway & Archibald 2008; Gathercole et al. 1999; Kohnert et al., 2006). Our findings suggest that NWR has little connection to language knowledge among both monolingual and bilingual children of this age, as there were nonsignificant correlations between language knowledge and this task. These results are in line with previous study describing no relation between NWR and language exposure in five-year-olds French-English bilinguals (Thordardottir & Brandeker, 2013). These results, however, conflict with other studies showing the relationship between vocabulary size and NWR performance. Lee and Gorman (2012), for example, found that vocabulary scores and NWR performance were correlated in all bilingual children with various first languages (e.g., Korean, Chinese, Spanish). One plausible explanation for these differences is the similarity and close relationship between the first and second languages of our bilinguals. Some studies, especially studies with French-English bilinguals in Canada, showed that some bilinguals performed on par with monolinguals in their first and second language (Smithson et al., 2014; Thordardottir, 2011). Previous studies have found that NWR highly depends on language proficiency both vocabulary and grammar (Archibald & Gathercole, 2006; Ellis Weismer et al., 2000; Thordardottir et al., 2010). Favoring this explanation, many studies suggested that although NWR are not related to long-term lexical knowledge, nonword recall is linked to familiarity with the phonotactic properties of the language in which NWR are performed (Gathercole, 1995; Gathercole et al., 1999; Kovács & Racsmány, 2008; Roodenrys & Hinton, 2002; van Bon & van der Pijl, 1997). That is the similarity between L1 and the language in which NWR is tested may be considered as an advantage. Another possible explanation for the mixed results in terms of its connection to language abilities is that there are different types of NWR measurements which are different in the rate of wordlikeliness and therefore to what extent the nonwords are related to real vocabulary (Thordardottir & Brandeker, 2013).

The results of this study showed strong intercorrelations between FDS, BDS, and NWR, suggesting that these tasks may tap on similar underlying abilities. These results are in line with some studies that have found that NWR is moderately to strongly associated with BDS and/or visuospatial working memory (Baniqued et al. 2013, Cleary et al. 2001). In the other words, our results suggested that the distinction between STM and WM is not in place at least among preschool children. This finding has been well established and supported in the previous studies (Davidson et al., 2006; Jensen et al., 2007; Krumm et al., 2009; Martinez et al., 2011). Likewise, a number of studies provided empirical evidence that STM and WM could possibly represent the same latent construct (Colom, Rebollo et al., 2006; Colom, Shih et al., 2006; Unsworth & Engle, 2006). Unsworth and Engle (2007), for instance, proposed simple and complex tasks measure the same cognitive process showing STM and WM are indistinguishable.

Naturally, the present study has a number of limitations. One of the main limitations is the close relationship between French and English languages. Moreover, our simultaneous bilinguals were exposed to their both languages in a sociocultural context that supported bilingualism. Therefore, more studies with different language groups and larger sample sizes are needed to expand the generalizability of these findings. Another limitation of this study is that our participants were likely from high-SES families. Although we did not measure SES directly, our recruitment approach likely targeted high-SES families. Previous studies have shown a significant association between SES and WM functioning (Noble et al., 2005; 2007). Thus, an interesting avenue for further research is investigating how SES might influence the relationship between VWM and language.

In conclusion, the results of this study suggest that bilingual preschoolers, despite their lower vocabulary scores in English, performed equivalently to monolinguals on two measures of verbal working memory (VWM). The notion of a critical threshold of language exposure before performing well on VWM tasks, as also suggested by Thordardottir (2020), emerges as a central theme in our discussion. However, our findings contribute additional nuances to this idea by shedding light on the comparable performance of bilinguals and monolinguals in VWM tasks despite differing vocabulary proficiency.

In terms of clinical implications, the understanding of a critical language exposure threshold could have significant ramifications for the design and interpretation of cognitive and language assessments for bilingual children. If language exposure proves to be a pivotal factor in determining VWM performance, clinicians might need to incorporate measures of language exposure when evaluating bilingual children. Furthermore, it's important to consider that the bilingual children in this study, despite their lower English vocabulary scores, were frequently exposed to English either from birth or very early on. This suggests that the age, volume and quality of language exposure, rather than simply the number of languages spoken, might play a crucial role in VWM performance.

More research is undoubtedly required to further elucidate these links and inform clinical practice. Specifically, longitudinal studies that monitor language exposure and VWM performance over time would be beneficial. Meanwhile, clinicians and researchers should remain cautious in interpreting the results of VWM tasks, particularly when comparing bilingual and monolingual children.

References

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech, Language, and Hearing Research, 38*(2), 403-414. https://doi.org/10.1044/jshr.3802.403

Adi-Japha, E., Berberich-Artzi, J., & Libnawi, A. (2010). Cognitive flexibility in drawings of bilingual children. *Child Development, 81*(5), 1356-1366. https://doi.org/10.1111/j.1467-8624.2010.01477.x

Allison, P. D. (2001). *Missing data. Sage University Papers Series on Quantitative Applications in the Social Sciences*, 07-136. Sage Publications. https://doi.org/10.4135/9781412985079

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20-29. https://doi.org/10.1016/j.jecp.2009.11.003

Alloway, T. P., & Archibald, L. (2008). Working memory and learning in children with developmental coordination disorder and specific language impairment. *Journal of Learning Disabilities, 41*(3), 251-262. https://doi.org/10.1177/0022219408315815

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences, 21*(1), 133–137. https://doi.org/10.1016/j.lindif.2010.09.013

Alloway, T. P., Gathercole, S. E., Adams, A. M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology, 23*(3), 417-426. https://doi.org/10.1348/026151005X26804

American Speech-Language-Hearing Association. (2004). Knowledge and skills needed by speech-language pathologists and audiologists to provide culturally and linguistically appropriate services. *ASHA Supplement,* 24, 152–158.

Antonijevic-Elliott, S., Lyons, R., O'Malley, M. P., Meir, N., Haman, E., Banasik, N., ... & Fitzmaurice, Y. (2020). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics, 34*(4), 293-311. https://doi.org/10.1080/02699206.2019.1637458

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders, 41*(6), 675-693. https://doi.org/10.1080/13682820500442602

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment.* Multilingual Matters.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417– 423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 47–90). Academic Press.

Bialystok, E. (2010). Global-local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology, 46*(1), 93. https://doi.org/10.1037/a0015466

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 859. https://doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., Luk, G., Peets, K. F., & Sujin, Y. A. N. G. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525-531. https://doi.org/10.1017/S1366728909990423

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology, 128,* 105-119. https://doi.org/10.1016/j.jecp.2014.06.007

Blom, E. & Boerma, T. (2017). Effects of language impairment and bilingualism across domains: Vocabulary, morphology and verbal memory. *Linguistic Approaches to Bilingualism, 7*(3-4), 277-300. https://doi.org/10.1075/lab.15018.blo

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

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research, 58*(6), 1747-1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 36*(4), 421-432. https://doi.org/10.1080/13682820110074971

Bouffier, M., Barbu, C., & Majerus, S. (2020). Verbal working memory but not attention is related to language proficiency: Evidence from multilingual speakers. *Psychologica Belgica, 60*(1), 270-293. https://doi.org/10.5334/pb.525

Buchsbaum, B. R., & D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex, 112*, 134-148. https://doi.org/10.1016/j.cortex.2018.11.010

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*(2), 282-298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research, 40*, 519-525. https://doi.org/10.1044/jslhr.4003.519

Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research, 59*(5), 1179-1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293

Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology, 89*(3), 183-213. https://doi.org/10.1016/j.jecp.2004.07.004

Cockcroft, K. (2016). A comparison between verbal working memory and vocabulary in bilingual and monolingual South African school beginners: Implications for bilingual language assessment. *International Journal of Bilingual Education and Bilingualism, 19*(1), 74-88. https://doi.org/10.1080/13670050.2014.964172

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition, 34*(1), 158-171. https://doi.org/10.3758/BF03193395

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory, 14*(7), 804-813. https://doi.org/10.1080/09658210600680020

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from ma-nipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Delcenserie, A., & Genesee, F. (2016). The effects of age of acquisition on verbal memory in bilinguals. *International Journal of Bilingualism, 21*(5), 600-616. https://doi.org/10.1177/1367006916639158

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

Dollaghan, C. & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language and Hearing Research, 41*, 1136-1146. https://doi.org/10.1044/jslhr.4105.1136

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-III*. American Guidance Service. https://doi.org/10.1037/t15145-000

Dunn, L. M., Thériault-Whalen, C. M. & Dunn, L. M. (1993). *Échelle de vocabulaire en images Peabody - Forme révisée*. Psycan.

Ebert, K. D., Pham, G., & Kohnert, K. (2014). Lexical profiles of bilingual children with primary language impairment. *Bilingualism: Language and Cognition, 17*(4), 766-783. https://doi.org/10.1017/S1366728913000825

Ellis Weismer, S. E., & Evans, J. L. (2002). The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders, 22*(3), 15-29. https://doi.org/10.1097/00011363-200205000-00004

Ellis Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43*(4), 865-878. https://doi.org/10.1044/jslhr.4304.865

Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research, 51*(6), 1580-1587. https://doi.org/10.1044/1092-4388(2008/07-0210)

Engel de Abreu, P. M. J., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology, 104*(4), 974. https://doi.org/10.1037/a0028390

Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19*(5), 529-537. https://doi.org/10.1080/09658211.2011.590504

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 972-992. https://doi.org/10.1037/0278-7393.18.5.972

Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*, 177-195. https://doi.org/10.1044/1092-4388(2007/015)

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fennell, C., & Lew-Williams, C. (2017). Early bilingual word learning. In G. Westermann & N. Mani (Eds.), *Early Word Learning* (pp. 110-122). Taylor and Francis. https://doi.org/10.4324/9781315730974

Fernandes, M. A., Craik, F., Bialystok, E., & Kreuger, S. (2007). Effects of bilingualism, aging, and semantic relatedness on memory under divided attention. *Canadian*

Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 61(2), 128-141. https://doi.org/10.1037/cjep2007014

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*(1), 83-94. https://doi.org/10.3758/BF03210559

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., Alloway, T. P., Kirkwood, H. J., Elliott, J. G., Holmes, J., & Hilton, K. A. (2008). Attentional and executive function behaviours in children with poor working memory. *Learning and Individual Differences, 18*(2), 214-223. https://doi.org/10.1016/j.lindif.2007.10.003

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336-360. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E. & Baddeley, A. D. (1996). *The children's test of nonword repetition.* The Psychological Corporation.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 84-95. https://doi.org/10.1037//0278-7393.25.1.84

Gathercole, S. E., Hitch, G. J., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*(6), 966-979. https://doi.org/10.1037/0012-1649.33.6.966

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*(5), 887-898. https://doi.org/10.1037/0012-1649.28.5.887

Gerton, B. K., Brown, T. T., Meyer-Lindenberg, A., Kohn, P., Holt, J. L., Olsen, R. K., & Berman, K. F. (2004). Shared and distinct neurophysiological components of the digits forward and backward tasks as revealed by functional neuroimaging. *Neuropsychologia, 42*(13), 1781-1787. https://doi.org/10.1016/j.neuropsychologia.2004.04.023

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 640-665. https://doi.org/10.1037/a0014981

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society, 13*(2), 197-208. https://doi.org/10.1017/S1355617707070038

Gray, S., Green, S., Alt, M., Hogan, T., Kuo, T., Brinkley, S., & Cowan, N. (2017). The structure of working memory in young children and its relation to intelligence. *Journal of Memory and Language, 92*, 183-201. https://doi.org/10.1016/j.jml.2016.06.004

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

Gutiérrez-Clellen, V. F., Calderón, J., & Ellis Weismer, S. (2004). Verbal working memory in bilingual children. *Journal of Speech, Language, and Hearing Research, 47*, 863-876. https://doi.org/10.1044/1092-4388(2004/064)

Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., Mieszkowska, K., Łuniewska, M., Kołak, J., Miękisz, A., Kacprzak, A., Banasik, N., & Foryś-Nogala, M. (2017). How does L1 and L2 exposure impact L1 performance in bilingual children? Evidence from Polish-English migrants to the United Kingdom. *Frontiers in Psychology, 8*, 1444. https://doi.org/10.3389/fpsyg.2017.01444

Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences, 30*(7), 317-324. https://doi.org/10.1016/j.tins.2007.05.001

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137-166. https://doi.org/10.1017/S0261444810000509

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition, 15*(3), 470-489. https://doi.org/10.1017/S1366728911000472

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice, 21*(1), 19-29. https://doi.org/10.1111/j.1540-5826.2006.00204.x

Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16(*2)*,* 155-172. https://doi.org/10.1017/S0142716400007062

Kovács, G., & Racsmány, M. (2008). Handling L2 input in phonological STM: The effect of non-L1 phonetic segments and non-L1 phonotactics on nonword repetition. *Language Learning, 58*(3)*,* 597-624. https://doi.org/10.1111/j.1467-9922.2008.00451.x

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18*(2)*,* 137-171. https://doi.org/10.1191/0267658302sr201oa

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37,* 347-364. https://doi.org/10.1016/j.intell.2009.02.003

Lee, H. J., Kim, Y. T., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International Journal of Speech-Language Pathology, 15*(4)*,* 375-382. https://doi.org/10.3109/17549507.2012.752866

Lee, S. A. S., & Gorman, B. K. (2012). Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism, 17*(4)*,* 479-495. https://doi.org/10.1177/1367006912438303

Liu, Q., & Liu, Z. (2021). A review on the relationship between bilingualism and working memory. *Open Journal of Modern Linguistics, 11*(2)*,* 121-134. https://doi.org/10.4236/ojml.2021.112010

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114*(1)*,* 29-41. https://doi.org/10.1016/j.cognition.2009.08.014

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language, 89*(1)*,* 76-82. https://doi.org/10.1016/S0093-934X(03)00300-6

Martinez, K., Burgaleta, M., Roman, F. J., Escorial, S., Shih, P. C., Quiroga, M. A., & Colom, R. (2011). Can fluid intelligence be reduced to 'simple' short-term storage? *Intelligence, 39,* 473-480. https://doi.org/10.1016/j.intell.2011.09.001

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*(1), 81-93. https://doi.org/10.1017/S1366728907003227

Masoura, E. V., & Gathercole, S. E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory, 13*(3-4), 422-429. https://doi.org/10.1080/09658210344000323

Meir, N., Walters, J., & Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguistic Approaches to Bilingualism, 7*(5), 514-553. https://doi.org/10.1075/lab.15007.mei

Miller G. A., Galanter E., Pribram K. H. (1960). *Plans and the structure of behavior*. Holt, Rinehart and Winston, Inc. https://doi.org/10.1037/10039-000

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language and Hearing Research, 48*, 61-78. https://doi.org/10.1044/1092-4388(2005/006)

Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism, 13*(5), 597-616. https://doi.org/10.1080/13670050.2010.488288

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science, 10*(4), 464-480. https://doi.org/10.1111/j.1467-7687.2007.00600.x

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science, 8*(1), 74-87. https://doi.org/10.1111/j.1467-7687.2005.00394.x

Nicoladis, E., & Mimovic, A. (2022). Validity of parents' classifications of bilingual children's language dominance. *International Journal of Bilingual Education and Bilingualism, 25*(6), 2293-2302. https://doi.org/10.1080/13670050.2020.1850635

Oetting, J. B. (2018). Prologue: Toward accurate identification of developmental language disorder within linguistically diverse schools. *Language, Speech, and Hearing Services in Schools, 49*(2), 213-217. https://doi.org/10.1044/2018_LSHSS-CLSLD-17-0156

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*(2), 191-230. https://doi.org/10.1017/S0142716407070117

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research, 56*(3), 971-981. https://doi.org/10.1044/1092-4388(2012/12-0050)

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology, 108*(1), 113-125. https://doi.org/10.1016/j.jecp.2010.07.011

Peña, E., Bedore, L., & Zlatic-Giunta, R. (2002). Use of a category generation task to assess vocabulary skills of Spanish-English bilinguals. *Journal of Speech, Language, and Hearing Research, 46*, 938-947. https://doi.org/10.1044/1092-4388(2002/076)

Pierce, L. J., Genesee, F., Delcenserie, A., & Morgan, G. (2017). Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics, 38*(6), 1265-1300. https://doi.org/10.1017/S0142716417000236

Pigdon, L., Willmott, C., Reilly, S., Conti-Ramsden, G., Gaser, C., Connelly, A., & Morgan, A. T. (2019). Grey matter volume in developmental speech and language disorder. *Brain Structure and Function, 224*(9), 3387-3398. https://doi.org/10.1007/s00429-019-01978-7

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*(2), 253-262. https://doi.org/10.1017/S1366728909990526

Richardson, J. T. (2007). Measures of short-term memory: a historical review. *Cortex, 43*(5), 635-650. https://doi.org/10.1016/S0010-9452(08)70493-3

Rispens, J., & Parigger, E. (2010). Non-word repetition in Dutch-speaking children with specific language impairment with and without reading problems. *British Journal of Developmental Psychology, 28*(1), 177-188. https://doi.org/10.1348/026151009X482633

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 29-33. https://doi.org/10.1037//0278-7393.28.1.29

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2-to 4-year-olds. *Journal of Speech, Language, and Hearing Research, 47*(1), 223-234. https://doi.org/10.1044/1092-4388(2004/019)

Schwering, S. C., & MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience, 14,* 68. https://doi.org/10.3389/fnhum.2020.00068

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A, 45*(1), 21-50. https://doi.org/10.1080/14640749208401314

Shallice, T., & Papagno, C. (2019). Impairments of auditory-verbal short-term memory: do selective deficits of the input phonological buffer exist? *Cortex, 112,* 107-121. https://doi.org/10.1016/j.cortex.2018.10.004

Shokrkon, A., & Nicoladis, E. (2021). Absence of a bilingual cognitive flexibility advantage: A replication study in preschoolers. *Plos One, 16*(8), e0255157. https://doi.org/10.1371/journal.pone.0255157

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 64*(9), 3578-3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Smithson, L., Paradis, J., & Nicoladis, E. (2014). Bilingualism and receptive vocabulary achievement: Could sociocultural context make a difference? *Bilingualism: Language and Cognition, 17*(4), 810-821. https://doi.org/10.1017/S1366728913000813

Statistics Canada. (2016). 2016 Census: Immigrant languages in Canada. Statistics Canada catalogue no: 11-627-M2017025. Retrieved from https://www150.statcan.gc.ca

Sullivan, M. D., Poarch, G. J., & Bialystok, E. (2018). Why is lexical retrieval slower for bilinguals? Evidence from picture naming. *Bilingualism: Language and Cognition, 21*(3), 479-488. https://doi.org/10.1017/S1366728917000694

Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language & Communication Disorders, 45*(4), 480-493. https://doi.org/10.3109/13682820903198058

Swanson, H. L., & Siegel, L. (2001). Elaborating on working memory and learning disabilities: A reply to the commentators. *Issues in Education, 7*(1), 107-129.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism, 15*(4), 426-445. https://doi.org/10.1177/1367006911403202

Thordardottir, E. (2020). Are background variables good predictors of need for L2 assistance in school? Effects of age, L1, amount, and timing of exposure on Icelandic language and nonword repetition scores. *International Journal of Bilingual Education and Bilingualism, 23*(4), 400-422. https://doi.org/10.1080/13670050.2017.1358695

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders, 46*(1), 1-16. https://doi.org/10.1016/j.jcomdis.2012.08.002

Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders, 4*(1), 1-21. https://doi.org/10.1080/14769670500215647

Thordardottir, E., Keheyia, E., Lessard, N., Sutton, A., & Trudeau, N. (2010). Typical performance on tests of language knowledge and language processing of French-speaking 5-year-olds. *Revue Canadienne D'orthophonie et D'audiologie, 34*(1), 5-16.

Thorn, A. S., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology: Section A, 52*(2), 303-324. https://doi.org/10.1080/713755823

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54*(1), 68-80. https://doi.org/10.1016/j.jml.2005.06.003

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6), 1038-1066. https://doi.org/10.1037/0033-2909.133.6.1038

U.S. Census Bureau. (2000). *Languages spoken at home.* Retrieved August 1, 2009 from http://www.census.gov.

U.S. Census Bureau. (2013). *American community survey reports.* Retrieved November 14, 2015, from https://www.census.gov/prod/2013pubs/acs-22.pdf.

van Bon, W. H., & van der Pijl, J. M. (1997). Effects of word length and wordlikeness on pseudoword repetition by poor and normal readers. *Applied Psycholinguistics, 18*(1), 101-114. https://doi.org/10.1017/S0142716400009899

van den Noort, M., Vermeire, K., Bosch, P., Staudte, H., Krajenbrink, T., Jaswetz, L., ... & Lim, S. (2019). A systematic review on the possible relationship between bilingualism, cognitive decline, and the onset of dementia. *Behavioral Sciences, 9*(7), 81. https://doi.org/10.3390/bs9070081

van Dun, K., & Mariën, P. (2016). Cerebellar-induced aphasia and related language disorders. In P. Mariën & M. Manto (Eds.), *The linguistic cerebellum* (pp. 107-133). Academic Press. https://doi.org/10.1016/B978-0-12-801608-4.00006-2

Wechsler, D. 2003. *Wechsler Intelligence Scale for Children-4th Edition (WISC-IV)*. Harcourt Assessment. https://doi.org/10.1037/t15174-000

Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology, 14*(3), 292-305. https://doi.org/10.3109/17549507.2011.636071

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology, 19*(4), 298-310. https://doi.org/10.1044/1058-0360(2010/09-0064)

Wodniecka, Z., Craik, F. I., Luo, L., & Bialystok, E. (2010). Does bilingualism help memory? Competing effects of verbal ability and executive control. *International Journal of Bilingual Education and Bilingualism, 13*(5), 575-595. https://doi.org/10.1080/13670050.2010.488287

Yoshida, H., Tran, D.N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning. *Frontiers in Psychology, 2*, 210. https://doi.org/10.3389/fpsyg.2011.00210

Zamuner, T. S., Gerken, L., & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language, 31*(3), 515-536. https://doi.org/10.1017/S0305000904006233

Data, Code and Materials Availability Statement

The raw data and analysis syntax are available on the Open Science Framework at https://osf.io/dhj39/

Ethics Statement

Ethics approval was obtained from the ethics committee of the University of Alberta. All participants gave informed written consent before taking part in the study.

Authorship and Contributorship Statement

Farzaneh Anjomshoae conceived of the study, designed the study, analyzed the data and wrote the first draft of the manuscript. **Elena Nicoladis** contributed to the design of the study, collected the data, and revised the manuscript. **Anahita Shokrkon** revised the manuscript. All authors approved the final version of the manuscript.

Appendix

Table A1. ***Spearman correlations between Peabody Picture Vocabulary Test Scores, forward and backward digit span, and nonword repetition.***

	Age	PPVT	FDS	BDS	NWR
Age	–	.296	.120	.468**	.392*
PPVT	.553**	–	.285	.070	.300
FDS	.352*	.431**	–	.112	.099
BDS	.528**	.499**	.437**	–	.406*
NWR	.326*	.112	.296	.340*	–

Note: Top of matrix above diagonal indicates correlations for monolinguals, bottom of matrix below diagonal indicates correlations for bilinguals; PPVT = Peabody Picture Vocabulary Test; FDS = forward digit span; BDS = backward digit span; NWR = non-word repetition; $^{**}p < .01$, $^{*}p < .05$

Table A2. ***Partial correlations (controlling for age) between verbal working memory tasks in monolinguals and bilinguals.***

	FDS	BDS	NWR
FDS	–	.027	.277
BDS	.300	–	.340*
NWR	.151	.205	–

Note: Top of matrix above diagonal indicates correlations for monolinguals, bottom of matrix below diagonal indicates correlations for bilinguals; FDS = forward digit span; BDS = backward digit span; NWR = non-word repetition; *$p < .05$

License

Quality of remotely-collected gaze data in autistic and nonspectrum children

Rhiannon Luyster
Emerson College, Boston, MA

Taylor Boyd
New York University, New York, NY

Amelia Steele
Emerson College, Boston, MA

Thuy Buonocore
Emerson College, Boston, MA

Catherine Sancimino
Icahn School of Medicine at Mount Sinai, New York, NY

Sudha Arunachalam
New York University, New York, NY

Abstract: Many developmentalists have shifted to remote research. This project uses secondary data to evaluate the quality of eye-gaze data from 30 autistic children and a language-matched sample of 30 nonspectrum children (mean ages 48 and 27 months, respectively). All children completed an experimenter-moderated preferential looking paradigm via Zoom. Frequency of co-occurring child and household events, rates of missing data, and percent agreement between gaze coders were assessed. Results indicated that co-occurring events were minimal, with no diagnostic group differences. Missing data rates were low overall and were unrelated to diagnostic group, age, or language level of participants; however, higher rates of co-occurring child behaviors were associated with higher rates of missing data. Agreement between coders for eye gaze data was comparable to in-lab studies. Results affirm the usefulness of remote, experimenter-moderated gaze-based research with autistic and nonspectrum children.

Keywords: autism; preferential looking; online research; methodology

Corresponding author(s): Rhiannon J. Luyster, Communication Sciences and Disorders, Emerson College, 120 Boylston St., Boston, MA 02216. Email: rhiannon_luyster@emerson.edu

ORCID ID(s): Rhiannon Luyster: https://orcid.org/0000-0001-8311-4772; Taylor Boyd: https://orcid.org/0009-0004-3123-7550; Amelia Steele: https://orcid.org/0009-0003-1033-9841; Thuy Buonocore: https://orcid.org/0009-0007-3060-7048; Catherine Sancimino: https://orcid.org/0000-0003-3151-7734; Sudha Arunachalam: https://orcid.org/0000-0003-4394-3626

Citation: Luyster, R., Boyd, T., Steele, A., Sancimino, C., Buonocore, T., & Arunachalam, S. (2025). Quality of remotely-collected gaze data in autistic and nonspectrum children. *Language Development Research*, *5*(3), 131–154. http://doi.org/10.34842/ldr2025-633

Introduction

With the COVID-19 pandemic, researchers devised new strategies for pursuing their work, including remote data collection via videoconferencing platforms (Tsuji et al., 2022). This approach appeared to yield similar results to in-person paradigms (Bánki et al., 2022; Chuey et al., 2021; Steffan et al., 2023) and offered unanticipated enrollment benefits for sample diversity (Shields et al., 2021, Ozernov-Palchik et al., 2022), making it likely to persist. These new opportunities are exciting but also pose potential challenges. When participants are at home, the experimenter has less control over the environment (Gijbels et al., 2021), which could lead to poorer quality data. In the current study, we used secondary data analysis to examine the quality of eye-gaze data collected remotely with young children on the autism spectrum and with nonspectrum children. Both groups completed an experimenter-moderated language learning task using a variant of the preferential looking paradigm (Golinkoff et al., 1987) on Zoom.[1]

The measure of interest in the current study is children's eye gaze as they looked at their video screen and heard an auditory prompt directing their attention to a particular image. Gaze was recorded using a webcam and later coded offline by trained coders. This paradigm, sometimes referred to as "intermodal preferential looking" or "looking while listening" (e.g., Fernald et al., 2008; Golinkoff et al., 1987) has been successfully used with autistic children in lab settings (e.g., Bebko et al, 2006; Ellis Weismer et al., 2016; Horvath et al., 2018; Venker et al., 2013) and in the home with experimenters bringing a portable setup (e.g., Goodwin et al., 2012; Naigles & Tovar, 2012; Swensen et al., 2007).

The central construct of this investigation—the quality of eye-gaze data collected remotely—requires consideration of different metrics of "eye-gaze data quality". In terms of eye-gaze quality, one important metric is missing data; that is, those moments when direction of gaze cannot be determined or when the child is looking offscreen. Missing data are inevitable, because blinking results in missing data. However, it can also occur because, for example, child participants may lean forward to look more closely at the screen, leaving their eyes outside the camera's range, or they may turn their heads to look at a caregiver. Some of these behaviors may be influenced by setting (i.e., lab-based vs. remote home-based) and diagnosis. For instance, Lapidow and colleagues (2021) noted that caregivers were more inclined to interact

[1] The terms autism, autism spectrum and autism spectrum disorder (ASD) will be used interchangeably. Moreover, in light of recent dialogue (e.g., Botha et al., 2021) around diverse preferences for person-first versus identity-first language, the terms "on the autism spectrum" and "autistic" will both be used to refer to individuals with a confirmed diagnosis of ASD per the DSM-5 (APA, 2013). Finally, rather than referring to the comparison sample as "typically developing," we will use the term "nonspectrum".

with their children during online (vs. lab-based) administrations. Somewhat surprisingly, then, collecting remote rather than lab-based data from children in manually coded gaze-tracking paradigms has not consistently been shown to substantially influence rates of missing data, at least for nonspectrum children (e.g., Scott & Schulz, 2017). For example, Morini and Blair (2021) enrolled nonspectrum preschoolers and reported that the number of analyzable trials was comparable across face-to-face and virtual settings (ranging from a mean difference of .1 to 1.6 trials across ages and trial types). Similarly, Bacon and colleagues (2021) used a looking-while-listening virtual platform with nonspectrum toddlers; they reported that data integrity was robust against internet quality and that the percentage of includable trials (88%) was comparable to previous lab-based rates (e.g., 66% to 78% in Venker et al., 2020).

We might expect that missing data might be more common in remote paradigms for autistic (vs. nonspectrum) children, however. Consider the fact, for instance, that missing data can result from movement, and autistic children may be particularly prone to movement-related data loss (e.g., Venker et al., 2020). Moreover, given suggestions that autistic children may, on one hand, find gaze-tracking paradigms particularly challenging due to the need to remain relatively still (Venker & Kover, 2015) but, on the other hand, may participate more easily in the predictable environment of a home-based study (Gijbels et al., 2021), it is particularly important to see if remotely collected data quality differs for autistic and nonspectrum children. Most previous studies with autistic children using preferential looking paradigms in the home have had experimenters physically present with the child during the study; this allowed study staff to ensure a consistent setup, monitor the environment for distractions, and support the child and caregiver in following directions (e.g., Jyotishi et al., 2017; Potrzeba et al., 2015; Tovar et al., 2015; but see Arunachalam et al., 2024 for a recent example of a study with fully remote task administration). With the fully remote task administration required during the pandemic, the experimenter has less knowledge of what is occurring in the home environment, including the details of the setup as well as what unrelated stimuli might be co-present. Thus, a new look at data quality with this population is warranted; in this study, we explore the rate of missing data in our remote paradigm, as well as whether this differs by diagnostic group (autistic vs. nonspectrum).

Another important consideration in evaluating the quality of eye-gaze data is interrater reliability. Manual coding of gaze from video generally yields lower track-loss rates compared to automatic eye-tracking, including for children on the autism spectrum (Haviland et al., 1996; Venker & Kover, 2015; Venker et al., 2020). When using manual coding, it is important to have multiple coders and to quantify their agreement (e.g., Fernald et al., 2008). In the home setting, where we have limited control over the precise visual angle between the child and the screen, as well as (in the current study) over the exact dimensions of the screen being used, it is likely that coders will have more difficulty determining whether a child is looking, for example, to the right side of the screen or to an object to the right of the device. Prior findings suggest that remote data may present specific difficulties for agreement: Morini and Blair (2021) reported that inter-rater reliability was lower for remote data vs. a face-to-face setting. We therefore expect that remote data collection may result in lower agreement than lab-based paradigms or in-home studies where the experimenter is present and brings their own setup. Inter-rater reliability can be quantified by measuring the percent of frames on which coders agree on a particular code for direction of gaze, as well as by Cohen's kappa; Cohen (1960) suggested that kappa values of .81 or higher indicate excellent agreement. In our lab's training process, we require coders to achieve a kappa of >.9 with the training standard before they can code independently. Nevertheless, because percent agreement is more commonly reported in studies using this paradigm, we report here on percent agreement.

Therefore, in this study we use secondary data analysis to examine the quality of manually coded gaze data gathered from autistic and nonspectrum preschool-aged children via a remote platform by reporting on (1) missing data and (2) percent agreement among gaze coders. Additionally, we reviewed the videos to determine the frequency of co-occurring events that might affect the quality of the gaze data and asked whether these were associated with missing data or percent agreement among coders.

Method

All recruitment and testing procedures were approved by the Biomedical Research Alliance of New York (BRANY), which provides IRB services for multi-site studies.

Participants

Participants contributing data to this secondary data analysis are a subset of those in a larger study. A US national sample of families was recruited for a remote study focusing on language learning in children on the autism spectrum. Families of autistic children were recruited through online advertisements, a specialized recruitment service, and the SPARK national autism research registry (Feliciano et al., 2018). Families of nonspectrum children were recruited through online advertisements, parent organization emails, and our own research participant databases.

Children on the autism spectrum were eligible to participate in the larger study if they were 36.0 to 71.9 months old, had a previous medical or educational diagnosis of autism spectrum disorder (ASD), and scored 12 points or higher on the Social Communication Questionnaire (SCQ; Rutter et al., 2003), originally published as the Autism Screening Questionnaire (ASQ, Berument et al., 1999). Nonspectrum children were eligible if they were aged 24 to 48 months (younger than the autistic group in order to match groups on language, see below), had no previous diagnoses that would affect language or cognition, had no immediate family members diagnosed with autism, and if they scored less than 12 points on the SCQ. Across both groups, children were not eligible to participate if their caregiver reported they (a) were born before 37 weeks of pregnancy, (b) had uncorrected vision or hearing impairments, (c) were colorblind or (d) heard English less than 70% of the time. With regards to the latter requirement, parents reported on their child's language exposure. The 70% cutoff is based on Cattani et al. (2014), who found that bilingual preschoolers perform as well as monolingual English-learning preschoolers on standardized language assessments if they have at least 60% exposure to English; here, a stricter criterion of 70% is used because the word learning tasks tap into processing abilities that go beyond the offline performance measured in standardized assessments.

Participant diagnostic status was confirmed using a multi-step process that was developed to be suitable for remote data collection. As mentioned above, SCQ scores were used as a screening tool for eligibility. SCQ validity studies indicate an optimum cutoff score of 15 for children 4 years and older (Berument et al., 1999; Rutter et al., 2003), however, subsequent research has identified a lower cutoff score of ≥12 to yield best sensitivity and specificity for children younger than age 4 years (Allen et al., 2007; Corsello et al., 2007; Wiggins et al., 2007). Thus, because of the younger age of many of the children in the present sample, the current study utilized a cutoff score of 12, requiring that children in the autistic group score 12 or higher and that nonspectrum children score below 12. This SCQ cutoff score also allowed the researchers to cast a wider net for recruitment of autistic children (given that we also had a licensed clinical psychologist, the fourth author, confirm diagnosis using all available data, as noted below).

Next, we gathered caregiver report information about intervention and diagnostic history using a questionnaire, including services provided in the school and the community (see data on OSF). Caregivers reported on their child's current autism-related symptoms using the Gilliam Autism Rating Scale–Third Edition (Gilliam, 2014). Caregivers also completed the Vineland Adaptive Behavior Scales, Third Edition (Vineland-3; Sparrow et al., 2016) to provide information about adaptive functioning skills; because autism is commonly associated with impairments in adaptive functioning, results of the Vineland-3 were reviewed to determine whether the Communication, Daily Living, and Socialization scales were consistent with autism. Finally, caregivers and children completed a 15-minute guided, semi-structured and video recorded interaction based on an adaptation of the Childhood Autism Rating Scale–Second edition (CARS-2; Schopler et al., 2010). Previous research published near the onset of the COVID-19 global pandemic demonstrated that the CARS-2 can be effectively adapted to a brief observation entitled "CARS-2-obs", with the examiner providing prompts to the caregiver while observing their interaction to identify child behaviors indicative of autism (Sanchez & Constantino, 2020). A licensed psychologist with advanced training and expertise in autism diagnosis (spanning research and clinical settings) reviewed all available clinical materials to confirm diagnostic status. Nine participants from the larger study (out of 156) were recruited for the autistic group but did not receive diagnostic confirmation of autism based on clinical judgment after review of all available data.

Participant Matching

For the present analyses, we identified two subgroups of children (drawn from the larger study) who were matched by gender and language ability. To ensure comparability with similar in-lab studies, we selected a sample size based on previous research, which typically used groups of 30 or fewer participants per group (e.g., Goodwin et al., 2012; Venker, 2019; Venker et al., 2013). Our two goals in identifying subgroups for the present analyses were to include children with a wide range of language abilities, given a previous finding that children with lower language abilities were more likely to look away from the screen in a similar paradigm (Bebko et al., 2006), and within that, to match children on gender and expressive vocabulary.

Therefore, we first aimed to identify—from the 147 participants with confirmed autism diagnoses in the larger study—a subset of 30 children on the autism spectrum to include for the present analyses. To ensure a wide range of language abilities, we first binned all children in the larger study based on total number of words produced on the MacArthur Bates Communicative Development Inventory (MCDI) Words and Sentences Long Form (which contains 680 vocabulary words; Fenson et al., 2006). Six bins were used: fewer than 100 words, 100-199 words, 200-299 words, 300-399 words, 400-499 words and 500 or more. From each bin, we selected at least one child (when more than one child was available, we used random selection) while maintaining the same approximate MCDI distribution in the subsample that we had in the larger sample. This process yielded a subgroup of 30 children on the autism spectrum (18 males, 12 females) aged 36 to 67 months (M = 48.73, SD = 8.93), all of whom had their diagnosis confirmed by the licensed psychologist according to the process outlined earlier.

Next, we matched each autistic participant to a nonspectrum participant from the larger study based on gender and vocabulary scores from the MCDI (within +/- 20 words). Although +/- 20 on the MCDI is a relatively large spread, it has been previously used for language-based matching (e.g., Naigles et al., 2016) and allows the inclusion of children with lower vocabulary scores. A stricter matching protocol would have excluded autistic children with lower vocabulary scores due to difficulty in finding exact matches. When more than one nonspectrum child who was a vocabulary- and gender-match was available, a single one was selected at random. This process was successful for all autistic participants except for two, for whom a vocabulary matched participant within 20 words could not be identified; we therefore selected the closest available same-gender nonspectrum match for these children (one pair was matched within 21 words and the other was matched within 44 words; see similar approach in Luyster & Lord, 2009). This matching process resulted in our second subgroup, comprising 30 nonspectrum children (18 males, 12 females) aged 24 to 34 months ($M = 27.37$, $SD = 3.10$).

As is common when using language-matched sampling for young autistic and nonautistic children (see Charman, 2004 for a discussion), the autistic and nonspectrum groups differed significantly on age ($t = 12.38$, $p < .001$) and SCQ ($t = 12.31$, $p < .001$) but not on MCDI ($t = .094$, $p = .93$). See Table 1. Caregivers reported on children's race and ethnicity using NIH categories as follows: 3 Black/African-American, 3 Asian, 43 White, 3 More than one race, 2 Prefer not to answer; 8 Hispanic or Latine, 52 Not Hispanic or Latine. We did not explicitly ask for information pertaining to socioeconomic status, but we note that to participate, families had to have a sufficiently strong internet connection to engage in a Zoom call and watch streaming videos on an appropriate device.[2]

On the day of the study, the caregiver and child logged onto a Zoom meeting with the experimenter. Children sat in front of a desktop, laptop, or tablet with a screen at least 5.5 inches by 8.5 inches. The child usually sat independently, but some children sat in their caregiver's lap. Parents were coached on an appropriate distance to have the child sit from the screen, but we did not require them to measure it. We discouraged parents from having the child hold anything during the study, but if the parent believed the child would be better able to sit still and participate while holding a toy or eating we allowed it. The experimenter conducted a warm-up, followed by a word-learning experiment (described below), and then a 15-minute guided play-based observation between the child and caregiver. The session lasted about 45 minutes and was recorded using Zoom. For this paper, we only report details of the procedure relevant to the current analyses of eye-gaze data quality during the word-learning experiment.

Word-Learning Experiment

The experiment comprised two word-learning trials on which the child was introduced to a new word (e.g., "modi") and then tested on the novel word's meaning. For the current paper, we focus on the "test phase" of each of the two trials, which were structured identically. Test phases consisted of 8-second videos. The 8-second video comprised three phases: Baseline (3 seconds), during which children viewed the images on the screen (as depicted in Figure 1) and heard a prompt designed to direct their attention to the screen ("Whoa, look!"), Query (2 seconds), during which the images disappeared, replaced by a large central fixation image, and children heard an auditory prompt to find the target (e.g., "Where's the modi?"), and Response (3 seconds) during which the images reappeared in the same locations as during Baseline

[2] From the larger study, there were 24 families who attended an orientation call but did not show up for the study visit.

and children heard an additional prompt (e.g., "Find the modi!"). We analyzed their gaze during Baseline and Response, as would typically be done in word learning experiments; Baseline provides a measure of children's a priori preference for the images and Response provides a measure of their preferences after being asked to find the referent. We did not analyze gaze during the Query phase for two reasons. First, this phase did not have images in the target locations, and second, in our experience, this phase is often when children are likely to look away from the screen (e.g., to share attention with their parent).

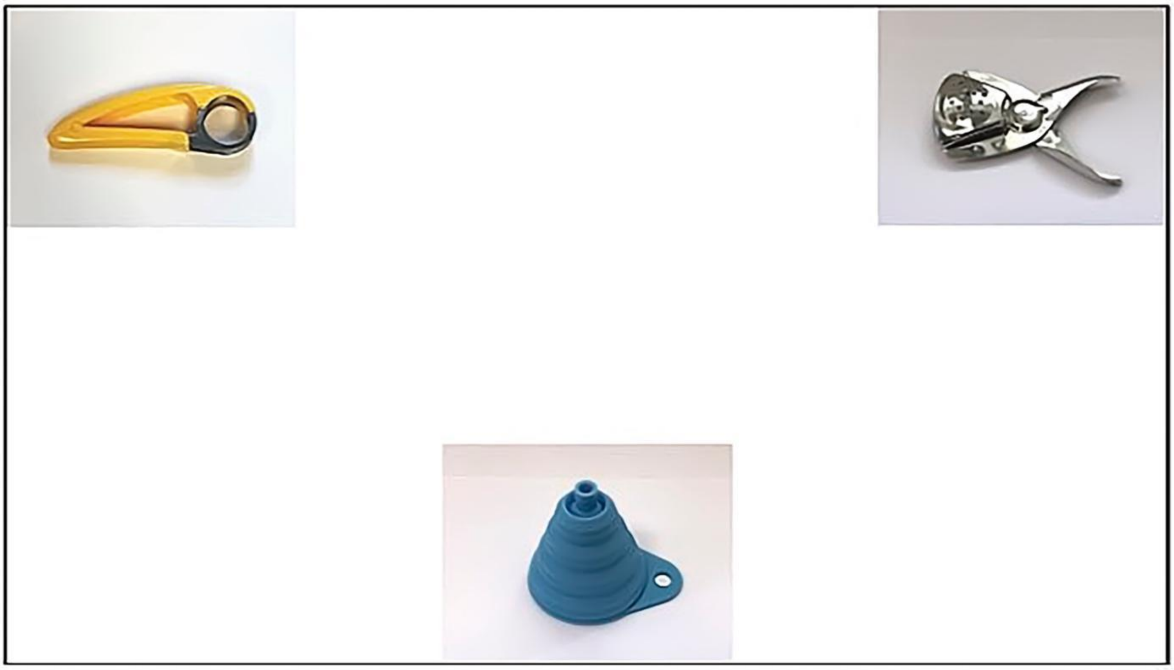

Figure 1. ***Example of visual stimuli during Baseline and Response phases***

Coding

We coded two types of variables: (1) co-occurring events, including both child behaviors and external household events, that might be expected to disrupt performance in an experimental task. (2) children's gaze behavior as the videos played, to assess both rates of missing data and inter-rater agreement.

Co-occurring events

We coded for child behaviors and external, household events (see Table 2 for definitions and guidelines used by coders). With respect to child behaviors, we coded for child vocalization or physically interacting with an object, and additionally, given Venker and Kover's (2015) suggestion that child behaviors associated with autism might lower the quality of eye-gaze data, we also coded for repetitive child sensorimotor movements that are characteristic of autism, such as rocking or hand flapping (e.g., Rutter, Le Couteur, & Lord, 2003).

Table 2. ***Coding definitions for co-occurring events***

Event Category	Coding Definition
Child sensorimotor movements	Includes whole body, torso, or arm/hand/finger movements. For example: rocking, hand flapping, peering through hands/fingers. These movements must include *voluntary* repetitive movements. Nail biting, hair twisting, thumb sucking and the like are all excluded. Chewing and drinking will not be coded; oral motor movements (e.g., popping lips, sticking out tongue, sucking thumb) in isolation (i.e., in the absence of other sensorimotor movements as listed) will not be coded. If this co-occurs with another category, code both.
Child vocalizations	Nonword vocalizations (e.g., laughter, jargoning) or speech (e.g., talking to caregiver, repeating audio from experimental stimuli). Making noises while chewing and drinking or breathing will not be coded. Do not count yawning, grunting (unless communicative), lip popping, sighing, raspberries. If this co-occurs with another category, code both.
Child physical distractions	Resulted from something the child was doing. Active involvement of/with physical object or agent resulting from child's behavior; for instance, child playing with a toy in hand or touching a computer keyboard. If the child is holding something or has something in their lap but they are not actively involved with it (meaning, they are not playing with it, looking with it, moving it around etc.), do not code. If this co-occurs with another category, code both.
External physical distractions	Sudden appearance/interruption by agent/physical object that (1) enters the child's visual field (e.g., sibling running in front of child) or (2) makes physical contact with the child (e.g., cat jumping on child's lap). These are not due to the child's behavior and do not include ongoing physical contact from the parent, who may be holding the child during the session. If this co-occurs with another category, code both.

With respect to external, household events, we coded for intrusions (e.g., from caregivers, pets or siblings) that entered the child's field of view or made physical contact with the child; this final category was only observed for one child in the data set, and so we did not analyze it quantitatively. We initially intended to include auditory distractions such as caregiver vocalizations or external noises (e.g., phone ringing, baby crying) in this last category, but Zoom recordings varied in how much of this external noise was filtered out by the software, and so we could not reliably determine whether these noises were present in the home for all videos.

Coding was done in 1-second bins for each of the 3 seconds in the Baseline period and 3 seconds in the Response period; each second was binary coded (i.e., presence or absence) for each of the four categories of co-occurring events. Videos were played in Adobe Premiere Pro and codes were recorded on a spreadsheet. All videos were coded for co-occurring events by two research assistants; inter-rater reliability was calculated for a randomly selected 20% of the sample. Percent agreement between the two raters was high for all four event categories: child sensorimotor movements (100%), child vocalizations (97.2%), child physical distractions (100%), external physical distractions (100%). After calculating inter-rater reliability metrics, disagreements were resolved via consensus between the same two coders.

Gaze Coding

Using standard procedures (e.g., Fernald et al., 2008), three trained research assistants who were naive to diagnosis independently coded the direction of children's gaze on the screen (top left, top right, center) from video recordings of the test phase at a rate of 30 frames/second.[3] The coders had to be able to hear the audio to determine the onset of Baseline and Response phases, but they did not know which object was the intended target or where it was located on the screen, both of which were counterbalanced across participants. Each video was coded by two of the three coders, who viewed videos using Adobe Premiere Pro software and recorded gaze codes on a spreadsheet. (Note that while most studies involve multiple coders on only a subset of trials to check reliability, we enlisted two coders for coding gaze on all videos because we were specifically interested in evaluating inter-rater agreement.) Missing data consisted of frames on which the eyes were closed (blinks), the child was looking outside of the areas of interest (e.g., looking off-screen, turning to look at a caregiver), or the child's eyes were not visible to the coder (e.g., blurry video, child was out of frame). The proportion of codes for these events, out of all of the 90 coded frames

[3] As in our prior work using this remote data collection (Arunachalam et al., 2024), we coded the videos using audio cues indicating the start of each phase within each trial to minimize any lag that might have accumulated during video playback. None of the included videos had an accumulated lag of more than 2 frames (66 ms) using this approach.

during each of the Baseline and Response phases of each test video, was calculated. Percent agreement to determine inter-rater reliability between the two coders were calculated for each trial using the "irr" package (Gamer, Lemon, Fellows, & Singh, 2019) in R version 4.1.2 (R Core Team, 2020).

Results

Co-occurring events

As described above, co-occurring events were categorized as child sensorimotor movements, child vocalizations, child physical distractions, or external physical distractions (see Figure 2).

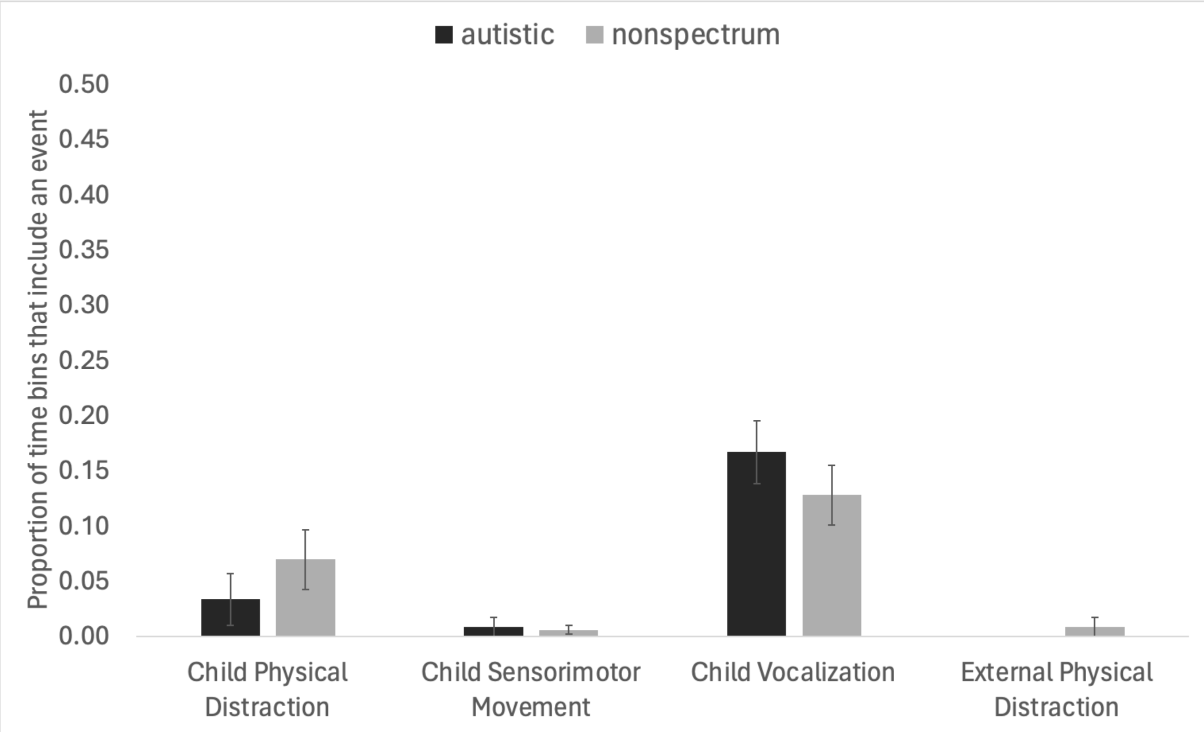

Figure 2. ***Proportion of 1-second time bins during which child behaviors or external physical distractions were observed across groups (autistic children, nonspectrum children) during the two coded phases of the trial (Baseline, Response).*** *Note:* The y-axis range is only depicted from 0 to 0.5 for readability. These events were coded across 6 seconds per trial (3 seconds of the Baseline phase, 3 seconds of the Response phase) over 60 trials per group (2 trials per child, 30 children per group).

Child vocalizations were the most common behaviors in both groups, followed by child physical distractions. Child sensorimotor movements were rare in both groups:

For 10 children in the autistic group and 9 children in the nonspectrum group, there were no child sensorimotor movements at all. As mentioned above, there was only one child who experienced an external event in the dataset, and so we did not analyze this category quantitatively. On half of the trials (n = 29 for the autistic group, n = 31 for the nonspectrum group), no events were coded at all. Given the small numbers of event occurrences, we used non-parametric two-tailed Mann-Whitney U tests for group differences; we found none (child sensorimotor movements: Mann-Whitney U = 450.5, $n_1 = n_2 = 30$, $p = 1.00$; child vocalizations: Mann-Whitney U = 530.5, $n_1 = n_2 = 30$, $p = 0.22$; child physical distractions: Mann-Whitney U = 406.0, $n_1 = n_2 = 30$, $p = 0.25$).

Missing Data

The mean proportion of frames with missing data during the Baseline phase was 0.097 (SD = 0.22, range 0-1) for the autistic group and 0.082 (SD = 0.20, range 0-1) for the nonspectrum group, and during the Response phase, it was 0.096 (SD = 0.20, range 0-1) for the autistic group and 0.059 (SD= 0.19, range 0-1) for the nonspectrum group. Missing data rates were severely right-skewed, and standard approaches to transform the data did not address this non-linearity; therefore, nonparametric approaches are more appropriate. We used quasibinomial nonparametric regression, or generalized additive models, with the "mgcv" package in R version 4.1.2 (R Core Team, 2020). (Note that the same results were obtained with parametric regression for all analyses, which are reported in our data repository.) We ran two generalized additive models, one for the Baseline phase and one for the Response phase, with missing data as the dependent variable and a fixed effect of diagnostic group (sum coded with the autistic group as +1 and the nonspectrum group as -1). For this and other regressions, we only report significant parameters of interest; full models are available at the OSF repository (https://osf.io/w9vmk/?view_only=dd1288d5ca014a0cbd0d472455c81c77). These analyses yielded no significant effects of diagnostic group (Baseline: $ß$ = 0.094, p = 0.70, deviance explained = 0.20%; Response: $ß$ = 0.26, p = 0.32, deviance explained = 1.49%).We then added age and MCDI scores (centered around their means); these were not highly correlated (R = 0.24) because of the heterogeneity of language abilities among the (chronologically older) autistic group. These analyses also yielded no significant effects for either the Baseline phase (Diagnostic group: $ß$ = -0.12, p = 0.82; Age: $ß$ = 0.019, p = 0.63; MCDI: $ß$ = -0.00094, p = 0.49; deviance explained = 0.85%) or the Response phase (Diagnostic group: $ß$ = 0.51, p = 0.31; Age: $ß$ = -0.026, p = 0.57; MCDI: $ß$ = -0.0016, p = 0.25; deviance explained = 5.66%), and we did not include these factors in subsequent missing data analyses.

To see if missing data rates were predicted by co-occurring child behaviors, we added to the simple models a fixed effect of the sum of the number of seconds (out of 3 seconds) during each phase (Baseline, Response) of each trial on which each of these behaviors were present (because 3 types of behaviors were measured during each of the 3 seconds, the range of values was 0-9), and its interaction with diagnostic group.

In the Baseline phase, this analysis yielded a significant simple effect of child behaviors (*ß* = 0.67, *p* = 0.00014), with a greater number of behaviors associated with higher rates of missing data, but no significant effect of group (*ß* = 0.85, *p* = 0.17) and no significant interaction (*ß* = -0.51, *p* = 0.13) (deviance explained = 18.3%). The same pattern obtained for the Response phase: a significant simple effect of child behaviors indicating that more behaviors was associated with higher rates of missing data (*ß* = 0.45, *p* = 0.0072), but no significant effect of group (*ß* = 0.22, *p* = 0.73) and no significant interaction (*ß* = 0.28, *p* = 0.40) (deviance explained = 10.4%).

Percent agreement among coders for gaze coding

We calculated percent agreement for gaze coding for each phase of each trial and participant separately and included those in similar analyses as for missing data. The mean percent agreement between gaze coders for the Baseline phase was 96.6% (*SD* = 8.1%) for the autistic group and 98.1% (*SD* = 8.1%) for the nonspectrum group; for the Response phase, it was slightly lower: 93.5% (*SD* = 16.0%) for the autistic group and 96.9% (*SD* = 8.3%) for the nonspectrum group. Percent agreement was left-skewed, so we used generalized additive models as above.

For both phases, this analysis did not yield a significant main effect of diagnostic group (Baseline: *ß* = -0.61, *p* = 0.33, deviance explained = 2.08%; Response: *ß* = -0.76, *p* = .14, deviance explained = 3.28%). We then added age and MCDI scores (centered around their means). For Baseline, this analysis yielded no significant main effects (group *ß* = 0.38, *p* = 0.76; age *ß* = -0.044, *p* = 0.34; MCDI *ß* = 0.00046, *p* = 0.79; deviance explained = 4.08%). For Response, there was still no significant main effect of group (*ß* = 0.84, *p* = 0.37), but there were significant effects of age (*ß* = -0.074, *p* = 0.035) and MCDI (*ß* = 0.0036, *p* = 0.0042) (deviance explained = 13.8%), indicating that there was higher agreement for younger children and children with higher MCDI scores.

Finally, we asked whether percent agreement was predicted by co-occurring child behaviors; we added to the simple model a fixed effect of the sum of the total number of seconds on each trial (out of 3 seconds) on which each of these behaviors was present for each trial (because three types of behaviors were measured during each of the 3 seconds, the range of values was 0-9), and its interaction with diagnostic group. This analysis yielded no significant simple effects and no significant interaction during either Baseline (behaviors *ß* = 0.089, *p* = 0.80; group *ß* = -0.49, *p* = 0.48; interaction *ß* = -0.27, *p* = 0.71; deviance explained = 2.43%) or Response (behaviors *ß* = -0.099, *p* = 0.63; group *ß* = -0.78, *p* = 0.21; interaction *ß* = 0.032, *p* = 0.94; deviance explained = 3.61%).

Discussion

The goal of this study was to explore the quality of remotely-collected eye-gaze data gathered from autistic and nonspectrum preschoolers. We quantified co-occurring events (both child and external/household) during brief Baseline and Responses phases of a word-learning task, and we tested the associations of co-occurring events with two common quality metrics (missing data and inter-rater reliability).

In our sample, sporadic co-occurring events—both child and external—were observed for many participants. As in lab-based settings, interruptions are inevitable during experimental sessions. However, these events were relatively infrequent and for half of the trials, these events did not occur at all. Moreover, children on the autism spectrum and nonspectrum children did not differ in rates of either child or external events. This is somewhat unexpected given that autistic children might be more prone to distraction and movement, and that the sensorimotor movements we coded for are characteristic of autism (Venker & Kover, 2015). It suggests that in a home-based remote testing condition, autistic children are not more likely to experience these interruptions compared to their nonspectrum peers. Moreover, our tallies indicated that external/household distractions were extremely rare, occurring for only one child (and affecting roughly .01 of time bins for nonspectrum children). This finding does not support previous suggestions that remote research may be particularly vulnerable to family interruptions and child attrition (Lapidow et al. 2021; Steffan et al., 2023). In our study, we believe that the pre-visit orientation video call that we provided may have helped caregivers create a focused environment for their child (Gijbels et al., 2021). Overall, then, these results attest to the suitability of curated remote-testing conditions.

Next, we explored missing data. We found that rates of missing data were relatively low and did not differ for children on the autism spectrum and nonspectrum children; this is in contrast to previous findings that autistic children look away from stimuli significantly more often than nonspectrum children (e.g., Tenenbaum et al., 2017). Moreover, given that many studies with autistic children of this age apply a criterion of >50% missing data when deciding whether to exclude children (e.g., Horvath et al., 2018; Venker et al., 2013; Venker, 2019; Venker et al., 2020), our exclusion rate on this basis would be just 4%—which is comparable to those studies for which exclusion rates range from approximately 5% (Horvath et al., 2018) to 16% (Venker et al., 2020). For subsequent analyses, we included even those children with high rates of missing data; these children would typically be excluded from analyses, but our reports of data quality are contingent on understanding how missing data is related to co-occurring events. Indeed, there was a significant association between co-occurring events and missing data across both the Baseline and Response phases. Therefore, even though overall rates of co-occurring events were quite low across groups, the frequency with which they occurred was associated with data loss. There was, however, no main or interaction effect of group, age or language level. This finding affirms the importance of the quality control measures that many researchers take in order to minimize distractions, whether in laboratory or remote settings. Our finding indicates that by minimizing both child-based or environmental artifacts, we can reduce data loss.

The percent agreement between coders who manually coded children's eye gaze (autistic = 94-97%; nonspectrum = 97-98%) was slightly lower than but similar to what has been reported for children on the spectrum or with other developmental conditions or language delays in lab-based settings or at-home studies in which the experimenter brings a portable setup: e.g., 98% in Venker et al. (2013) and Venker et al. (2021); 97% in Ellis Weismer et al. (2016); 93-99% for pre-term and full-term toddlers in Loi et al. (2017); 98% in Naigles et al. (2011). This was a somewhat surprising outcome for us given that we had substantially less control over factors that would influence coders' judgments, such as distance from the screen (which affects visual angle) and dimensions of the device's screen. Moreover, the agreement between raters was not detrimentally affected by child behaviors or household events.

We did find, however, that gaze coding agreement was higher for younger children and those with higher MCDI scores during the Response phase. In other words, raters were less reliable when coding the children who were older and/or had more pronounced developmental (or at least language) delays. This finding is particularly intriguing in light of the fact that—due to the language-matched nature of our sample—the autistic group was older than the nonautistic group. We are not certain of an explanation for this variability in reliability. Our results suggest that agreement was not related to child behaviors, so it is unlikely to be caused by differences in regulation or externalizing behaviors. An alternative explanation might be that—perhaps related to language delay—these children had less clearly defined gaze patterns when asked to identify an object, perhaps doing more exploratory scanning than directed gaze, leading to lower coder agreement. This interpretation is consistent with the fact that we found an effect of age and MCDI only in the Response phase, and not during the Baseline period. Although analyses addressing whether or not these children successfully learned the word, as measured by gaze during the Response phase, is beyond the scope of this paper, future analyses might help to support or disprove our hypothesized interpretation.

There are certainly benefits to remote research including increased familiarity (and perhaps comfort) of the home environment, reduced barriers for study visits, and broader inclusion of diverse samples, and our work here suggest that there are relatively few disruptions arising from co-occurring events in remote research designs. Nevertheless, there are other disadvantages of at-home studies that researchers should consider. One notable difference from lab-based studies is that in the lab, we can keep the surrounding environment free from material and visual distractions. In the home, we did ask caregivers to try to identify a space without a lot of clutter, but there were likely other objects nearby that could have attracted children's attention. In coding eye gaze, coders may have inferred that the child's gaze was directed to one side of the screen when in fact it was directed to something just beside the screen. This lack of precision in coding is a disadvantage of at-home studies, although the relatively high percent agreement among coders somewhat mitigates this concern. Another potential challenge is the difficulty of verifying that the child is viewing the stimuli as intended (Tompkins, 2022). While we believe we substantially minimized this concern by (1) offering a pre-visit orientation video call, (2) having the experimenter present during data collection, (3) checking in frequently with the caregiver during transitions from one part of the procedure to the next and (4) asking caregivers to turn on the "do not disturb" function on their devices, it is certainly possible that for example, colors appeared differently than we intended or that distracting notifications popped up on participants' screens.

Several important limitations of our work should be noted. First, in recruiting for this study, we explained to caregivers that children would be asked to sit in front of their home computer for the duration of the task; families who agreed to enroll were likely self-selected based on the likelihood that their child could meet the study demands. Therefore, the children (both on and off the spectrum) enrolled do not necessarily exhibit the full range of developmental heterogeneity observed at these ages. Second, our word-learning task was conducted synchronously; before and after the task, the child was interacting directly with an experimenter. These results may not generalize to other types of experimental paradigms that are less interactive or are asynchronous/unattended (e.g., Scott & Schulz, 2017). Third, because Zoom software filtered out background noise, we were not able to assess the frequency of household auditory distractions such as a phone ringing or baby crying. Fourth, this study was limited to families who had access to computer/tablet with Zoom software and a stable internet connection. Although upwards of 90% of American families have access to these resources as of 2021 (US Census Bureau, 2024), it is certainly possible that the patterns reported here might differ in the remaining 10% of families and/or in families who feel distrust for technology use, particularly in a research context (Beaton et al., 2017). Finally, due to the COVID-19 pandemic and the cessation of in-person data collection in our labs, we did not compare these results to an in-person version of this same task. Instead, we drew inferences from published in-lab studies, and—in doing so—we also want to recognize some differences between our paradigm and those in-lab studies; for example, our task was modeled after an in-person study that presented three test objects, while eye-tracking tasks often present only two. We do not expect this choice to substantially affect our conclusions about overall quality of remotely collected data because our primary questions of interest in the current paper did not concern whether children looked at a target or distractor but rather how easy it was to assess whether and where they were looking. However, it may explain the slightly decreased percent agreement statistics as compared to in-lab studies because there were more possible codes to choose from. Our study also included a relatively small number of trials; while word learning studies commonly present only one trial (e.g., Dautriche et al., 2014; Yuan et al., 2012) or 2-4 trials (e.g., Horvath et al., 2018; Gliga et al., 2012; Naigles et al., 2011), other studies using similar paradigms with familiar/known words often have many more trials, including some of the studies we cite above as reference points (e.g., Venker et al., 2013; Venker, 2019; Venker et al., 2020). Our results are most straightforwardly relevant for other studies with similar task demands and may not generalize to other paradigms.

In sum, our findings suggest that—for both autistic and nonspectrum children—the data gathered from a remote gaze-data paradigm are characterized by minimal missing data and adequate agreement between coders. Child and household factors were noted (and the former were more frequent than the latter), and although the quality of gaze data was reduced by co-occurring events during the session, these events were generally infrequent. In a broader sense, the current study allowed us to test whether autistic and nonspectrum children differ from each other in remote gaze-based studies, which fills a crucial gap missing from prior work. We conclude that experimenter-moderated remote data collection offers a promising alternative to lab-based settings for manually-coded gaze paradigms for both autistic and nonspectrum children.

References

Allen, C.W., Silove, N., Williams, K. et al. (2007). Validity of the Social Communication Questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders, 37,* 1272–1278. https://doi.org/10.1007/s10803-006-0279-7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Arunachalam, S., Steele, A., Pelletier, T., & Luyster, R. (2024). Do focused interests support word learning? A study with autistic and nonautistic children. *Autism Research, 17*(5), 955-971. https://doi.org/10.1002/aur.3121

Bacon, D., Weaver, H., & Saffran, J. (2021). A framework for online experimenter-moderated looking-time studies assessing infants' linguistic knowledge. *Frontiers in Psychology*, 4078. https://doi.org/10.3389/fpsyg.2021.703839

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam-and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933. https://doi.org/10.3389/fpsyg.2021.733933

Beaton, B., Perley, D., George, C., & O'Donnell, S. (2017). Engaging remote marginalized communities using appropriate online research methods. *The Sage Handbook of Online Research Methods*, 563-577.

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry, 47*(1), 88-98. https://doi.org/10.1111/j.1469-7610.2005.01443.x

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry, 175,* 444-51. https://doi.org/10.1192/bjp.175.5.444

Botha, M., Hanlon, J., & Williams, G. L. (2021). Does language matter? Identity-first versus person-first language use in autism research: A response to Vivanti. *Journal of Autism and Developmental Disorders, 53*(2), 870-878. https://doi.org/10.1007/s10803-020-04858-w

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

Charman, T. (2004). Matching preschool children with autism spectrum disorders and comparison children for language ability: Methodological challenges. *Journal of Autism and Developmental Disorders, 34*, 59-64. https://doi.org/10.1023/B:JADD.0000018075.77941.60

Chuey, A., Asaba, M., Bridgers, S., Carrillo, B., Dietz, G., Garcia, T., Leonard, J. A., Liu, S., Merrick, M., Radwan, S., Stegall, J., Velez, N., Woo, B., Wu, Y., Zhou, X. J., Frank, M. C., & Gweon, H. (2021). Moderated online data-collection for developmental research: Methods and replications. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.734398

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L., & Lord, C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, 48*(9), 932–940. https://doi.org/10.1111/j.1469-7610.2007.01762.x [dataset] Arunachalam, S., & Luyster, R. J.; 2022; Quality of remotely-collected gaze data in children with and without autism; Open Science Framework https://osf.io/w9vmk/?view_only=dd1288d5ca014a0cbd0d472455c81c77

Dautriche, I., Cristia, A., Brusini, P., Yuan, S., Fisher, C., & Christophe, A. (2014). Toddlers default to canonical surface-to-meaning mapping when learning verbs. *Child Development, 85*(3), 1168-1180. https://doi.org/10.1111/cdev.12164

Ellis Weismer, S., Haebig, E., Edwards, J., Saffran, J., & Venker, C. E. (2016). Lexical processing in toddlers with ASD: Does weak central coherence play a role? *Journal of Autism and Developmental Disorders, 46*(12), 3755–3769. https://doi.org/10.1007/s10803-016-2926-y

Feliciano, P., Daniels, A. M., Snyder, L. G., Beaumont, A., Camba, A., Esler, A., ... & Brewster, S. J. (2018). SPARK: a US cohort of 50,000 families to accelerate autism research. *Neuron, 97*(3), 488-493. https://doi.org/10.1016/j.neuron.2018.01.015

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006). MacArthur-Bates Communicative Development Inventories, Second edition. *PsycTESTS Dataset*. https://doi.org/10.1037/t11538-000

Fernald, A. E., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening. *Language Acquisition and Language Disorders,* 97–135. https://doi.org/10.1075/lald.44.06fer

Gamer, M., Lemon, J., & Fellows, I., & Singh, P. (2019). *Irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Gijbels, L., Cai, R., Donnelly, P. M., & Kuhl, P. K. (2021). Designing virtual, moderated studies of early childhood development. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.740290

Gilliam, J. E. (2014). Gilliam autism rating scale–Third edition (GARS-3). *Austin, TX: Pro-Ed.*

Gliga, T., Elsabbagh, M., Hudry, K., Charman, T., Johnson, M. H., & BASIS team. (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development, 83*(3), 926-938. https://doi.org/10.1111/j.1467-8624.2012.01750.x

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child*

Language, 14(1), 23-45. https://doi.org/10.1017/S030500090001271X

Goodwin, A., Fein, D., & Naigles, L. R. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. *Autism Research, 5*(2), 109–123. https://doi.org/10.1002/aur.1220

Haviland, J. M., Walker-Andrews, A. S., Huffman, L. R., Toci, L., & Alton, K. (1996). Intermodal perception of emotional expressions by children with autism. *Journal of Developmental and Physical Disabilities, 8*(1), 77–88. https://doi.org/10.1007/bf02578441

Horvath, S., McDermott, E., Reilly, K., & Arunachalam, S. (2018). Acquisition of verb meaning from syntactic distribution in preschoolers with autism spectrum disorder. *Language, Speech, and Hearing Services in Schools, 49*(3S), 668–680. https://doi.org/10.1044/2018_lshss-stlt1-17-0126

Jyotishi, M., Fein, D., & Naigles, L. (2017). Investigating the grammatical and pragmatic origins of wh-questions in children with autism spectrum disorders. *Frontiers in Psychology, 8*, 319. https://doi.org/10.3389/fpsyg.2017.00319

Lapidow, E., Tandon, T., Goddu, M., & Walker, C. M. (2021). A tale of three platforms: Investigating preschoolers' second-order inferences using in-person, zoom, and Lookit methodologies. *Frontiers in Psychology, 12*, 731404. https://doi.org/10.3389/fpsyg.2021.731404

Loi, E. C., Marchman, V. A., Fernald, A., & Feldman, H. M. (2017). Using eye movements to assess language comprehension in toddlers born preterm and full term. *The Journal of Pediatrics, 180*, 124-129. https://doi.org/10.1016/j.jpeds.2016.10.004

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism Diagnostic Observation Schedule, Second Edition. *Torrance, CA: Western Psychological Services.*

Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology, 45*(6), 1774. https://doi.org/10.1037/a0016223

Morini, G., & Blair, M. (2021). Webcams, songs, and vocabulary learning: A comparison of in-person and remote data collection as a way of moving forward with child-language research. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.702819

Naigles, L. R., Cheng, M., Rattanasone, N. X., Tek, S., Khetrapal, N., Fein, D., & Demuth, K. (2016). "You're telling me!" The prevalence and predictors of pronoun reversals in children with autism spectrum disorders and typical development. *Research in Autism Spectrum Disorders*, *27*, 11-20. https://doi.org/10.1016/j.rasd.2016.03.008

Naigles, L. R., Kelty, E., Jaffery, R., & Fein, D. (2011). Abstractness and continuity in the syntactic development of young children with autism. *Autism Research, 4*(6), 422-437. https://doi.org/10.1002/aur.223

Naigles, L. R., & Tovar, A. T. (2012). Portable intermodal preferential looking (IPL): Investigating language comprehension in typically developing toddlers and young children with autism. *JoVE (Journal of Visualized Experiments), 70,* e4331. https://doi.org/10.3791/4331

Ozernov-Palchik, O., Olson, H. A., Arechiga, X. M., Kentala, H., Solorio-Fielder, J. L., Wang, K. L., ... & Gabrieli, J. D. (2022). Implementing remote developmental research: A case study of a randomized controlled trial language intervention during COVID-19. *Frontiers in Psychology, 12*, 734375. https://doi.org/10.3389/fpsyg.2021.734375

Potrzeba, E. R., Fein, D., & Naigles, L. (2015). Investigating the shape bias in typically developing children and children with autism spectrum disorders. *Frontiers in Psychology*, *6*, 446. https://doi.org/10.3389/fpsyg.2015.00446

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Rutter, M., Bailey, A., & Lord, C. (2003). The Social Communication Questionnaire (SCQ). *Torrance, CA: Western Psychological Services*.

Rutter, M., Le Couteur, A., & Lord, C. (2003). Autism Diagnostic Interview-Revised (*ADI-R*). *Torrance, CA: Western Psychological Services*.

Sanchez, M. J., & Constantino, J. N. (2020). Expediting clinician assessment in the diagnosis of autism spectrum disorder. *Developmental Medicine & Child Neurology*, *62*(7), 806–812. https://doi.org/10.1111/dmcn.14530

Schopler, E., Van Bourgondien, M., Wellman, G., & Love, S. (2010). The Childhood Autism Rating Scale 2nd edition (CARS-2). *Torrance, CA: Western Psychological Services*.

Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind*, *1*(1), 4-14. https://doi.org/10.1162/OPMI_a_00002

Shields, M. M., McGinnis, M. N., & Selmeczy, D. (2021). Remote research methods: Considerations for work with children. *Frontiers in Psychology, 12*, 703706. https://doi.org/10.3389/fpsyg.2021.703706

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). Vineland Adaptive Behavior Scales, Third Edition (Vineland-3). San Antonio, TX: Pearson.

Steffan, A., Zimmer, L., Arias-Trejo, N., Bohn, M., Ben, R. D., Flores-Coronado, M. A., et al. (2023). Validation of an open source, remote web-based eye-tracking method (WebGazer) for research in early childhood. *PsyArXiv*. https://doi.org/10.1111/infa.12564

Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of language acquisition in children with autism: Evidence from preferential looking. *Child Development*, *78*(2), 542–557. https://doi.org/10.1111/j.1467-8624.2007.01022.x

Tenenbaum, E. J., Amso, D., Righi, G., & Sheinkopf, S. J. (2017). Attempting to "increase intake from the input": Attention and word learning in children with autism. *Journal of Autism and Developmental Disorders*, *47*(6), 1791–1805. https://doi.org/10.1007/s10803-017-3098-0

Tompkins, D. (2022). What do our participants really see during unmoderated remote studies? *International Congress of Infant Studies (ICIS)*. https://infantstudies.org/what-do-our-participants-really-see-during-unmoderated-remote-studies/

Tovar, A. T., Fein, D., & Naigles, L. R. (2015). Grammatical aspect is a strength in the language comprehension of young children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, *58*(2), 301-310. https://doi.org/10.1044/2014_JSLHR-L-13-0257

Tsuji, S., Amso, D., Cusack, R., Kirkham, N., & Oakes, L. M. (2022). Editorial: Empirical research at a distance: New Methods for Developmental Science. *Frontiers in Psychology, 13,* 1-5. https://doi.org/10.3389/fpsyg.2022.938995

US Census Bureau. (2024). *Computer and Internet Use in the United States: 2021*. United States Census Bureau. https://www.census.gov/newsroom/press-releases/2024/computer-internet-use-2021.html

Venker, C. E. (2019). Cross-situational and ostensive word learning in children with and without autism spectrum disorder. *Cognition*, *183,* 181–191. https://doi.org/10.1016/j.cognition.2018.10.025

Venker, C. E., Eernisse, E. R., Saffran, J. R., & Weismer, S. E. (2013). Individual differences in the real-time comprehension of children with ASD. *Autism Research, 6*(5), 417–432. https://doi.org/10.1002/aur.1304

Venker, C. E., & Kover, S. T. (2015). An open conversation on using eye-gaze methods in studies of neurodevelopmental disorders. *Journal of Speech, Language, and Hearing Research, 58*(6), 1719–1732. https://doi.org/10.1044/2015_jslhr-l-14-0304

Venker, C. E., Mathée, J., Neumann, D., Edwards, J., Saffran, J., & Ellis Weismer, S. (2021). Competing perceptual salience in a visual word recognition task differentially affects children with and without autism spectrum disorder. *Autism Research, 14*(6), 1147-1162. DOI: https://doi.org/10.1002/aur.2457

Venker, C. E., Pomper, R., Mahr, T., Edwards, J., Saffran, J., & Ellis Weismer, S. (2020). Comparing automatic eye tracking and manual gaze coding methods in young children with autism spectrum disorder. *Autism Research, 13*(2), 271–283. https://doi.org/10.1002/aur.2225

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The utility of the Social Communication Questionnaire in screening for autism in children referred for early intervention. *Focus on Autism and Other Developmental Disabilities, 22*(1), 33-38. https://doi.org/10.1177/10883576070220010401

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development, 83*(4), 1382-1399. https://doi.org/10.1111/j.1467-8624.2012.01783.x

Data, Code and Materials Availability Statement

The data and code reported on here are available in Open Science Framework at: https://osf.io/w9vmk/?view_only=dd1288d5ca014a0cbd0d472455c81c77. The methods and analyses were not preregistered.

Ethics Statement

Ethics approval was obtained from the Biomedical Research Alliance of New York (BRANY). All participants gave informed written consent before taking part in the study.

Authorship and Contributorship Statement

Rhiannon Luyster and **Sudha Arunachalam** conceived of the study. **Taylor Boyd**, **Amelia Steele**, **Thuy Buonocore**, and **Catherine Sancimino** made substantial contributions to data collection, coding, analysis, and interpretation. **Rhiannon Luyster**

and **Sudha Arunachalam** drafted the manuscript. All authors contributed to and approved the final version.

Acknowledgements

This work was supported by the National Institutes of Health (NIH R01DC017131) to Sudha Arunachalam and Rhiannon Luyster. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would also like to express our gratitude to María Cobo Nieto and Taina Hernandez McShane for their help with data collection and coding. We would also like to thank the many children and families who shared their time with us as part of this study, including all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population dataset by applying at https://base.sfari.org. We are also grateful to our Autistic Advisory Board for their insights into the autistic experience.

License

Effects of reduced exposure to societal language on vocabulary and morphological knowledge of bilingual children

Anat Prior
Gal Pedael
Edmond J. Safra Brain Research Center for the Study of Learning Disabilities
Faculty of Education, University of Haifa, Israel

Abstract: Children's vocabulary and morphological knowledge both arise from exposure to their surrounding language, albeit through different learning mechanisms. Vocabulary is driven mostly by exposure to specific words, namely token exposure, whereas knowledge of morphological regularities also arises from cumulative exposure to language patterns, namely type exposure. Here we examine the impact of the reduced exposure of bilingual children to the societal language, Hebrew, on their vocabulary and morphological knowledge. The study included 148 preschool children (half bilingual) who performed a productive vocabulary task, two inflection and two derivation tasks. One of the inflectional tasks used pseudo-words in order to examine abstract morphological knowledge while neutralizing lexical knowledge. Overall, bilingual children showed lower performance than monolingual peers across both vocabulary and morphological tasks. Importantly, error analyses, tapping into participants' ability to utilize morphological knowledge in the absence of lexical representations, revealed equivalent performance of bilingual and monolingual children in inflection, and small differences in derivation. Methodologically, these results highlight the importance of de-coupling lexical and morphological knowledge, especially when studying bilingual individuals. Theoretically, the current findings suggest that the acquisition of morphological regularities, driven mostly by type exposure, is more resilient than the acquisition of lexical knowledge, driven by token exposure, in the face of reduced exposure associated with bilingualism.

Keywords: bilingual, token, type, vocabulary, morphology.

Corresponding author(s): Anat Prior, Department of Learning Disabilities and Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, 199 Abba Hushi Blvd., Haifa, Israel. Email: aprior@edu.haifa.ac.il

ORCID ID(s): Anat Prior: https://orcid.org/0000-0001-8026-2825

Citation: Prior, A. & Pedael, G. (2025). Effects of reduced exposure to societal language on vocabulary and morphological knowledge of bilingual children. *Language Development Research*, *5*(3), 155–190. http://doi.org/10.34842/ldr2025-770

Introduction

Bilingual children grow up being exposed to two languages. In the case of minority or heritage language speakers, the home environment usually supports the minority language, whereas children engage with the majority, or societal, language outside the home, and often in the school environment (Armon-Lotem et al., 2019; Paradis, 2023). Because bilingual children divide their time between the two languages, their exposure to the societal language is reduced in comparison to monolingual peers, who are exposed exclusively to the majority language across different social contexts. Exposure is a driving force of language acquisition in children, and thus bilingual children acquire the societal language at a slower rate than monolingual peers (Pearson et al., 1994; Uccelli & Páez, 2007). Such findings have been demonstrated for vocabulary (In English: Bialystok et al., 2010; Hoff, 2021; in Hebrew: Prior et al., 2014), morphology (In English: Kieffer & Box, 2013; Nicoladis et al., 2007, for a meta-analysis see Bratlie et al., 2022) and syntax (In Greek: Andreou & Tsimpli, 2020; in English: Chondrogianni & Marinis, 2011).

However, reduced exposure to the societal language might not impact all domains of linguistic knowledge in the same manner. Specifically, lexical knowledge, which requires repeated encounters with each individual word (token), might be more strongly affected than morphological knowledge, which requires accumulating a critical mass of encounters with linguistic regularities across different items (type exposure; Bybee, 2007). Because specific words appear less frequently in the language than do morphological structures that are shared by many words, the impact of reduced exposure is likely to be greater on the former than the latter. Further, the acquisition of such linguistic regularities is likely influenced by the morphological structure of the societal language, and the frequency and consistency of each morphological regularity.

The current study investigates the inflectional and derivational morphological knowledge in the societal language of bilingual Hebrew speaking 5-6 year-old children, in comparison with their monolingual peers. A careful understanding of the consequences of reduced exposure for vocabulary and morphology can inform theories of the interplay between item specific knowledge and knowledge of regularities (e.g., Ramirez et al., 2014; Sparks & Deacon, 2015). In addition, in light of the central role of morphological knowledge in literacy acquisition in Hebrew (Share & Bar On, 2019), identifying gaps in the morphological knowledge of young bilingual children is critical for developing effective instruction to optimally prepare them for school entry.

Usage-Based Models of Language Acquisition

Children learn language through daily exposure and social interactions, according to usage-based models (Beckner et al., 2009; Lieven et al., 2003; Tomasello, 2001). Through such exposure to language, children learn both specific linguistic units (words) and linguistic regularities (morpho-syntax; Ellis et al., 2015). Importantly, though both facets of knowledge are driven by the input that children are exposed to, different aspects of their exposure contribute to each of them (Bybee, 2007; Fejzo, 2021). Thus, word knowledge is acquired most directly through token frequency, namely the number of exposures or encounters with a specific lexical item. Morphological knowledge, on the other hand, is linked to type frequency, namely the number of encounters with different words sharing a linguistic pattern (Ravid, 2019b; see also Michaly & Prior, 2025; Shahar-Yames et al., 2018). In both cases, researchers posit that there is a critical mass, or threshold, of exposure that is necessary before the linguistic form is successfully acquired by the learner (e.g. Fejzo, 2021; Marchman & Bates, 1994). However, critical exposure to morphological regularities, at least those encountered relatively often in the language, is likely to accrue more quickly than the critical necessary exposure to individual tokens.

Token and type frequency are of course not independent of each other. With greater linguistic input, the learner can generalize more morphological patterns by linking them to specific items in their lexicon (Shahar-Yames et al., 2018). Abstract categories are gradually learned from the items children have been exposed to, based on frequency in the input (Ashkenazi et al., 2020; Bybee, 2007). Nonetheless, as stated above, reduced linguistic input, as is the case for many bilingual children, may differentially affect lexical and morphological knowledge. Here, we investigate this question in the context of Hebrew, a language with an exceptionally rich morphology.

Hebrew Morphology

Derivational Morphology

Hebrew is characterized by a productive and complex morphology. Most Hebrew words have a morphological root, which consists of three consonants and carries the main semantic meaning (Bolozky, 2007; Schwarzwald, 2002). Root morphemes are embedded in nominal or verbal pattern morphemes, which provide the lexical category of the word. The Hebrew lexicon is based mainly on the non-linear combinations of consonantal root and affixal patterns (Ravid, Ashkenazi et al., 2016). The system includes seven verb patterns and approximately 100 noun patterns. Adjectives are formed using specific nominal or verbal patterns. For example, the root *g-d-l* combines with verbal patterns: CaCaC to form the verb *gadal* 'grew up' and hiCCiC to form the verb *higdil* 'enlarge'. The same root also combines with nominal pattern: CCiCa to form the noun *gdila* 'growth' and CiCuC to form the noun *gidul* 'growth/tumor'. Finally, the same root also combines with the nominal/adjectival pattern CaCoC to form the adjective *gadol* 'large' (Bolozky, 2007; Schwazwald, 2002). This process creates a family of distinct words, all derived from the same root (Ashkenazi et al., 2016; Ravid, Ashkenazi et al., 2016). Importantly, not all roots combine exhaustively with all patterns. For example, a combination of the root *g-d-l* with the passive verb pattern niCCaC, which would create the form *nigdal*, is not a word in the Hebrew lexicon.

Acquisition of the Hebrew derivational verb system starts as early as age two (Berman, 1985, 2016; Ravid, 2019a). Between ages 3-5 years, children occasionally combine a consonantal root into an inappropriate pattern, showing growing awareness of the verbal morphological system, but incomplete knowledge of all lexical forms (Berman, 2003). For example, children may combine the root *p-r-k* with the niCCaC pattern and say *nifrak* instead of *hitparek* 'fall to pieces' in hitCaCeC pattern. Between the ages of 5-6 years, children acquire the full verbal pattern system (Ben-Zvi & Levie, 2016; Berman, 1985). Adjective acquisition has a more protracted developmental trajectory (Ravid, Bar-On et al., 2016). Awareness of roots increases with development, and schooling plays an important role in this process. In contrast, pattern awareness emerges later, towards adolescence, and plays a major role in Hebrew word reading and spelling (Ben-Zvi & Levie, 2016; Share & Bar-On, 2018; Ravid, 2011).

In the current study, we probe children's derivational knowledge of two structures: Deriving a verb from a noun, and deriving an adjective from a verb. In the verb derivation task, children were presented with a sentence frame including a noun, and had to then identify the root morpheme and use it to derive the appropriate verb, using one of the 3 active patterns. As described above, roots do not combine exhaustively with patterns, such that for each item there was a single correct response. In the adjective derivation task, children were presented with a sentence frame including a verb, and had to then identify the root morpheme and use it to derive the appropriate resultative adjective, using one of 3 passive verbal patterns (which also denote resultative adjectives), each corresponding to an active verbal pattern. Here as well, roots do not combine exhaustively with patterns, so each item only had one correct response.

Inflectional Morphology

Inflectional morphemes indicate different grammatical properties of words such as tense, person, gender and number. The inflectional morphology of Hebrew is mostly transparent and systematic, across the nominal, verbal and adjectival systems (Bolozky, 2007; Schwarzwald, 2002), but there are still some exceptions. For example, in the nominal system pluralization suffixes differ by the grammatical gender of the word – "*im*" for masculine, and "*ot*" for feminine. However, there are also exceptions to the rule. For example, the Hebrew singular noun *kir* 'wall' is a masculine noun but it takes the plural feminine suffix *ot*: *kirot* 'walls' (Armon-Lotem & Reznick, 2022).

Modern Hebrew contains over 200 masculine nouns that take the *ot* suffix and about 50 feminine nouns that take the *im* suffix, compared with tens of thousands of nouns with gender-linked plurals (Schwartz et al., 2009). Another reason for pluralization irregularity are morpho-phonological alterations of the stem. Thus, for example, the plural form of *simla* 'dress', a feminine noun, is not *simlot* but rather *smalot* 'dresses' (which includes a stem change). Children who are unfamiliar with the specific item tend to produce regularization errors when inflecting such words (Schwartz et al., 2009). Verb inflection in Hebrew includes suffixes marking gender and number (identical with those applied in the nominal system), and systematic pattern changes marking person and tense. Here too exceptions arise from morpho-phonological alterations.

Children generally acquire regular structures before irregular structures. In addition, in Hebrew the masculine form is acquired before the feminine form (Armon-Lotem & Reznick, 2022). Irregular forms are subject to frequency effects, as their memorization depends on opportunities for learning (Schwartz et al., 2009). The acquisition of inflectional morphology is dependent on development of the content-word and function-word lexicon, and on children's developing understanding of syntactic-semantic relations. Hebrew speaking children start marking inflections toward the end of the second year of life (Berman, 1985; 2016).

In the current study, we examine two inflection processes in Hebrew – noun pluralization and verb inflection (for person, number and tense). The noun pluralization task focused mainly on irregular inflections, namely words including a gender atypical suffix, a stem change, or both, because by the age of 6 Hebrew speaking children have mastered the regular pluralization of nouns. The verb inflection task required children to change the verb's person, gender or tense in order to fit a syntactic frame. Because the task utilized a pseudo-root in Hebrew (*š-l-z*), all inflection processes were fully regular.

Effects of Reduced Exposure on Vocabulary and Morphological Knowledge

Bilingual children, because of dividing their exposure across two languages, usually have smaller vocabularies in each of their language relative to monolingual peers (Bialystok et al., 2010; Hoff, 2021). This pattern has also been documented for bilingual Hebrew speaking children (Altman et al., 2017; Armon-Lotem et al., 2011; Katzir et al., 2019; Michaly & Prior, 2025; Schwartz, 2014; Schwartz & Katzir, 2012; Shahar-Yames et al., 2018) and adolescents (Prior et al., 2014). A recent meta-analysis reports that bilingual children speaking various societal languages have lower morphological knowledge, of both inflection and derivation, than monolingual peers (Bratlie et al., 2022). However, research regarding the morphological development of bilingual Hebrew speaking children is more limited.

Monolingual Hebrew speaking children learn to use morphological structures and to make generalizations that aid in learning new words from around 2 years of age (Ben-Zvi & Levie, 2016; Berman, 2016). Morphological learning of both inflection and derivation is interwoven with lexical growth (Ravid, 2019a). A study conducted among Hebrew-speaking toddlers around the age of two found that lexical learning in Hebrew is morphologically oriented, such that children's learning of verb inflection and derivation is coupled with the development of the verb lexicon (Ashkenazi et al., 2020). Bidirectional links between vocabulary and morphological knowledge have also been documented for bilingual Hebrew speaking elementary-school aged children (Michaly & Prior, 2025; Shahar-Yames et al., 2018).

Inflectional morphology is a highly regular and frequent system that children acquire early (Kuo & Anderson, 2006). Nevertheless, in inflectional morphology tasks in English, bilingual children with diverse language backgrounds demonstrate lower performance than monolingual children (e.g. Rattansone & Demuth, 2023). However, in Hebrew, several studies show that Russian-Hebrew bilingual children (age 3 to 8) demonstrate a rapid acquisition of regular plural inflections (Reznick & Armon-Lotem, 2022; Schwartz et al., 2009; 2014).

These findings suggest that bilingual children reach the 'critical mass' of exposure to the type frequency of pluralization in Hebrew. In contrast, monolingual children are more accurate than bilingual children in applying irregular pluralization suffixes (Reznick & Armon-Lotem, 2022; Schwartz et al., 2009; 2014), which require token exposure to the specific lexical unit. In irregular cases, children cannot rely on knowledge built through frequency of type exposure to the regular pattern, because it does not apply. Thus, bilingual children who have less exposure to the societal language, find the production of irregular forms especially challenging.

Research on the derivational knowledge of bilingual Hebrew speaking children is more limited. Altman and colleagues (2017) report that Russian-Hebrew bilingual 5-6 year olds made fewer derivationally driven errors than monolingual peers in a language production task, thus demonstrating weaker derivational knowledge in Hebrew. Michaly and Prior (2025) investigated Hebrew speaking monolingual children and Russian-Hebrew bilingual children in 2nd and 4th grade and found that the two groups demonstrated equal derivational knowledge in comprehension tasks, but here as well bilingual children had weaker derivational knowledge in language production tasks. Finally, a study comparing monolingual and bilingual Hebrew speaking 5th graders (Shahar-Yames et al., 2018) found that bilinguals had lower performance compared to monolinguals on morphological derivation tasks including real words, which require lexical knowledge. However, bilinguals and monolinguals performed equally well in tasks with pseudo-words, which require abstract morphological knowledge that does not depend on lexical knowledge. Importantly, in all these studies bilingual children consistently had smaller Hebrew vocabularies than monolingual children.

The Current Study

Here we investigate the Hebrew lexical and morphological knowledge of bilingual preschool children, who have reduced exposure to the Hebrew language, compared to Hebrew monolingual peers. A main question of interest is to better understand the impact of reduced exposure to the societal language on acquiring linguistic knowledge driven by token frequency vs. that driven by type frequency. We address this issue in four complementary ways. First, we compare the accuracy of monolingual and bilingual children in vocabulary, inflectional morphology and derivational morphology. Second, we report correlations between vocabulary knowledge (driven by token exposure) and morphological knowledge (driven by both token and type exposure). Third, we compare the performance of monolingual and bilingual children on inflection of real irregular words (driven mostly by token exposure) vs. non-words (driven exclusively by type exposure). Finally, we report detailed error analyses, documenting to what degree monolingual and bilingual children recruit inflectional and derivational morphological knowledge (driven by type exposure) even when they are unfamiliar with a specific lexical item (driven by token exposure).

We predict that bilingual children will have lower vocabulary scores than monolingual children, as has been reported in many previous studies (e.g., Hoff, 2021; Michaly & Prior, 2025). We also predict that bilingual children will be less accurate than monolingual children in tasks including real words (one inflectional task, and two derivational tasks). Finally, we hypothesize that group differences will be reduced or eliminated in a non-word inflection task and in the error analyses. This is because bilingual children's performance on real words can be negatively impacted both by their smaller vocabulary knowledge (token) and by their smaller morphological knowledge (type), but performance on non-words only depends on morphological knowledge, which we argue should show smaller group differences. These last two predictions are based on previous findings in Hebrew (Shahar-Yames et al., 2018) and in other languages (Bratlie et al., 2022).

Method

The study described in this work is part of the Safra Longitudinal Study, funded by the Edmond J. Safra Brain Research Center for the Study of Learning Disabilities. As part of the longitudinal study, each child was tested individually on a large battery of linguistic, numeric and cognitive tasks. In the current manuscript we only analyze the tasks assessing lexical and morphological knowledge.

Participants

The longitudinal study received Ethics approval by the Chief Scientist of the Israeli Ministry of Education and by the IRB at the University of Haifa. Letters describing the study were distributed to parents in 122 kindergarten classes in the north of Israel. Data were then collected from children whose parents gave informed consent for their participation in the study, and who willingly cooperated with the research assistants. The longitudinal sample included 1,157 Hebrew-speaking children.

The initial sample for the current study included all children identified as bilingual among the participants of the longitudinal study (*n*=148), and a matched number of monolingual children. Bilingual children were identified based on parent reports that a language different than Hebrew was used in the home. Monolingual children were selected such that for each bilingual child, a monolingual child of the same gender was selected from the same kindergarten class. If such a match was not available, a chid of the opposite gender was selected. The rationale of this procedure was to create two groups that are closely matched on their language exposure in the educational setting (the same kindergarten teachers) and on socio-economic status (residing in the same neighbourhoods, and as validated by measures of parental education, see below). At the time of testing, all children attended kindergarten schools where the language of instruction was Hebrew.

More detailed language background questionnaires (see below) were distributed to the parents of all bilingual children at the end of kindergarten, so that we could report detailed sample characteristics (as recommend e.g. by DeBruin, 2019; Prior & van Hell, 2021). However, only ~50% of the parents (*n*=74; 40 males) completed these. We therefore decided to reduce the sample only to those children for whom we had detailed information about their language environment and retained a matched number of monolingual children (*n*=74; 32 males) according to the same procedure described above. All the results and analyses reported in this paper are based only on this reduced sample, with 74 children per group. Based on a G*Power calculation, this reduced sample size still allows us to detect a medium effect size (0.6, which has frequently been reported in previous research) with a power of .97. However, we also analyzed the full sample, with 148 children per group, and found the same pattern of results, with a few slight differences. The performance of the wider sample is presented in Appendix B.

Most of the bilingual children in the sample spoke Russian as a home language (*n*=50). Other home languages include Amharic (*n*=8), English (*n*=6), Arabic (*n*=2), and one speaker each of Hungarian, Italian, French, Georgian, German, Japanese and Portuguese. Participant characteristics are presented in Table 1. The groups were well matched on important background variables, including age, average family income, parental education, parental reports of children's attentional profile, and home literacy indices.

Table 1: ***Participant characteristics***

	Monolinguals		Bilinguals	
	N	*M* (*SD*)	*N*	*M* (*SD*)
Age (years)	59	6.13 (0.50)	66	6.27 (0.51)
Paternal education (scale 1-6)	52	3.35 (1.24)	60	3.77 (1.14)
Maternal education (scale 1-6)	60	3.67 (1.03)	66	3.71 (1.17)
Number of Siblings	69	2.67 (0.97)	74	2.43 (0.92)
Average family income (scale 1-5)	70	3.31 (0.65)	72	3.17 (0.76)
Attention average (scale 1-2)	73	1.77 (0.24)	74	1.79 (0.20)
Number of adult books at home (scale 1-5)	71	.267 (1.50)	73	3.22 (2.59)
Number of children's books at home (scale 1-5)	70	3.47 (1.08)	75	3.38 (0.91)
Frequency of reading stories at home (scale 1-5)	73	3.73 (1.01)	75	3.95 (0.89)

For all variables, group comparisons p >.1. See Appendix A for information on scales. Note that not all background information was available for all children.

Measures

Parent Questionnaires

Demographic questionnaire: included questions about family education, income and home environment (see information in Table 1, and Appendix A).

Language background questionnaire: A questionnaire completed by parents of bilingual children. It includes questions about children's exposure to their two languages, children's and parents' language proficiency, and patterns of family communication (see Table 2).

Non-Verbal Working Memory – Corsi Blocks

Working memory was assessed using a non-verbal task in which participants had to remember a sequence of spatial locations in two different conditions. In the forward condition, children were asked to reproduce a sequence of locations in the same order that it was presented to them, and in the backward condition they were asked to reproduce the sequence in the opposite order. Each condition included 6 blocks, and the length of the sequence increased by one for each consecutive block. Each block (sequence length) included 2 items, for a total of 12 items. The reliability (Cronbach's alpha) of the task in the longitudinal sample was .81, and in the current sample was .77.

Table 2: ***Language characteristics of bilingual families (N=74)***

	M (*SD*)
Mother Hebrew proficiency	3.90 (1.4)
Mother other-language proficiency	3.79 (1.7)
Father Hebrew proficiency	4.06 (1.3)
Father other-language proficiency	3.79 (1.7)
Child Hebrew proficiency	3.99 (0.8)
Child other-language proficiency	3.18 (1.5)
Child percent of exposure to Hebrew	52% (20)
Child age of exposure to Hebrew (years)	2.70 (2.1)

Parental language proficiency is based on self-rating across talking, reading and writing in each language, on a scale of 0 (non-existent) – 5 (excellent). Child language proficiency is based on parental ratings averaged across talking and understanding, on a scale of 0-5, as above.

Language Tasks

The current data were collected as part of a large-scale longitudinal study, assessing a wide range of child abilities (including early literacy, early numeracy, memory and executive functions). Thus, of necessity, the language tasks administered had to be short, in order to fit within this wide battery. Full testing materials are available on https://osf.io/q8hfn/?view_only=eddfd5d9e7d34417a64e939a2695218b

Hebrew vocabulary knowledge was assessed using a picture naming test, consisting of 14 items, all depicting nouns (Goralnik, 1995). Children were presented with one picture at a time, and requested to state its name in Hebrew. Accuracy was coded online. The reliability (Cronbach's alpha) of the task in the longitudinal sample was .84, and in the current sample was .79. Because bilingual children spoke eleven different home languages, it was unfeasible to test their vocabulary knowledge in their home language as well.

Morphological tasks. Morphological knowledge was assessed using four tasks: two measuring inflectional morphology and two measuring derivational morphology. Before each task, children completed two example items, on which they received feedback. Then, the test items were read to the children without further explanation and feedback. For all morphological tasks, the experimenter documented the child's response in writing and also coded it online as being correct or incorrect. We first analyzed children's overall accuracy in each of the tasks, transformed to percent correct due to the differences in number of items across tasks. Second, for the non-word inflection task and both derivation tasks, we coded offline the types of morphological information retained in children's answers, when they did not give the expected (correct) response. This partial information coding scheme is described below, for each task.

Inflectional Morphology

Noun Pluralization (Cohen-Mimran et al., 2018b; adapted from Lavie, 2006, and Yegev, 2001). The task includes 15 items. The examiner presents a picture of a single object and says its Hebrew name. The examiner then points to the image of several objects of the same kind and asks the child to say the plural Hebrew name. The items are shown to participants in succession, followed by a spoken sentence. For example, "This is a kadur (ball). There are many __________ (kadurim, 'balls')". Of the 15 items, 1 takes a regular inflection, 8 take an irregular inflection, and 12 involve a stem change. This task was coded for overall accuracy, with 1 point given for each correct response. The reliability (Cronbach's alpha) of the task in the longitudinal sample was .77, and in the current sample was .79.

Non-Word Verb Inflection (Shalev-Laifer et al., 2013). This task consists of 10 items. The examiner reads a sentence including a verb created by combining the pseudo-root š-l-z with an existing verbal-pattern, inflected for tense, number and person. The children were requested to use the same root to complete a second sentence by using the correct inflection to create a pseudo-word that fits the morpho-syntactic context. All pseudo-words were based on the same pseudo-root (š-l-z) and the missing word included a change in tense or in person. For example, "Yesterday he šalaz, and yesterday she __________ (šalza)" – person change from masculine to feminine; or "Now you šolez, and tomorrow you __________ (tišloz)" – tense change, from present to future.

The task was scored twice: The first score is the absolute accuracy, namely 1 point for each correct answer. This score was used in the accuracy analyses and the cross-task comparisons. The reliability (Cronbach's alpha) for the accuracy coding in the longitudinal sample was .71, and in the current sample was .67.

The second score gave credit for partial morphological knowledge reflected in responses, and was used in the error analyses conducted for each task separately. The partial knowledge score relied on a detailed analysis, with one point given for each of the following: use of the same root as the stimulus sentence (root), use of the same verb pattern as the stimulus sentence (pattern), inflection in the required person (person), inflection in the required tense (tense; see Appendix C for examples). Thus, the partial score could range from 0-4 points.

Derivational Morphology

Verb Derivation (Cohen-Mimran et al., 2018c, adapted from Novogrodsky & Kreiser, 2015). This task consists of 8 items. The children were instructed to complete a sentence with a suitable verb, derived from a presented Hebrew noun. The verbs required using one of the three active patterns in Hebrew – CaCaC, CiCeC, or hiCCiC (two items also allowed for using the reciprocal pattern, hitCaCeC). For example, "What do we do with the tseva (color)? With the tseva __________ (tsov'im 'we color')".

The task was scored twice: The first score is the absolute accuracy, namely 1 point for each correct answer. This score was used in the accuracy analyses and the cross-task comparisons. The reliability (Cronbach's alpha) for the accuracy coding in the longitudinal sample was .74, and in the current sample was .76.

The second score gave credit for partial morphological knowledge reflected in responses, and was used in the error analyses conducted for each task separately. The partial knowledge score relied on a detailed analysis, with one point given for use of the same root as the stimulus sentence (root), and for use of one of the three possible verb patterns (pattern; see Appendix C for examples), thus it could range from 0-2.

Adjective Derivation (Cohen-Mimran et al., 2018a, adapted from Yegev, 2001). The task consists of 10 items. The examiner said a sentence describing a picture and the children were instructed to complete a sentence, by using the verb from the first sentence to create a suitable adjective, describing the result of the action (see Table 3). For example, "sidru ([they] organized) the books. Now the books are __________ (mesudarim, 'organized')".

Table 3: ***Hebrew resultative adjectives, mapping active verb patterns to the passive adjectival form***

Active pattern	Adjectival pattern
CaCaC – *katav* ([he] wrote)	CaCuC – *katuv* (written)
CiCeC – *sider* ([he] arranged)	meCuCaC – *mesudar* (arranged)
hiCCiC – *histir* ([he] hid)	muCCaC – *mustar* (hidden)

The task was scored twice: The first score is the absolute accuracy, namely 1 point for each correct answer. The reliability (Cronbach's alpha) for the accuracy coding in the longitudinal sample was .74, and in the current sample was .75. The second score gave credit for partial morphological knowledge reflected in responses. The partial knowledge score awarded one point for use of the same root as the stimulus sentence (root) and one point for use of one of the possible resultative adjective patterns (pattern; see Appendix C for examples).

Procedure

Children were tested by trained research assistants in a quiet room in their school. The entire battery of the longitudinal study was administered over 3 individual sessions with each child (1-3 days apart), each lasting approximately 30 minutes. Of the measures reported here, the working memory (forward and backward), real word inflection and non-word inflection tasks were administered in the first session; Vocabulary and verb derivation tasks were administered in the second session and the adjective derivation task was administered in the third session. In each session, the tasks were administered in the order listed here, with additional tasks (not analyzed here) interleaved between them.

Parental demographic questionnaires (hard copy) were distributed to parents who gave consent to their children's participation in the study, in parallel with the children completing the in-school testing sessions. The language background questionnaires were distributed electronically to the parents of bilingual children, identified on the basis of information provided by parents in the demographic questionnaire. These were completed by the parents during the summer after children graduated from kindergarten, or during the first few months of their enrolment in first grade.

Analysis Approach

The performance of monolingual and bilingual children was compared using MANOVA, one-way and repeated measures Analyses of variance using SPSS. All dependent variables were normally distributed (Skewness values ranged from –1.05 to 0.14; Kurtosis values ranged from –0.97 to 0.37).

Results

All experimental data is available at
https://osf.io/q8hfn/?view_only=eddfd5d9e7d34417a64e939a2695218b

As a first step we compared the performance accuracy of monolingual and bilingual children across the different tasks, using a MANOVA. Monolingual children were more accurate than bilingual children in all language tasks (vocabulary and morphology), but the groups had equal performance in the non-linguistic tasks (Table 4).

Table 4: ***Mean percent correct (SD) for experimental tasks, by language group***

		Monolingual (N=74)	Bilingual (N=74)	Comparison
Working memory	Forward	42.2 (13.8)	42.1 (15.4)	$F(1, 148) = .002$, $p = .965$, $\eta_p^2 = .00$
	Backward	29.9 (20.2)	31.1 (19.3)	$F(1, 148) = .136$, $p =. 713$, $\eta_p^2 =.001$
Vocabulary		74.8 (20.4)	58.5 (23.3)	$F(1, 148) = 20.5$, $p <.001$, $\eta_p^2 = .123$
Morphological Inflection	Real words, noun pluralization	72.7 (19.2)	57.0 (22.4)	$F(1, 148) = 21.1$, $p <.001$, $\eta_p^2 = .125$
	Non word, verb inflection	51.5 (23.6)	43.9 (22.6)	$F (1, 148) = 3.98$, $p =.048$, $\eta_p^2 = .026$
Morphological Derivation	Verb	59.2 (26.5)	34.8 (25.3)	$F(1, 148) = 32.9$, $p <.001$, $\eta_p^2 = .183$
	Adjective	50.1 (24.5)	42.1 (25.9)	$F(1, 148) = 15.8$, $p <.001$, $\eta_p^2 = .097$

Before analyzing performance in each morphological task independently we also wished to know to what extent the morphological tasks are correlated with each other, namely, do they tap into a single construct. Thus, in each group of speakers, we examined the correlations between performance in the four morphological tasks and in the vocabulary task by running Pearson correlation analyses. In both language groups, the three morphological tasks that included real words (Noun plural inflection, Verb Derivation, Adjective Derivation), were moderately and significantly positively correlated with each other, and with the vocabulary task. Across all three morphological tasks, children were required to recruit specific lexical knowledge with varying morphological knowledge. The final morphological task, pseudo-word verb inflection, which required only pure morphological knowledge and does not require lexical knowledge, was less strongly (though still significantly) correlated with the vocabulary measure and the three remaining morphological tasks. This pattern was especially evident among the bilingual children (Table 5).

Table 5: ***Correlations between morphological tasks by group, Monolinguals (n=74) below the diagonal and bilinguals (n=74) above the diagonal***

	1	2	3	4	5
1. Vocabulary		.652**	.358*	.692**	.637**
2. Noun Plural inflection	.609**		.304*	.680**	.616**
3. Pseudo-Word verb inflection	.419**	.526**		.325*	.348*
4. Verb derivation	.682**	.597**	.400**		.722**
5. Adjective derivation	.647**	.473**	.302*	.623**	

*$p < .01$, **$p < .001$

Comparing Performance Across Real Word and Non-Word Inflection

In order to examine the extent of the difference between the language groups in inflecting real words and non-words, we used a two-way repeated measures ANOVA, with group (Monolingual, Bilingual) as a between participant factor, and word type (Real word, Non-word) as a within participant factor (Figure 1). Monolinguals were more accurate than bilinguals across both tasks ($F(1,147) = 14.8$, $MSE = 680.2$, $p < .001$, $\eta_p^2 = .091$), and accuracy was higher for inflecting real words than non-words ($F(1,147) = 76.5$, $MSE = 287.5$, $p < .001$, $\eta_p^2 = .342$). Importantly, the interaction between group and task type was also significant, ($F(1,147) = 4.29$, $MSE = 287.5$, $p = .04$, $\eta_p^2 = .03$), because group differences were larger for real words than for non-words, though both differences were significant as demonstrated by post-hoc comparisons ($p < .001$ for real words and $p = .048$ for non-words).

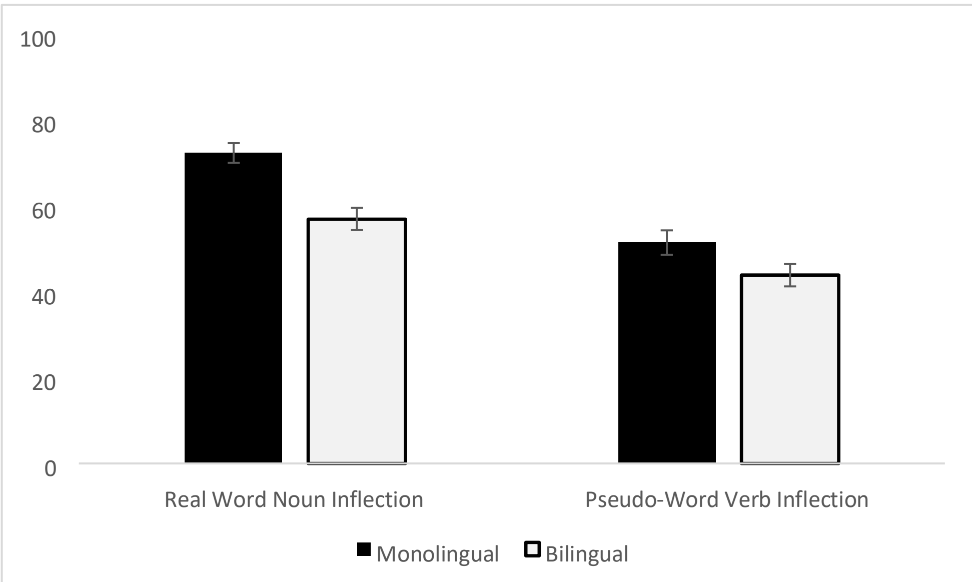

Figure1 : ***Accuracy in inflecting real nouns and pseudo-verbs, by group.***

Comparing Performance Across Verb and Adjective Derivation

To test whether there are differences between the two derivation tasks, we compared the absolute performance in the verb and adjective derivation tasks, with a two-way repeated measures ANOVA with group (Monolingual, Bilingual) as a between participant factor, and derivation type (Verb, Adjective) as a within participant factor (Figure 2). Monolinguals were more accurate than bilinguals across both tasks ($F(1,144) = 25.2$, $MSE = 1055$, $p < .001$, $\eta_p^2 = .149$), and accuracy was higher for deriving verbs than for deriving adjectives ($F(1,144) = 8.01$, $MSE = 214$, $p = .005$, $\eta_p^2 = .053$). Importantly, the interaction between group and task type was also significant, ($F(1,144) = 5.78$, $MSE = 213$, $p = .018$, $\eta_p^2 = .04$). Follow up comparisons demonstrated that whereas monolingual children had higher accuracy rates in the verb derivation than in the adjective derivation task ($t(72) = 3.4$, $p = .001$), the bilingual children showed no significant differences between the tasks ($t(74) = 0.34$, $p = .733$).

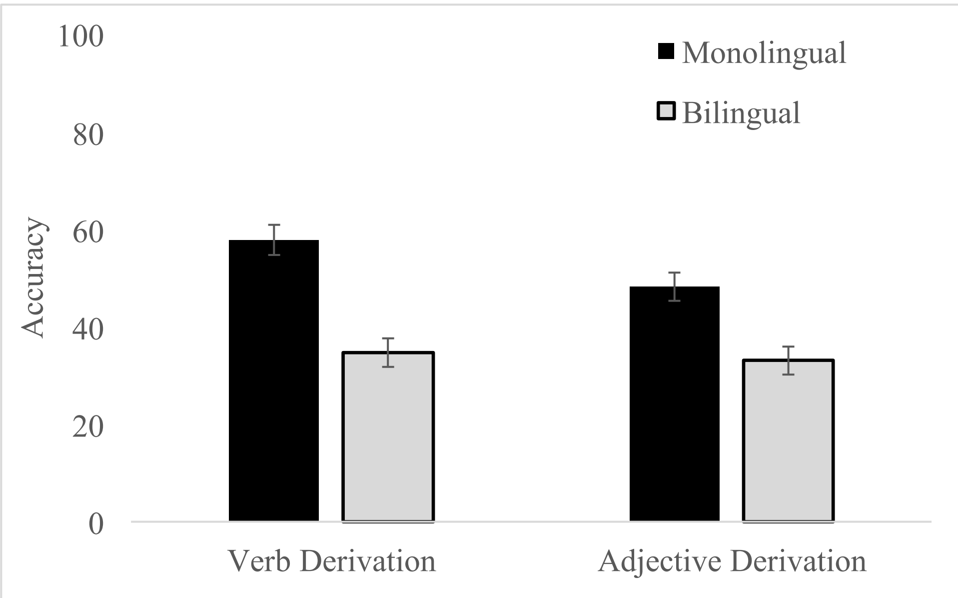

Figure 2: ***Accuracy in deriving verbs and adjectives, by group.***

Partial Knowledge Analyses

Pseudo-Word Verb Inflection

Our main question of interest here was which type of morphological knowledge children with different language backgrounds rely on when inflecting unfamiliar pseudowords. Due to the relatively lower correlations of this task with vocabulary knowledge, the partial knowledge score is informative of children's abstract morphological knowledge. Because monolingual children were overall more accurate than bilingual children (see Table 4), we transformed the partial knowledge score to percentages. Thus, for each child we coded for the incorrectly answered items, what percent of responses preserved different types of morphological information. This allowed us to overcome the difference in basic performance and to test which type of knowledge was more accessible to children in the two groups (see examples in Appendix C1).

To this end, percentages of preserved knowledge were analyzed using a two-way repeated measures ANOVA, with group as a between-participants factor (Monolingual, Bilingual) and knowledge type as a within participant factor (Root, Pattern, Person, Tense). The main effect of group was not significant ($F < 1$), demonstrating the monolingual and bilingual children were equally able to recruit different types of morphological knowledge. The main effect of knowledge type was significant [$F(3,426) = 59.62$ $MSE = .065$, $p < .001$, $\eta_p^2 = .296$]. Participants showed the highest level of accuracy in retaining root information, ($M = 74$, $SD = 3.5$), followed by correct person inflection ($M = 50$, $SD = 0.02$) and correct tense inflection ($M = 43$, $SD = 3.0$). Children found it most difficult to preserve accurate pattern information ($M = 36.5$, $SD = 0.02$; see Figure 3). The interaction between group and error type was not significant ($F < 1$). Thus, when relying on pure morphological knowledge for inflecting pseudo-verbs, monolingual and bilingual children showed the exact same pattern of performance.

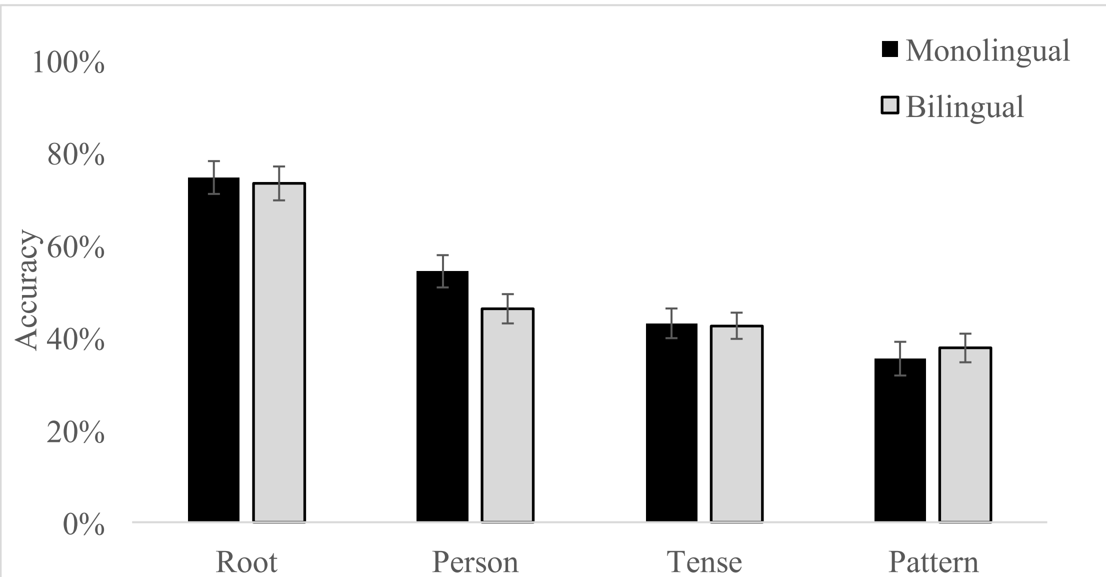

Figure 3: *Preservation of partial knowledge in pseudo-word verb inflection, by group.*

Verb Derivation

In analyzing this task, we examined children's ability to derive real verbs from a given Hebrew noun. As reported above, monolingual children had higher scores on this task than bilingual children, when comparing simple accuracy rates (Table 4). Here, our main interest focused on the partial knowledge scores, to better understand what morphological knowledge children in both groups were able to access in attempting to produce verbs in Hebrew. We transformed the partial knowledge score to percentages, thus for each child we coded what types of morphological knowledge were preserved when he or she did not provide the fully correct expected response. This allowed us to examine whether children's responses were due to a lack of awareness of the roots, by using a word from another morphological family in an accurate pattern (e.g., said *xotxim* 'cut' in root *x-t*-x, CaCaC pattern, instead of *soxtim* 'squeeze' in root *s-x*-t, CaCaC pattern), or due to a lack of specific lexical knowledge by producing an incorrect combination of the correct root in a possible verbal pattern (e.g., said *masxitim* in root *s-x*-t, hiCCiC pattern, instead of *soxtim* 'squeeze' in root *s-x*-t, CaCaC pattern. See further examples in Appendix C2).

To this end, the partial knowledge scores were analyzed using a two-way repeated measures ANOVA, with group as a between-participants factor (Monolingual, Bilingual) and knowledge type as a within participant factor (Root, Pattern). The main effect of group was significant ($F(1,142) = 11.04$, $MSE = .073$, $p = .001$, $\eta_p^2 = .72$), because monolingual children managed to express more correct morphological information even when they made errors ($M = 45.8$) compared to bilingual children ($M = 35.2$). The main effect of knowledge type was also significant ($F(1,142) = 404.06$, $MSE = .056$, $p < .001$, $\eta_p^2 = .74$). Follow up analyses revealed that participants most easily expressed morphological knowledge in choosing an appropriate pattern ($M = 68.4$, $SD = 3.6$), but found it more difficult to preserve root information ($M = 12.6$, $SD = 2.2$; see Figure 4). Finally, the interaction between group and knowledge type was not significant ($F(1,142) = 3.042$, $MSE = .056$, $p = .083$, $\eta_p^2 = .021$).

This pattern of results shows that in most cases the children adopted a lexical strategy, that is they produced an existing verb in an appropriate pattern, which fits semantically, but does not use the required root (e.g., with the noun drum (*tof*), children responded with *menagnim* 'play an instrument' instead of *metofefim,* 'beat'). In a minority of the cases, where children retained the root in their response, they did indeed use a morphological strategy, whereby they incorporated a required root in a possible verbal-pattern (e.g., with the noun *masxeta* 'juicer' they produced the verb *masxitim* in the hiCCiC pattern, which is not a lexical item in Hebrew, instead of *soxtim* 'squeeze' in the CaCaC pattern, which does exist in the Hebrew lexicon) This pattern was common to both monolingual and bilingual children.

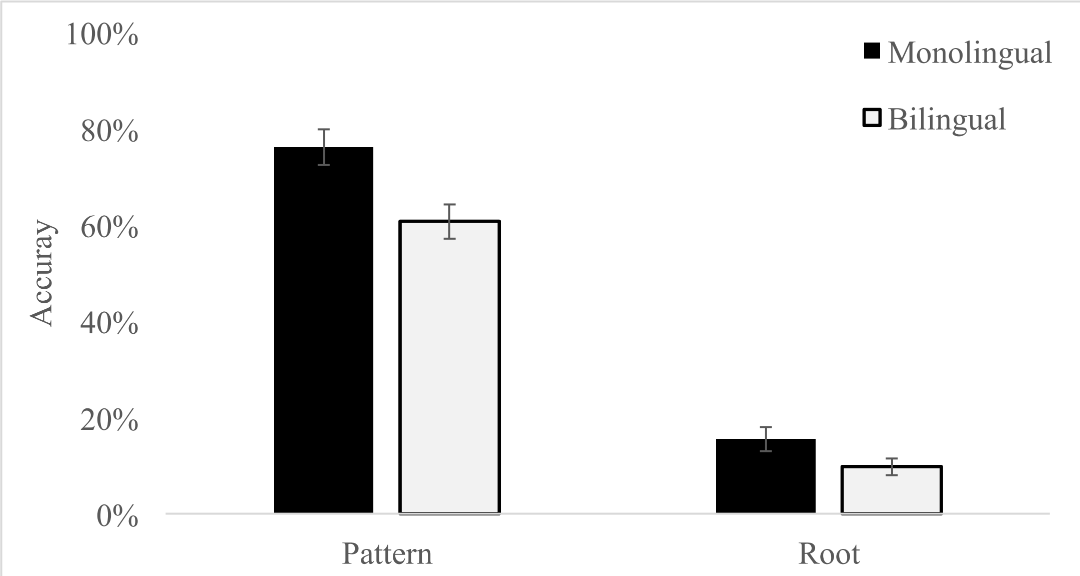

Figure 4: ***Preservation of partial knowledge in verb derivation, by group***

Adjective Derivation

In analyzing this task, we examined children's ability to morphologically derive real adjectives from given Hebrew verbs. Our main question of interest here was what types of morphological knowledge would be accessible to children in both groups when producing adjectives in Hebrew. We transformed the partial knowledge score to percentages, thus for each child we coded the percentage of errors in which knowledge of each type was preserved. This allowed us to examine whether the errors were due to a lack of awareness of the roots, as when children used a word from another morphological family (e.g., said *mušlam* 'perfect' in root *š-l-m* instead of *murkav* 'put together' in root *r-x-v* in muCCaC pattern), or due lack of awareness of specific lexical knowledge by producing an incorrect combination of the correct root in a possible verbal pattern, but not the accurate pattern (e.g., used the correct root *r-x-v* , but embedded it CaCuC pattern and said *raxuv* , which is not a lexical item in Hebrew, instead of using the muCCaC pattern to give the correct response of *murkav* , which is an existing word in Hebrew. See further examples in Appendix C3).

The data were analyzed using a two-way repeated measures ANOVA, with group as a between-participants factor (Monolingual, Bilingual) and knowledge type as a within participant factor (Root, Pattern). The main effect of group was not significant (F(1,142) = 2.31, MSE = 0.12, p = .131, η_p^2 = .016). The main effect of knowledge type was marginally significant (F(1,142) = 3.55, MSE = 0.061, p = .62, η_p^2 = .024), because children were somewhat more likely to choose a possible passive pattern (M =53.6, SD = 3.3), than to retain root information (M = 48.4, SD = 3.6). The interaction between group and error type was significant (F(1,142) = 3.96, MSE = 0.061, p = .048, η_p^2 = .027). To follow up on this interaction, we conducted independent samples t-tests for each knowledge type separately. These revealed that children from both groups were similarly likely to preserve the root (t(142) = 0.062, p = .95), but monolingual children were more likely than bilingual children to produce a possible passive pattern (t(142) =2.49, p =.14; Figure 5). As explained above, such pattern preservation mostly constitutes a lexical strategy, in which children select an alternative adjective, which uses a resultative pattern and is semantically appropriate, but which is not derived from the same root.

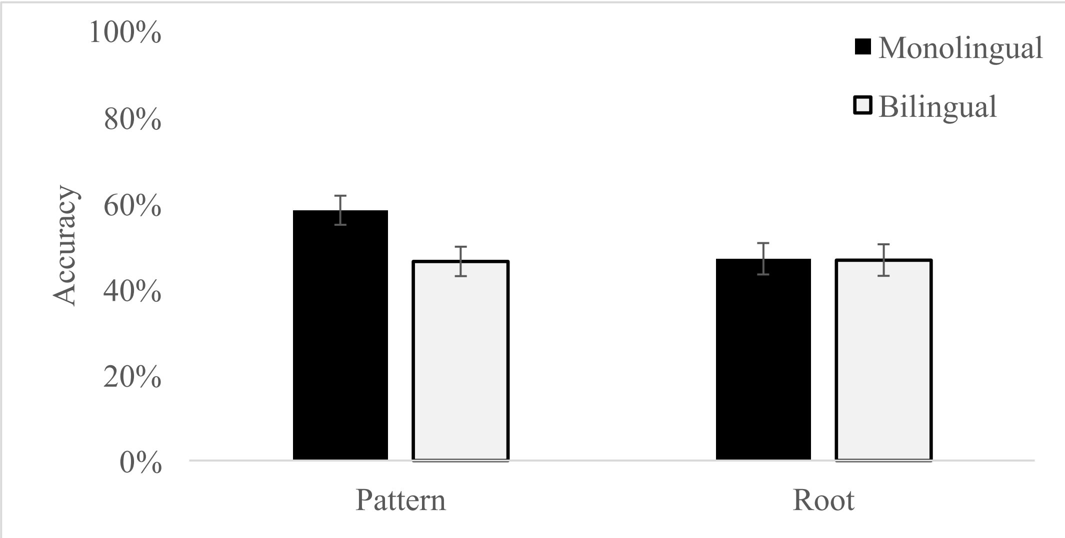

Figure 5: ***Preservation of partial knowledge in adjective derivation, by group***

Discussion

The present study examined monolingual and bilingual preschool children's knowledge of their societal language, Hebrew. As in many previous studies (e.g. Armon-Lotem et al., 2019; Hoff, 2021; Schwartz & Katzir, 2012) the bilingual children in the current sample were exposed to the societal language about 50% of the time, according to parental reports. A main goal of the study was to better understand the impact of the reduced exposure of bilingual children to the societal language on their acquisition of linguistic knowledge driven by token frequency, namely vocabulary, vs. that driven by type frequency, namely morphology. Monolinguals were significantly more accurate than bilinguals in all morphology tasks and in a vocabulary task, highlighting the critical role of reduced exposure on bilingual language development. Importantly, however, careful analyses suggest that such reduced exposure has a stronger impact on token-based knowledge than on type-based knowledge. In addition, when controlling for the contribution of lexical knowledge, bilingual children were indistinguishable from monolinguals in their knowledge of Hebrew inflectional morphology but still showed small gaps in their knowledge of derivational morphology.

Children's language acquisition is driven by their exposure to the language around them (Bybee, 2007; Tomasello, 2001). Importantly, this exposure supports children both in learning specific words and in reaching generalizations about morpho-syntactic rules. In the current study, bilingual children had smaller Hebrew vocabularies than monolingual children, and also had lower performance in inflectional and derivational morphology tasks. These findings align with previous studies describing gaps in the vocabulary (in general Hoff, 2021; and in Hebrew, Altman et al., 2017; Shahar-Yames et al., 2018) and morphological knowledge (in general: Bratlie et al., 2022; in Hebrew: Michaly & Prior, 2025; Reznick & Armon-Lotem, 2022) of bilingual children.

The vocabulary and morphology tasks were strongly and positively correlated for both monolingual and bilingual children (with the exception of the Pseudo-word inflection task, which was only moderately correlated with the other tasks, more on this below). This finding again supports the notion that the acquisition of vocabulary and morphology are closely intertwined (Fejzo, 2021; Nicoladis et al., 2007; Ravid, 2006), and specifically that morphological knowledge is driven by both token and type exposure. Thus, as children's lexicon expands, they may find it easier to extract morphological regularities and systematic representations of inflections and derivations. At the same time, children's growing morphological knowledge can support vocabulary expansion and scaffold learning new words (Bybee, 2007).

However, in terms of being able to tease apart the contributions of type and token exposure to the acquisition of morphological regularities, and specifically to be able to examine more closely the impact of reduced exposure on this process, this close coupling is a hindrance. We addressed this issue in two ways, by including a task with pseudo-words and by looking at error patterns.

Inflection

Examining correlations between the study tasks, a weaker correlation was observed between the pseudo-verb inflection task and the rest of the tasks. Thus, pseudo-verb inflection was only moderately correlated with vocabulary, for both monolinguals and bilinguals. Similarly, pseudo-verb inflection was again only moderately correlated with the remaining morphological tasks (strongly correlated among themselves), which all involved morphological manipulation (inflection or derivation) of real vocabulary. This indicates that a task with pseudo-words more strongly relies on abstract morphological representations, and recruits lexical knowledge to a lesser degree (for similar findings see Shahar-Yames et al., 2018).

Bilinguals were less accurate than monolinguals when inflecting pseudo-verbs, but the effect was much smaller than in all other morphological tasks (see Table 4 and Figure 1), and specifically smaller than the group difference evident in the inflection of real nouns. This pattern suggests that as early as age 5, the gaps between bilingual and monolingual children in knowledge of inflection regularities in Hebrew, driven by type exposure, are smaller than the gaps evidently driven by token exposure and lexical knowledge. To wit, the group difference in the real noun inflection task, which included mostly words with irregular plural inflections, were much more pronounced. These findings align well with previous studies showing equal performance of bilingual and monolingual children on regular inflections, concurrently with group differences in irregular inflections for Hebrew (Schwartz et al., 2009, 2014; Reznick & Armon-Lotem, 2022), as well as other languages (e.g. English: Paradis et al., 2011; Rattanasone & Demuth, 2023).

This conclusion is further strengthened by analyzing the error patterns in pseudo-verb inflection, where there was no evidence for group differences. Namely, when they failed to correctly inflect the pseudo-verb, monolingual and bilingual children exhibited the exact same use of their existing morphological knowledge. Children were most likely to retain correct root information, and erred most often in not retaining the correct verbal pattern in their response. This pattern aligns with the primacy of the root over the pattern in the acquisition of Hebrew morphology (Ravid, Ashkenazi et al., 2016), though note that the facility with retaining the root might be to some degree driven by the fact that all items in this task shared the same pseudo-root. Monolingual and bilingual children were again equally likely to exhibit correct person and tense information in their responses.

Thus, before elementary school entry, bilingual Hebrew speaking children seem to have mostly reached the type exposure threshold necessary for accurate representation of the highly regular inflection system of Hebrew (Marchman & Bates, 1994). Of note, these same bilingual children have significantly lower vocabulary knowledge then their monolingual peers. These results clearly demonstrate the differential impact of reduced exposure to the societal language. In our case, children were exposed to Hebrew about 50% of the time on average, over 3.5 years. Whereas this reduction has a significant negative impact on knowledge extracted from token exposure, it did not similarly influence highly regular and consistent inflectional knowledge extracted from cumulative type exposure.

The finding that bilingual children master the regular inflection system relatively quickly has implication for instruction as well as assessment. In terms of readiness for elementary school, it seems that instructional efforts should not focus on inflectional forms, since these are mostly well established in 5-6 year-old children. However, bilingual children who demonstrate significant difficulties in correctly inflecting regular forms despite being exposed to Hebrew at least 50% of the time, might be at risk for language delay, and should thus undergo more detailed assessment.

Derivation

Monolinguals were more accurate than bilinguals in deriving verbs and adjectives in the current study. Both derivation tasks were highly correlated with vocabulary knowledge, suggesting that accurate performance relies to some extent on lexical as well as morphological knowledge. Indeed, because Hebrew roots do not combine exhaustively with the active verb patterns and passive adjective patterns tested here (Schwartzwald, 2002), producing a correct response was more likely if children were familiar with the target lexical item. Here again we see that smaller exposure to the societal language negatively impacts performance that relies on token exposure.

Analyzing the error patterns reveals a more complex picture. Examining error patterns shows us what children are capable of doing when they manifestly do not have the specific lexical knowledge required. Thus, we can tap into the abstract morphological representations that are available to them, gained exclusively through type exposure and generalization. When children from both groups were unfamiliar with the correct response in the verb derivation task, they predominantly produced an alternative verb that was semantically appropriate. Such responses used one of the possible verb patterns, but did not use the target root (Figure 4). This finding suggests that both monolingual and bilingual children have good representations of the active verb patterns, extracted based on type exposure, and aligns with the expected developmental stages of Hebrew speaking children (Ashkenazi et al., 2016; Ravid, 2019a). Notably, monolingual children produced such pattern-preserving responses significantly more often than did bilingual children.

One interpretation is that bilingual children were less successful in extracting such abstract morphological knowledge due to their reduced exposure to the language, and specifically reduced type exposure. However, we wish to argue that this observed group difference might at least partially be driven by gaps in lexical knowledge as well. Specifically, producing a verb derived from a different root can be characterized as representing a lexical response strategy, one that relies on retrieving an appropriate word from the lexicon and not necessarily on completing a morphologically driven derivation process. Because bilingual children have reduced token exposure and smaller Hebrew vocabularies, it is likely that in some cases such an alternative was not available to them.

In a minority of the cases (about 15%), children did use the target root to derive a verb using one of the possible patterns (though not the expected one), demonstrating morphologically driven processing. Bilingual children tended to do this less often than monolingual children, though the group difference was only marginal. Due to the small percent of responses in this category, as well as the weak evidence of group difference, we can only cautiously suggest that it might indicate that bilingual children indeed have less stable representations of the verbal morphology tested here, as a result of reduced type exposure. This aligns with the findings of Altman and colleagues (2017) who also reported fewer morphologically motivated errors in bilingual than in monolingual children of the same age group tested here, and with those of Michaly and Prior (2025) showing smaller derivational knowledge in bilingual 2nd and 4th graders, relative to monolingual peers.

Error analysis of the adjective derivation task showed some similarities to the verb derivation task. When unfamiliar with the target adjective, here as well children from both groups predominantly produced a semantically appropriate adjective, derived in one of the three possible patterns, but not using the target root. As observed for verbs, monolingual children were significantly more likely to do this than bilingual children, indicating more stable morphological representations of adjective morphology, larger vocabularies, or both. In contrast, root-preserving responses were more prevalent in the adjective derivation task (~45%) than in the verb derivation task, and importantly were equally likely for monolingual and bilingual children. Children might have resorted to morphologically driven processing more often when producing adjectives than verbs because the adjective lexicon is smaller (Ben Zvi & Levie, 2016; Ravid, Bar-On et al., 2016) and they might have been less successful in retrieving an appropriate lexical alternative. Critically, bilingual and monolingual children were equally able to use the target root in an adjectival pattern, suggesting that they might not differ in their abstract morphological knowledge.

Taken together, these results clearly support the notion that bilingual children have fewer lexical resources at their disposal, due to reduced token exposure. It is less clear, however, whether bilingual children also have less-well established morphological derivational knowledge, namely have not amassed sufficient type exposure to meet the necessary threshold (Marchman & Bates, 1994). Results from the verb derivation task weakly suggest that this might be the case, but performance in the adjective derivation task demonstrates equal performance across groups.

Given the central role of derivational morphology in supporting Hebrew reading (Share & Bar-On, 2018) and writing (Ravid, 2011), we suggest that school readiness interventions with bilingual children should incorporate morphological components. Such activities could act to diminish the gaps in derivational knowledge observed here, and also provide scaffolding for expanding bilingual children's vocabulary knowledge, which is smaller than that of monolingual peers. Similarly, in light of the reciprocal relations between vocabulary and morphology, activities aimed at expanding bilingual children's exposure to Hebrew and enriching their vocabulary could also arguably benefit children's ability to extract morphological regularities.

Limitations and Future Research

The final sample analyzed here included only half of the language production data collected, due to difficulties in receiving adequate background information from the bilingual families. In addition, because the current study was part of a large-scale longitudinal study, per force the assessment tasks were rather short, though they did show good reliability. Future research could therefore study specific inflection and derivation structures in greater depth, to achieve a more nuanced picture of acquisition patterns in bilingual children.

The bilingual children studied here spoke a wide variety of home languages. Whereas this is definitely a strength of the current study in providing good generalizability of the results, it does mean that we were not able to objectively assess children's proficiency in their home language (we relied exclusively on parental reports) nor to study specific patterns of cross-language influence from different languages onto Hebrew (see e.g. Meir et al., 2017).

Conclusions

We demonstrate the impact of reduced exposure to the societal language on bilingual children's knowledge of that language, and show how it interacts with token and type based learning mechanisms. Token based performance in vocabulary and morphological tasks was lower in bilinguals than in monolinguals. However, when probing children's ability to utilize morphological knowledge in the absence of lexical representations we found equivalent performance of bilingual and monolingual children in inflection, and small differences in derivation. These results highlight the importance of research methods that can distinguish between lexical and morphological knowledge, especially when studying bilingual individuals (Shahar-Yames et al., 2018). Acquiring derivational morphology is a more protracted process than acquiring inflectional morphology in Hebrew (Ben Zvi & Levie, 2016; Berman, 2003). Therefore, whereas bilingual children were able to use their morphological knowledge in the inflection system, they had not yet reached the threshold of exposure that would allow them to do so in the derivational system, which is less systematic and regular in Hebrew. Taken together, the current findings suggest that the acquisition of morphological regularities, driven mostly by type exposure, is more resilient in the face of reduced exposure associated with bilingualism than token-based lexical learning.

References

Altman, C., Goldstein, T., & Armon-Lotem, S. (2017). Quantitative and qualitative differences in the lexical knowledge of monolingual and bilingual children on the LITMUS-CLT task. *Clinical Linguistics & Phonetics,* 31(11-12), 931-954. https://doi.org/10.1080/02699206.2017.1312533

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: Author. https://doi.org/10.1176/appi.books.9780890425596

Andreou, M. & Tsimpli, I.M (2020). Bilingualism, biliteracy and syntactic complexity: The role of crosslinguistic influence and cognitive skills. In A. Cardinaletti, C. Branchini, G. Giusti, & F. Volpato (Eds.), *Language acquisition, processing and bilingualism: Selected papers from the Romance Turn VII* (pp. 132–159). Cambridge Scholars Publishing.

Armon-Lotem, S., Meir, N., De Houwer, A., & Ortega, L. (2019). The nature of exposure and input in early bilingualism. In A. De Houwer & L. Ortega (Eds.), *The Cambridge handbook of bilingualism,* 193-212. https://doi.org/10.1017/9781316831922.011

Armon-Lotem, S., Walters, J., & Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool children. *Linguistic Approaches to Bilingualism, 1(3),* 291–317. https://doi.org/10.1075/lab.1.3.04arm

Ashkenazi, O., Gillis, S., & Ravid, D. (2020). Input–output relations in Hebrew verb acquisition at the morpho-lexical interface. *Journal of Child Language, 47(3),* 509-532. https://doi.org/10.1017/S0305000919000540

Ashkenazi, O., Ravid, D., & Gillis, S. (2016). Breaking into the Hebrew verb system: A learning problem. *First Language,* 36(5), 505-524. https://doi.org/10.1177/0142723716648865

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59,* 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

Ben-Zvi, G., & Levie, R. (2016). Development of Hebrew derivational morphology from preschool to adolescence. In Ruth A. Berman (Ed.), *Acquisition and development of Hebrew: From infancy to adolescence* (pp. 135-173). John Benjamins. https://doi.org/10.1075/tilar.19.05ben

Berman, R. A. (1985). Acquisition of Hebrew. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 255–371). Erlbaum. https://doi.org/10.4324/9781315802541-4

Berman, R. A. (2003). Children's lexical innovations: Developmental perspectives on Hebrew verb-structure. In J. Shimron (Ed.), *Language processing and language acquisition in a root-based morphology* (pp. 243–91). Benjamins. https://doi.org/10.1075/lald.28.13ber

Berman, R. A. (2016). Typology, acquisition, and development: The view from Israeli Hebrew. In R. A. Berman (Ed.), *Acquisition and development of Hebrew: Infancy to adolescence* (pp. 1–38). Benjamins. https://doi.org/10.1075/tilar.19.01ber

Bialystok, E., Luk, G., Peets, K. F., & Sujin, Y. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bolozky, S. (2007). Israeli Hebrew morphology. In A. S. Kaye (Ed.), *Morphologies of Asia and Africa (including the Caucasus)* (pp. 283–308). Eisenbrauns. https://doi.org/10.5325/j.ctv1bxh537.19

Bratlie, S. S., Brinchmann, E. I., Melby-Lervåg, M., & Torkildsen, J. von K. (2022). Morphology—A gateway to advanced language: Meta-analysis of morphological knowledge in language-minority children. *Review of Educational Research*, *92*(4), 614–650. https://doi.org/10.3102/00346543211073186

Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, *1*(3), 318-345. https://doi.org/10.1075/lab.1.3.05cho

Cohen-Mimran, R., Gott, D., Reznik-Nevet, L. & Share D. L. (2018a). *Adjective Derivation test***.** Unpublished test, University of Haifa (in Hebrew).

Cohen-Mimran, R., Gott, D., Reznik-Nevet, L. & Share D. L. (2018b). *Plural inflection test***.** Unpublished test, University of Haifa (in Hebrew).

Cohen-Mimran, R., Gott, D., Reznik-Nevet, L. & Share D. L. (2018c). *Verb Derivation test*. Unpublished test, University of Haifa, (in Hebrew).

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences, 9(3),* 33. https://doi.org/10.3390/bs9030033

Ellis, N., O'Donnell, M., & Römer, U. (2015). Usage-based language learning. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 163–80). Wiley. https://doi.org/10.1002/9781118346136.ch7

Fejzo, A. (2021). The contribution of morphological awareness to vocabulary among L1 and L2 French-speaking 4th-graders. *Reading and Writing, 34,* 659–679. https://doi.org/10.1007/s11145-020-10084-8

Goralnik, E. (1995). *Goralnik screening test for Hebrew.* Matan (in Hebrew).

Hoff, E. (2021). Why bilingual development is not easy. In J. Lockman (Ed.). *Advances in child development and behavior* (Vol. 61, pp. 129–167). Elsevier. https://doi.org/10.1016/bs.acdb.2021.03.002

Katzir,T. Lipka, O. Prior, A. & Shany, M. (2019). A multi-component approach to reading comprehension: Teacher training from mapping and intervention. *Report to the Chief Scientist of the Ministry of Education: The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities.* University of Haifa (in Hebrew).

Kieffer, M. J., & Box, C. D. (2013). Derivational morphological awareness, academic vocabulary, and reading comprehension in Spanish-speaking language minority learners and their classmates. *Learning and Individual Differences, 24,* 168–175. https://doi.org/10.1016/j.lindif.2012.12.017

Lavie, N. (2006). Hebrew noun plurals in development. Unpublished master's thesis, Tel Aviv University, Israel (in Hebrew).

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*(2), 333–370. https://doi.org/10.1017/S0305000903005592

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*(2), 339–366. https://doi.org/10.1017/S0305000900009302

Meir, N., Walters, J., & Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguistic Approaches to Bilingualism, 7(5),* 514-553. https://doi.org/10.1075/lab.15007.mei

Michaly, T., & Prior, A. (2025). Development of derivational morphological knowledge in monolingual and bilingual children: Effects of modality and lexicality. *Journal of Child Language, 52*(3), 532-557. https://doi.org/10.1017/S0305000924000126

Nicoladis, E., Palmer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science, 10*(2), 237-254. https://doi.org/10.1111/j.1467-7687.2007.00582.x

Novogrodsky, R., & Kreiser, V. (2015). What can errors tell us about specific language impairment deficits? Semantic and morphological cuing in a sentence completion task. *Clinical Linguistics and Phonetics*, 29(11), 812–825. https://doi.org/10.3109/02699206.2015.1051239

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language, 50*(4), 793-817. https://doi.org/10.1017/S0305000922000708

Paradis, J., Nicoladis, E., Crago, M., & Genesee, F. (2011). Bilingual children's acquisition of the past tense: A usage-based approach. *Journal of Child Language, 38*(3), 554-578. https://doi.org/10.1017/S0305000910000218

Pearson, B. Z., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning, 44*(4), 617-653. https://doi.org/10.1111/j.1467-1770.1994.tb00633.x

Prior, A., & Markus, E. (2014). Morphological activation in sentence context: When the root prevails over the meaning. *Language, Cognition and Neuroscience*, 29(9), 1180-1188. https://doi.org/10.1080/23273798.2014.920511

Prior, A., & van Hell, J. G. (2021). Including bilingual participants in cognitive research. In W. Francis (Ed), *Bilingualism across the lifespan* (pp. 47-65). Routledge. https://doi.org/10.4324/9781315143996-5

Ramirez, G., Chen, X. I., Geva, E., & Luo, Y. (2011). Morphological awareness and word reading in English language learners: Evidence from Spanish-and Chinese-speaking children. *Applied Psycholinguistics, 32*(3), 601–618. https://doi.org/10.1017/S0142716411000233

Rattanasone, N. X., & Demuth, K. (2023). Produced, but not 'productive': Mandarin-speaking pre-schoolers' challenges acquiring L2 English plural morphology. *Journal of Child Language, 50*(3), 581-609. https://doi.org/10.1017/S0305000921000969

Ravid, D.D (2006). Word-level morphology: A psycholinguistic perspective on linear formation in Hebrew nominals. *Morphology,* 16(1), 127-148. https://doi.org/10.1007/s11525-006-0006-2

Ravid, D. D. (2011). *Spelling morphology: The psycholinguistics of Hebrew spelling* (Vol. 3). Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-0588-8_13

Ravid, D.D. (2019a). Chapter 8: Derivation. In Ruth A. Berman (Ed.), *Usage-based studies in Modern Hebrew: Background, morpho-lexicon, and syntax* (pp. 203-265). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.210.09rav

Ravid, D.D. (2019b). First-Language acquisition of morphology. In *Oxford Research Encyclopedia of Linguistics.* https://doi.org/10.1093/acrefore/9780199384655.013.603

Ravid, D., Ashkenazi, O., Levie, R., Zadok, G. B., Grunwald, T., Bratslavsky, R., & Gillis, S. (2016). Foundations of the early root category. In Ruth A. Berman (Ed.), *Acquisition and development of Hebrew: From infancy to adolescence* (pp. 95-134). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tilar.19.04rav

Ravid, D., Bar-On, A., Levie, R., & Douani, O. (2016). Hebrew adjective lexicons in developmental perspective: Subjective register and morphology. *The Mental Lexicon, 11*(3), 401-428. https://doi.org/10.1075/z.238.05rav

Reznick, J., & Armon-Lotem, S. (2022). Children's command of plural marking on Hebrew nouns: Evidence from Russian-Hebrew bilingual acquisition. In Levie, R., Bar-On, A., Ashkenazi, O., Dattner, E., & Brandes, G. (Eds.). *Developing Language and Literacy: Studies in Honor of Dorit Diskin Ravid* (pp. 573-596). Springer International Publishing. https://doi.org/10.1007/978-3-030-99891-2_23

Schwartz, M., & Katzir, T. (2012). Depth of lexical knowledge among bilingual children: The impact of schooling. *Reading and Writing, 25*(8), 1947–1971. https://doi.org/10.1007/s11145-011-9308-9

Schwartz, M., Kozminsky, E., & Leikin, M. (2009). Delayed acquisition of irregular inflectional morphology in Hebrew in early sequential bilingualism. *International Journal of Bilingualism,* 13(4), 501-522. https://doi.org/10.1177/1367006909353238

Schwartz, M., Nir, B., Leikin, M., Levie, R., & Ravid, D. (2014). The acquisition of noun plurals among early sequential Russian-Hebrew speaking bilinguals: A longitudinal multiple case study. *Heritage Language Journal, 11*(2), 151–185. https://doi.org/10.46538/hlj.11.2.3

Schwarzwald, O. R. (2002). *Modern Hebrew morphology*. Tel Aviv: Open University

Shahar-Yames, D., Eviatar, Z., & Prior, A. (2018). Separability of lexical and morphological knowledge: Evidence from language minority children. *Frontiers in Psychology*, 9, 163. https://doi.org/10.3389/fpsyg.2018.00163

Shalev-Laifer, C., Reznik-Nevet, L., & Shere, D. (2013). *Verb inflection test SHALAZ,* Unpublished test, University of Haifa (in Hebrew).

Share, D. L., & Bar-On, A. (2018). Learning to read a Semitic abjad: The triplex model of Hebrew reading development. *Journal of Learning Disabilities*, *51*(5), 444-453. https://doi.org/10.1177/0022219417718198

Sparks, E., & Deacon, S. H. (2015). Morphological awareness and vocabulary acquisition: A longitudinal examination of their relationship in English-speaking children. *Applied Psycholinguistics*, *36*(2), 299–321. https://doi.org/10.1017/S0142716413000246

Tomasello, M. (2001). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics, 11*, 61-82. https://doi.org/10.1515/cogl.2001.012

Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech & Hearing Services in Schools, 38(3)*, 225-236. https://doi.org/10.1044/0161-1461(2007/024)

Yegev, I. (2001). Preschool language skills: Comparing cleft palate children to healthy children their age. Unpublished MA Thesis, Tel Aviv University (in Hebrew).

Data, Code and Materials Availability Statement

All study materials and data are available at: https://osf.io/q8hfn/?view_only=eddfd5d9e7d34417a64e939a2695218b

Ethics Statement

The study received Ethics approval by the Chief Scientist of the Israeli Ministry of Education and by the IRB at the University of Haifa. Parents gave informed consent for their children's participation in the study, and data were collected only from children who willingly cooperated with the research assistants.

Authorship and Contributorship Statement

Anat Prior and **Gal Pedael** contributed to conception and design of the work, analysis and interpretation. **Gal Pedael** drafted the initial manuscript, and both authors were involved in revision and final approval. Both authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

Data collection was funded by the Edmond J. Safra Brain Research Center for the Study of Learning Disabilities. Gal Padeal received an excellence scholarship from the University of Haifa.

Appendix A: Scales of the Parental Demographic Questionnaire

Average family income: Average of reports for father and mother. Scale 1-5: 1-no income, 2-below average, 3-around average, 4-above average, 5-far above average.

Parental education, reported separately for mother and father. Scale 1-6: 1- high school without matriculation, 2-high school with matriculation, 3-diploma studies, 4-BA, 5-MA, 6-Ph. D or higher

Number of (adult/children) books at home Scale 1-5: 1 – 0-10, 2 – 11-20, 3 – 21-50, 4 – 50-100, 5 – over 100)

How often do you read stories to your child? Scale 1-5: 1-never, 2-once a month, 3-once a week, 4-several times a week, 5-every day

Attention score: Hebrew translation of the criteria from the DSM 5 (APA, 2013). The score reported here is the average of 18 statements about attention completed by parents about their children. “Yes” responses were coded as 1, and “no” responses were coded as 2. Thus, lower scores indicate more attentional difficulties.

Appendix B: Full Sample Data

Table B1: ***Participant characteristics, Full Sample***

	Monolinguals		Bilinguals	
	N	M (SD)	N	M (SD)
Age (years)	120	6.13 (.04)	133	7.01 (0.85)
Paternal education (scale 1-6)	111	3.50 (1.31)	111	3.68 (1.27)
Maternal education (scale 1-6)	120	3.76 (1.19)	123	3.60 (1.26)
Number of Siblings	140	2.76 (0.80)	140	2.41 (0.81)
Average family income (scale 1-5)	142	3.32 (0.65)	137	3.12 (0.85)
Attention average (scale 1-2)	148	1.77 (0.24)	140	1.77 (0.23)
Number of adult books at home (scale 1-5)	143	2.78 (1.51)	127	30.5 (2.19)
Number of children's books at home (scale 1-5)	144	3.49 (1.02)	136	3.36 (1.03)
Frequency of reading stories at home (scale 1-5)	148	3.79 (1.07)	140	3.92 (0.96)

For all variables, group comparisons p>.1. Note that not all background information was available for all children.

Table B2: ***Mean percent correct (SD) for experimental tasks by language group, Full sample***

		Monolingual (N=145)	Bilingual (N=145)	Comparison
Working memory	Forward	41.9 (13.7)	40.3 (16.5)	$F(1, 288) = .71$, $p = .402$, $\eta_p^2 = .002$
	Backward	29.8 (19.1)	29.3 (19.4)	$F(1, 288) = .04$, $p = .840$, $\eta_p^2 = .00$
Vocabulary		76.8 (17.5)	57.1 (26.2)	$F(1, 288) = 56.4$, $p < .001$, $\eta_p^2 = .164$
Morphological Inflection	Real words, noun pluralization	13.6 (17.6)	56.4 (23.3)	$F(1, 288) = 49.8$, $p <.001$, $\eta_p^2 = .148$
	Non word, verb inflection	50.9 (23.3)	42.7 (23.5)	$F(1, 288) = 8.9$, $p = .003$, $\eta_p^2 = .030$
Morphological Derivation	Verb	60.3 (24.1)	36.1 (25.8)	$F(1, 288) = 68.1$, $p < .001$, $\eta_p^2 = .191$
	Adjective	52.9 (24.9)	35.1 (26.2)	$F(1, 288) = 35.1$, $p <.001$, $\eta^2 = .109$

Appendix C: Examples of Error Coding

Non-Word Verb Inflection

The task was scored twice: The first score is the absolute accuracy, namely 1 point for each correct answer. The second score gave credit for partial morphological knowledge reflected in responses. The partial knowledge score relied on a detailed analysis, with one point given for each of the following: use of the same root as the stimulus sentence (root), use of the same verb pattern as the stimulus sentence (pattern), inflection in the required person (person), inflection in the required tense (tense; see Table C1 for examples). Responses that did not preserve the root were still given credit for the other criteria because children still performed the morphological inflection. Responses in which there was an error in the affixes of the verbal-pattern and of the tense (e.g., said *šolazet* instead of *šolezet*), did not receive credit. The criteria of person and tense were coded and received credit only if the verbal-pattern (binyan) exists in Hebrew (whether it was accurate in the present context or not).

Table C1: ***Examples of error analysis and partial scores in Non-word Verb inflection***

Error type (prompt)	Correct answer	Child response	Root	Pattern	Person	Tense
No root preservation (*šolez*)	*šolezet*	*šolexet*	0	1	1	1
Error in verbal pattern affix (*šilez*)	*šilza*	*šilaza*	1	0	0	0
Error using an existing verbal-pattern in Hebrew (*eštalez*)	*tištalez*	*išloz*	1	0	0	1
Error using a verbal-pattern that does not exist in Hebrew (*mešalezet*)	*tešalez*	*šzelt*	0	0	0	0

Verb Derivation

The task was scored twice: The first score is the absolute accuracy, namely 1 point for each correct answer. The second score gave credit for partial morphological knowledge reflected in responses. The partial knowledge score relied on a detailed analysis, with one point given for use of the same root as the stimulus sentence (root), and for use of one of the three possible resultative verb patterns (pattern; see Table C2). Responses not using the requested lexical category (e.g. using the infinitive *le-hadbik* 'to paste' instead of the inflected verb *madbikim* 'pastes') but retaining the root and an accurate verbal pattern, received points on both criteria. Similarly, responses inflected for person (e.g., using *soxetet* in the feminine singular, instead of *soxtim* in masculine plural) also received credit for both criteria. Credit for root preservation was given only if it was fully preserved, but not if it was partially represented in the children's response (e.g., said *metofim* instead of *metofefim* 'beat'). If it was not possible to determine based on the transcription whether the response was a noun (*masrek* 'comb') or a verb (*mesarek* 'to brush'), children were given credit for the pattern. Responses including phonological mistakes (e.g., said *metopef* instead of *metofef*), were accepted as correct.

Table C2: ***Examples of error analysis and partial scores in verb derivation***

Error type (prompt)	Correct answer	Child response	Root	Pattern
No root preservation – Lexical strategy (*masxeta* 'juicer')	*soxtim*	*xotxim*	0	1
A suitable root combination in a possible pattern – Morphological strategy (*masxeta* 'juicer')	*soxtim*	*masxitim*	1	1
Using another lexical category (<u>*devek*</u> 'glue')	*madbikim*	*lehadbik*	1	1

Adjective Derivation

The task was scored twice: The first score is the absolute accuracy, namely 1 point for each correct answer. The second score gave credit for partial morphological knowledge reflected in responses. The partial knowledge score relied on a detailed analysis, with one point given for use of the same root as the stimulus sentence (root) and one point given for use of one of the possible resultative adjective patterns (pattern; see Table C3 for examples). Responses including an error in gender or person (e.g., using the masculine adjective *taluy* instead of feminine adjective *tluya*), received credit for both criteria. Responses including phonological mistakes (e.g., pronouncing the word 'broken' as s*habur* instead of s*havur*) were accepted as correct.

Table C3: ***Examples of error analysis and partial scores in adjective derivation***

Error type (prompt)	Correct answer	Child response	Root	Pattern
Correct root with a possible pattern, but not the accurate response (*hirkivu* 'put together')	*murkav*	*raxuv*	1	1
Root not preserved (*na'alu* 'locked')	*naul*	*sagur*	0	1
Incorrect root in a possible pattern (*tiknu* 'fix')	*metukan*	*muxan*	0	1
Person disruption only (*talu* 'hang')	*tluya*	*taluy*	1	1

License

Child-directed speech in Ku Waru and Nungon (Papua New Guinea)

Hannah S. Sarvasy
Western Sydney University, Australia

Alan Rumsey
Josua Dahmen
Australian National University, Australia

John Onga
Kailge Community, Papua New Guinea

Stephanie Yam
University of Cologne, Germany

Abstract: It is still unknown whether parents in all societies make special speech adjustments when speaking to infants and small children. Researchers in a range of disciplines continue to cite the Kaluli community of Papua New Guinea (PNG) as evidence that not all societies make such adjustments in child-directed speech (CDS), but until recently there have been few modern, quantitative analyses of prosodic, phonological or morphosyntactic features of CDS for any language of PNG. Now, however, a solid body of research on CDS in PNG communities attests to widespread adjustments in CDS to toddlers and preschoolers, especially in the two languages on which we conduct firsthand research: Ku Waru and Nungon. Here, we present the state-of-the-art in current understanding of special features of CDS in Ku Waru and Nungon, in the domains of prosody, phonology, lexicon, and morphosyntax. Nungon CDS has higher mean pitch and greater pitch range than adult-directed speech (ADS), while Ku Waru results are less conclusive in this direction. CDS in both languages features optional modification of consonants that makes them sound similar to early child productions, while Nungon CDS vowels are not hyper-articulated relative to ADS vowels. Both languages are described by native speakers as utilizing a medium-sized set of special baby-talk lexical items, and these have variable distribution relative to ADS lexical counterparts in corpora. CDS in Nungon, but not as clearly in Ku Waru, shows evidence of morphosyntactic "fine-tuning" to child production abilities. Nungon CDS features an unusual morphosyntactic alteration that arguably makes sentences longer and more syntactically complex but simplifies words morphologically. Overall, the possible modifications available for CDS in both languages constitute less a coherent "register" that speakers may slip into or out of, but more a menu of optional features, some apparently binary and some measured in terms of degree, which may be applied in conjunction with each other or separately, and which adults often apply variably within a single recording session.

Keywords: child-directed speech; Papuan languages; Nungon; Ku Waru.

Corresponding author: Hannah S. Sarvasy, MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith NSW 2751, Australia. Email: h.sarvasy@westernsydney.edu.au

ORCID IDs: Hannah Sarvasy: https://orcid.org/0000-0002-9551-480X; Alan Rumsey: https://orcid.org/0000-0002-9710-4695; Josua Dahmen: https://orcid.org/0000-0002-1015-2353

Citation: Sarvasy, H. S., Rumsey, A., Dahmen, J., Onga, J., & Yam, S. (2025). Child-directed speech in Ku Waru and Nungon (Papua New Guinea). *Language Development Research, 5*(3), 191–244. http://doi.org/10.34842/ldr2025-857

Introduction

In most of the world's communities for which the linguistic features of child–caregiver interactions have been studied, child-directed speech (CDS) is reported to display special acoustic, lexical, prosodic, and/or morphosyntactic features that differ from those of adult-directed speech (ADS) (Soderstrom, 2007; Weinstein & Baldwin, 2024, *inter alia*). Perhaps the most famous reported exception is the Kaluli-speaking community (Bosavi[1] language family) of Papua New Guinea (PNG) which continues to be frequently cited as evidence against the universality of special CDS features (e.g., Duranti, 2009; Johnson et al., 2023; Rosenberg et al., 2004; Rowe, 2008; Soderstrom, 2007; *inter alia*). Indeed, until recently there were few modern, quantitative analyses of prosodic, phonological or morphosyntactic features of CDS for any language of PNG, including Kaluli. But now, a growing body of research on CDS in communities of PNG attests to widespread use of at least some special CDS features. In this section, we use the term "CDS" to refer to both the general notion of speech directed to children (of any age), and to refer in some cases specifically to speech directed to children older than 1;0. Indeed, the youngest children in almost all studies of CDS in PNG have been older than 1;0 (Hellwig et al., forthcoming).

Features of CDS have been examined to varying degrees for the nine languages of PNG in which child language acquisition has been studied (for more information on what is known about child language acquisition in Papuan languages of PNG, see Hellwig et al., forthcoming). The best-known accounts of CDS in a PNG language come from Schieffelin's pioneering research on Kaluli (Schieffelin, 1985, 1990), but these lack details on some facets of CDS, with no measurements of pitch or vowel acoustics, for instance. Goldman (1986, 1987) describes aspects of the Huli (Engan) baby-talk lexicon while reporting, without details, the presence of prosodic and pragmatic adjustments in Huli CDS that align with cross-linguistic norms. San Roque reports possible consonant substitution, prompting routines, and a prevalence of "where" questions in Duna (Bogaia) CDS (Rumsey et al., 2013; San Roque, 2008, 2016). As part of a broader study of language shift from Tayap (isolate) to the Papua New Guinean lingua franca Tok Pisin (descended from a variety of Pacific Pidgin English), Kulick (1980, 1992; Kulick & Terrell, 2019) reports on some aspects of mixed Tayap/Tok Pisin CDS. Boer et al.'s (2022) study of child production of Tok Pisin consonants in children's speech lacks reference to CDS.

Recent quantitative research into linguistic features of CDS in PNG has focused on four languages: Ku Waru (Chimbu-Wahgi), Nungon (Finisterre), Qaqet (Baining), and Yélî Dnye (isolate). Previous work on Ku Waru CDS investigated phonetics/phonology (Rumsey, 2017) and morphosyntax (Rumsey et al., 2020); in this paper, we present

[1] Throughout this paper, the family to which a language belongs is given in parentheses the first time the language is mentioned.

new studies of prosody and the baby-talk lexicon in Ku Waru CDS. Previous work on Nungon CDS focused on phonetics and prosody (Sarvasy et al., 2019, 2022), morphosyntax (Sarvasy, 2019, 2020, 2021, 2022, 2023b), and pragmatics (Sarvasy, 2022, 2023c). While a baby-talk CDS lexicon for Nungon was mentioned in the Nungon reference grammar (Sarvasy, 2017) and in several previous studies focusing on other topics, the distributions of Nungon baby-talk lexical items were never examined: this will be done here. Frye's monograph (2022) on Qaqet (Baining) CDS is the only book on linguistic characteristics of CDS in a language of PNG; Frye investigates mean length of utterance, disfluencies, mean pitch and pitch range, speaking rate, attention-directing, responses to child errors, and baby-talk lexicon. Marisa Casillas is the only researcher to date who has used day-long recordings with a PNG speech community, that of Yélî Dnye (Bergelson et al., 2023; Bunce et al., 2024; Casillas et al., 2021); she has used these to assess CDS quantity, rather than its linguistic features (her studies are also the principal ones in PNG that involved children younger than 1;0 alongside older children).

Passing mention of lexical and discourse aspects of CDS occurs in reference grammars of additional PNG languages, such as Mali (Baining; Stebbins, 2011, pp. 28–29) and Manambu (Ndu; Aikhenvald, 2008, pp. 44, 138). Discussion of facets of how adults interact with small children in PNG societies is also scattered throughout the language socialization and anthropological linguistics literature (San Roque & Schieffelin, forthcoming).

In this paper, we present an overview of special CDS adjustments in multiple linguistic domains in Ku Waru and Nungon. We survey our published work on CDS phonetics and phonology and morphosyntax, and present new data and analyses on baby-talk lexicons and aspects of Ku Waru prosody and morphosyntax. Overall, the picture is of a complex constellation of adjustments in CDS of both languages, with structural similarities and divergences, and wide variation in application within and across individuals.

In the next section, we provide background information on the two languages and our recording methods. Each of the following four sections presents a comparative overview of known features of an area of CDS (Prosody, Phonology, Lexicon, and Morphosyntax) in both languages. We then summarize and discuss the evidence, with further comparison to Papuan languages Duna, Huli, Kaluli, Mali, Manambu, and Qaqet, and close with a brief conclusion.

Language Background and Methods

Ku Waru

Ku Waru is a regional speech variety within a large dialect continuum in PNG's Western Highlands Province (for details about the grammar of Ku Waru, see Merlan & Rumsey, 1991, 322–343; Rumsey et al., 2020, 3–7). It has about 15,000 speakers and is the first language of everyone born in the village of Kailge, where Alan Rumsey's fieldwork has been carried out in collaboration with Francesca Merlan and John Onga, who is himself a native Ku Waru speaker. Nearly everyone at Kailge under the age of 65 also speaks Tok Pisin (a largely English-lexified pidgin/creole, which is PNG's main lingua franca, see Introduction). They primarily use Tok Pisin when travelling to other language regions, rather than in Kailge. Most children in Kailge these days learn Tok Pisin as a second language and alternate between it and Ku Waru to some extent even at Kailge (for details, see Rumsey, 2014).

The Ku Waru data that we draw on here are taken mainly from a corpus that consists of transcripts of interaction between five target children and their caregivers, audio and video recorded between 2013 and 2016 by John Onga and another Ku Waru-speaking local collaborator, Andrew Noma, using Olympus LS14 digital audio recorders and Canon HFM52 digital video recorders (see Rumsey et al., 2024). The children ranged in age between 1;4 and 4;9 at the time of recording. The selection of focus children was based on interviews with a large number of parents, and the need to find children with birth dates spaced 6–7 months apart so as maximize the overall age range within the initial 2-year period when most of the recordings were made. The recordings were made monthly, for 60–65 minutes.

The article also draws on audio recordings made by Onga and Noma during 2004–2007 (on Uher cassette recorders with Audio-Technica ATR25 microphones) involving two focus children (one of them John Onga's son Jesi). Those recordings were made at longer intervals than the ones in the 2013 and 2016 study—generally of about 3 months. For both corpora, in line with the caregivers' preferences, almost all of the recordings were made in their houses, during the daytime, talking around the hearth or while engaging in indoor work activities.

After completing the recording sessions, Onga and Noma transcribed their respective recordings by hand and translated them into English. Their transcripts were later typed by Appen Language Services into a plain-text format. They were then processed by corpus managers Tom Honeyman and Charlotte van Tongeren in OpenRefine to fix regular scribe and typist spelling errors and inconsistencies. All of the transcripts in the 2013–2016 corpus were entered into ELAN (Max Planck Institute for Psycholinguistics, 2024), and time-aligned with the associated media files. The ELAN files with the transcripts, and media files, are freely available to interested researchers in the

PARADISEC Catalog (https://catalog.paradisec.org.au/), in collection AR3.

Nungon

Nungon is the umbrella term for the six southern village-lects of an oval-shaped, thirteen-village, dialect continuum in and around the Uruwa River valley in the Saruwaged Mountains, Morobe Province, PNG (see the Nungon reference grammar for background on the area and grammar of the Nungon dialects: Sarvasy, 2017). The Uruwa River valley dialects differ greatly in phonology and lexicon (with 60–88% cognacy rates; Wegmann, 1994); the northern dialects also differ from the southern dialects in the morphological formation of the remote future tense (Wegmann, 1994). Each Nungon village-lect has no more than about 400 speakers in the region and the diaspora; in total, there are about 1,000 speakers of Nungon dialects. The Nungon-speaking area is not accessible by road, and Tok Pisin is less present there than in the Ku Waru-speaking area.

The Nungon corpus, recorded in just the Towet village dialect, includes three longitudinal studies, capturing data from nine children aged 1;1–5;10; it totals about 182 hours of digitally transcribed child–caregiver interactions (Sarvasy, 2023a). Study 1 (2015–2017) involved five children, aged 1;13 to 3;10 at the study's outset, audio- and video-recorded, using Zoom H5n recorders and a Canon digital camera with video capability for 1 hour per month over 2 years (17 months for the youngest child). Study 2 (2019) involved three additional children, aged 2;4–2;8 at the outset, each recorded for 4 hours within 1 week per month, for 5 months. Study 3 (2023) involved one additional child, aged 2;1 at the outset, recorded for four half-hour sessions each month, over 3 months.

All recordings were digitally transcribed in the village, using solar-powered laptops, by Towet Nungon speakers Lyn Ögate, Stanly Girip, James Jio, Nathalyne Ögate, Tabitha James, Yöngwenwen Hessy, and Böiwa Ögate. Transcriptions used Mid-CHAT format (MacWhinney, 2000) and were completed using a word-processor and the software CLAN (MacWhinney, 2000).

Prosody

Perhaps one of the most immediately salient characteristics of CDS in many languages is its prosodic modifications relative to ADS: cross-linguistically, these typically involve higher overall pitch, greater pitch range, distinctive prosodic contours, slower speech, shorter utterances, longer pauses, and a more reliable positioning of pauses at phrase boundaries, compared to ADS (Cruttenden, 1994; Fernald, 1989; Katz et al., 1996; Soderstrom, 2007; Stern et al., 1983; Wang et al., 2015).

We focus here on one of the most widely studied and perceptually salient features of CDS, cross-linguistically: higher mean pitch and greater pitch range. Perceptually higher pitch in CDS was reported incidentally for the PNG language Kaluli in Schieffelin's groundbreaking and influential work; for instance, when a mother and her daughter (aged 2;6) engage in teasing word play together, they both are said to maintain "high pitch" (Schieffelin, 1990, pp. 109–110). Schieffelin also notes that mothers hold pre-verbal infants to face themselves or others and speak "for them," using a high-pitched, nasalized register similar to the one used to address dogs (1990, p. 71–72; see also Schieffelin, 1979, pp. 106–108).[2] The relationship in pitch between IDS/CDS and pet-directed speech has also been noted for English-speaking communities (Gergely et al., 2017; Jeannin et al., 2017; Xu et al., 2013).

Fernald et al.'s (1989) acoustic analyses of mother's speech to their children in six language varieties (American English, British English, German, French, Italian, and Japanese) revealed a higher fundamental frequency and wider pitch range than in speech directed to other adults by the same speakers. Others corroborated these results for some of the languages included in that study, such as American English (Garnica 1977), British English (Shute & Wheldall, 1989), German (Fernald & Simon, 1984), and Japanese (Amano et al., 2006), and for other languages, like Mandarin Chinese (Grieser & Kuhl 1988; Liu et al., 2009), Cantonese (Wang et al., 2021), Dutch (Benders et al., 2021), and Australian English and Thai (Kitamura et al., 2001), among others. Among these, however, pitch range in CDS has been shown to be constrained to a degree in tonal languages (Mandarin, Cantonese, Thai) or pitch-accented languages (Japanese). For instance, Kitamura et al. (2001) suggest that Thai mothers "may restrict pitch excursions in order not to disrupt tonal information" (p. 386).

[2] There is also a very small amount of recent data, but without language-specific analyses on pitch, in infant-directed speech (IDS) in Enga (Engan), the Papuan language of PNG with the most speakers. As part of a cross-cultural experiment that aimed to assess the universality of various prosodic features of infant-directed speech and song (Hilton et al., 2022), anthropologist Pauline Wiessner recorded six female speakers of Enga each producing one very brief (3–16 seconds long) spoken snatch, and one very brief sung snatch (8–13 seconds long), in response to the prompt that they should simulate calming a "fussy" infant (usually, in the presence of an actual infant). Five of the six speakers also recorded similarly brief adult-directed spoken and sung snatches. However, this very small amount of data (13–27 seconds of IDS per speaker) in very noisy settings and without Enga transcriptions is unfortunately insufficient to draw conclusions about prosodic and phonological features of Enga IDS. Further, even if there were more data and of better quality, we would be sceptical about drawing conclusions about differences between IDS and ADS based on this study, due to the mismatch in affect between the highly emotionally "marked" IDS elicited (to calm an agitated infant) and the emotionally neutral ADS (normal conversation and song—not, for instance, to calm an agitated adult). Wiessner also notes that some of the "songs" sung to the infants were actually magic charms, hinting at a fascinating facet of Enga speech to children that is yet to be fully described (Pauline Wiessner, p.c., 2025).

K'iche' Mayan-speaking adults famously were shown to not make significant pitch modifications in CDS compared to ADS, even sometimes using a lower pitch register when addressing children (Bernstein Ratner & Pye, 1984). Bernstein Ratner and Pye (1984) point out that in K'iche', high pitch serves as a signal of deference to addressees of high status. In contrast, young children "come last in the Mayan age-grading system" (Bernstein Ratner & Pye, 1984, p. 521) and do not belong to the category of interlocutors with whom high-pitched voice is used.

Here, we focus on mean pitch and pitch range in Ku Waru and Nungon CDS.

CDS Prosody: Ku Waru

In order to check whether the general cross-linguistic tendency for higher mean pitch and greater pitch range also manifests in Ku Waru CDS, we conducted a small-scale acoustic study of the differences in mean pitch and pitch range between Ku Waru CDS and ADS using the existing Ku Waru child language corpus and interviews with parents who featured in that corpus.

Data Collection

The CDS data come from the Ku Waru child language corpus, while the ADS data (due to a lack of suitable recordings of ordinary adult conversations in Ku Waru) are from interview sessions that had been recorded for another purpose: Ku Waru speaker Onga asked parents questions while acting as an interpreter in the presence of an English-speaking Australian researcher (Lauren Reed). From these data sources, we selected three women who featured in both CDS and ADS corpora. The children whom these women were addressing in the CDS portion were aged 2;4 (a boy), 2;10 (a girl) and 3;3 (a girl) at the time of recording in July 2013. We note that the data sample used for this prosodic analysis was comparatively small and we did not conduct any prosodic analyses of speech by male participants. Further, our ADS data for these speakers was not truly conversational, since it was from interviews with a foreign researcher and non-Ku Waru speaker present, to whom the women might have in effect been directing their speech.

Speech Sampling

We sampled 50 utterances of both ADS and CDS for each of the women. Utterances were defined as segments of a speaker's continuous speech that are separated by more than 300 ms of non-speech by the same speaker (cf. Woolard et al., 2023; Kitamura et al., 2001). To broaden the range of contexts for data sampling, one set of 25 consecutive utterances were extracted from the beginning of each recording, and a second set of 25 consecutive utterances were taken from halfway into the recording (cf. Kitamura et al., 2001), recording length permitting. Utterances consisting of only one word or a standalone non-lexical vocalization, and any other types of interference that might affect the fundamental frequency (overlapping talk, extra-linguistic noise and clicks, laughter, whispering and creaky voice) were excluded from the speech sampling. All utterances were extracted from Ftable using Praat (Boersma & Weenink, 2025), resulting in a total of 300 sound files (3 caregivers × 2 contexts × 50 utterances).

Pitch Measurement and Analysis

We used Praat's automatic measurement tool to measure the mean F_0 (fundamental frequency), minimum F_0 and maximum F_0 (see also Kitamura et al., 2001; Frye, 2022) for each of the 300 utterances sampled. To minimize pitch-tracing errors, the pitch trace of each sound file was visually and aurally inspected in the Praat sound editor. An expanded pitch range of 100–600 Hz was set to account for both Praat's recommended pitch range of 100–500 Hz for analyzing adult female voices and the higher pitch range that has been found in other studies of women's CDS, which can reach up to 600 Hz (see Fernald & Simon, 1984). All other pitch parameters were left at Praat's default settings. To facilitate data interpretation and more accurately reflect the logarithmic character of human pitch perception, all fundamental frequency values were subsequently converted to a chromatic scale with twelve semitones (the same scale that is used for conventional music notation). The mean pitch was calculated with a reference value of 100 Hz, using the formula: semitones = $12 \times \log_2(\text{max}_{F0} \div 100)$. The pitch range was determined by subtracting the minimum pitch from the maximum pitch, using the formula: semitones = $12 \times \log_2(\text{max}_{F0} \div \text{min}_{F0})$. The results were analyzed in RStudio (Posit team, 2024).

Results

Mean Pitch in Ku Waru CDS. Figure 1 provides a visual summary of the distribution of mean pitch data across the three women. The greatest difference in mean pitch is found in speech addressed to the youngest child, aged 2;4, by the woman Naldi[3]. The linear regression model showed the CDS of this speaker to have a significant pitch difference compared to the corresponding ADS dataset (estimate = 2.33; standard error = 0.49; $p < 0.00001$). For the two other speakers, the median pitch of the analyzed CDS speech was slightly higher (Annie) or lower (Saina) than the corresponding ADS, but these differences were not statistically significant. However, it is worth noting that for all three speakers the highest-pitched utterances all occurred in the CDS datasets.

[3] Ku Waru children and parents are referred to by first names here.

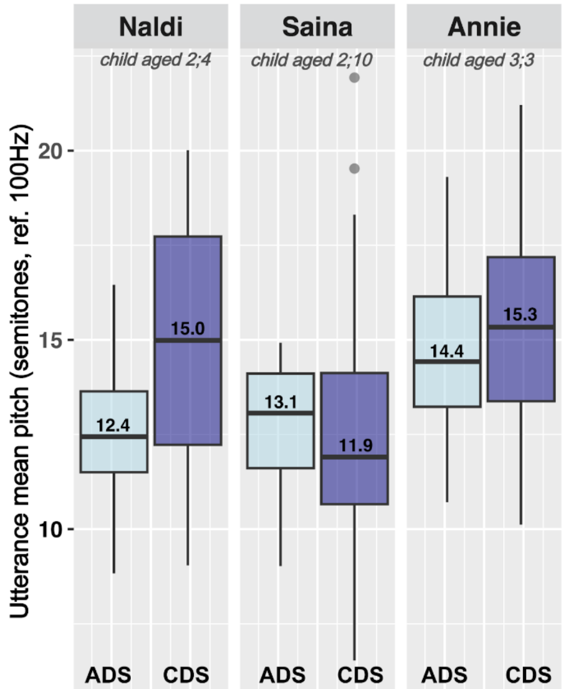

Figure 1. ***Utterance mean pitch in Ku Waru ADS and CDS.***

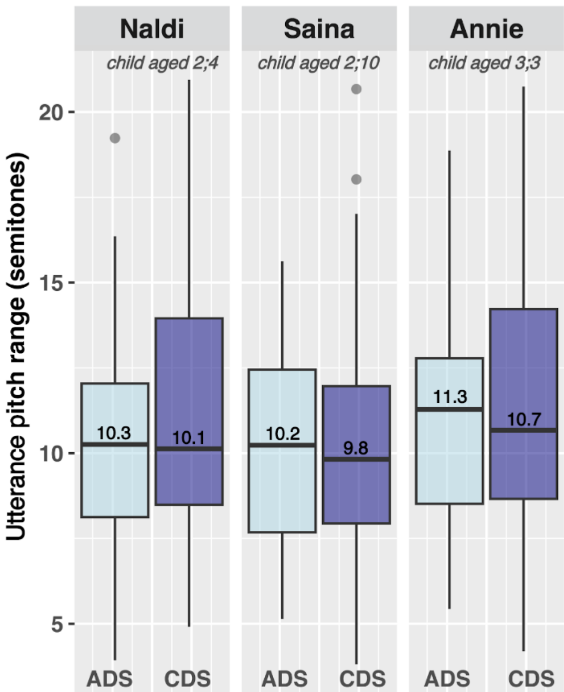

Figure 2. ***Utterance pitch range for Ku Waru ADS and CDS.***

Pitch Range in Ku Waru CDS. To examine potential differences in pitch range between CDS and ADS in Ku Waru, we extracted and compared the pitch range of the CDS and ADS datasets, shown in Figure 2. For all three speakers, the ADS median utterance pitch range was slightly greater for ADS than for CDS—a difference of 0.2 semitones (Naldi), 0.4 semitones (Saina) and 0.6 semitones (Annie), respectively. However, for all three speakers, a linear regression model showed that these differences in pitch range were not significant (estimate = 0.52; standard error = 0.40; p = 0.195). While the median pitch range was slightly greater in ADS for all three speakers, it is worth pointing out that for all three speakers, the utterances with the greatest pitch ranges all occurred in CDS.

CDS Prosody: Nungon

For CDS in Nungon, Sarvasy et al. (2022) examined mean pitch and pitch range in vowel tokens across three types of speech: a) CDS to 2-year-olds (2;2–2;9) and 3-year-olds (3;0–3;10), b) conversational ADS in native Nungon speaker pairs, and c) monologue narratives with an adult, female, non-native Nungon speaker as primary interlocutor. The results showed that pitch in these vowel tokens differed between CDS (to both 2- and 3-year-olds) and conversational ADS, such that both men and women used higher mean pitch and an increased pitch range in CDS than in ADS. Women's and men's use of pitch differed only in the monologual recordings with the non-native Nungon speaker as interlocutor; that was where women showed the greatest mean pitch, even higher than in CDS, while men's highest mean pitch was in CDS, followed by the monologues, then by conversational ADS. Women's pitch range in the monologues was similar to the range in their CDS, and these were both higher than the range in conversational ADS; men's pitch range was greatest in CDS, while their ranges in conversational ADS and the monologues were similar. This suggests that, in terms of pitch, Nungon-speaking men treat CDS in a different way than either conversational ADS or more performative monologue recording sessions, while for women, the monologue recording sessions were akin to CDS and even elicited a higher mean pitch than CDS (see Sarvasy et al., 2022, for a more detailed discussion).

Comparative CDS Prosody Summary

Our knowledge of mean pitch and pitch range in Ku Waru and Nungon CDS stems from two different methodological approaches. The new Ku Waru study we present here follows Frye (2022) in using the utterance as the unit for pitch analyses, while Sarvasy et al. (2022) used individual vowels as the unit of analysis. The ADS data used as a counterpart to CDS for Ku Waru here stem from not free conversations, but interviews with a foreign, non-Ku Waru speaker and Ku Waru-speaking interpreter;

they are thus imperfect counterparts to the Ku Waru conversational CDS. The conversational ADS data used by Sarvasy et al. (2022) are dyadic conversations in Nungon; Sarvasy et al. (2022) also, however, examined a second type of ADS: monologues recorded with a non-native Nungon speaker as primary interlocutor.

Our results also differ in that Sarvasy et al. (2022) found Nungon corroboration for the cross-linguistic tendency for CDS to feature both higher mean pitch and greater pitch range than conversational ADS, while the present analyses of Ku Waru CDS and ADS show significantly higher mean pitch in CDS for only one of three speakers, and no significant difference in pitch range between CDS and ADS for any of the speakers. The study on Nungon by Sarvasy et al. (2022) suggests that non-conversational ADS, especially with a non-native speaker present, may exhibit higher pitch than conversational ADS, and (for Nungon-speaking women, but not men) greater pitch range. Thus, it would be ideal to investigate mean pitch and pitch range in Ku Waru CDS using truly conversational ADS samples in the future.

Phonology

Along with prosody, phonetic and phonological adjustments in CDS can be highly salient and are well-attested cross-linguistically. Widely attested modifications include many that make CDS sound similar to early child productions: consonant substitutions, consonant cluster reductions, long-distance consonant assimilation (or "consonant harmony"), and an abundance of CVCV-structure words (Cruttenden, 1994). Some voice-onset time (VOT) modifications in CDS have been shown to reduce the acoustic contrast between two phonemes, as in Swedish (Sundberg & Lacerda, 1999), perhaps again, approximating child productions. Another type of attested modifications does not necessarily make CDS sound child-like: instead, they have been hypothesized to have a didactic purpose and serve to scaffold language acquisition. Among these are voice-onset time (VOT) modifications that serve to accentuate phonemic differences in Hakka and English CDS (Cheng, 2014, Cristia, 2010, Moslin, 1979), and vowel hyper-articulation (Sarvasy et al., 2022).

The reduction in CDS of phonemic contrasts between consonants is not universal, but it is well-attested, cross-linguistically. Recent research has shown reduced acoustic contrast in CDS for the four-way voicing distinction in Nepali stops in CDS to 10–18-month-old children (Benders et al., 2019), and for sibilants in Danish CDS to 19–20-month-old children (Bohn, 2013). Further systematic modification of consonants in CDS relative to ADS is attested for languages in Australia: Arandic (Turpin et al., 2014) and Warlpiri (Laughren, 1984). In fact, the CDS register for those languages has been analyzed as involving reduced phonological contrasts and co-occurrences, and a smaller phonological inventory, than ADS. Turpin et al. (2014, p. 51) report that Arandic adults describe the CDS register as mimicking the way children speak. Indeed, Turpin and colleagues distinguish two phases of Arandic CDS, "early" and "late," each with distinct phonological patterns; the early phase has the most neutralization of phonological contrasts, and its phoneme inventory can be represented as a set of nine segments—much reduced from the full 27 of Arandic ADS (Turpin et al., 2014, pp. 56–57). Some contrasts, such as between lamino-palatal and lamino-dental sounds, are neutralized in both phases of the Arandic CDS register.

For vowels, multiple studies have shown that the triangle in acoustic space formed by the three cardinal vowels /i/, /u/ and /a/ is larger in CDS than in ADS; this has been shown to be the case for, among others, American, Australian and British English (Burnham et al., 2002; Kuhl et al., 1997; Uther et al., 2007; but see Green et al., 2010), Mandarin (Liu et al., 2003), Spanish (García-Sierra et al., 2021), Swedish (Kuhl et al., 1997; but see Van de Weijer, 2001), and Japanese (Andruski et al., 1999; but see Martin et al., 2015; Miyazawa et al., 2017). There is some debate as to whether this should be considered "hyper-articulation," with the presumed aim of being understood and/or teaching correct speech sounds, or whether it stems from some other factors (Cristia, 2013; Cristia & Seidl, 2014). Indeed, the opposite—smaller vowel triangles in CDS, compared to ADS, or "hypo-articulation"—has been found for CDS in some other languages, such as Cantonese (Rattanasone et al., 2013), Dutch (Benders, 2013), German (Audibert & Falk, 2018), and Norwegian (Englund & Behne, 2006).

Here, we will focus on contrast reduction among consonants in Ku Waru CDS (making CDS sound more child-like) and show that Nungon CDS has an absence of vowel hyper-articulation (so lacks this presumably didactic element). We summarize our previous work in this area, which happened to differ in foci between Ku Waru (consonants) and Nungon (vowels).

CDS Phonology: Ku Waru

Our discussion of Ku Waru CDS phonology focuses on key consonants that have markedly different realizations in CDS than in ADS, namely the palatal consonants /ny/ and /ly/ in relation to apico-alveolar counterparts, the lateral consonants, and the apico-alveolar rhotic tap/trill /r/. The Ku Waru phonemic consonant inventory (ADS) is shown in Table 1. Characters shown without brackets are the ones used in the practical orthography for Ku Waru in this article. Accompanying some of the characters, in square brackets, are phonetic representations of the main allophones of the corresponding phonemes, with the most frequently occurring ones on the left.

In Ku Waru CDS, we see a general pattern by which sounds with various places of articulation are replaced by counterparts with an alveolar place of articulation, especially: palatals and velars, as well as the post-alveolar retroflex flap.

Table 1. ***Ku Waru phonemic consonant inventory***

Stops, glides and rhotic consonants

	Labial	Alveolar	Palatal	Velar
Plain stop	p	t		k
Fricative		s		
Prenasalized stop	b [mb], [mp]	d [nd], [nt]	j [ɲd͡ʒ]	g [ŋg]
Nasal	m	n	ny [ɲ]	ng [ŋ]
Continuant	w		y	
Rhotic		r [ɾ], [r]		

Lateral consonants

	Post-alveolar retroflex flap	Alveolar continuant	Palatal continuant	Prestopped velar
Lateral	rlt [ɺ̢]	l	ly [ʎ], [ʎ̥]	gl [ᶢ͡ʟ], [ᵏ͡ʟ̥]

Nasals

Until the age of about 3;0, Ku Waru-speaking children do not produce distinct forms for the palatal and alveolar nasals. Instead, they pronounce the palatal nasals as alveolar ones. Adults sometimes do likewise when talking to young children. Examples of this are shown in Table 2.

Table 2. ***Examples of replacement of Ku Waru palatal nasal consonants by alveolars in the speech of young children's speech and in some CDS by adults***

Phonemic form	Adult pronunciation	Children's form and occasional CDS form	Meaning
nyim	[ɲɪm]	[nɪm]	he/she said
manya	[maɲʌ]	[manʌ]	down

Laterals

Just as with nasals, until the age of about 3;0, Ku Waru-speaking children do not produce distinct forms for the palatal and alveolar laterals, instead pronouncing the palatal nasals as alveolar ones. And as with nasals, adults sometimes do likewise when talking to young children. Examples are shown are in Table 3.

Of the four Ku Waru lateral consonants, the typologically unusual and phonetically complex velar lateral /gl/ is the most frequently occurring lateral consonant (Rumsey, 2017). The alveolar /l/ appears to have come into the phonemic inventory of adult Ku Waru only since the arrival into the region of the pidgin-cum-creole Tok Pisin, after the 1930s. This is evident from the fact that in Ku Waru ADS this sound only occurs in loan words from Tok Pisin. A minimal quadruplet exists in which the four lateral consonants contrast in word-medial position (*korlta* [koɭʌ] 'chicken' / *kolya* [koʎʌ] 'place' / *kogla* [koᶢ͡ʟʌ] 'cry' / *kola* [kolʌ] 'cola drink'), and other near-minimal contrasting forms exist.

Table 3. ***Examples of replacement of Ku Waru palatal lateral consonants by alveolars in the speech of young children's speech and in some CDS by adults***

Phonemic form	Adult pronunciation	Children's form and occasional CDS form	Meaning
lyim	[ʎɪm]	[lɪm]	he/she took
ilyi	[ɪʎi]	[ɪli]	this
paly	[paʎ̥]	[pal]	all

Although alveolar /l/ does not occur in the adult Ku Waru lexicon in non-loan words, children use it in place of /gl/ and /lk/, and so do some adults in their CDS, as seen in Table 4.

Table 4. ***Examples of replacement of Ku Waru pre-stopped velar lateral consonant by alveolar lateral in the speech of young children and in some CDS***

Phonemic form	Adult pronunciation	Children's form and occasional CDS form	Meaning
ogla	[oᶢ͡ʟʌ]	[olʌ]	up
manya mogla	[maɲʌ moᶢʟʌ]	[manʌ molʌ]	sit down!
mogl	[moᵏ͡ʟ̥]	[mol]	no

Most children do not produce adult-like versions of the [ᶢ͡ʟ]/[ᵏ͡ʟ̥] sound until they are 5–6 years old. In the meantime, as alternative pronunciations of it they use not only [l] as shown above, but also [k], [g] and [t], and later [ɣ] and [x]. Interestingly, adults and older children when speaking to children in our corpus never use [k], [g], [ɣ], or [x] as imitative baby-talk realizations of [ᶢ͡ʟ], only [t] and, far more often, [l].

Rhotic Tap/Trill

Up until the age of 2;6–3;0, Ku Waru-speaking children and some adults in their CDS use alveolar [l] in place of /r/ (replacement of rhotics by laterals, stops and glides is widely attested in child speech, cross-linguistically; Tomić & Mildner, 2015). Examples are in Table 5.

Table 5. ***Examples of replacement of Ku Waru rhotic [r] with alveolar lateral [l] in the speech of young children and in some CDS by adults***

Phonemic form	Adult pronunciation	Children's form and occasional CDS form	Meaning
rais	[rais]	[lais] ~ [lait] ~ [las]	rice
kera	[kerʌ]	[kelʌ]	bird
kar	[kar]	[kal]	motor vehicle

Alternatively, intervocalic /r/ is sometimes pronounced as [t] in the speech of Ku Waru children and in some CDS. An example is [utu] for *uru* 'sleep'.

Alveolar Consonants in Place of Velars

As is widely attested in child-language studies around the world (McAllister Byun, 2012), especially before about 2;6, Ku Waru children sometimes use alveolar consonants in place of velar ones, and so do some adults in their CDS. Examples are in Table 6.

Table 6. ***Examples of replacement of Ku Waru velar consonants with alveolars in the speech of young children and in some CDS***

Phonemic form	Adult pronunciation	Children's form and occasional CDS form	Meaning
ga	[ŋga]	[da] ~ [nda]	sweet potato
kim	[kim]	[tim]	a girl's name
le	[ᶢ͡ʟe]	[te]	excrement

In this section we have presented examples of six consonant mutations that sometimes occur in Ku Waru CDS: ɲ > n, ʎ/ʎ̥ > l, ᶢ͡ʟ/ᵏ͡ʟ̥ > l, r > l and r > t, and velars > alveolars. It is striking that all of these mutations involve replacement of sounds with other places of articulation by alveolar sounds. In the Lexicon section, we will compare that set of mutations with the ones that are observed within a set of what 11 Ku Waru parents identify as baby-talk lexical items (*kang kel-nga ung-ma* 'words of or for small children').

CDS Phonology: Nungon

We will briefly summarize generalizations about consonants in Nungon CDS, then sum up what we know about Nungon CDS vowel phonetics.

Nungon CDS Consonants

In Nungon CDS, consonant replacement is attested, and at least one baby-talk lexical item involves a non-phonemic sound (cf. Manambu extra-phonemic bilabial trill in one baby-talk term; Aikhenvald, 2008, p. 44).

In the Kotet Nungon dialect, word- and syllable-final /k/ are realized as glottal stops, and there is no rhotic, such that the counterpart to /r/ in this dialect is /l/ or /d/. The Nungon child–caregiver conversation data all come from the Towet village dialect, which lacks a glottal stop and lateral /l/, except in inter-dialect loans, like the flowering tree name *longgö longgö*. Speakers of Towet Nungon sometimes say that the Kotet dialect, with /l/ instead of /r/ and with the glottal stop for /k/ sounds "childish," and in fact, CDS and CS (child speech) in Towet Nungon optionally exhibit both of these features (Sarvasy, 2019). Towet adults sometimes replace word- and syllable-final /k/ with glottal stops, and /r/ with /l/ or the palatal glide /y/, when speaking to small children. This is widely attested in the Towet Nungon child–caregiver corpus, but its use has not yet been investigated systematically to find, for example, correlations between these features in the children's speech and their occurrence in CDS.

In the Lexicon section, we will see that the Nungon baby-talk lexicon includes at least one word with a non-phonemic sound: baby-talk [ɣɪ] 'pitpit', 'wild sugar cane' (counterpart to adult *dee*) begins with a voiced velar fricative, which never occurs word-initially in adult Nungon, where it is solely an intervocalic allophone of velar stops (Sarvasy, 2017).

Nungon CDS Vowels

Towet Nungon has six phonemic vowels (two front, three back, and one low-mid) and contrastive vowel length. For instance, *yo-nga* [jɔ.ŋa] means 'speaking,' but *yoo-nga* [jɔːŋa] means 'taking them.' The acoustics of the six phonemic vowels in Towet Nungon adult speech were investigated in Sarvasy (2017) and then with a more detailed, quantitative treatment in Sarvasy et al. (2020).

Sarvasy et al. (2022) investigated whether Towet Nungon vowels in CDS are hyper-articulated relative to those in adult–adult paired conversations, and adult monologues produced with a non-native (researcher) primary interlocutor. Vowel tokens were extracted from three datasets: child–caregiver conversations (10 speakers, five men and five women: 1,580 vowel tokens), adult dyadic conversations (eight speakers, four men and four women: 718 vowel tokens), and adult monologue narratives (eight speakers, five women and three men: 1,507 vowel tokens). Four speakers (two men and two women) featured in all three datasets, while seven (three women and four men) featured in both child–caregiver and adult dyadic conversations, and five (two women and three men) featured in both the child–caregiver and adult monologue narratives.

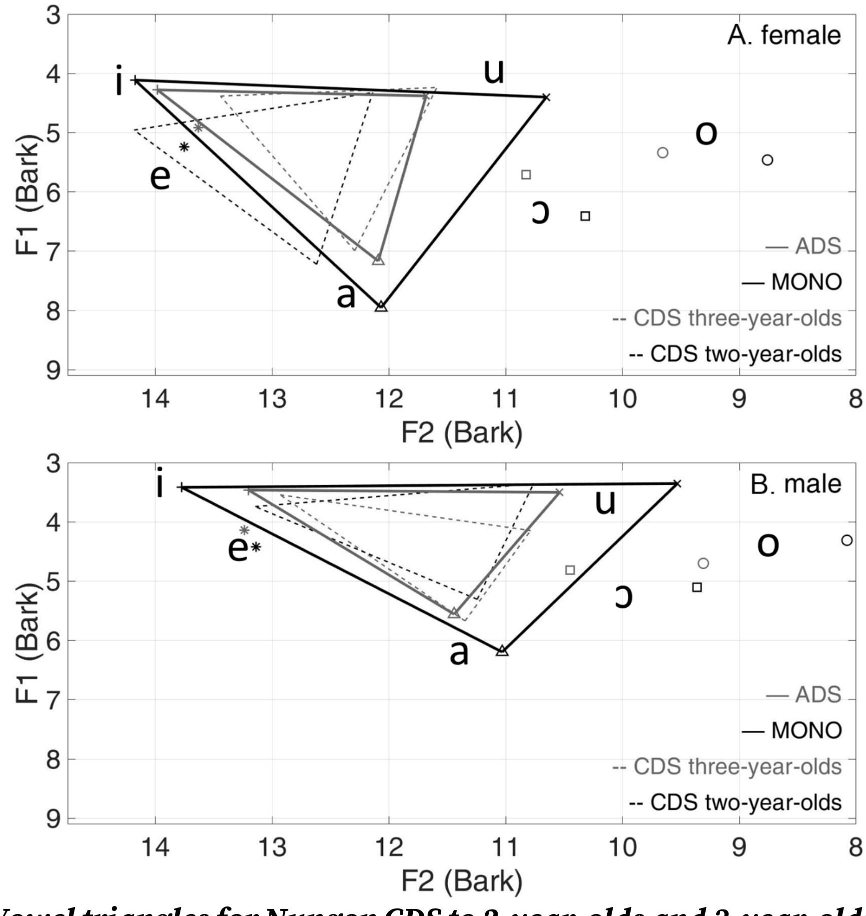

Figure 3. ***Vowel triangles for Nungon CDS to 2-year-olds and 3-year-olds, conversational ADS, and monologues with a non-native speaker interlocutor (Sarvasy et al., 2022, p. 12).***

We assessed vowel hyper-articulation using the "vowel triangle" technique, by which the area of the triangle formed by connecting the vowels /i/, /u/ and /a/ in an F1/F2 acoustic space graph is calculated (following García-Sierra et al., 2021). These triangles are shown in Figure 3.

There was no significant difference in vowel triangle size between CDS (to either 2-year-olds or 3-year-olds) and the adult–adult paired conversations (which we took to exemplify conversational ADS), but the vowel triangles in the monologues with a foreign, non-native Nungon speaker as primarily interlocutor/recorder were significantly larger than those of both CDS and conversational ADS (Sarvasy et al., 2022). Thus, Towet Nungon CDS to children aged two and three does not feature vowel hyper-articulation relative to conversational ADS.

We found that only one feature, vowel duration, differed in CDS directed to 2-year-olds versus CDS directed to 3-year-olds. That is, both men and women produced vowels of longer duration when speaking to 2-year-olds than they did when speaking to 3-year-olds and other local adults. The vowel durations in the ADS monologues were similar to those in the CDS to the 3-year-olds.

CDS Phonology Comparative Summary

We have seen that in both Ku Waru and Nungon, there is some alteration of consonants in CDS that makes them closer to the productions of small children. In Ku Waru, this even entails substituting a non-native Ku Waru phoneme, the alveolar [l] (found in adult speech only in Tok Pisin loan words), for all three other Ku Waru lateral phonemes (retroflex, palatal and pre-stopped velar). The Ku Waru consonant replacements we discussed generally involve use of alveolar sounds instead of sounds with other places of articulation: palatals to alveolars, velars to alveolars, the rhotic to alveolar lateral or stop, and the reduction of contrasts among laterals such that the alveolar lateral is used in place of the others. In contrast, in Nungon CDS of Towet village, while the alveolar lateral is one of the common replacements for the rhotic, the replacement of velar oral or nasal stops by alveolar ones is unattested in Nungon CDS. Towet Nungon CDS further exhibits a consonant replacement unattested in Ku Waru CDS: the replacement of word-final /k/ with glottal stops. This shows clearly that, while there are cross-linguistic patterns, the specifics of consonant replacement in CDS are necessarily language-specific.

We also saw that in Nungon CDS, vowels are not hyper-articulated relative to conversational ADS. Instead, monologues directed at a non-native Nungon speaking interlocutor did feature vowel hyper-articulation. A Ku Waru vowel acoustics study remains a desideratum for future research.

Lexicon

A central aspect of children's early language development is the acquisition of the core lexicon of their language. CDS figures in that process in various ways (Clark, 1993)—for example, through particular CDS distributions of vocabulary that differ from those in ADS. Here, our treatment will be limited to "nursery" or "baby-talk" lexical items that are used mainly or entirely in children's speech and in adults' CDS.

One of the most striking features of CDS in many languages is the presence of special "nursery" or "baby-talk" words used only when speaking to children and by children themselves, most of which have different ADS counterparts. An example in English is baby-talk *tummy* for ADS *stomach.* Baby-talk lexical items can be found in languages as typologically and geographically diverse as Romāni (Réger & Gleason, 1991), Inuktitut (Crago & Allen, 1997) and Japanese (Mazuka et al., 2008), among many others. Ferguson (1964, 1977) and then Haynes and Cooper (1986) surveyed reports on selected languages from around the world and found that most had special baby-talk lexicons. But Ferguson's sample included no Pacific languages, and Haynes and Cooper's (1986) included only one of them (the Oceanic language Palauan)—a severe under-representation, given that about a quarter of the world's languages are spoken in the region.

Languages of Oceania still remain greatly underrepresented in child language studies, but it was from one of them—Kaluli—that the first strong challenge to the universality of baby-talk lexicons emerged. One of Schieffelin's findings was that "Kaluli use no baby-talk lexicon, for they said (when I asked about it) that to do so would result in a child sounding babyish, which was clearly undesirable and counterproductive" (Ochs & Schieffelin, 1984, p. 293). That finding has been widely cited as a counterexample to the presumed universality of baby-talk lexicons. However, a less-cited point is that an extensive set of baby-talk lexical items was described for a neighboring language, Huli (Goldman, 1986, 1991).

Before we turn to our own new data on baby-talk lexical items in Ku Waru and Nungon CDS, we note that our methodology expands on that of Ferguson (1977) and Haynes and Cooper (1986), which primarily used word lists, without noting variation or frequency in actual CDS. We present word lists along with frequencies in corpora, following Crago and Allen (1997) on Inuktitut, Réger and Gleason (1991) on Romāni, and Ota et al. (2018; see also Ota & Skarabela, 2016, 2018) on English.

CDS Lexicon: Ku Waru

Unlike Schieffelin's (1990) report for Kaluli, Rumsey and Onga have found that every Ku Waru speaking adult or older child whom they have asked about the matter says that there are distinct words that are used by and to small children and has readily provided examples of what they consider to be such words. Further, again unlike the Kaluli, Ku Waru people say it is good for parents to use those words to small children sometimes, “to help them understand and learn to talk”. In order to investigate these local understandings of CDS systematically, Onga interviewed 11 Ku Waru-speaking parents about them. Speaking to the parents in Ku Waru, he asked them the equivalent of the following questions:

1. Do small children have their own ways of talking that are different from adults’?
2. If they do, what do you call their way of talking?
3. What words of that kind can you think of?
4. Do adults and older children sometimes use those same words when talking to small children?
5. If they do, why?
6. Is it good or bad for them to do that?
7. Why?

All 11 interviewees answered yes to question 1. The most frequent answer to question 2 was *kangabola kel-ma-nga ung-ma,* which can be translated either as ‘small children’s words’ or ‘words for small children’ (hereafter “baby-talk” or “BT”). When asked for examples of such words, as in question 3, all interviewees readily provided them, in sets containing from 11 to 17 items, with a median of 13. All of the words that were included in any of those sets are shown in the leftmost column of Table 7, along with what was said to be the equivalent Ku Waru adult word in column 2, and our English gloss of it in column 3. Column 4 shows the number of lists on which each of the words was included. That number can be taken as a rough measure of the extent to which the form in question figures in Ku Waru people’s shared understandings or stereotypes about how small children talk, and how others talk to them.

Whereas columns 1–4 provide information about what Ku Waru adult caregivers take to be typical of baby-talk, columns 5–8 provide information about the extent to which the actual adult CDS in the Ku Waru child–caregiver interaction corpus includes those words (for this purpose, individuals over the age of ten were treated as adults). For interpreting the figures in columns 6 and 7, it is important to take account of the fact that the children represented in the corpus did not talk as much as the adults. Rather, out of a total of 1,365,614 word tokens in the corpus, 469,820, or about 34.4%, were spoken by children, while 895,798, or about 65.6%, were spoken by adults. In order to take account of that difference, a weighting of 1.907 was assigned to the numbers in column 7 to compute the weighted ratio of baby-talk word tokens spoken by adults vs. children in column 8.

Table 7. ***Forms identified by Ku Waru parents as "small children's words" and their frequency of use by children and their adult interlocutors in our corpus***

Baby-talk form	Adult word in Ku Waru	English gloss	No. of parental BT word lists that contain the BT word	Tokens of BT word in the corpus	Tokens of BT word by adults in CDS	Tokens of BT word by children	Weighted ratio of incidence of BT word in CDS vs. children's speech (in %)
da	*ga*	sweet potato	6	4	3	1	61–39
pipi	*kera*	bird	6	230	142	88	46–54
dit/det	*kar*	car/truck	5	112	63	49	40–60
lais/lait/las	*rais*	rice	5	43	3	40	4–96
tali/talip	*kalyip*	peanut	5	0	0	0	NA
neta	*neka*	red pandanus	4	0	0	0	NA
balut/palut	*balus*	plane	3	7	0	7	0–100
tenapa	*kenapa*	corn	3	0	0	0	NA
tim/timu	*kim/kimu*	green vegetable	3	3	2	1	51–49
yayi	*kuru*	evil spirit / devil	3	1028	668	360	49–51
gu	*lku*	house	2	382	2	380	0–100
kal	*kar*	car/truck	2	94	12	82	7–93
kela	*kera*	bird	2	83	9	74	6–94
lawa	*plawa*	flour/flower	2	5	1	4	12–88
nela	*neka*	red pandanus	2	0	0	0	NA
pawa	*flawa*	flour/flower	2	42	3	39	4–96
tait	*rais*	rice	2	0	0	0	NA
te	*gle*	excreta	2	5	2	3	26–74
toti	*sosis*	sausage	2	0	0	0	NA
tu	*ku*	money/stone	2	0	0	0	NA
tun	*kung*	pig	2	4	0	4	0–100
ulu	*uru*	sleep	2	1	0	1	0–100
utu	*uru*	sleep	2	113	82	31	58–42
bil	*bi*	pen/pencil	1	4	4	0	100–0
deti	*tata/lapa*	father	1	144	69	75	40–60
lu	*lku*	house	1	19	5	14	16–84

Baby-talk form	Adult word in Ku Waru	English gloss	No. of parental BT word lists that contain the BT word	Tokens of BT word in the corpus	Tokens of BT word by adults in CDS	Tokens of BT word by children	Weighted ratio of incidence of BT word in CDS vs. children's speech (in %)
miyau	*lopa pus*	cat	1	32	20	13	45–55
mun mun	*kalyimbi*	moon	1	0	0	0	NA
nana	---	small/cute/baby thing	1	2514	1671	843	51–49
nunu	*keri to- no-*	kiss	1	34	29	5	75–25
pepe	*pekpek*	excreta	1	49	32	17	50–50
pola	*pora*	finish	1	14	5	9	23–77
tera	*kera*	bird	1	0	0	0	NA
wala	*wara*	water	1	117	78	39	51–49

The following points are evident from the data in Table 7:

1. There were considerable differences among the 11 parents as to which words they cited in the interviews as typical baby-talk ones. Nonetheless, a clear majority of the words cited by any of the parents (24/34) were cited by at least one other parent.
2. There is a disparity between the frequency with which the interviewees cite given words as typical baby-talk and how often those words actually occur in the corpus, either within adults' speech or the children's. The most extreme examples of this are: *da* for *ga* 'sweet potato,' which was cited by six of the 11 parents but only occurs four times in the corpus; *tali* or *talip* for *kalyip*, 'peanut,' cited by five parents, which never occurs in the corpus; and *neta* for *neka* 'red pandanus,' cited by five of the parents, which also never occurs in the corpus. This disparity is probably related to the fact that the substitution of alveolar stops for velar stops occurs mainly in CDS to children under about 2;6, whereas a large majority of our samples are from interactions with older children.
3. Conversely, several of the most frequently occurring words in Table 7 are among those cited by a small minority of the interviewees. Examples are the following, each of which was offered as baby-talk by only one of the 11 parents: *pepe* for *pekpek,* 'excreta' (Tok Pisin), with 49 instances; *wala* for *wara* 'water' (Tok Pisin), with 117 instances, *deti* (Tok Pisin) for *tata* or *glapa* 'father', with 144 instances; and *nana* 'small/cute/baby thing', with a whopping 2514 instances. For *pepe, wala,* and *deti*, it is possible that most interviewees declined to list these because they are based on

Tok Pisin words (*pekpek*, *wara* and *dedi*). Consistent with the parents' characterization of the words listed in the table (as used both by small children and by adults when speaking to them), among the 25 words with any attested instances in the corpus, all but three (*balut* 'plane,' *tun* 'pig,' and *ulu* 'sleep') are used by both children and their adult interlocutors.

4. There are wide disparities in the weighted ratios of uses of putative baby-talk words by children *vs.* their use by the adult interlocutors. This variation occurs in both directions, with some of those words being used far more often by the children and some by the adults. For the words that occur only a few times in the corpus, this variation is not significant. So, limiting the count to the putative baby-talk words that occur ten or more times in the corpus, we can see the adult-to-child usage ratio ranges from 75–25 for *nunu* 'kiss' (n=34) to .03/99.7 (rounded to 0/100 in the table) for *gu* 'house' (n=382).
5. In the Phonology section, we discussed six consonant mutations that have been observed in Ku Waru CDS: ɲ > n, ʎ/ʎ̥ > l, ᶢ͡ʟ/ᵏ͡ʟ̥ > l, r > l, r > t, and velar stops > alveolar stops. Only four of those six mutations are present among the items provided by the interviewees: ʎ/ʎ̥ > l (*tali* 'peanut'), r > l (*kal* 'car,' *kela* 'bird,' *ulu* 'sleep,' *wala* 'water'), r > t (*tait* 'rice,' *utu* 'sleep') , and velar stops > alveolar stops (*da* 'sweet potato,' *tali* 'peanut,' *neta* 'red pandanus,' *tenapa* 'corn,' *tim* 'green vegetable,' *te* 'excreta,' *tera* 'bird,' *tu* 'money/stone,' *tun* 'pig').
6. Although we are discussing Ku Waru CDS phonology and lexicon in separate sections in this paper, some of the items in Table 7 suggest that the distinction between the two is better seen as a cline rather than a categorical difference. As shown in the Phonology section, parents are aware of the systematic differences between children's pronunciation of a wide range of Ku Waru consonants and that of adults and sometimes match their pronunciation of those consonants to that of children.

The data above exhibit a disparity between what a community may acknowledge as appropriate words to use with children and actual practice. In the Ku Waru case, there is extreme variability in the extent to which adults identify and use baby-talk forms. We also note the absence of ɲ > n or ᶢ͡ʟ/ᵏ͡ʟ̥ > l mutations among the words cited by the interviewees. But there are also many differences that Ku Waru parents *are* aware of, and these play a part, both in how they speak to children, and in the lexical items they point to as stereotypic instances of it, or CDS shibboleths.

CDS Lexicon: Nungon

In her first nine months of immersion fieldwork on Nungon grammar, Sarvasy (2017) found that adults agreed widely that there were baby-talk lexical items in Nungon. As reported further in Sarvasy (2019), these have varying word classes, origins, and semantics. A partial list is in Table 8.

Table 8. ***A selection of Nungon baby-talk lexical items with adult counterparts***

Baby-talk form	Adult word (Towet Nungon)	English gloss	Origin of baby-talk form
bui	*hup*	chicken	the call by which Nungon speakers summon chickens
buu	*yup* (Nungon), *barus* (Tok Pisin)	airplane	onomatopoeia
dada	*dat*	(older) sibling	phonological modification
dudu	*murong*	genitals	unknown
ede	*dogu*	ghost, spirit	unknown
mama	*mak*	mother	phonological modification
nana	*tanak*	food	modification
[ɣɪ]	*dee*	pitpit	unknown
nauk	*yamuk*	water	phonological modification
nou=ma	*unom=ma*	bogeyman	phonological modification
nunu	*i-no-ng yo-*	kiss	unknown
papa	*nan*	father	Tok Pisin?
tutu	*mum*	breast(milk)	Tok Pisin?
adö	*amök*	sit	phonological modification
aing	*yii-*	bite	onomatopoeia?
dai	*duo-*	sleep	Tok Pisin?
ding!	*di-*	burn/danger	Co-opting of existing (ADS) form
hoit	*honggit-*	grab	phonological modification
pait	*emo-*	fight	Tok Pisin
purik	*iwan-*	turn	unknown
toik	*towi-*	fix	phonological modification

Of the Nungon baby-talk forms in Table 8, some, like those for 'water,' 'food,' 'airplane,' etc. are used as nouns. Others, like 'bogeyman' and 'genitals,' can be used somewhat like adjectives, indicating that something is related or suggestive of a bogeyman, or that someone is revealing their genitals. These functions are not also found with the adult counterparts; for instance, the adult *murong* is strictly a noun, referring to the genitals, and someone who is dressed immodestly can be called *murong-ni* 'genitals-ADJ,' not *murong* without an adjectivizer. Still others have verbs as their non-baby-talk counterparts, like those for 'turn,' 'grab,' 'sleep,' and 'bite'—but the baby-talk forms do not inflect directly, like their non-baby-talk counterparts. Children use these as stand-alone utterances, but in CDS, these are most often used in a complex predicate construction with an auxiliary verb *to-* 'do,' as in example (1):

(1) Aing to-wangka-k.
bite.BT do-NFUT.SG-3SG
'It will bite.' [Lit.: 'It will do biting.']

The adult way of expressing 'it will bite' requires the speaker to indicate, through an obligatory prefix on 'bite,' the person and number of the bitten argument, as in (2):

(2) Ge-i-wangka-k.
2SG.O-bite-NFUT.SG-3SG
'It will bite you.'

At least two baby-talk forms do not have direct adult counterparts. When something is dangerous to touch, adults tell children *ding!* This looks like the form of the verb *di-* 'burn' that occurs in multi-verb predicates, *di-ng.* But in CDS, *ding!* is used for both very hot things and for things that could be otherwise harmful—such as stinging tree leaves (here, this appears similar to Huli; Goldman, 1986, p. 200 [1.6, 1.8]). Such a warning is not used with adults. Also, the widely used baby-talk *dada* 'sibling' is used for both same- and different-sex siblings, but in non-baby-talk, the terms *naat* 'different-sex sibling,' *oruk* 'brother of male,' and *daa* 'sister of female' are used. There is a rare non-baby-talk form *dat* 'sibling,' attested in one song, but this is not used in everyday parlance.

Children learning Nungon vary in their own use of baby-talk verbs (Sarvasy, 2023b; this remains to be investigated for baby-talk nouns). The child TO produced 1–4 tokens of baby-talk verbs in each 1-hour recording session from age 2;3 through 2;7 (she was recorded for 1 hour per month), while the child AB produced only two baby-talk verb tokens in this period with a similar recording schedule of 1 hour per month, and only in one of his five recording sessions at these ages. The child MK produced only one baby-talk verb token across all 15 1-hour recording sessions she participated in from age 2;4 through 2;7, and the child MF produced no baby-talk verb tokens in any of her six recording sessions at 2;4 and 2;5 (Sarvasy, 2023b).

This variation possibly relates to variation in use of baby-talk forms in Nungon CDS, illustrated in Table 9 (verbs) and Table 10 (nouns).

Table 9. ***Baby-talk verb and adult counterpart token counts (adult form token counts in parentheses) in Nungon CDS***

Child, age (# of sessions)	Parent	*aing* 'bite' (adult *yii-*)	*toik* 'arrange' (adult *towi-*)	*purik* 'turn' (adult *iwan-*)	*hoit* 'grab' (adult *honggit-*)	*adö* 'sit' (adult *amök*)	*pait* 'fight' (adult *emo-*)	*dai* 'sleep' (adult *duo-*)	*ding* 'burn' (adult *di-*)
TO, 2;4	mother	1 (2)	0 (1)	2 (0)	2 (3)	0 (0)	0 (0)	1 (0)	0 (0)
TO, 2;4	father	1 (1)	0 (2)	0 (1)	1 (15)	0 (2)	0 (0)	3 (0)	0 (0)
TO, 2;5	mother	2 (0)	0 (0)	0 (0)	4 (1)	3 (2)	0 (0)	1 (0)	0 (0)
TO, 2;6	mother	0 (0)	0 (0)	1 (0)	3 (1)	0 (2)	0 (0)	0 (0)	0 (0)
TO, 2;7	mother	2 (0)	0 (2)	3 (1)	4 (3)	1 (4)	1 (0)	0 (0)	0 (0)
AB 2;4–2;7 (4 sess.)	father, mother	0 (3)	0 (14)	0 (1)	0 (13)	0 (10)	0 (1)	1 (7)	0 (1)
MK 2;4–2;7 (15 sess.)	father, mother	0 (39)	0 (30)	0 (22)	3 (144)	0 (27)	0 (0)	0 (25)	0 (5)
MF 2;4–2;5 (6 sess.)	father	0 (23)	0 (24)	0 (5)	0 (47)	0 (26)	0 (0)	2 (24)	0 (1)

Table 9 shows the distributions of the eight baby-talk verbs from Table 8 in Nungon CDS to four children aged 2;4–2;7 (just 2;4–2;5 for MF); note that two of the verbs do not occur in CDS in this sample at all. The child TO's parents produce a range of baby-talk verb types, and multiple tokens, in each recording session at 2;4, 2;5, 2;6, and 2;7 (in the table, there is just one 1-hour session at each age; the mother's and father's tokens are separated for the session at 2;4, but this was a single session in which both parents featured). This contrasts greatly with AB's, MK's, and MF's parents when those children were of the same age; for these children, multiple sessions are represented within a single row in the table, since there are so few tokens of baby-talk verbs in these parents' CDS. (MF and MK are from a 2019 study with dense recording protocols, so these children were generally recorded for four 1-hour sessions during 1 week each month; at 2;6, MK was recorded for only three 1-hour sessions, and MF was recorded for only two sessions at age 2;5.)

While TO's parents produce at least one token of all forms except *toik* and *ding* in these sessions, the parents of AB, MK and MF are attested as producing tokens of only two different forms during this period: *hoit* and *dai.* Yet this is clearly not an artefact of the conversational content: parents of all three children produced the adult counterparts to all baby-talk verbs with generally robust token numbers except *pait* 'fight' in the speech of the parents of MK and MF

Similarly, TO's parents produced more baby-talk noun tokens than the other three children's parents in the same period, as seen in Table 10. Here, the nouns *nauk* 'water,' *buu* 'airplane,' *nana* 'food,' and *dada* 'sibling' are counted in CDS per session (for TO) and across multiple sessions in the study period (AB, MF, and MK). TO's parents produce these multiple times per session, and in general, with higher frequency than the baby-talk verbs in Table 9.

Table 10. ***Baby-talk noun and adult counterpart token counts (adult form token counts in parentheses) in Nungon CDS***

Child, age (# of sessions)	Parent	*nauk* 'water' (adult *yamuk*)	*buu* 'airplane' (adult *balus*)	*nana* 'food' (adult *tanak*)	*dada* 'sibling' (adult: *naat, daa, oruk*)
TO, 2;4	mother	3 (6)	3 (6)	10 (8)	12 (0)
TO, 2;4	father	0 (2)	1 (5)	2 (0)	6 (0)
TO, 2;5	mother	2 (2)	6 (12)	9 (0)	12 (0)
TO, 2;6	mother	0 (4)	2 (0)	2 (0)	1 (1)
TO, 2;7	mother	10 (0)	12 (17)	0 (2)	7 (4)
AB, 2;4–2;7 (4 sessions)	father, mother	0 (36)	1 (5)	0 (1)	10 (20)
MK, 2;4–2;7 (15 sessions)	father, mother	0 (136)	0 (108)	1 (84)	3 (7)
MF, 2;4–2;5 (6 sessions)	father	0 (21)	0 (11)	0 (20)	16 (3)

In Table 10, baby-talk noun tokens are consistently present in CDS to TO between 2;4 and 2;7. In contrast, the other parents produce only one baby-talk token of *buu* 'airplane' and one of *nana* 'food' in the study period, with 0 tokens of *nauk*, but they all produce more tokens of the endearing kin term *dada* 'sibling' than of these other nouns. This is clearly not because they never speak of water, airplanes, or food, as indicated through the token numbers for the adult counterparts to these nouns (Table 10). It appears that the kin baby-talk term *dada* either features more heavily and widely in use in CDS than the three inanimate nouns in Table 10, and/or its use persists in CDS even when parents have stopped using baby-talk terms for inanimate objects and substances. The prevalent use of baby-talk *dada* 'sibling' is likely related to the affective implications of its use to refer to a relationship between children (reminiscent of Schieffelin's discussion of the importance of a particular sibling relationship for Kaluli children; Schieffelin, 1990).

Finally, Tables 9 and 10 show that even parents who use baby-talk verbs and nouns relatively frequently, like TO's parents, do not use these in CDS to the exclusion of adult counterparts: TO's parents generally use a combination of baby-talk terms and the adult counterparts in a single session.

Lexicon Comparative Summary

The new data we presented in this section show that Ku Waru CDS and Nungon CDS involve the use of special baby-talk lexical items that are recognized as such by speakers, but that distribution of these items in child–caregiver interaction corpora is mixed. We further showed that: a) for Ku Waru, at least, there is high variability in the baby-talk terms listed by individual parents, and b) for Nungon, at least, variation in degree of use of particular baby-talk verbs and nouns is not an artefact of conversation topic.

Morphosyntax

Cross-linguistically, utterances in CDS are expected to feature fewer errors and to be shorter and simpler than in ADS, although fine-tuning of complexity and length to children's abilities is also attested (Snow, 1995; Soderstrom, 2007). Only rarely is CDS reported to involve caregiver production of morphological forms that would be unacceptable in ADS. For instance, an Inuktitut CDS sample involved a relatively small percentage of verbs (10%) and nouns (20%) that bore no affixes; in Inuktitut ADS, all verbs and nouns always bear affixes (Crago & Allen, 1997). Here, we first focus on evidence for and against morphosyntactic fine-tuning in Ku Waru and Nungon CDS, and then on morphological and syntactic structures that would be unacceptable in ADS but are used in CDS in these languages.

Fine-Tuning: Clause Chaining and Multi-Verb Predicates in CDS

In the first areas of morphosyntax considered here, clause chaining and multi-verb predicates, we consider whether there is evidence for fine-tuning of morphosyntactic complexity in Ku Waru and Nungon CDS to children as they develop.

Papuan languages like Ku Waru and Nungon are known for a special sentence type that is not present in the grammars of Western European languages: clause chains (Sarvasy, 2024, forthcoming; Sarvasy & Aikhenvald, 2024a, 2024b). In a clause chain, one or more clauses with verbal predicates that are under-specified for, usually, tense and often other grammatical categories, combine with a single clause, usually the last, of which the verbal predicate bears full specification for all the grammatical categories marked on the verbs of independent clauses in the language. A Ku Waru clause chain is provided in (3).

(3)

olyo	med	maket-ma-nga	**pu-p**	kalyip	baim	**te-p**
1PL	down.there	market-PL-GEN	go-NF.1	peanut	buy	do-NF.1

no-b	pilawa	**lyi-p**	**no-b**	**pu-mulu**
eat-NF.1	flour.balls	get-NF.1	eat-NF.1	go-FUT.1PL

'We'll go down to the markets and buy peanuts and eat them and get some flour balls and eat them and then we'll go.' [Lit.: 'As for us, going down to the markets, buying peanuts, eating (them), getting flour balls, eating (them), we will go (away again).'] (Rumsey et al., 2020, p. 4)

Example (3) is a clause chain sentence containing six clauses (in this clause chain, three of the clauses, *no-b*, *no-b*, and *pu-mulu*, comprise just a verbal predicate, with no explicit arguments). All of the clauses except the last one have a verbal predicate that lacks tense and subject number marking and bears only the first-person "non-final" verb suffix *-p/-b*. Only the last clause, *pu-mulu* 'we will go,' could function as an independent clause; *pu-mulu* bears inflection for future tense and for plural number as well as first person; these categories are understood to apply to the entire preceding clause chain.

In both Ku Waru and Nungon, clause chains contain a further fascinating feature: "switch-reference marking" (Haiman & Munro, 1983). With switch-reference marking, the non-final (also called "medial") verbal predicates in clause chains must be marked in a binary fashion according to whether the subject of the upcoming clause will be co-referential with the subject of the present clause, or not. This is illustrated in (4a) and (4b), two short Nungon clause chains addressed to TO, 2;7:

(4a)	o	babiya-ya	ngo	**to-nga**	babiya	bög-in
	CONJ	book-2SG.POSS	PROX	SG.O.take-MV.SS	book	house-LOC

	ong-i-roc-ma	ngo.
	go-IRR-2SG-RF	PROX

'And taking this book of yours, you will go to school, here.'
(Mother to TO, 2;7, activity: teasing)

(4b)	ogo	nungon	**na-una**	**aa-ha-rok**?
	SAME.LEVEL.NEAR	what	eat-DS.3SG	see-PRES.SG-2SG

'Over there, it having eaten what, do you see it?'
(Mother to TO, 2;7, activity: looking at a picturebook)

Both (4a) and (4b) are minimal clause chains of just two clauses: the non-final clause with verbal predicate that lacks tense marking, and the final clause with verbal predicate bearing full specification for tense and subject information. But the verbal predicate of the non-final clause in (4a), *to-nga*, is marked for "same-subject," since its (unstated) subject, 2SG, is shared by the following, final clause. In contrast, the verbal predicate of the non-final clause in (4b), *na-una,* is marked for "different-subject," since its (unstated) subject (a crocodile pictured in a book) differs from the subject of the following clause, 2SG.

We are beginning to know more about how clause chaining features in CDS in languages that have this sentence type (Sarvasy & Choi, 2020a, 2020b), and data from Ku Waru and Nungon have played a central role in this recent expansion in knowledge, along with Japanese (Clancy, 2020, 2024), Korean (Choi, 2020, forthcoming; Sarvasy & Choi, 2024), and Turkish (Aksu-Koç & Ögel-Balaban, 2020, 2024). Clause chains are robustly present in CDS, like discourse more generally, in clause chaining languages; in Korean and Nungon CDS, for instance, the non-final clause types used in clause chains are consistently more frequent than other non-independent clause types, like relative or complement clauses (Sarvasy & Choi, 2024).

Clause chains contain multiple verbs, and these are predicates of different clauses within the chains, hence clause chains are complex, multi-clause sentences. Another type of complex morphosyntactic structure involving multiple verbs that is common in languages of PNG involves a single clause with a complex predicate. This is the multi-verb predicate, which encompasses both the more narrowly defined “serial verb construction,” which cannot contain any markers of dependence on the verbs in it (Aikhenvald, 2018) and, more broadly, any predicate type that contains more than one verb (Sarvasy, 2021a). The acquisition of multi-verb predicates is under-studied (Sarvasy, 2021b), and highly language-specific, but there is some evidence from Nungon that children’s expansion of use of multi-verb predicates corresponds to an expansion in use of clause chains (Sarvasy, 2021). In the remainder of this section, we consider whether CDS in Ku Waru and Nungon shows evidence of age-related fine-tuning in terms of morphosyntactic complexity.

Clause Chaining in Ku Waru CDS

Rumsey et al. (2020) is a detailed study of early clause chains in Ku Waru children’s speech from age 1;8 to 4;11. In a corpus of 32,760 reviewed utterances from those samples, we found that two-clause chains first appear in the data from the four target children between the ages of 1;9 and 2;10. Philip produces his first two-clause chains at 2;07, and Jacklyn at 2;10. Across the samples including all four target children, three-clause chains first appear between 2;11 and 3;3; four-clause ones at 3;10 (from Philip, the only child in sample older than 3;08); and five-clause chains at 4;07 (from Philip). For the two children for whom we have data beyond the age of 3:00, Jacklyn and Philip, there is a turning point at around that age (2;10 for Jacklyn and 3;0 for Philip), after which there at are least some child-produced clause chains in every session; clause chains occur in 23% of Jacklyn’s verb-bearing utterances by 3;3 and in 19% of Philip’s by 3;6.

To investigate the extent of fine-tuning of clause chains in Ku Waru caregivers’ CDS with reference to that corpus, we counted clause chains in both child and adult speech and tabulated the incidence of clause chains of various lengths in the speech of caregivers to Philip at six ages between 2;5 and 4;9 and to Jacklyn at five ages between 2;8 and 3;4, and compared it to those of the children at the same sessions.

Table 11. ***Incidence and length of clause chains in samples from two Ku Waru speaking children and their caregivers***

Child	Age	Number of clause chains produced by child	Average length of clause chains produced by child	Longest clause chain produced by child	Number of clause chains addressed to child by caregivers	Average length of clause chains addressed to child by caregivers	Caregivers' longest clause chain
Philip	2;05	0	NA	NA	117	2.28	4
	2;11	0	NA	NA	98	2.27	4
	3;05	19	2.00	2	85	2.31	5
	3;10	49	2.20	4	96	2.36	5
	4;03	55	2.07	3	131	2.39	6
	4;09	96	2.17	4	133	2.33	6
Jacklyn	2;08	0	NA	NA	41	2.07	3
	2;11	21	2.10	3	64	2.16	4
	3;01	22	2.09	3	47	2.23	6
	3;02	18	2.00	2	32	2.19	4
	3;04	8	2.13	3	22	2.09	3

With respect to the numbers of clause chains per session, Table 11 shows little evidence for fine-tuning by the caregivers of either child. Philip's caregivers use large numbers of them in the 2;05 and 2;11 sessions, where Philip uses none. In the 3;05 and 3;10 sessions the caregivers use slightly fewer clause chains than they did at 2;11, while Philip's usage climbs markedly, from 19 to 49. Between 4;03 and 4;09 it increases by another 57%, while the caregivers' increases by only 1.2%. The average and maximum *lengths* of clause chains used by Philip's caregivers show a smoother rise across the sample period than do clause chain counts. In that respect, they resemble the rise in Philip's numbers of chains, but not the average and maximum lengths of his clause chains, which show more erratic developmental trajectories.

It is harder to find evidence of fine-tuning as related to clause chaining in CDS by Jacklyn's caregivers, probably because of the much smaller period for which data are available. Table 11 shows that Jacklyn's caregivers produce fewer clause chains per session than Philip's; we attribute this to two factors. First, the interactional style used by the caregivers in these sessions tends to involve firing simple, non-clause chain, questions at the child, and second, use of Tok Pisin (a non-clause chaining language) by both caregivers and the child is much higher in Jacklyn's sessions in general than in Philip's sessions (Rumsey, 2014); use of the non-clause chaining Tok Pisin could "prime" discourse participants to use fewer or shorter clause chains in their Ku Waru utterances.

Based on the same samples that are drawn on in the above discussion of frequency and length of clause chains, Rumsey et al. (2020, pp. 20–26) found little age-based structural simplification of clause chains in CDS in three other respects:
1) in the *kinds* of clauses used by the adults vs. children;
2) in the complexity of event structure within the clause chains;
3) in the incidence of a certain optional morphological simplification of medial verbs.

Note, however, that all of these findings pertain to interactions with children of 2;5 and above. Further work would be required in order to determine the extent to which they are valid for interactions with younger children.

Clause Chaining in Nungon CDS

Like Ku Waru, Nungon features clause chaining, but with two main differences: formal and distributional. First, Nungon medial verbs are formally simpler than Ku Waru medial verbs, in that they do not indicate temporal simultaneity or sequentiality. Second, in terms of switch-reference marking, Nungon discourse involves roughly equal proportions of same-subject (4a) and different-subject (4b) medial verbs, while Ku Waru discourse shows a strong preference for only same-subject medial verbs (as in example (3); Rumsey et al., 2020; Sarvasy, 2020). This distributional difference bears out in the proportions of same- and different-subject-marked medial verbs in Ku Waru and Nungon CDS, and is reflected in a major difference in the development of different-subject marking among children learning Ku Waru and Nungon: Ku Waru-speaking children experience a major delay before producing clause chains with different subjects, while Nungon-speaking children do not (Sarvasy, 2020; Sarvasy & Choi, 2020b, 2024).

For Nungon, the question of whether parents fine-tune their clause chaining to their children's development was investigated in a somewhat different way from Ku Waru (see previous section), in that clause chain length in Nungon CDS was not directly investigated. Sarvasy (2020) simply quantified the proportion of a speaker's total verbs per recording session that were medial (non-final) verbs. This was assessed for three children, Abraham, Niumen, and TO, and three of their parents: Niumen's mother, and TO's father and mother. A higher proportion of medial verbs suggests longer and/or more frequent clause chains. Results are shown in Figure 4.

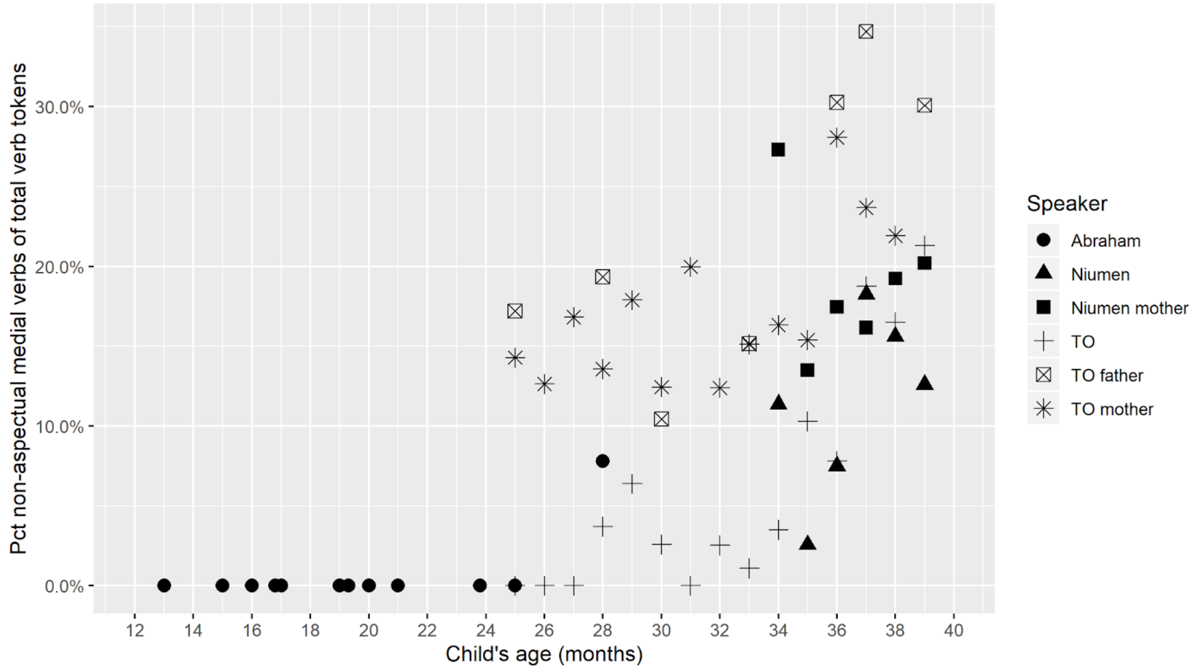

Figure 4. ***Percentage of total verb tokens per session that were medial verbs, per speaker, Nungon sessions (Sarvasy, 2020, p. 9).***

Figure 4 shows that from the earliest developmental stages for which we have Nungon CDS data—2;01 for TO's parents and 2;10 for Niumen's mother—medial verbs are present in the parents' CDS.

Both parents of TO show an increase in proportional use of medial verbs (relative to total verbs) from when TO is 36 months old (3;00), which is around the age (starting at 35 months) when TO herself begins to show a steeper increasing trend in medial verb proportions; this could be related to fine-tuning of parental speech. Similarly, Niumen's mother's proportional medial verb use shows an overall increasing trend from 2;11 through 3;3. (Niumen's session at 2;10 may be considered an outlier, as the first session of the longitudinal study for him and his mother, involving less speech overall from Niumen's mother than the other sessions, and involving the presence of additional interviewers, including Sarvasy, unlike other sessions.) Niumen's mother shows her highest proportion of medial verbs in the final session at 3;3. Although Niumen's medial verb proportions do not climb as steadily upward as TO's, his three highest medial verb proportions occur in the last three sessions of the study period.

In sum, Nungon parental proportional use of medial verbs appears to broadly keep pace with children's use, in a kind of fine-tuning. For TO, there is an extended period between 2;1 and 3;00 during which the proportional use of medial verbs by both TO and her parents stays relatively low and stable. We have no data on Niumen or his mother's productions before 2;10, but overall, both his and his mother's proportional medial verb use show generally increasing trends from 2;11 on. Parents thus appear to be tailoring their speech, in terms of the use of medial verbs, at least, to the perceived communicative abilities of their children. As seen in Sarvasy (2020), these two children show increases in clause chain numbers and number of clauses per chain by around 3;1, so it makes sense that their medial verb proportions increase around this age (since clause chains contain minimally one medial verb).

Multi-Verb Predicates in Nungon CDS

Nungon features multi-verb predicates (or multi-verb constructions, MVCs): serial verb-like constructions in which multiple verb roots, sometimes bearing dependent suffixes, combine in a single predicate. These differ from clause chains, in which medial verbs serve as predicates of separate clauses. (In Ku Waru, there is no morphological difference between medial verbs and the verbs used in MVCs; Rumsey et al., 2020; in Nungon, there is a clear morphological difference.)

Sarvasy (2021) examined MVCs in the speech of Niumen (called NN there) and TO, and their parents. There, the overall frequency of MVCs was assessed through the ratio of MVCs to total utterances. Results are in Figure 5.

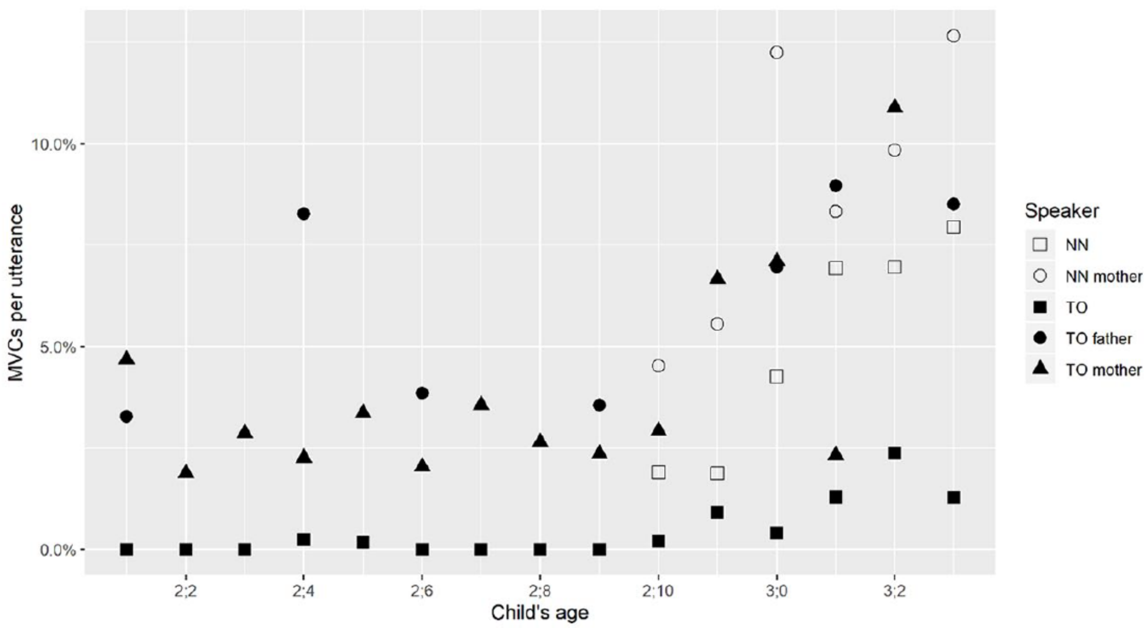

Figure 5. ***Multi-verb predicates per utterance, by speaker (Sarvasy, 2021, p. 494).***

Sarvasy (2021) interpreted Figure 5 to suggest a maintenance in parental MVC proportions while the children produce relatively few MVCs themselves (through about age 2;10), then a jump in adult use of MVCs around age 2;11 and continuing increasing trend thereafter. If true, this would fit with the finding above (interpretation of Figure 4, medial verb proportions results) that parents' clause chaining stays at a relatively steady level until about age three, when children's clause chain production accelerates; then there is a jump in parental clause chain production as well, and ongoing increasing trend.

Special Alterations to CDS

Although baby-talk lexicons are widespread, it is thought to be relatively rare for CDS to contain morphosyntactic structures that are unacceptable in ADS. But this is evident in two areas of Ku Waru and Nungon CDS: a special baby-talk form of the ergative case suffix in Ku Waru CDS, and a special morphosyntactic expansion of simplex sentences in Nungon CDS.

The Ku Waru CDS Ergative Marker Form

As in many of the world's languages (Dixon, 1979), Ku Waru grammar treats the subjects of transitive verbs differently from those of intransitive verbs. Transitive verb subjects receive an ergative case suffix *-ni/-n*, under certain circumstances (for the relevant conditioning factors, see Rumsey, 2010). This marking is seen in (5):

(5)	Abak-n	marasin	tu-ru-m
	man's.name-ERG	fertilizer	hit/do-RP-3SG
	'Abak pumped the fertilizer.'		

In a study based on two of the same longitudinal runs as for the clause chain study discussed above, Rumsey (in Rumsey et al., 2013) found that ergative marking was first attested from one of the children, Jesi, at 1;10 and at 2;4 for the other, Enita, but that there were very few instances of it until 2;5 for Jesi and 2;9 for Enita. Despite the children's sparse production of ergative marking early on, their caregivers invariably include ergative marking where appropriate from the earliest speech samples for the children (1;9 for Jesi and 1;8 for Enita), at about the same rate as in their ADS.

However, among the ergative-marked tokens in Ku Waru child speech and CDS are some that use a distinct baby-talk form of the ergative marker: *-na/-ne*. An example from CDS is (6), spoken by John Onga to his son Jesi at the age of 3;11.

(6)	Alan-nga	redi-na	ung	nyi-kim
	Alan-GEN	radio-ERG	words/talk	say-PPR-3SG
	'Alan's radio is talking.'			

If this were produced in ADS, the ergative marker on *redi* 'radio' would take the form *-n*. In our samples, Jesi and Enita use both the *-na/-ne* variant and the *-ni/-n* one, with a preponderance of the latter, from the earliest stage at which they used any ergative marking.[4] Their caregivers also use mainly the *-ni/-n* variant in their CDS, with occasional instances of *-na/-ne*, as in (6).

Nungon CDS Morphosyntactic Expansions

Nungon speakers have been shown to use a special morphosyntactic expansion construction in CDS to small children by which any lexical verb (e.g., 'die') can be used in nominalized form with the verb *to-* 'do' as auxiliary (e.g., 'do dying'), in a way that is never done by adults in ADS. This "nominalization plus do auxiliary" construction was discussed initially for just TO's parents in Sarvasy (2019) and then established as a community-wide practice in Sarvasy (2023b). It is illustrated in (7):

(7)	Father:	Obö-k	to-wangka-k;	na-mo-k	to-i.
		break-NMZ	do-NFUT.SG-3SG	1SG.O-give-NMZ	do-IMP.2SG
		'It will do breaking; do giving it to me.' (AB 1;5; Sarvasy, 2023b)			

In (7), the verbs *obö-* 'break' and *na-mo-* 'give me' are nominalized through the addition of an uninflecting *-k* to the verb root; they form a complex predicate with the verb *to-* 'do,' which carries the tense, mood, and subject person/number inflections for the whole predicate. In standard Nungon ADS, this would never be expressed in this way; rather, the two lexical verbs would inflect directly, as:

(8)	Obö-wangka-k;	na-mo-hi.
	break-NFUT.SG-3SG	1SG.O-give-IMP.2SG
	'It will break; give it to me.'	

The special CDS nominalization with 'do' auxiliary construction appears to be heavily present in the CDS of some parents to children in their second year of life, and is largely phased out in CDS by about 3;3. For instance, one parent uses this more than normal inflected verbs when his child is 1;5, but by the time that child is 2;3, the father uses almost all normal inflected verbs (Sarvasy, 2023b).

Some children may seize on the nominalized verb forms used in this construction and then use them as an optional "default" verb form; this was the case for TO (described in Sarvasy, 2019), but not for three other children studied in the same age range (Sarvasy, 2023b).

[4] For the precise figures see Rumsey et al. (2013, p. 157).

Nungon verbs have dozens of inflectional forms, depending on tense (there are five grammatical tenses), mood, modality, and subject person/number (including singular, dual, and plural number values). Beyond this, verbs belong to seven different classes, based on the forms of the inflectional suffixes used with them. Class membership is lexically determined: it cannot be predicted on the basis of phonological characteristics of the verb roots. This makes for much morphological complexity for the child to master.

When a verb is nominalized, this simplifies things in two ways: first, some verb classes share a nominalizing suffix, such that there are fewer nominalizing suffixes than verb classes (and consonant-final verb roots do not receive any suffix to function as nominalized). Second, nominalized verbs do not inflect for tense, mood, modality, or subject person/number; they bear an unchanging nominalizing suffix. The Nungon CDS nominalization plus 'do' auxiliary thus could be argued to scaffold child learning by presenting children with just one verb's set of inflections to learn (the verb *to-* 'do') first, while other verbs are packaged as simple nominalizations. But all parents observed thus far use this construction alongside directly inflected verbs; the questions of how and when they choose to use the nominalization plus 'do' auxiliary construction, and at what point their speech illustrates the full range of inflectional possibilities for all verb classes, remain open for further study.

Morphosyntax Comparative Summary

We found some evidence of fine-tuning of morphosyntactic complexity in Ku Waru and Nungon CDS. For instance, in Ku Waru, there is a slight increase in overall clause chain length in CDS to older children, and the proportion of all clauses that are chained is greater in ADS than in CDS in general, suggesting that at some point adults begin to address older children using more clause chains. In Nungon, adults begin to use more chained clauses and more multi-verb predicates in speech to children from about the age of 3;0, which is also when Nungon-speaking children begin to use longer and more complex clause chains, and more multi-verb predicates.

Both languages have a morphosyntactic CDS feature that is unacceptable in ADS. In Ku Waru CDS, there is a special baby-talk form of the ergative suffix, while in Nungon CDS, parents optionally expand simplex predicates into more morphosyntactically complex auxiliary constructions.

Discussion

We organize our discussion here according to the four areas of language presented above.

Prosody

Our prosodic analysis of CDS in Ku Waru and Nungon compared the mean pitch and the pitch range with the ADS equivalents in both languages. The Nungon data involved children aged 2;2–3;10, while the children in the Ku Waru dataset were aged 2;4–3;3 at the time of recording. In the Nungon dataset, and for one speaker of Ku Waru (addressing a child aged 2;4), the mean pitch of CDS was found to be significantly higher than that of ADS. These results are in line with the widely reported cross-linguistic tendency for higher pitch in CDS, reported without measurements for Huli CDS (Goldman, 1986) and attested but without statistical significance for Qaqet CDS (Frye, 2022). While two Ku Waru speakers showed no significant mean pitch difference between ADS and CDS in the analyzed dataset, their highest-pitched utterances all occurred in CDS, not ADS. For Kaluli, Schieffelin (1990, pp. 102–106, 108–110) remarks on the use of higher pitch and "exaggerated intonation and descent in melody" in playful interaction between an older sibling and a younger one, and in "teasing" between a mother and child, but does not explicitly link this to the child-directed nature of the interactions.

We found a slight difference between Ku Waru and Nungon pitch range in CDS *vs.* ADS. In Nungon, CDS pitch range was significantly greater than that of ADS across sexes—a result that aligns with general CDS expectations (Soderstrom, 2007) and also mirrors the findings reported for Qaqet (Frye, 2022). Ku Waru, on the other hand, showed no significant difference in pitch range between ADS and CDS in the dataset. As with mean pitch, however, the Ku Waru sampled utterances with the greatest pitch ranges were all CDS, not ADS.

Phonology

In both Ku Waru and Nungon, adults optionally modify consonants in CDS. In both languages, these modifications result in productions that resemble early child productions. It seems clear that the adult modifications mimic child productions, rather than the other way around. For Ku Waru, for instance, children consistently have difficulty producing palatal nasals and laterals and the velar lateral phoneme until about age 2;6 or older (much older for the velar lateral), and so CDS optionally involves the replacement of these sounds with their alveolar equivalents, as occurs in children's speech. Children learning Ku Waru also produce an alveolar /l/ or /t/ instead of /r/ and sometimes replace velar stops with alveolar ones, as adults occasionally do in Ku Waru CDS. In Nungon CDS, there are fewer documented consonant alterations than in Ku Waru: the prominent ones discussed in this paper are the replacement of /r/ with /l/ or the glide /y/, and the replacement of word-final /k/ with a glottal stop. It is noteworthy that CDS in both Ku Waru and Nungon involve the production of consonant segments that are not part of the adult phoneme inventory yet are associated with early child speech production. These are the alveolar /l/ in Ku Waru (except in words derived from Tok Pisin); and the /l/, glottal stop, and (word-initial) voiced velar fricative in the Towet dialect of Nungon.

We can draw a distinction between the phenomenon of optional systematic consonant alteration, and that of extra-phonemic sounds that obligatorily occur in just a particular baby-talk lexical item. The first type covers all the alterations described for Ku Waru CDS (and CS), and the general replacement of /r/ and word-final /k/ in Towet Nungon CDS. Such generalized replacements are also reported for CDS in the Papuan language Duna (Duna-Bogaia; San Roque & Schieffelin, forthcoming). In contrast, just the Towet Nungon baby-talk word for 'pitpit' begins with a voiced velar fricative, which is not a true phoneme of adult Nungon. This is not a widespread feature of Nungon CDS: it is confined to this baby-talk lexical item. Similarly, just one baby-talk lexical item in the East Sepik language Manambu begins with a sound that is extra-phonemic for adults: a bilabial trill (Aikhenvald, 2008, p. 44).

We have not yet studied the acoustics of vowels in Ku Waru ADS or CDS. The Nungon CDS vowel triangle has an equivalent size to that of Nungon ADS in adult dyadic conversation, so there is no evidence of vowel hyper-articulation in Nungon CDS. But vowel durations are overall longer in Nungon CDS to 2-year-olds than 3-year-olds, as is found in CDS in many other languages (Soderstrom, 2007).

Lexicon

Ku Waru and Nungon speakers have acknowledged widely used sets of "baby-talk" lexical items, like speakers of the Papuan languages Huli (Goldman, 1986), Mali (Baining; Stebbins, 2011, pp. 28–29), and Manambu (Aikhenvald, 2008, pp. 44, 138). Ku Waru and Nungon baby-talk terms include some with transparent origins, arising through: (a) modification of adult forms (as in only a small minority of Huli baby-talk forms; Goldman 1986, pp. 199–200), (b) onomatopoeia, or (c) borrowing from Tok Pisin (which, as in Manambu, is seen as an "easy" language for children; Aikhenvald, 2008, pp. 615, 617, possibly boding ill for local language maintenance). Multiple forms in both Ku Waru and Nungon also exhibit "full lexical replacement" (like most Huli baby-talk words; Goldman, 1986, p. 199), showing no relation to the counterpart adult form, or having no single-word adult counterpart. We have no ready explanation for why the two languages, possibly related only in the deep past, share a baby-talk term for 'kiss': *nunu*.

We presented new data on baby-talk lexical items for both Ku Waru and Nungon that go beyond general word lists to examine corpus distributions and individual variation. For Ku Waru, the 11 adults consulted by John Onga agreed that Ku Waru speakers use baby-talk lexical items, but when asked to list these, their lists diverged to a large degree. Further, the most commonly cited baby-talk words were not necessarily the ones used most commonly in CDS in the Ku Waru child–caregiver speech corpus.

For Nungon, child–caregiver pairings differed in the frequencies of baby-talk lexical items for the age range 2;4–2;7. For both baby-talk nouns and verbs, we used comparative counts of adult and baby-talk lexical counterparts to show that such differences were not an artefact of conversation topic. Further, caregivers who did use baby-talk terms tended to use both the baby-talk terms and their adult counterparts in CDS in the same sessions.

Our findings on the existence, local awareness of, and widespread use of baby-talk lexical items in Ku Waru and Nungon CDS diverge from what has been reported for Kaluli (Schieffelin, 1990) and Qaqet (where adults interviewed in Tok Pisin provided just three examples of Qaqet baby-talk lexical items, despite the more robust use of baby-talk lexical items in closely-related Mali; Frye, 2022, pp. 149–150; Stebbins, 2011, pp. 28–29). But the evidence given against baby-talk lexicons for Kaluli seems to stem mainly from adult responses to questioning, while for Qaqet, Frye relies on the occurrence of baby-talk forms in retellings of the silent "Pear Story" video (Chafe, 1980), rather than in naturalistic corpus data. Cross-linguistically, Pear Story retellings are not known for containing baby-talk terms or being emblematic of CDS styles, so it is possible that the failure to produce baby-talk terms here is an artefact of the study design.

Morphosyntax

Our discussion of Ku Waru and Nungon morphosyntax focused on two areas: fine-tuning of morphosyntactic complexity to children's developmental stages, and non-ADS-like structures. The aspect of morphosyntax that we discussed for both languages was clause chaining—a type of complex sentence with one or more non-embedded dependent clauses that is widespread in Papuan languages. Ku Waru and Nungon CDS show similarities and differences in clause chaining. In both, clause chains are present in CDS from the earliest sampled children's ages, and in the children's speech between 2;0 and 3;0, albeit in much lower proportions. In both Ku Waru and Nungon there is a turning point at around 3;0, after which the children's clause chains become more frequent. But ADS differs across the two studies in that in Nungon the children's turning point coincides with an increase in the proportion of clause chains in the parents' ADS, which has the effect of keeping it higher than the children's, whereas in Ku Waru CDS the proportion of clause chains is already high by 2;5 and does not show a noticeable increase thereafter. Nungon multi-verb constructions (MVCs) in child speech and CDS show a similar pattern, with increases in both from about 2;11.

The apparent Nungon fine-tuning in number of clause chains in CDS versus the lesser degree of it in Ku Waru CDS is consistent with another aspect of our findings concerning Ku Waru CDS after 2;5: its low degree of child-age-based structural simplification of CDS with respect to clause type, event structure, and morphology.

The last two areas of morphosyntax that we examined show a further contrast between Ku Waru and Nungon CDS, in the extent to which CDS utterances comprise acceptable ADS-like utterances. Ku Waru adults consistently use well-formed, adult-like utterances with regard to ergative marking: there is no indication that they sometimes omit ergative marking, even in CDS to children who do not yet use ergative marking reliably. But they do sometimes use a distinct baby-talk form of the ergative marker in the same environments where they use the adult form in their ADS. Nungon adults do something that is cross-linguistically unusual (or at least poorly attested): (a) they optionally produce longer and more syntactically complex utterances in CDS than in the ADS counterparts, bucking expectations that CDS utilize shorter and syntactically simpler sentences than ADS (Soderstrom, 2007), and (b) further, these expanded CDS utterances, while not strictly "ungrammatical," are nonsensical and unattested in ADS. Here, Nungon CDS appears to sacrifice brevity and syntactic simplicity for the sake of morphological simplicity.

Conclusion

We have shown that speech directed to toddlers and preschool-aged children in Ku Waru and Nungon can have special qualities, different from those of conversational ADS, in multiple linguistic domains. It is important to remember that the features we describe have variable distributions within and across speakers at any given time point. Overall, the possible modifications available for CDS in both languages constitute less a coherent "register" that speakers may slip into or out of, but more a menu of optional features, some apparently binary and some measured in terms of degree, which may be applied in conjunction with each other or separately, and which adults often apply variably within a single recording session. Our data thus confirm earlier researchers' doubts about the notion of "register" (see references in Weinstein & Baldwin, 2024) while confirming the use of adjustments in CDS for two Papuan languages.

References

Aikhenvald, A. Y. (2008). *The Manambu language of East Sepik, Papua New Guinea*. Oxford University Press. https://doi.org/10.1093/oso/9780199539819.001.0001

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *The Journal of the Acoustical Society of America, 119*(3), 1636–1647. https://doi.org/10.1121/1.2161443

Audibert, N., & Falk, S. (2018). Vowel space and f0 characteristics of infant-directed singing and speech. In *Proceedings of the 19th International Conference on Speech Prosody* (pp. 153–157). https://doi.org/10.21437/SpeechProsody.2018-31

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics, 5*(4), 339–350. https://doi.org/10.1016/S0095-4470(19)31204-5

Benders, T., Pokharel, S., & Demuth, K. (2019). Hypo-articulation of the four-way voicing contrast in Nepali infant-directed speech. *Language Learning and Development, 15*(3), 232–254. https://doi.org/10.1080/15475441.2019.1577139

Benders, T., StGeorge, J., & Fletcher, R. (2021). Infant-directed speech by Dutch fathers: Increased pitch variability within and across utterances. *Language Learning and Development*, *17*(3), 292–325. https://doi.org/10.1080/15475441.2021.1876698

Bernstein Ratner, N., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language, 11*(3), 515–522. https://doi.org/10.1017/S0305000900005924

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C. F., Ramírez-Esparza, N., R. Hamrick, L., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., Benetti, L., Alphen, P. V., & Cristia, A. (2023). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences, 120*(52), Article e2300671120. https://doi.org/10.1073/pnas.2300671120

Boer, J., Claessen, M., & Williams, C. (2022). Acquisition of Tok Pisin phonology in the multilingual highlands of Papua New Guinea. *International Journal of Speech-Language Pathology, 24*(3), 283–293. https://doi.org/10.1080/17549507.2022.2080271

Boersma, P. & Weenink, D. (2025). *Praat* (Version 6.4.27) [Computer software]. http://www.praat.org/

Bohn, O. S. (2013). Acoustic characteristics of Danish infant directed speech. In *Proceedings of Meetings on Acoustics ICA2013, 19*(1), Article 060055. https://doi.org/10.1121/1.4798488

Buchan, H. (2012). *Phonetic variation in Gurindji Kriol and northern Australian English: A longitudinal study of fricatives in maternal speech* [Doctoral dissertation]. University of Wollongong. http://ro.uow.edu.au/theses/3789

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435–1435. https://doi.org/10.1126/science.1069587

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language, 48*(4), 792–814. https://doi.org/10.1017/S0305000920000549

Chafe, W. L. (Ed.). (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Praeger.

Cheng, M.-C. (2014). Exploration of the phonetic difference in stops between Hakka infant-directed speech and adult-directed speech. *Concentric: Studies in Linguistics 40*(1), 1–35. https://doi.org/10.6241/concentric.ling.40.1.01

Choi, S. (2020). Development of clause chaining in Korean. *Frontiers in Psychology, 11*, 256. https://doi.org/10.3389/fpsyg.2020.00256

Clancy, P. M. (2020). To link or not to link: Clause chaining in Japanese narratives. *Frontiers in Psychology, 10*, Article 3008. https://doi.org/10.3389/fpsyg.2019.03008

Clancy, P. M. (2024). Clause chains and intonation units in Japanese narratives. In H. S. Sarvasy & A. Y. Aikhenvald (Eds.), *Clause chaining in the languages of the world* (pp. 702–736). Oxford University Press. https://doi.org/10.1093/oso/9780198870319.003.0026

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

Clark, E. (2016). *First language acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9781316534175

Crago, M. B., & Allen, S. E. M. (199). Linguistic and cultural aspects of simplicity and complexity in Inuktitut (Eskimo) child-directed speech. In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 91–102). Cascadilla Press.

Cristia, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *The Journal of the Acoustical Society of America, 128*(1), 424–434. https://doi.org/10.1121/1.3436529

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass, 7*(3), 157–170.

https://doi.org/10.1111/lnc3.12015

Cristia, A., & Seidl, A. (2014). The hyper-articulation hypothesis of infant-directed speech. *Journal of Child Language, 41*(4), 913–934. https://doi.org/10.1017/S0305000912000669

Cruttenden, A. (1994). Phonetic and prosodic aspects of baby talk. In B. J. Richards & C. Gallaway (Eds.), *Input and interaction in language acquisition* (pp. 135–152). Cambridge University Press. https://doi.org/10.1017/CBO9780511620690.008

Dilley, L. C., Millett, A. L., McAuley, J. D., & Bergeson, T. R. (2014). Phonetic variation in consonants in infant-directed and adult-directed speech: The case of regressive place assimilation in word-final alveolar stops. *Journal of Child Language, 41*(1), 155–175. https://doi.org/10.1017/S0305000912000670

Dixon, R. M. W. (1979). Ergativity. *Language 55*(1), 59–138. https://doi.org/10.2307/412519

Duranti, A. (2009). Linguistic anthropology: History, ideas, and issues. In A. Duranti (Ed.), *Linguistic anthropology: A reader* (2nd ed., pp. 1–59). Wiley-Blackwell.

Englund, K., & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development, 15*(2), 139–160. https://doi.org/10.1002/icd.445

Fais, L., Kajikawa, S., Amano, S., & Werker, J. F. (2010). Now you hear it, now you don't: Vowel devoicing in Japanese infant-directed speech. *Journal of Child Language, 37*(2), 319–340. https://doi.org/10.1017/S0305000909009556

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66*, 103–114. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00060

Ferguson, C. A. (1977). Baby talk as a simplified register. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 209–235). Cambridge University Press.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 60(6), 1497–1510. https://doi.org/10.1111/j.1467-8624.1989.tb04020.x

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Foley, W. A. (1986). *The Papuan languages of New Guinea*. Cambridge University Press. https://doi.org/10.1017/S002222670001166X

Frye, H. (2022). *Child-directed speech in Qaqet: A language of east New Britain, Papua New Guinea*. ANU Press. https://doi.org/10.22459/CDSQ.2022

García-Sierra, A., Ramírez-Esparza, N., Wig, N., & Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: Testing neural commitment using the positive-MMR. *Brain and Language, 212*, Article 104890. https://doi.org/10.1016/j.bandl.2020.104890

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 63–88). Cambridge University Press.

Gergely, A., Faragó, T., Galambos, Á., & Topál, J. (2017). Differential effects of speech situations on mothers' and fathers' infant-directed and dog-directed speech: An acoustic analysis. *Scientific Reports, 7*(1), Article 13739. https://doi.org/10.1038/s41598-017-13883-2

Goldman, L. R. (1986). Anatomical terms in Huli names and games. *Mankind, 16*(3), 190–208. https://doi.org/10.1111/j.1835-9310.1986.tb00734.x

Goldman, L. R. (1987). Ethnographic interpretations of parent-child discourse in Huli. *Journal of Child Language, 14*(3), 447–466. https://doi.org/10.1017/S0305000900010230

Goldman, L. R. (1991). Language acquisition - Bambi B. Schieffelin, The give and take of everyday life: Language socialization of Kaluli children. Cambridge University Press, 1990. Pp. viii+ 278 [Review]. *Language in Society, 20*(4), 663–668. https://doi.org/10.1017/S0047404500016821

Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research, 53*(6), 1529–1542. https://doi.org/10.1044/1092-4388(2010/09-0005)

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24*(1), 14–20. https://doi.org/10.1037/0012-1649.24.1.14

Haynes, L. & Cooper, R. (1986). A note on Ferguson's proposed baby-talk universals.

In J. Fishman (Ed.), *The Fergusonian impact: In honor of Charles A. Ferguson on the occasion of his 65th birthday. Volume 1: From phonology to society* (pp. 127–134). De Gruyter Mouton. https://doi.org/10.1515/9783110873641-011

Hellwig, B., Sarvasy, H. S., & Casillas, M. (forthcoming). Language acquisition. In N. Evans & S. Fedden (Eds.), *Oxford guide to the Papuan languages*. Oxford University Press.

Jeannin, S., Gilbert, C., Amy, M., & Leboucher, G. (2017). Pet-directed speech draws adult dogs' attention more efficiently than adult-directed speech. *Scientific Reports, 7*(1), Article 4980. https://doi.org/10.1038/s41598-017-04671-z

Johnson, A., Elliot, M., & Allen, S. E. (2023). Morphological simplification in Inuktitut child-directed speech. In *Proceedings of the 25th Workshop on Structure and Constituency in Languages of the Americas* (pp. 103–115). UBC Working Papers in Linguistics.

Katz, G. S., Cohn, J. F., & Moore, C. A. (1996). A combination of vocal f0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development, 67*(1), 205–217. https://doi.org/10.1111/j.1467-8624.1996.tb01729.x

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

Kik, A., Adamec, M., Aikhenvald, A. Y., Bajzekova, J., Baro, N., Bowern, C., Colwell, R. K., Drozd, P., Duda, P., Ibalim, S., Jorge, L. R., Mogina, J., Ruli, B., Sam, K., Sarvasy, H. S., Saulei, S., Weiblen, G. D., Zrzavy, J., & Novotny, V. (2021). Language and ethnobiological skills decline precipitously in Papua New Guinea, the world's most linguistically diverse nation. *Proceedings of the National Academy of Sciences, 118*(22), Article e2100096118. https://doi.org/10.1073/pnas.2100096118

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development, 24*(4), 372–392. https://doi.org/10.1016/S0163-6383(02)00086-3

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sunderg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kulick, D. (1980). Gapun/Tayap recordings. Collection DK1 at catalog.paradisec.org.au. https://doi.org/10.4225/72/56E979D6D23EA

Kulick, D. (1992). *Language shift and cultural reproduction: Socialization, self, and syncretism in a Papua New Guinea village*. Cambridge University Press.

Kulick, D., & Terrell, A. (2019). *A grammar and dictionary of Tayap: The life and death of a Papuan language*. De Gruyter. https://doi.org/10.1515/9781501512209

Laughren, M. (1984). Warlpiri baby talk. *Australian Journal of Linguistics 4*(1), 73–88. https://doi.org/10.1080/07268608408599321

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skill. *Developmental Science, 6*, F1–F10. https://doi.org/10.1111/1467-7687.00275

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study. *Journal of Child Language, 36*(4), 909–922. https://doi.org/10.1017/S030500090800929X

Longacre, R. E. (2007). Sentences as combinations of clauses. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 372–420). Cambridge University Press. https://doi.org/10.1017/CBO9780511619434.007

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Lawrence Erlbaum Associates.

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.) *Child Phonology, Volume 2* (pp. 173–184). Academic Press.

Max Planck Institute for Psycholinguistics. (2024). *ELAN* (Version 6.9) [Computer software]. https://archive.mpi.nl/tla/elan

Mazuka, R., Kondo, T., & Hayashi, A. (2008). Japanese mothers' use of specialized vocabulary in infant-directed speech: infant-directed vocabulary in Japanese. In N. Masataka (Ed.), *The origins of language: Unraveling evolutionary forces* (pp. 39–58). Springer. https://doi.org/10.1007/978-4-431-79102-7_4

McAllister Byun, T. (2012). Positional velar fronting: An updated articulatory account. *Journal of Child Language 39*(5), 1043–1076.

https://doi.org/10.1017/S0305000911000468

Merlan, F., & Rumsey, A. (1991). *Ku Waru: Language and segmentary politics in the western Nebilyer Valley*. Cambridge University Press. https://doi.org/10.1017/CBO9780511518218

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In R. Shweder & R. Levine (Eds.), *Culture theory: Essays on mind, self and emotion* (pp. 276–320). Cambridge University Press.

Ota, M., & Skarabela, B. (2016). Reduplicated words are easier to learn. *Language Learning and Development, 12*(4), 380–397. https://doi.org/10.1080/15475441.2016.1165100

Ota, M., & Skarabela, B. (2018). Reduplication facilitates early word segmentation. *Journal of Child Language, 45*(1), 204–218. https://doi.org/10.1017/S0305000916000660

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why choo-choo is better than train: The role of register-specific words in early vocabulary growth. *Cognitive Science, 42*(6), 1974–1999. https://doi.org/10.1111/cogs.12628

Posit team. (2024). *RStudio: Integrated development environment for R*. [Computer software]. http://www.posit.co/

Réger, Z., & Gleason, J. B. (1991). Romāni child-directed speech and children's language among gypsies in Hungary. *Language in Society 20*(4), 601–617. https://doi.org/10.1017/S0047404500016742

Rosenberg, K. R., Golinkoff, R. M., & Zosh, J. M. (2004). Did australopithecines (or early Homo) sling? *Behavioral and Brain Sciences, 27*(4), 522–522. https://doi.org/10.1017/S0140525X04430118

Rowe, M. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*, 185–205. https://doi.org/10.1017/S0305000907008343

Rumsey, A. (2014). Bilingual language learning and the translation of worlds in the New Guinea Highlands and beyond. *HAU: Journal of Ethnographic Theory 4*(2), 119–140. https://doi.org/10.14318/hau4.2.006

Rumsey, A. (2017). Dependency and relative determination in children's language acquisition: The case of Ku Waru. In N. J. Enfield (Ed.), *Dependencies in language: On* *the causal ontology of linguistic systems* (pp. 97–114). Language Science Press.

Rumsey, A., Reed, L., & Merlan, F. (2020). Ku Waru clause chaining and the acquisition of complex syntax. *Frontiers in Communication, 5,* Article 19. https://doi.org/10.3389/fcomm.2020.00019

Rumsey, A., San Roque, L., & Schieffelin, B. (2013). The acquisition of ergative marking in Kaluli, Ku Waru and Duna (Trans New Guinea). In E. L. Bavin & S. Stoll (Eds.), *The acquisition of ergativity* (pp. 133–182). John Benjamins. https://doi.org/10.1075/tilar.9.06rum

Rumsey, A., Merlan, F., Onga, J., Noma, A., Kuri, S., Reed, L. W., Devitt, D., Gonzalez Ochoa, S., van Tongeren, C., Honeyman, T., Yam, S., Ross, M. A., Kalyan, S., Peck, N., Kashima, E., Hendy, C., Cutfield, S., Strong, S., Creaghe, N., & Wang, J. (2024). *Ku Waru Child Language Socialization Study*. https://sites.google.com/view/kwclss/home

San Roque, L. (2008). *An introduction to Duna grammar* [Doctoral dissertation]. Australian National University.

San Roque, L. (2016). 'Where' questions and their responses in Duna (Papua New Guinea). *Open Linguistics 2*(1), 85–104. https://doi.org/10.1515/opli-2016-0005

San Roque, L., & Schieffelin, B. (forthcoming). Language socialisation in the Papuan context. In N. Evans & S. Fedden (Eds.), *Oxford guide to the Papuan languages*. Oxford University Press.

Sarvasy, H. S. (2017). *A grammar of Nungon: A Papuan language of northeast New Guinea*. Brill. https://doi.org/10.1163/9789004340107

Sarvasy, H. S. (2019). The root nominal stage: a case study of early Nungon verbs. *Journal of Child Language 46*(6), 1073–1101. https://doi.org/10.1017/S0305000919000357

Sarvasy, H. S. (2020). Acquisition of clause chaining in Nungon. *Frontiers in Psychology 11*, Article 1456. https://doi.org/10.3389/fpsyg.2020.01456

Sarvasy, H. S. (2021). Acquisition of multi-verb predicates in Nungon. *First Language, 41*(4). https://doi.org/10.1177/0142723720938568

Sarvasy, H. S. (2022). The near future tense in child and child-directed Nungon speech: A case study. In A. Storch & R. M. W. Dixon (Eds.), *The art of language* (pp.

275–287). Brill. https://doi.org/10.1163/9789004510395_018

Sarvasy, H. S. (2023a). Nungon child speech corpus. Western Sydney University Research Management Repository. https://doi.org/10.26183/essm-1v90

Sarvasy, H. S. (2023b). Early verb production in Nungon. *Frontiers in Psychology 14*, Article 1241447. https://doi.org/10.3389/fpsyg.2023.1241447

Sarvasy, H. S. (2023c). Verbatim narrative prompting in Nungon. In A. Y. Aikhenvald, R. Bradshaw, L. Ciucci, & P. Wangdi (Eds.), *Celebrating Indigenous voice* (pp. 121–141). De Gruyter. https://doi.org/10.1515/9783110789836-006

Sarvasy, H. S. (2024). Clause chains in Finisterre Papuan languages. In H. S. Sarvasy & A. Y. Aikhenvald (Eds.), *Clause chaining in the languages of the world* (pp. 187–229). Oxford University Press. https://doi.org/10.1093/oso/9780198870319.003.0007

Sarvasy, H. S. (forthcoming). Clause chains in Papuan languages. In N. Evans & S. Fedden (Eds.), *Oxford guide to the Papuan languages*. Oxford University Press.

Sarvasy, H. S., & Choi, S. (2020a). Beyond the two-clause sentence: Acquisition of clause chaining in six languages. *Frontiers in Psychology 11*, Article 1586. https://doi.org/10.3389/fpsyg.2020.01586

Sarvasy, H. S., & Choi, S. (Eds.). (2020b). *Acquisition of clause chaining*. Frontiers Media SA. https://doi.org/10.3389/978-2-88966-291-3

Sarvasy, H. S., & Choi, S. (2024). Acquisition of clause chaining. In H. S. Sarvasy & A. Y. Aikhenvald (Eds.), *Clause chaining in the languages of the world* (pp. 60–96). Oxford University Press. https://doi.org/10.1093/oso/9780198870319.003.0003

Sarvasy, H. S., Elvin, J., Li, W., & Escudero, P. (2019). An acoustic analysis of Nungon vowels in child-versus adult-directed speech. In *Proceedings of the 19th International Congress of Phonetic Sciences Melbourne* (pp. 3155-3159).

Sarvasy, H. S., Elvin, J., Li, W., & Escudero, P. (2020). An acoustic phonetic description of Nungon vowels. *Journal of the Acoustical Society of America 147*(4), 2891–2900. https://doi.org/10.1121/10.0001003

Sarvasy, H. S., Li, W., Elvin, J., & Escudero, P. (2022). Vowel acoustics of Nungon child-directed speech, adult dyadic conversation, and foreigner-directed monologues. *Frontiers in Psychology 13*, Article 805447. https://doi.org/10.3389/fpsyg.2022.805447

Schieffelin, B. B. (1979). Getting it together: An ethnographic approach to the study of the development of communicative competence. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 73–108). Academic Press.

Schieffelin, B. B. (1985). The acquisition of Kaluli. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Volume 1* (pp. 525–593). Erlbaum. https://doi.org/10.4324/9781315802541-6

Schieffelin, B. B. (1990). *The give and take of everyday life: Language socialization of Kaluli children*. Cambridge University Press.

Snow, C. E. (1995). Issues in the study of input: Fine-tuning, universality, individual and developmental differences and necessary causes. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 180–193). Basil Blackwell. https://doi.org/10.1111/b.9780631203124.1996.00007.x

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Stebbins, T. (2011). *Mali (Baining) grammar*. The Australian National University.

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language, 10*(1), 1–15. https://doi.org/10.1017/S0305000900005092

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica, 56*(3–4), 186–199. https://doi.org/10.1159/000028450

Tomić, D., & Mildner, V. (2015). Development of /r/ in Croatian. In *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 0716.1–0716.5). The University of Glasgow.

Turpin, M., Demuth, K., & Campbell, A. N. (2014). Phonological aspects of Arandic baby talk. In R. Pensalfini, M. Turpin, & D. Guillemin (Eds.), *Language description informed by theory* (Studies in Language Companion Series, Vol. 147, pp. 49–80). John Benjamins. https://doi.org/10.1075/slcs.147.04tur

Wang, Y., Seidl, A., & Cristia, A. (2015). Acoustic-phonetic differences between infant-and adult-directed speech: The role of stress and utterance position. *Journal of Child Language, 42*(4), 821–842. https://doi.org/10.1017/S0305000914000439

Wang, L., Kalashnikova, M., Kager, R., Lai, R., & Wong, P. C. M. (2021). Lexical and prosodic pitch modifications in Cantonese infant-directed speech. *Journal of Child Language, 48*(6), 1235–1261. https://doi.org/10.1017/S0305000920000707

Wegmann, U. (1994). *Yau-Uruwa dialect survey* [Unpublished manuscript]. Summer Institute of Linguistics.

Weinstein, N., & Baldwin, D. (2024). Reification of infant-directed speech? Exploring assumptions shaping infant-directed speech research. *Culture & Psychology, 30*(1), 216–242. https://doi.org/10.1177/1354067X221147683

Woolard, A., Benders, T., Campbell, L. E., Whalen, O. M., Mallise, C., Karayanidis, F., Barker, D., Murphy, V. E., Tait, J., Gibson, P., Korostenski, L., & Lane, A. E. (2023). The relationship between pitch contours in infant-directed speech and early signs of autism in infancy. *Infant Behavior and Development, 72*, Article 101860. https://doi.org/10.1016/j.infbeh.2023.101860

Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog- and infant-directed speech. *Anthrozoös, 26*(3), 373–380. https://doi.org/10.2752/175303713X13697429463592

Abbreviations

1, 2, 3	first, second, third person
ADS	adult-directed speech
BT	baby-talk
CDS	child-directed speech
CONJ	conjunction
CS	child speech
DEF	definite
DS	different-subject
DU	dual
ERG	ergative
FUT	future
GEN	genitive
IMP	immediate imperative
IRR	irrealis
LOC	locative
MV	medial verb form
NA	not applicable
NF	non-final
NFUT	near future

NMZ	nominalizer
O	object
PL	plural
PNG	Papua New Guinea
PPR	present progressive
PRES	present
PRF	perfective
PROX	proximal
Q	polar question
RF	remote future
RP	remote past
SG	singular
SS	same-subject

Data, Code and Materials Availability Statement

The Editor, Ben Ambridge, granted (on 4 September 2024) an exemption to data-sharing for the Ku Waru corpus, on the basis that this corpus is already available, after free sign-up, at the PARADISEC site (https://catalog.paradisec.org.au/collections/AR3). The Editor also granted an exemption to data-sharing for the Nungon corpus, on the basis that these are child–parent interactions from a community who are largely unfamiliar with the workings of the internet, large-language-models etc., and who are therefore unable to provide fully-informed consent for data sharing. Applications for access to the corpus should be sent to the first author. The data for the new analyses reported here (.wav, ELAN, PRAT and spreadsheet files) as well as the analysis code are available at https://osf.io/zmchn/.

Ethics Statement

Data discussed here were collected under the Australian National University Human Ethics approval 2013/055 (Ku Waru, Nungon) and Western Sydney University Human Ethics approval H13536 (Nungon). Written consent was obtained from families who participated.

Authorship and Contributorship Statement

Hannah Sarvasy proposed the comparative study, analyzed Nungon lexical data, drafted general and Nungon data sections, and contributed to revisions. **Alan Rumsey** analyzed Ku Waru lexical data, drafted general and Ku Waru data sections, and contributed to revisions. **Josua Dahmen** analyzed Ku Waru prosodic data, contributed to drafting the general prosody and Ku Waru prosody sections, and contributed to revisions. **John Onga** collected and recorded Ku Waru lexical perception data. **Stephanie**

Yam analyzed Ku Waru prosodic data and contributed to drafting the general prosody and Ku Waru prosody sections.

Acknowledgments

For their very helpful feedback on drafts of this article we thank Francesca Merlan and Lauren Reed. Thanks to Sam Passmore for providing statistics advice for the study of Ku Waru prosody. Financial support for the research on which this article draws was provided by the Australian Research Council, under grants DE180101609 and CE140100041 for Sarvasy, and DP130101655, CE140100041, and DP220100971 to Rumsey. We extend our heartfelt thanks for the support and hospitality of Nungon speakers in Towet village and of Ku Waru speakers in the Kailge community, where research was carried out.

A vocabulary checklist for early lexical development in Tseltal

Marisa Casillas
University of Chicago, USA

Ruthe Foushee
The New School for Social Science, USA

Humbertina Gómez Pérez
Juan Méndez Girón
Gilles Polian
Centro de Investigaciones y Estudios Superiores en Antropología Social, Sureste, Mexico

Kennedy Casey
Princeton University, USA

Penelope Brown
Max Planck Institute for Psycholinguistics, The Netherlands

Abstract: This study presents a checklist for measuring the expressive vocabularies of young children acquiring the Mayan language Tseltal. Adapted from the vocabulary checklist of US English MacArthur-Bates Communicative Development Inventories and tested with 84 Tseltal-acquiring children (9–23 months), this Tseltal checklist shows desired within- and across-subcategory variability in age of acquisition, expected age- and gender-related change, and typical patterns of relative over- and under-representation in the most cross-linguistically stable domains. We discuss potential uses of the checklist and the next steps in its future development.

Keywords: vocabulary checklist; parent report; Tseltal; Mayan

Corresponding author(s): Correspondence concerning this article should be addressed to Marisa Casillas, 1101 E 58th St, Chicago, IL 60637. E-mail: mcasillas@uchicago.edu

ORCID ID(s): Marisa Casillas: https://orcid.org/0000-0001-5417-0505; Ruthe Foushee: https://orcid.org/0000-0002-8361-1725; Gilles Polian: https://orcid.org/0000-0002-0642-2875; Kennedy Casey: https://orcid.org/0000-0001-5326-3080

Citation: Casillas, M., Foushee, R., Gómez Pérez, H., Méndez Girón, J., Polian, G., Casey, K., & Brown, P. (2025). A vocabulary checklist for early lexical development in Tseltal. *Language Development Research*, *5*(3), 245–275. http://doi.org/10.34842/ldr2025-862

Introduction

Vocabulary checklists, such as the MacArthur-Bates Communicative Development Inventories (CDI, Marchman et al., 2023), are efficient and reliable tools for gleaning general measures of children's lexical and grammatical development (Fenson et al., 1994; Frank et al., 2021; Marchman & Dale, 2023). Rather than measuring a child's actual lexical inventory or current morphosyntactic knowledge, these tools generate a quick snapshot of children's overall language development that—combined with other information sources (e.g., interviews, longitudinal data, other assessments)—can help identify delays in language development.

Further, in understudied and underresourced language communities, checklists can help to rapidly map out some typical pathways for lexical development (e.g., Alcock et al., 2015; Southwood et al., 2021; Vogt et al., 2015; Weber et al., 2018), particularly when they complement observational and experimental data in the creation of language acquisition sketches (Hellwig et al., 2021). Such sketches can be used to establish typical development milestones and learning materials for clinical and educational professionals who work in the language community and who support language maintenance. Such documentation may be especially useful for communities experiencing language endangerment—often due to colonization—when there is a strong interest in heritage language maintenance. Heritage language loss has been closely tied to the loss of cultural knowledge (Bromham et al., 2020; Kik et al., 2021; Maffi, 2005), and child language development is an appropriate target for projects seeking to more broadly fortify cultural identity in colonized societies and minoritized communities.

The semi-standardized format of vocabulary checklists also makes cross-linguistic developmental comparison possible at an unprecedentedly wide scope: as we write this, researchers have begun the process of adapting the CDI into at least 117 language varieties, and CDI *data* is freely available from over 92k children sampled across 42 language varieties on the Stanford Wordbank archive (https://wordbank.stanford.edu/). Such large datasets can be used to identify trends in lexical development that crosscut structurally and culturally diverse language communities as well as clear points of variation in development that afford new insights into the human cognitive system (Braginsky et al., 2019; Frank et al., 2021).

Adaptation, Not Translation

Originally developed for US English-acquiring children, the CDI has been authorized for adaptation in 117 other language varieties, many of which now have available adaptations (https://mb-cdi.stanford.edu/adaptations.html). The CDI adaptation materials speak of "adaptation" rather than "translation" of an instrument because the specific items on one list may not be culturally or linguistically appropriate for another.

For example, the relevant common animals, foods/drinks, household objects, and more, can be expected to differ between geographically and economically diverse language communities. Even when an item is relevant in multiple language communities, its meaning may substantially differ between languages, leading to false equivalences at the item level. That said, highly overlapping conceptual categories can sometimes be identified across diverse lexicons ("unilemmas"), in which case they can provide unique insights into cross-linguistic conceptual development; an affordance particularly useful when studying children learning more than one language (Byers-Heinlein et al., 2024; Tan et al., 2024).

Another important consideration for CDI adaptation is when and how to use inflected forms on the checklist. For example, in the language we focus on presently—Tseltal—bare transitive verb stems are ungrammatical. There is no single inflected verb form for transitive or intransitive verbs that is so common or so representative of the major inflectional paradigms as a whole that it can confidently be used to assess vocabulary development.

Lastly, while item adaptation is discussed extensively in the CDI's online materials, adaptations in the manner of data collection are less often addressed. The CDI was originally intended as a fillable paper form that could be given to parents at a lab visit or via mail. This format has been seamlessly adapted to online data collection via secure webforms (deMayo et al., 2021). However, for populations in which the primary caregivers are not comfortably literate in the language under study, in which a more conversational interaction is normatively appropriate, and/or in which primary childcare is divided among multiple family members, the basic format of one participant filling out a written form is going to fall short of accurately capturing information about child development.

Study Population

The present study focuses on Tseltal-speaking families with small children. Tseltal is a Mayan language spoken by more than 500k people in southeastern Mexico. Most Tseltal-speaking communities sit in central and northeastern Chiapas. There are three primary dialect areas (north, central, and south); we focus here on the central geolect, specifically the variant spoken in the Tenejapan municipality. Linguistically, Tseltal has several typological features that are understudied in the language development literature at large, including ergative-absolutive alignment, verb-first basic word order, an agglutinating and mildly polysynthetic morphology, a large inventory of "specific" verbs, an absolute frame of reference spatial system, ejective consonants, and more (Brown, 1997, 1998a, 2008; De León, 1999a, 1999b, 2001; Polian, 2013, 2017).

Tenejapa is a rural municipality in which subsistence farming (e.g., corn, beans, squash, potatoes, coffee, and more) is a primary way of life, often supported by one or more other sources of income (e.g., taxi driving, labor outside of the community, etc.). Most children in the sampled communities grow up primarily hearing and speaking Tseltal, though they may commonly encounter Spanish in some village areas (e.g., at public events on school grounds) and on television and the radio. Many Spanish words have been borrowed into the contemporary Tseltal lexicon, and share similar, if not identical, phonological forms with their original Spanish forms (e.g., *wakax* /wɑkɑʃ/ (Tseltal) vs. *vaca* (Spanish) "cow"; *tijeras* (Tseltal[1] and Spanish) "scissors"). Until they are walking reliably, around their first birthdays, Tseltal children spend most of their waking day in a sling worn by their mother. Sometimes another (typically female) relative carries the baby instead. Even after they are walking, young Tseltal children tend to stick close to their primary caregiver, and are often carried while they sleep or while the mother goes about her business in and out of the home.

Thus, in our targeted age group, children spend the majority of their waking time in very close proximity to their primary female caregiver(s). Many Tseltal children grow up in multigenerational households that include the child's parents and paternal family (father's parents and father's brothers' wives and children). Children and adults share some load in caring for young infants when the mother is not available, and it is common for married women to have 3–5 children; therefore our respondents have a great deal of experience caring for and observing children by the time they themselves reach motherhood. While Tseltal has previously been characterized as non-child-centered (Brown, 1998b, 2011; Brown & Gaskins, 2014), infants' primary caregivers—in most cases, their mothers—have immense amounts of exposure to their children's vocalizations from carrying and caring for them most of the day, most days. These caregivers are therefore extremely well positioned to report on what their children say.

Notably, there is a great deal of variation in the number of years of school Tseltal women currently complete, and even those who complete secondary education have very rare opportunities to read and write in Tseltal after the first two years of primary school (the language of schooling is Spanish). Therefore, any Tseltal checklist must be conducted via verbal interview. This technique has been used with vocabulary checklists before when interviewers cannot expect fluent literacy in the language(s) being examined (Alcock et al., 2015; Hamadani et al., 2010; O'Shannessy et al., 2024; Prado et al., 2018; Southwood et al., 2021; Vogt et al., 2015; Weber et al., 2018).

[1] 20 years ago, this borrowing of Spanish "tijeras" was pronounced more like *texerex* /teʃeɾeʃ/ in Tseltal. However, the conventional Spanish form has since taken its place. There are many such cases in the everyday Tseltal lexicon, which affects the responses given in a vocabulary checklist task.

Method

Our methodological approach strikes a balance between the MacArthur-Bates CDI guidelines for checklist adaptation and our team's own immediate research priorities. We set out to create a vocabulary checklist that would be structured and collected such that our data could be used in large-scale cross-linguistic analyses (e.g., Wordbank) while also maintaining sensitivity to local Tseltal accounts of typical early word knowledge—Tseltal early noun and verb development is the target of our current work involving this questionnaire.

Our methods thus follow similar guidelines to what is suggested for a "Level 2" adaptation of the vocabulary checklist of the CDI, with three important exceptions: (1) Our present research questions motivated a close look at an age group (12–20 months) on the border of the CDI: Words & Gestures (W&G; 8–18 months) and CDI: Words & Sentences (W&S; 16–30 months), so our adaptation initially included lexical and grammatical items from the US English W&S checklist *and* gesture items from the US English W&G checklist; (2) we only measured expressive item knowledge—and for most participants we additionally collected a reported phonological form of children's word productions; (3) we paused the inclusion and further development of grammatical items after getting uninterpretable data from a handful of early participants.

As described below, we incorporate *local* accounts of typical early word knowledge by using our transcribed corpus of Tseltal child speech and, via an iterative process of checklist refinement, asking early participants to tell us any words their children say that were not included in the list they had just completed. Following the CDI adaptation guidelines, the organization of items into semantic categories (e.g., "Animal names", "Food or drink") mostly aligns with the groupings suggested by the CDI board (which originate in groupings from the US English CDI checklist). However, some exceptions are noted below. Moreover, because "semantic category" is simply a type of metadata for each item, we note that the items can easily be subset and/or re-categorized into more locally meaningful categories by anyone wishing to use these data in the future—one of many reasons why it is essential that we share all data for all items as openly as possible.

Participants

We collected checklist data from 84 Tseltal-acquiring children between ages 9 and 23 months (mean = 16.03; 38 girls and 46 boys). This sample size for checklist responses is within the typical range for prior CDI adaptations at the pre-validation and pre-norming stage; however, we note that sample sizes for studies targeting larger age ranges and studies associated with public health research can be much higher (e.g., our unsystematic review reveals a range of N = 36–2,418[2]). Personal communication with CDI board members verified that there is no standard sample size for creating a preliminary adaptation, and that a more primary issue at this stage was identifying a range of easier and harder words within each checklist subcategory. We thus aimed for 75 participants for our preliminary checklist dataset, ultimately exceeding that goal with a final sample of 84 participants. Based on our past experience in these communities, we had assessed a sample size of 75 to be feasible during a two-week bout of intensive testing during which we simultaneously ran other studies.

To varying extents, all children also had direct and overheard exposure to Spanish, and sometimes other Mayan languages (e.g., Tsotsil) depending on the composition of their household. However, all children were reported by their primary caregiver(s) to hear Tseltal most of the time, to be typically developing, and to have no known problems in hearing, seeing, or speaking. Language background and typical development were thus determined by simply asking the primary caregiver: (a) whether they spoke to the child only in Tseltal or also in Spanish or other languages (and, if the latter, how often) and (b) whether the child had any problems with hearing, seeing, or talking, or any other problems in developing. When caregivers reported exposure to a second language, they typically gave a verbal description of quantity or context (e.g., "sometimes" or "with his uncle"). Using this information, the experimenters (typically HGP and MC) verified that each child was Tseltal dominant (i.e., vast majority input in Tseltal) before proceeding. All participants resided in the mountainous Tenejapa municipality of central Chiapas. Checklists were collected in 12 rural villages within that municipality, between June and October 2023. The study plans were pre-approved by the University of Chicago Institutional Review Board (IRB23-0244).

Checklist Items

Initial Item Adjustments

The vocabulary checklist reported here was developed over six iterations across the 84 participants. As a starting point (Version 0), we reviewed the checklist used in Brown, Gentner, and Braun (2005), which was based on an adaptation of the US English MacArthur-Bates CDI (Fenson et al., 1994; Marchman & Dale, 2023; Marchman et al., 2023). The original Brown et al. (2005) Tseltal checklist contained 613 items across 19 categories taken from the English CDI vocabulary checklist at that time, listed in

[2] N = 36 for Australian Aboriginal English/Kriol/Other in Jones et al., 2020; N = 50/58 for Kiswahili/Kirigiama in Alcock et al., 2015; N = 100 for Czech in Jarůšková et al., 2024; N = 110/115/105/98 for Afrikaans/isiXhosa/South African English/Xitsonga in Southwood et al., 2021; N = 241 for Wolof in Weber et al., 2018; N = 566 for Changana/Ronga/Portuguese/Other in Vogt et al., 2015; N = 29/869 for Chichewa/Chiyao and Krobo/Ewe/Twi/English in Prado et al., 2018; and N = 2418 for Bangladeshi in Hamadani et al., 2010.

order as: Sound effects and animal sounds (11); Animal names (37); Vehicles (8); Toys (9); Food or drink (46); Clothing (21); Body parts (22); Furniture and rooms (32); Small household items (49); Outside things and places to go (42); People (34); Games/routines (21); Action words (185); Words about time (13); Descriptive words (39); Pronouns (17); Question words (8); Prepositions and location (15); Quantifiers (4). These categories are slightly different from those featured in the US English CDI checklist, with Brown's adaptations for Tseltal including: (a) combining Outside and Places to Go, given that many relevant places to go are primarily outdoors or semi-outdoors; (b) considering all verbs under Action words, given that "Helping verb" is not a sensible linguistic category in Tseltal; and (c) removing the small and mixed category of Connecting Words, which in Tseltal could only include one subordinator, one subordinator that also acts as a definite article, one coordinator that also acts as a preposition, and two borrowed conjunctions from Spanish. In our adaptation, we respect these category decisions by Brown, which still largely align with the CDI's standards.

In the early 2000s, Penny Brown interviewed the families of 5 young Tenejapan Tseltal-acquiring children to document their vocabulary production and comprehension. The study was discontinued after the first 5 participants because the interviews lasted 2 or more hours and because the experimenter (PB) maintained doubts about the children's true word production and comprehension, compared to what was reported on the form (Brown, personal communication). We ran through this entire checklist, via interview, with the mother of an apparently linguistically advanced 18-month-old boy. Most words on that initial checklist were presented as grammatical, bare lexical stems, but verbs were presented in the incompletive, first person singular form (e.g., *ya x-ben-on* /jɑ ʃbenon/ (I walk)) and body parts were presented with a first person singular possessive morpheme (e.g., *j-k'ab* /xk'ɑb/ (my-hand)). Following checklist completion, we reviewed candidate words that were not on the list, and also words that were on the list but were unlikely to be useful in tracking young children's vocabulary in that mother's experience. This participant's checklist was considered pilot data; it is not included in the primary analyses of the 84 children.

In Version 1 of the checklist, using the pilot data from the advanced 18-month-old child, we cut the list down to 212 words and 9 gestures. We only asked about whether the child *produced* the word or not (i.e., we did not ask about comprehension). We also exploratorily asked about what kinds of sentences and errors the child was producing. These questions about sentences and errors elicited very different types of responses among the first three caregivers tested ($N = 3$; e.g., regarding errors: "none", "not talking enough to say", and "calling family members by the wrong name"). However, the other (word and gesture) caregiver responses appeared to function more reliably. In the process of conducting the first three interviews, we found it useful to provide multiple forms for each verb, namely: (a) a fully inflected incompletive first person form, (b) the same form without the aspect markers, and (c) the bare verb stem (e.g., for *tsak* /tsɑk/ (to take/grab) we would give the options "*ya j-tsak, j-tsak, o tsak*?" /jɑ xtsɑk/,

/xtsɑk/, or /tsɑk/). Caregivers were asked to respond 'yes' in the case that their child said any of the forms listed. The bare verb stem for transitive verbs (e.g., *tsak*) is ungrammatical, but such forms are typically found in the spontaneous speech of Tseltal-acquiring children (Brown, 1997), as they are in children learning other Mayan (De León, 1999b) and some polysynthetic languages (for a review, see Kelly et al., 2014). For each verb, caregivers were thus asked to report which of the options (if any) their child produced—in many cases, caregivers reported the bare stems.

In Versions 2 (267 words; 9 gestures; *N* = 5) and 3 (263 words; 9 gestures; *N* = 7) we removed the sentence type and error questions, standardly added the three verb forms described above for all verb items, and added more words, especially harder words that were likely to be known only by older toddlers. We removed some words that were considered to be old fashioned, or which were homophonous with other items on the list (e.g., *ja'* /xɑʔ/ is both the word for water and a demonstrative; we retained the former). Version 4 (263 words; 9 gestures; *N* = 3) maintained the same words as Version 3, but we changed how we asked about each item. Instead of simply asking whether the child produced each item or not, we now asked: "Does your child say this word? If so, *how* do they say it?" We would then write an impressionistic orthographic transcription of what the caregiver produced. For example, for *xawin* /ʃɑwin/ (cat), a typical response was: "'yes' the child says it" and they say it like "*win* /win/".[3] With this additional question, administrations of the checklist still typically only lasted 10–20 minutes in duration, and caregivers seemed to overall enjoy doing impressions of their young child's productions. In Versions 5 (273 words; 9 gestures; *N* = 20) and 6 (299 words; 9 gestures; *N* = 46), we continued adding harder words and missing words typical of early production, in addition to making minor changes to item order and categorization (e.g., *ton* /ton/ (rock) was moved from the "Toys" category to the "Outside things and places to go" category). We kept the same gesture items from Versions 1–6, but the gesture labels were reworded for greater clarity in Version 5 (the interviewer typically demonstrated the gesture, rendering the intended meaning clear in all cases).

Final Item Selection

As a final step, and following Alcock et al. (2015), we used the collected data to pare the list back down to ~250 items that include a range of earlier- and later-produced words within each sub-category of the checklist (e.g., Animal names; Vehicles; Toys; Food or drink; etc.). This process needed to be completed manually, and so was pre-

[3] To demonstrate the diversity of reported productions here, the unique reported productions of *xawin* (/ʃɑwin/) among the 84 participants were: *xawin, chawin, xamin, xiwin, xa, xaw, xux, waw, win, wixwix,* and *meumeu*. Onomatopoeic form substitutions like "meumeu" were marked, but noted as a different form, as were word forms for the same referent in another language (e.g., *gato*, from Spanish).

registered on OSF (https://osf.io/z8hdk) to mitigate bias in item selection. Our process for item selection followed Alcock et al. (2015)'s description fairly closely: First, find items that reach 50% production at any month—these will be the core words for the list. Then, add the 20 least-frequently produced words among those that were known by at least 20% of children at any age. Then, add the 20 least-frequently produced words among those that were known by at least 5% of children at any age. These are the hard items—the ones we only expect older children to produce.

We then generated the mean age of acquisition (AoA) for each item on the full list of 299 words and systematically reviewed the currently included words for a relatively even distribution in the 12–20-month-old age range and within each sub-category (e.g., Animal names; Vehicles; Toys; Food or drink; etc.). When the AoA distribution appeared uneven, we added, removed, or swapped out items to improve the representation of easier and harder words. In this process we also ensured that the verbs were somewhat balanced in transitivity, and that the transitive verbs included both specific and general verbs from Brown (1998a)'s list.[4] We also checked that the nouns were fairly balanced in concreteness, animacy, and ability to be handled. We kept an eye out for near synonyms and removed them unless each individual item was separately motivated (e.g., *bistuk* and *bi yu'un* both mean something like English "why", so we removed the latter; quantifiers *bayel* and *uts* both mean something like English "a lot", but the latter can also be inflected as an adjective). Where possible, we tried to keep mini sets of words within categories that are of theoretical interest (e.g., kinship terms, spatial terms, etc.). This process of scanning, swapping, and re-checking each sub-category and across age took many iterations.

This final list of 251 words and 9 gestures was checked with our co-authors, which include a native speaker of Tenejapan Tseltal, a near-native speaker of Tenejapan Tseltal (who is a native speaker of the closely related language Tsotsil), and a linguist specializing in Tseltal. The final list was accepted for current analysis, but cases of potential "missing" items (e.g., lower-frequency household items and animals) were noted to be tested in future versions of the checklist. The 251 words include 231 words from Brown et al.'s (2005) CDI-inspired checklist. The checklist items thereby overlap substantially with the US English Words and Sentences CDI, including 113 of the 639 unique unilemma concepts within the comparable categories of the US English Words and Sentences CDI.[5] They are divided among sub-categories as follows: Sound effects

[4] This distinction is relevant for transitive verbs which are either very restricted in the patients they take (heavy/specific verbs: e.g., *we'* /weʔ/ "eat-tortilla" and *top'* /top'/ "shatter") or which are instead very open (light/general: e.g., *ak'* /ɑk'/ "give/put" and *ai'y* /ɑʔi/ "see/hear/perceive").

[5] The 251 items also include 30 concepts not included in Wordbank's current unilemma inventory: flea, louse, buzzard, VW Beetle, commuter pickup truck, cold cornmeal (beverage), warm cornmeal (beverage), sombrero (hat type), wrap (clothing type), Tseltal skirt (clothing type), hammock, metal roofing material, stirring stick, peso, milpa (mixed-plant field), namesake, little one (term of endearment),

and animal sounds (10); Animal names (16); Vehicles (8); Toys (5); Food and drink (21); Clothing (10); Body parts (14); Furniture and rooms (13); Small household items (19); Outside things and places to go (15); People (12); Games and routines (14); Action words (41); Words about time (6); Descriptive words (15); Pronouns (10); Question words (6); Prepositions and location (10); Quantifiers (6).

Figure 1. ***Participants were verbally interviewed by native and near-native speakers who most often noted responses on paper copies of the checklist.***

Procedure

Participants were recruited via word of mouth. As illustrated in Figure 1, participants were either interviewed by appointment, in their home, or came to visit the interviewer(s) in another location during open testing hours (e.g., by a local school, in a neighbor's house, etc.). Most participants signaled to the interviewer in advance when they would be able to meet, via verbal agreement, direct/text message, or phone call. When the interview began, participants were first engaged in a consent process that described the context for the research study, their right to stop at any time, and their compensation, among other topics (see Appendix A for the full points covered). Consent was sought in a series of informally phrased points, with wording varying

Mrs (honorific type), Mr (honorific type), older sister or father's side aunt/cousin (kin type), older brother or older first cousin (kin type or honorific), older brother of a female (kin type), walk-on (greeting), here-take (presentational word), okay/fine/agreed, let's go, sound to call chickens, perceive/experience, uphill, and downhill.

slightly from participant to participant to ensure a more conversational flow. In our experience, this more interactional form of consent is more effective in eliciting questions and demonstrations of understanding from participants.

Once participants consented to participate, we began with the instructions. For each word, participants were asked to indicate if their child says (or used to say) that word and—for participant 16 onward (Version 4 onward)—how their child says it. During recruitment we asked to interview the child's primary caregiver, which is typically the mother in this community. However, in practice, our interviews often *additionally* included aunts, grandmothers, and older siblings who had spent significant time with the child and who were present during our interview period. Mothers were the primary interviewees, but attending family members sometimes offered their opinion on whether the child said a word or not, sometimes in response to a bid from the mother and sometimes spontaneously—ultimately we always accepted the mother's final judgment.

Participant responses were recorded in real time. When there were two interviewers present, one focused on talking with the caregivers and one on writing down responses. When there was only one interviewer present, they were responsible for both talking and writing. Consent and interviews were conducted by native or near-native Tseltal speakers who reside in one of the testing villages (HGP or another team member). Foreign researchers (RF, MC) served only as second interviewers, noting down responses as they were given. Interviewers were typically able to complete this entire checklist interview process in 10–20 minutes. We additionally note that the brevity of this interview made it easy to combine checklist data collection with experiment-based data collection in the same short visit (typically 20–40 minutes).

Most data were collected on paper copies of the list, but early versions were directly typed into a spreadsheet on a laptop, and a handful of sessions were collected via pdf markup on a tablet computer. Any checklists collected by our local, independent interviewer (HGP) were photographed and sent via encrypted message to our primary analysts (MC, KC, RF) for digital entry into the project database. Database entries were quality checked (MC, KC) prior to analysis.

Results

Our aim in the present paper is to test whether the checklist functions as expected, as an instrument designed to map variation in typical lexical development among Tseltal-acquiring children. We divide our analyses into three parts: (a) age of acquisition checks, (b) age-related change, and (c) relative representation across checklist categories. In all of the analyses below we use children's *conceptual* vocabulary (i.e., include an item if they are reported to say it in Spanish rather than the provided Tseltal item). The vast majority of reported productions—99.10%—aligned with the provided

Tseltal wordforms, or a referentially acceptable alternative in the speech community (e.g., substituting "darkness" for "night"; see Appendix C). Child wordforms reported in Tseltal varied between individuals, ranging from 92.59%–100% of items with Tseltal responses (mean = 99.02%). If we use Tseltal-only responses rather than all responses, it makes no qualitative difference in any of the findings reported. We exclude reported productions that don't map well onto the intended target item, including: associated words (e.g., "bite" for "snake"), overgeneralizations ("car" for "taxi"), and onomatopoeia used in place of object labels (e.g., "moo" for "cow"). The produced forms for the excluded items are nearly always captured by another item in the checklist (e.g., "car" for "car") and make up 0.66% (less than 1%) of the checklist responses we gathered.

Note that because the checklist was developed in versioned waves, the 251 final items vary in the number of times they were assessed; 219 items have data from all 84 children, 5 items have data from 76 children, 25 from 46 children, and 2 from 38 children. We do not impute missing data for any of the 251 items in the analyses below. Instead, we base proportions by item and by participant on the total data available for each case.

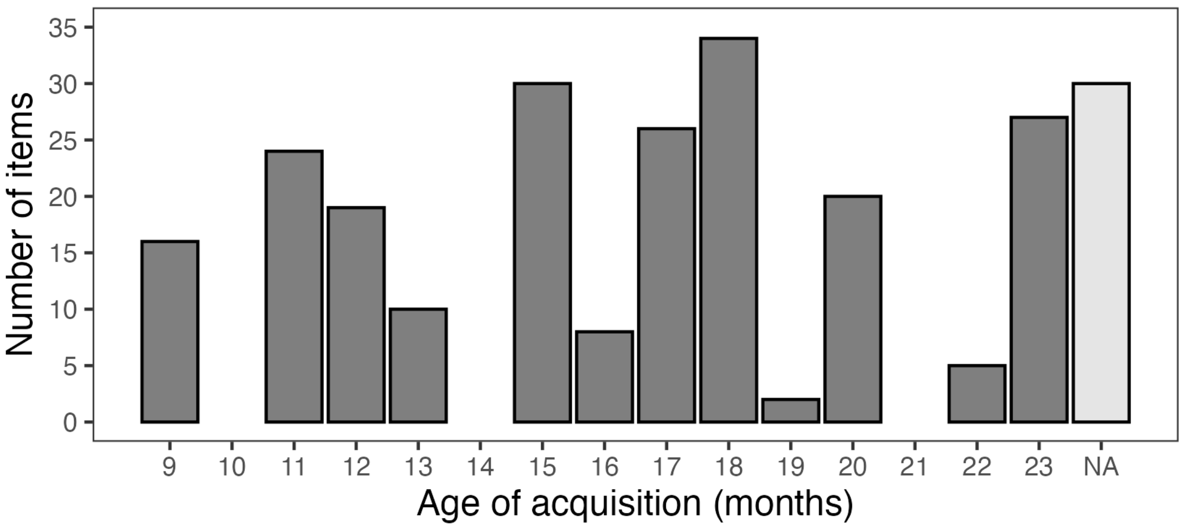

Figure 2. ***Distribution of age of acquisition—the first age with ≥ 50% of participants producing the item—over all items. Items listed in the "NA" bar (far right) do not reach ≥ 50% production at any observed age.***

Age of Acquisition Checks

By design, the checklist should include a range of earlier- and later-acquired words within and across all sub-categories. In the methods, above, we described how we attempted to achieve this distribution of word difficulties. Here we report on our success in doing so. For each of the 251 words on the final checklist, we define the age of acquisition (AoA) as the first age at which at least 50% of the participants were reported to produce the word.

The distribution of AoA over the entire checklist is shown in Figure 2. Indeed, we see a reasonably balanced distribution of AoAs between 9 and 23 months, the ages tested. Note that 11.9% ($N = 30$) items did not achieve 50% production at any age. This is partly attributable to having relatively little data for children older than 20 months ($N = 3$ children); based on collaborator discussion, we predict that many of these words would have an AoA before 24 months with more data collection (see Appendix B for further consideration on this issue).

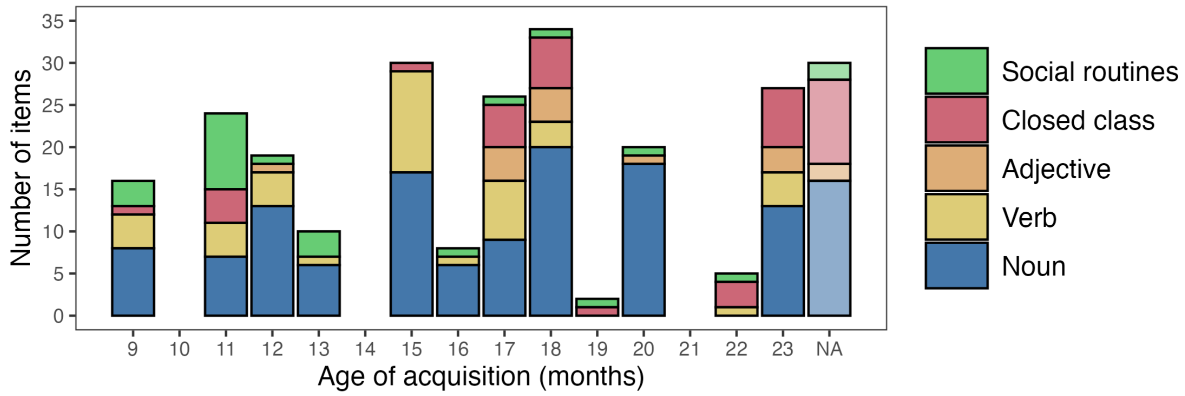

Figure 3. ***Distribution of age of acquisition over all items, separated by major syntactic types (Noun, Verb, Adjective, Closed class, Social routines). Items listed in the "NA" bar (far right) do not reach ≥ 50% production at any observed age.***

In Figure 3 we look at the data from a similar perspective, only this time breaking the data down by major syntactic types on the checklist, including three open-class categories (Noun, Verb, Adjective), a closed-class category (Closed class; e.g., pronouns, quantifiers), and a category for fixed expressions associated with everyday games and routines (Social routines). Again, we see that there are early, middle, and late AoAs within each type, though some categories are more limited in their spread than others. For example, the first adjective AoA does not occur until 12 months. Other categories are slightly unbalanced in their AoA distribution. For example, there is a cluster of social routine items acquired at 11 months. These slightly asymmetrical distributions are expected, considering that some word types (e.g., social routines) are typically acquired earlier or later than others (e.g., adjectives) given differences in salience, frequency, conceptual complexity, etc. (Arunachalam & Waxman, 2010; Bates et al., 1994; Braginsky et al., 2019; Frank et al., 2021; Gentner, 2006). In general, however, the items in the present checklist meet the aim of including a range of relatively easier and harder items within each sub-category.

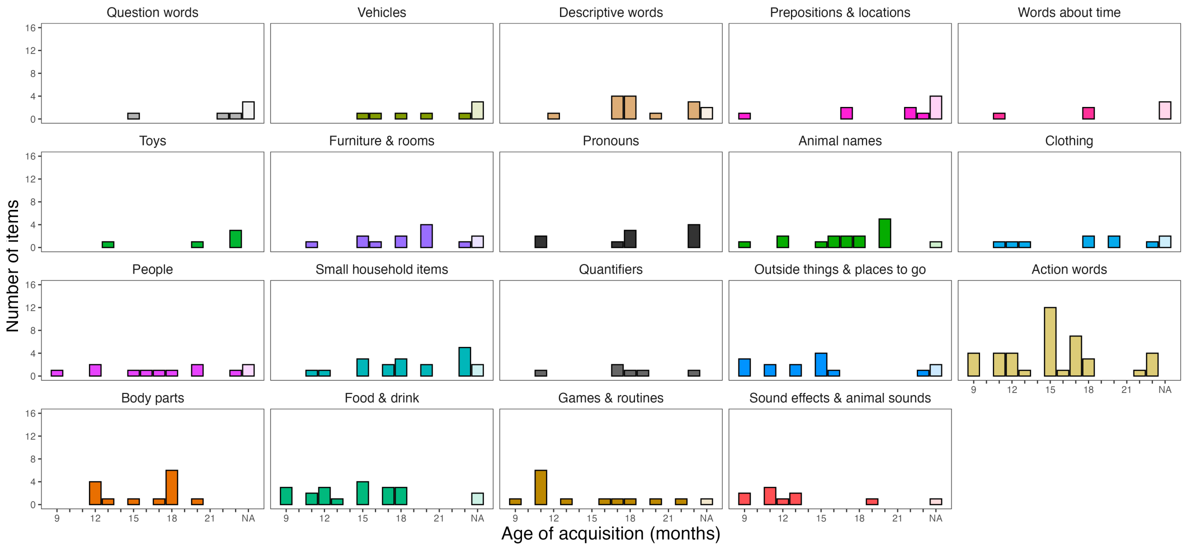

Figure 4. ***Distribution of age of acquisition over all items within each content sub-category of the list, sorted from the latest acquired (upper left; Question words) to the earliest acquired (lower right; Sound effects & animal sounds).***

One final check on AoA distributions comes from the semantic sub-categories standardly associated with CDI checklists, such as: Food and drink, Vehicles, Clothing, and more. Figure 4 shows the AoA distributions for items within each content sub-category. Similar to Figure 3, general skew across the observed age range is more apparent for some categories than others (e.g., Question words tend to be acquired later; Food or drink words tend to be acquired earlier). But even in the smallest sub-categories (e.g., 6 or fewer items in each in the categories: Question words, Words about time, Toys, and Quantifiers) there is a clear spread in AoA.

Overall, we find that the checklist effectively achieves its aim of including easier and harder words within each major sub-category and across the checklist as a whole.

Age-Related Change

Another checklist outcome worth assessing is age-related change. If the checklist is working as expected, we are very likely to see an increase in productive vocabulary size with age, particularly: evidence for an acceleration in word production starting around 18 months (Fenson et al., 1994; Marchman & Dale, 2023), and (more tentatively) larger vocabularies for female than male children (Kachergis, Francis, & Frank, 2023; Mayor & Plunkett, 2011). To test these predictions, we fit Generalized Additive Model for Location, Scale and Shape (GAMLSS) models (Rigby &

Stasinopoulos, 2005) to generate approximate percentiles for overall vocabulary size across age, overall (Figure 5) and for female versus male children (Figure 6; see also Jackson-Maldonado et al., 2024). These models are restricted to assume monotonic growth, such that vocabulary size strictly increases across age.[6]

In the overall data, the typical trajectory—at the 50th percentile—clearly suggests an acceleration in word production shortly after the first birthday. As children get older, and larger vocabulary sizes become more likely, we also see greater reported variability in observed vocabulary sizes, with an estimated spread between children in the 10th and 90th percentile of 200 checklist words by 24 months.

When we divide the data by child sex and examine the 50th percentile trajectories, we see that, numerically, female children are consistently reported to have larger vocabularies. However, this difference is small and non-significant, providing no clear evidence for early sex-based vocabulary differences in Tseltal.

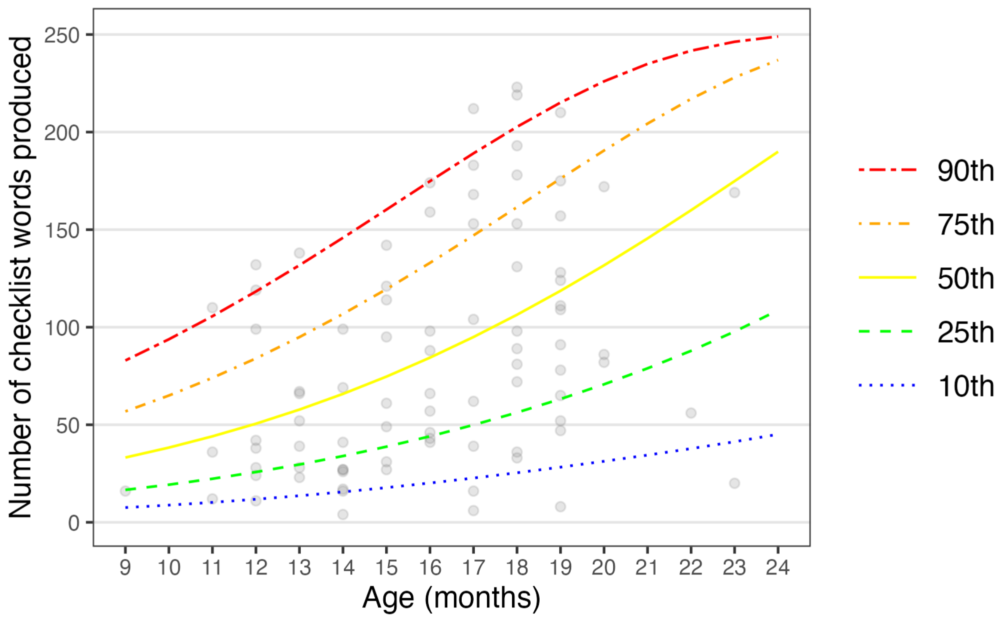

Figure 5. ***Number of checklist items produced across age by individual children (gray dots), showing percentiles (10th, 25th, 50th, 75th, and 90th) for lexical production.***

[6] gamlss(produces ~ pbm(age, lambda = 10000), sigma.formula = ~ pbm(age, lambda = 10000), family = BE, data = vocab.data)

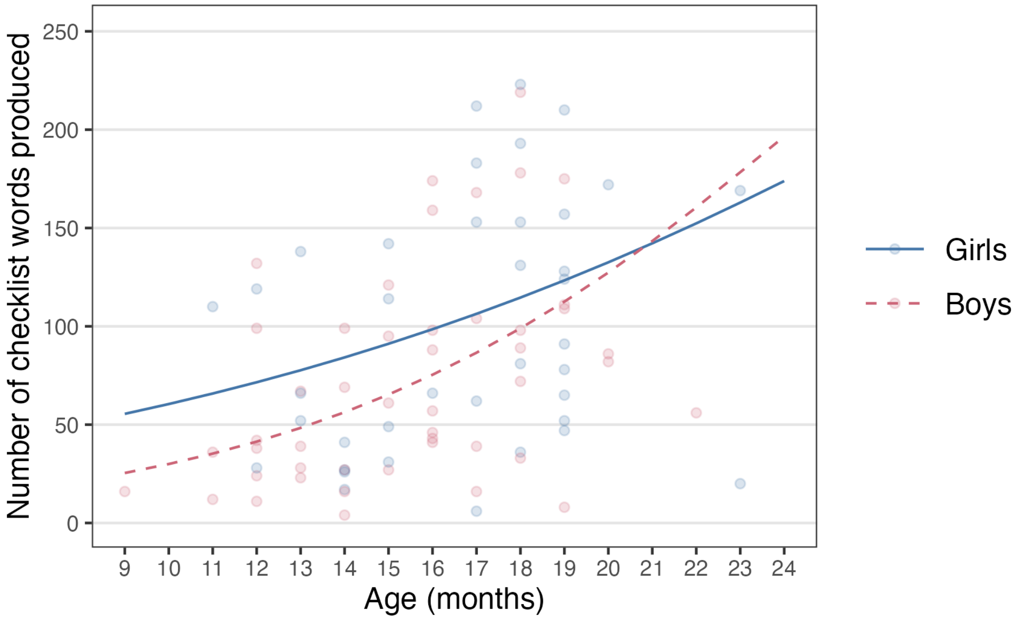

Figure 6. ***Number of checklist items produced across age by individual male (red) and female (blue) children, showing the estimated trajectory for the 50th percentile of lexical production in each group.***

In sum, age- and gender-related differences in reported vocabulary size, as measured by the checklist, largely accord with patterns observed previously in the literature using the checklists in other CDI instruments (Fenson et al., 1994; Jackson-Maldonado et al., 2024; Kachergis et al., 2023; Marchman & Dale, 2023; Mayor & Plunkett, 2011).

Relative Representation Across Checklist Categories

While we do aim to have a range of easier- and harder-to-acquire items within each sub-category of the checklist, we can expect systematic differences in word learning between categories due to their salience, conceptual complexity, and more (Arunachalam & Waxman, 2010; Bates et al., 1994; Braginsky et al., 2019; Frank et al., 2021; Gentner, 2006). Our final analysis thus sketches preliminary evidence for variability across the checklist's content categories, such as Animals, Small household items, Food or drink, and more.

Following Braginsky et al. (2019), we use a measure of relative representation to understand whether children's production of words within each sub-category is greater or lesser than we would expect on the basis of random, unbiased development. The analysis makes use of the idea that, if lexical development is unbiased, children should know words in any category proportionally to their overall word knowledge. So, a child who produces 50% of the words on the checklist should, on average, produce 50% of the words within any given sub-category. If that child produces more than 50% of the words in a sub-category, we can consider that category to be relatively over-represented. If the child produces less than 50% of the words in a sub-category, it would be relatively under-represented.

We can make some broad predictions for this analysis based on work from the Wordbank team (Braginsky et al., 2019; Frank et al., 2021). Namely, while sub-category rankings vary across languages, some domains show consistent over-representation in development (Sounds, Games & routines, and Body parts) while others show consistent under-representation (Places and Time words, Frank et al., 2021). The Tseltal data are consistent with this prediction (Figure 7). Sounds, Games & routines, and Body parts make three of the four most over-represented categories. Spatial and Time words are within the six most under-represented categories.

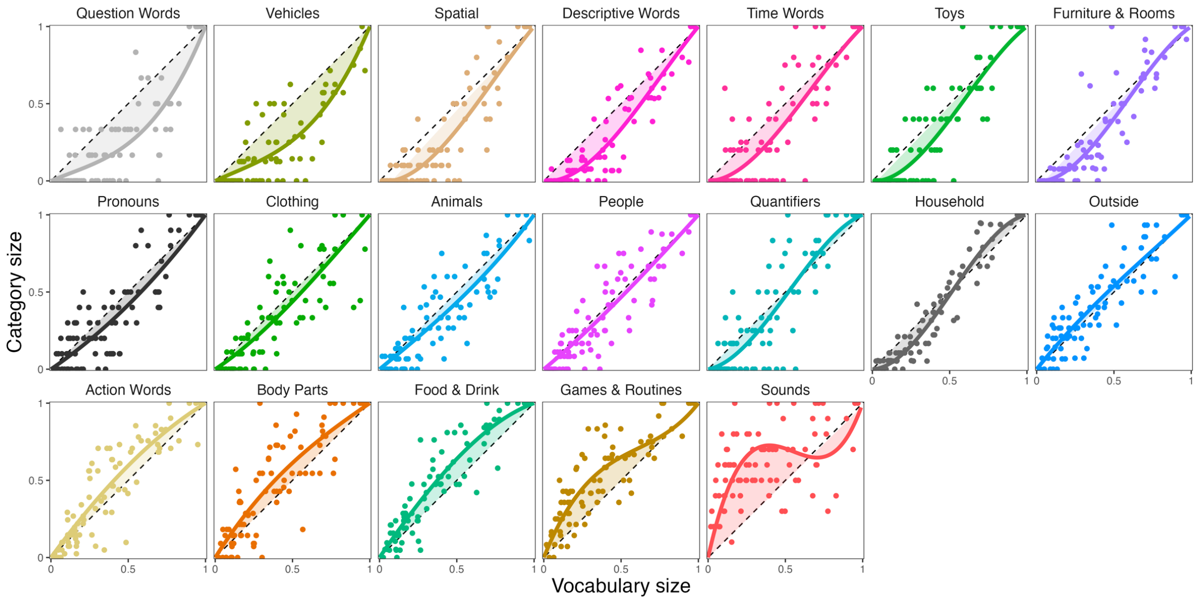

Figure 7. ***Distribution of age of acquisition over all items within each content sub-category of the list, sorted from the latest acquired (upper left; Question words) to the earliest acquired (lower right; Sound effects & animal sounds).***

We can assess the extent of bias in learning by measuring the area between the diagonal (unbiased learning) and the fitted line (observed data). Again, following Braginsky et al. (2019), we randomly sub-sampled and measured this area 1,000 times to create bootstrapped 95% confidence intervals around the estimated bias size. The resulting estimates and confidence intervals are shown in Figure 8. Sub-categories with effect size distributions overlapping with zero show no evidence for bias in learning; those below zero show evidence for under-representation, and those above zero for over-representation.

The bootstrapping analysis suggests a significant overrepresentation of Sounds, Games & Routines, and Body parts, among other categories, as well as a significant underrepresentation of Time words and Spatial words. The pattern accords well with the cross-linguistic predictions based on empirical observations from Wordbank (Frank et al., 2021).

In brief, variability across checklist sub-categories accords with the most consistent patterns found in prior work on the vocabulary checklists of CDI instruments.

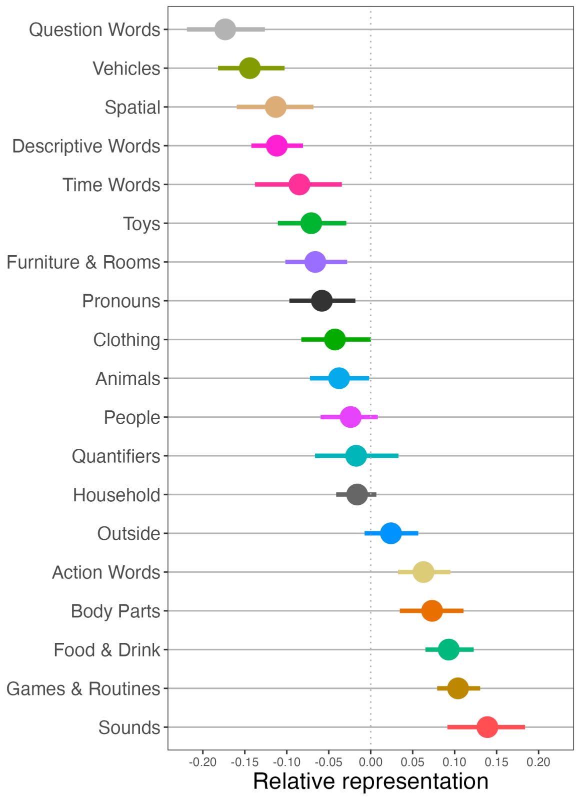

Figure 8. Bootstrapped relative representation effect sizes (x-axis) for word production across each content sub-category of the checklist (y-axis).

Discussion

This paper documents the creation of the first validated Tseltal vocabulary checklist. Based on a checklist first developed by Penelope Brown and Dedre Gentner, we collected data from 84 participants in an iterative development process that resulted in a 251-word checklist. Through this iterative process we were able to develop an inventory of items that are partially shared with other checklists but also reflect local caregiver reports on Tseltal children's early word production. Following CDI adaptation guidelines, these items are organized into standard CDI semantic categories, but could equally be rearranged and re-sampled to create more linguistically and contextually organized semantic categories for Tseltal children, depending on the relevant research question (e.g., based on shared semantic roots across syntactic classes, like *lo'* /loʔ/ "eat-soft" and *lo'bal* /loʔ.bɑl/ "banana" [literally: eat-soft-NOMINALIZER]). The list and its associated dataset are therefore ready to be used and further developed for both comparative and language-specific work on Tseltal.

The checklist displays reasonable variability within and across categories in age of acquisition, replicates basic patterns of age- and sex-related change, and demonstrates expected over- and under-representation patterns in the checklist's sub-categories given past findings (Bates et al., 1994; Braginsky et al., 2019; Fenson et al., 1994; Frank et al., 2021; Gentner, 2006; Jackson-Maldonado et al., 2024; Kachergis et al., 2023; Marchman & Dale, 2023; Mayor & Plunkett, 2011). In this brief discussion, we review the benefits of creating such an instrument for Tseltal before considering the most important next steps for continued instrument development.

Potential Uses of This Instrument

Vocabulary checklists are multi-purpose tools; we hope that this vocabulary checklist for Tseltal will be of use to a variety of professionals working in the fields of Tseltal (and perhaps more broadly, Mayan) child development. We have here tentatively concluded that the checklist broadly behaves like the vocabulary checklists of other CDI instruments, implying that it is likely useful for generating a lexically grounded snapshot of children's overall language development (Fenson et al., 1994; Frank et al., 2021; Marchman & Dale, 2023). The checklist may therefore, in the future, prove helpful in designing educational materials and identifying delays in linguistic development. We would, however, very strongly warn against using this checklist on its own as the foundation for decisions about intervention or education. Substantial further work that includes longitudinal data, test-retest reliability estimates, external validation, and more, will be required before the instrument can be treated as a reliable clinical tool. Furthermore, adequate application and further development of the instrument, where it concerns clinical treatment of children and clinician training, will necessarily involve the integration of observational data and interviews, among other data sources.

All that being said, these preliminary data collected using the checklist already begin to outline the distinct patterns in lexical development—along with expected sources of variability—that characterize Tseltal language development. These patterns can be leveraged to inspire language learning materials aimed at fortifying indigenous language maintenance efforts. The same patterns can help speak to the human capacity for language learning at large; similarities and differences in Tseltal lexical development may help us infer the adaptive capacities that underlie language learning across the diverse developmental milieux in which language is acquired.

The realization of this potential will only be achieved if the anonymized data, documentation, and analysis tools are freely and openly available to community members, clinical and educational professionals, and researchers. Sharing these aspects of the project redundantly and making sure they are well and securely disseminated is possible thanks to resources such as the Open Science Framework, the Wordbank archive, GitHub, and the CIESAS website (see Appendix A for links to each resource). Fully open materials will help ensure the healthy further development of this instrument over time.

Next Steps

The analyses in the present paper suggest that the checklist is, basically, functional in its current form. However, there are a number of clear future directions to take to ensure its usefulness and to further establish its validity (see Jarůšková et al., 2023). Regarding usefulness, we still wonder whether we have missed important words. Our ideal items are highly informative as developmental indices and additionally carry information about some linguistic or cultural feature that informs our stories of how Tseltal children develop (in particular) or our theories of human language cognition (in general). One example along these lines would be small lexical sets of spatial terms or kinship terms, which have setting-specific importance and also directly bear on theories of cognition (Bates et al., 1994; Clark & Sengul, 1978; Gentner, 2006; Mitchell & Jordan, 2021). There is also the important matter of ensuring that these items make sense across the major dialects of Tseltal.

In addition to new words and word substitutions, it would be useful to make two further structural changes to the checklist. First, we tried initially to ask about word combinations and errors, but our preliminary adaptation of these questions elicited highly variable response types. It is worth trying again, in a future iteration, to ask about morphosyntactic development. In a language with such a rich inflectional morphology (Polian, 2013), and with many aspects of morphosyntactic development well-captured in observational work (Brown, 1997, 2008; De León, 1999b, 1999a, 2001), there is a clear utility for a quick, rough measure of grammatical development. The second structural change would be to separate the checklist into two instruments: one aimed at younger children (akin to the Words and Gestures CDI instrument) and one aimed at older children (akin to the Words and Sentences CDI instrument). Our present checklist is aimed somewhere between these two traditional checklist populations—from the time just before first words to first word combinations. Our present age sample reflects the current needs of our research team, which is focused on a bigger project concerning lexical development in 12–20-month-olds (note that here we have allowed data collection up until 24 months). However in the long term it would be useful to have separate instruments as have been used in most other CDI adaptations.

Finally, there is a great deal more we could do to validate the instrument, internally and externally. Future steps should include longitudinal data collection, test-retest reliability measures, independent vocabulary measures, and more. Along with these efforts will come another necessary addition: much more data from many more children. These validation efforts are key to our interpretation—and thus application—of the checklist data. Should the checklist be used for clinical interventions, it will become especially urgent to establish these validity measures, in collaboration with clinicians, educational professionals, participant families, and other stakeholders. To better scale in these circumstances we may also need to consider a compromise between written parental report (the traditional method) and spoken parental interview (our current method). Following our Australian colleagues, we could consider a digital survey that features sound files for each word and an intuitive data-entry interface (O'Shannessy et al., 2024).

Conclusion

We present a preliminary vocabulary checklist for tracking the lexical development of children acquiring Tseltal as their primary language. The checklist displays many of the expected patterns for the vocabulary checklists of instruments based on the MacArthur-Bates CDI. We discuss important avenues for further development in the future.

References

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language*, *42*(4), 763–785. https://doi.org/10.1017/S0305000914000403

Arunachalam, S., & Waxman, S. R. (2010). Language and conceptual development. *WIREs Cognitive Science*, *1*(4), 548–558. https://doi.org/10.1002/wcs.37

Bates, E., Marchman, V. A., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*(1), 85–123. https://doi.org/10.1017/S0305000900008680

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

Bromham, L., Hua, X., Algy, C., & Meakins, F. (2020). Language endangerment: A multidimensional analysis of risk factors. *Journal of Language Evolution, 5*(1), 75–91. https://doi.org/10.1093/jole/lzaa002

Brown, P. (1997). Isolating the CVC root in Tzeltal Mayan: A study of children's first verbs. In E. V. Clark (Ed.), *Proceedings of the 28th annual Child Language Research Forum* (pp. 41–52). CSLI/University of Chicago Press.

Brown, P. (1998a). Children's first verbs in Tzeltal: Evidence for an early verb category. *Linguistics, 36*(4), 713–754. https://doi.org/10.1515/ling.1998.36.4.713

Brown, P. (1998b). Conversational structure and language acquisition: The role of repetition in Tzeltal. *Journal of Linguistic Anthropology, 8*(2), 197–221. https://doi.org/10.1525/jlin.1998.8.2.197

Brown, P. (2008). Verb specificity and argument realization in Tzeltal child language. In M. Bowerman & P. Brown (Eds.), *Crosslinguistic perspectives on argument structure: Implications for learnability* (pp. 167–189). Erlbaum.

Brown, P. (2011). The cultural organization of attention. In *The handbook of language socialization* (pp. 29–55). Wiley Online Library. https://doi.org/10.1002/9781444342901.ch2

Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. In *Cambridge handbook of linguistic anthropology* (pp. 187–226). Cambridge University Press. https://doi.org/10.1017/CBO9781139342872.010

Brown, P., Gentner, D., & Braun, K. &. (2005). *Initial acquisition of nouns and verbs in Tzeltal speaking children [Paper presentation]*. Tenth International Congress for the Study of Child Language, Berlin, Germany.

Byers-Heinlein, K., Gonzalez-Barrero, A. M., Schott, E., & Killam, H. (2024). Sometimes larger, sometimes smaller: Measuring vocabulary in monolingual and bilingual infants and toddlers. *First Language, 44*(1), 74–95. https://doi.org/10.1177/01427237231204167

CDI Advisory Board. (2015). Adaptations in other languages. Retrieved from https://mb-cdi.stanford.edu/adaptations.html

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language, 5*(3), 457–475. https://doi.org/10.1017/S0305000900002099

De León, L. (1999a). Verb roots and caregiver speech in early Tzotzil Mayan acquisition. In B. Box, D. Jurafsky, & L. Michaelis (Eds.), *Cognition, discourse, and function* (pp. 99–119). CSLI/University of Chicago Press.

De León, L. (1999b). Verbs in Tzotzil (Mayan) early syntactic development. *International Journal of Bilingualism, 3*(2-3), 219–240. https://doi.org/10.1177/13670069990030020601

De León, L. (2001). Finding the richest path: The role of language and cognition in the acquisition of verticality in Tzotzil. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 544–565). Cambridge University Press. https://doi.org/10.1017/CBO9780511620669.020

deMayo, B. E., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C. F., Frank, M. C., & Marchman, V. (2021). Web-CDI: A system for online administration of the MacArthur-ates communicative development inventories. *Language Development Research, 1*(1), 55–98. https://doi.org/10.34842/kr8e-w591

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), Serial no. 242, i–185. https://doi.org/10.2307/1166093

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project*. MIT Press. https://doi.org/10.7551/mitpress/11577.001.0001

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.003.0022

Hamadani, J. D., Baker-Henningham, H., Tofail, F., Mehrin, F., Huda, S. N., & Grantham-McGregor, S. M. (2010). Validity and reliability of mothers' reports of language development in 1-year-old children in a large-scale survey in Bangladesh. *Food and Nutrition Bulletin, 31*(2), S198–S206. https://doi.org/10.1177/15648265100312S212

Hellwig, B., Defina, R., Kidd, E., Allen, S. E. M., Davidson, L., & Kelly, B. F. (2021). Child language documentation: The sketch acquisition project. In G. Haig, S. Schnell, & F. Seifart (Eds.), *Doing corpus-based typology with spoken language data: State of the art* (pp. 29–58). University of Hawai'i Press.

Información Cultural, R. N. de. (2020). Tseltal. Gobierno de México. Retrieved from https://sic.gob.mx/ficha.php?table=inali_li&table_id=50

Jackson-Maldonado, D., Friend, M., Marchman, V. A., Weisleder, A., Auza, A., Conboy, B., Rubio-Codina, M., & Dale, P. S. (2024). The MacArthur Inventario del Desarrollo de Habilidades Comunicativas III: A measure of language development in Spanish-speaking two- to four-year-olds. *Language Development Research, 4*(1), 361-398. https://doi.org/10.34842/12fb-qz93

Jarůšková, L., Smolík, F., Chládková, K., Oceláková, Z., & Paillereau, N. (2023). How to build a communicative development inventory: Insights from 43 adaptations. *Journal of Speech, Language, and Hearing Research, 66*(6), 2095–2117. https://doi.org/10.1044/2023_JSLHR-22-00591

Kachergis, G., Francis, N., & Frank, M. C. (2023). Estimating demographic bias on tests of children's early vocabulary. *Topics in Cognitive Science, 15*(2), 303–314. https://doi.org/10.1111/tops.12635

Kelly, B., Wigglesworth, G., Nordlinger, R., & Blythe, J. (2014). The acquisition of polysynthetic languages. *Language and Linguistics Compass, 8*(2), 51–64. https://doi.org/10.1111/lnc3.12062

Kik, A., Adamec, M., Aikhenvald, A. Y., Bajzekova, J., Baro, N., Bowern, C., ... Novotny, V. (2021). Language and ethnobiological skills decline precipitously in Papua New Guinea, the world's most linguistically diverse nation. *Proceedings of the National Academy of Sciences, 118*(22), e2100096118. https://doi.org/10.1073/pnas.2100096118

Maffi, L. (2005). Linguistic, cultural, and biological diversity. *Annual Review of Anthropology, 34*(1), 599–617. https://doi.org/10.1146/annurev.anthro.34.081804.120437

Marchman, V. A., & Dale, P. S. (2023). The MacArthur-Bates communicative development inventories: Updates from the CDI advisory board. *Frontiers in Psychology, 14*, 1170303. https://doi.org/10.3389/fpsyg.2023.1170303

Marchman, V. A., Dale, P. S., & Fenson, L. (2023). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual, 3rd edition*. Brookes. https://doi.org/10.3389/fpsyg.2023.1170303

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science, 14*(4), 769–785. https://doi.org/10.1111/j.1467-7687.2010.01024.x

Mitchell, A., & Jordan, F. M. (2021). The ontogeny of kinship categorization. *Journal of Cognition and Culture, 21*(1-2), 152–177. https://doi.org/10.1163/15685373-12340101

O'Shannessy, C., Davis, V., Foster, D., Bartlett, J., & Nelson, A. (2024). Development of a CDI for multilingual children speaking Indigenous languages in Central Australia. International Association for the Study of Child Language (Prague, Czechia). Retrieved from https://little-kids-learning-languages.net/

Polian, G. (2013). *Gramática del Tseltal de Oxchuc [Grammar of Oxchuk Tseltal]*. Centro de Investigaciones y Estudios Superiores en Antropología Social.

Polian, G. (2017). Tseltal and Tsotsil. In J. Aissen, N. England, & R. Zavala Maldonado (Eds.), *The Mayan languages* (pp. 610–647). Routledge. https://doi.org/10.4324/9781315192345-22

Prado, E. L., Phuka, J., Ocansey, E., Maleta, K., Ashorn, P., Ashorn, U., ... Dewey, K. G. (2018). A method to develop vocabulary checklists in new languages and their validity to assess early language development. *Journal of Health, Population and Nutrition, 37*(13), 1–10. https://doi.org/10.1186/s41043-018-0145-1

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Applied Statistics, 54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Southwood, F., White, M. J., Brookes, H., Pascoe, M., Ndhambi, M., Yalala, S., ... Alcock, N. (2021). Sociocultural factors affecting vocabulary development in young South African children. *Frontiers in Psychology, 12*, 642315. https://doi.org/10.3389/fpsyg.2021.642315

Tan, A. W. M., Marchman, V. A., & Frank, M. C. (2024). The role of translation equivalents in bilingual word learning. *Developmental Science, 27*(4), e13476. https://doi.org/10.1111/desc.13476

Vogt, P., Mastin, J. D., & Aussems, S. (2015). Early vocabulary development in rural and urban Mozambique. *Child Development Research, 1*, 89195. https://doi.org/10.1155/2015/189195

Weber, A. M., Marchman, V. A., Diop, Y., & Fernald, A. (2018). Validity of caregiver-report measures of language skill for Wolof-learning infants and toddlers living in rural African villages. *Journal of Child Language, 45*(4), 939–958. https://doi.org/10.1017/S0305000917000605

Data, Code and Materials Availability Statement

All data, code, and materials are available from https://osf.io/g2spx/.

Ethics Statement

Ethics approval was obtained from the IRB of the University of Chicago (IRB23-0244). All participants gave informed verbal consent before taking part in the study, as described in the Procedure subsection of the Methods section.

Authorship and Contributorship Statement

Penelope Brown created the first ever Tseltal-adapted checklist, prior to this article. Collectively, **Humbertina Gómez Pérez**, **Juan Méndez Girón**, **Ruthe Foushee**, **Gilles Polian**, and **Marisa Casillas** iteratively revised the checklist. **Humbertina Gómez Pérez**, **Ruthe Foushee**, and **Marisa Casillas** collected the data, aided by those mentioned in the Acknowledgements. **Kennedy Casey** and **Marisa Casillas** digitally pre-processed the data. **Kennedy Casey**, **Ruthe Foushee**, and **Marisa Casillas** conducted and checked the analyses. **Marisa Casillas** wrote the first draft of the analyses. All authors commented on revisions of the manuscript thereafter. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

This work was supported by an NSF CAREER award to MC (2238609). We are indebted to Sebastiana Velasco Gutierrez, Micaela Velasco Gutierrez, Antonieta López Santiz, and the participating families who contributed to this dataset. Special thanks go to Mika Braginsky, Mike Frank, Adriana Weisleder, and Virginia Marchman for advice on analyses. Special thanks, too, to Philip Dale for advice on describing our checklist in relation to the CDI. We are deeply grateful to Dedre Gentner and Penny Brown for sharing their original adapted Tseltal checklist. This work was partly presented at IASCL and BUCLD 2024.

Appendix A: Links to Further Materials

Find the current and archived checklist, instructions, anonymized data, scripts, and more at one of the links below.

1) Open Science Framework (OSF) repositories:
   a) https://osf.io/z8hdk contains...
      i) **Method preregistration:** the pre-registered methods for developing the Tseltal checklist reported in this paper (up to version 6).
   b) https://osf.io/g2spx/ contains versioned archives of...
      i) **Data:** the anonymized data and participant for the children's 84 checklists reported on in the current paper,
      ii) **Scripts:** all associated analysis scripts (including what is required to reproduce this document),
      iii) **Checklists:** blank printable and editable copies of the most current checklist version(s) (in pdf, docx, and xlsx),
      iv) **Instructions:** blank printable and editable copies of the most current checklist instructions (in pdf and docx),
      v) **Consent example:** a pdf copy of the consent page we used for verbal consent in the currently collected data,
      vi) **Study metadata:** a general description of the study and reference to this paper for more details,
      vii) **Contact information:** up-to-date contact information for those who have follow-up questions
2) The CIESAS website (https://sureste.ciesas.edu.mx/polian-gilles/) contains archives of...
   a) **Checklists:** blank printable and editable copies of the most current checklist version(s) (in pdf, docx, and xlsx),
   b) **Instructions:** blank printable and editable copies of the most current checklist instructions (in pdf and docx),
   c) **Consent example:** a pdf copy of the consent page we used for verbal consent in the currently collected data,
   d) **Study metadata:** a general description of the study and reference to this paper for more details,
   e) **Contact information:** up-to-date contact information for those who have follow-up questions
3) The WordBank repository (https://wordbank.stanford.edu/) contains...
   a) **Data:** the anonymized data and participant for the children's 84 checklists reported on in the current paper,
   b) **Study metadata:** a general description of the study and reference to this paper for more details,

Contact information: up-to-date contact information for those who have follow-up question

Appendix B: Alternative Age of Acquisition Estimates

We were unable to establish an AoA based on proportional production (≥ 0.5 production) for 11.9% ($N = 30$) of the items on the checklist. As one reviewer pointed out, we can alternatively use binomial regression to estimate age of acquisition for all the items on our checklist, including the 11.9% that yielded no AoA in the current sample. We ran a logistic mixed-effects regression of whether or not a child produced an item (1/0) that included a fixed effect of child age in months (numeric) and a random effect of checklist item (factor). We then used the ggeffects package in R (Lüdecke, 2018) to estimate an AoA for each item. Below we plot the AoA distributions, which range from 5 months (unrealistic) to 31 months, with AoAs for most words sitting between 12 and 30 months. Peak acquisition rates for this list were estimated to be between 18 and 24 months.

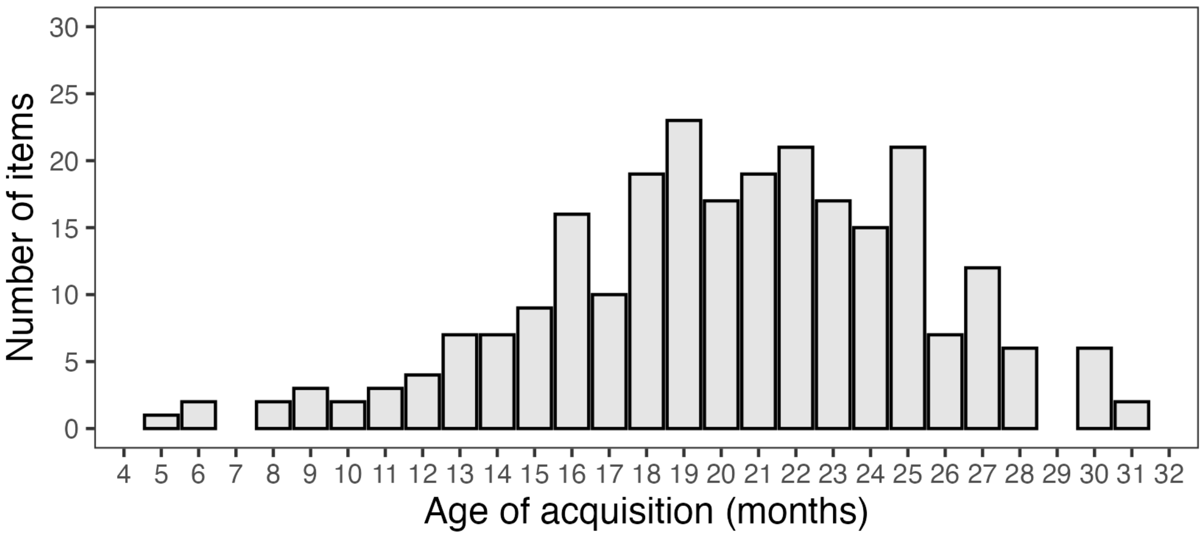

Figure SM1. Distribution of model-estimated age of acquisition—the first age with ≥ 50% of a sample estimated to produce the item—over all items.

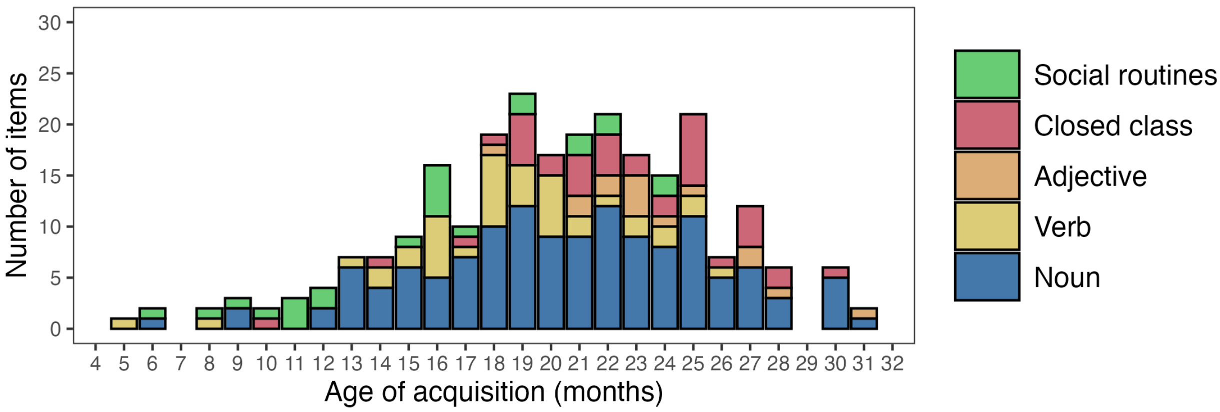

Figure SM2. Distribution of model-estimated age of acquisition over all items, separated by major syntactic types (Noun, Verb, Adjective, Closed class, Social routines).

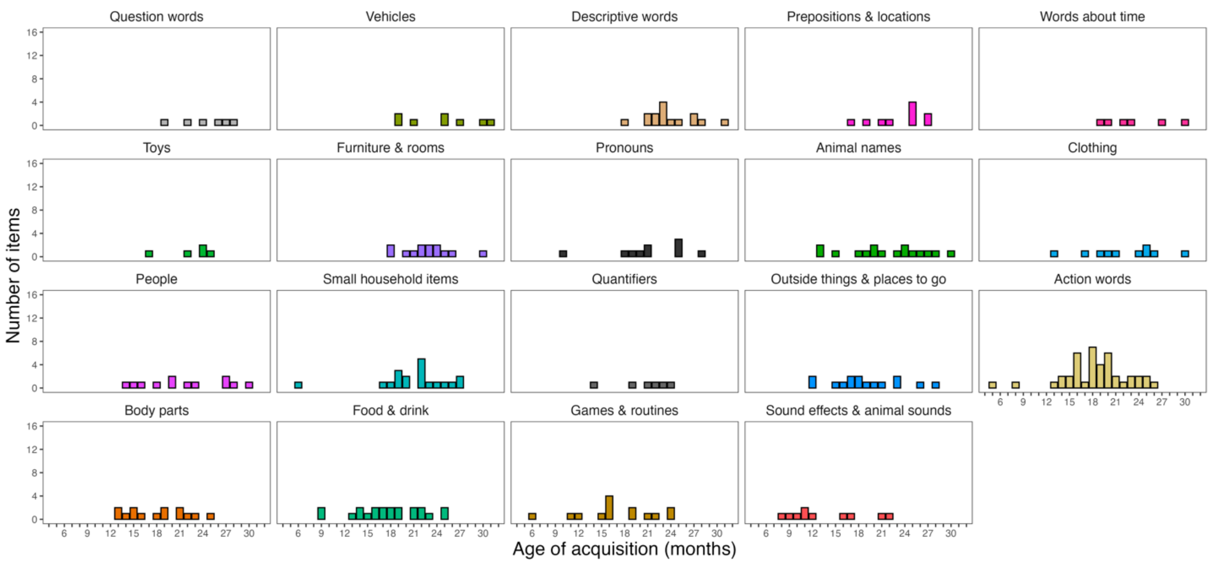

Figure SM3. Distribution of model-estimated age of acquisition over all items within each content sub-category of the list, sorted from the latest acquired (upper left; Question words) to the earliest acquired (lower right; Sound effects & animal sounds).

Appendix C: Data Pre-Processing and Marking of Alternatives

Participant responses were recorded by an experimenter in real time, either digitally or on paper (see Methods for details). These individual responses were then copied over (by MC or KC) to a single spreadsheet that we use for analysis. We passed through the spreadsheet many times to identify potential errors that emerged during data recording or data transfer. Experimenters are moving fast through the list during the interview, and so—while the vast majority of responses are sensible and interpretable—there were occasional inconsistencies in how responses were recorded. First, and most commonly, an onomatopoeic form was accepted in place of its associated noun label (e.g., "moo" for "cow"). We trained experimenters to verbally accept these responses verbally but to not write them down, considering that these onomatopoeic forms were already captured in the Sounds category. Second, mothers reported children producing a wide variety of alternative referring forms children use in place of the list item we asked about (e.g., "darkness" for "night"). To address these inconsistencies adequately, we created a marking system:

- **Canonical:** word forms in Tseltal and/or relating directly to the standard forms offered in the list (which sometimes come from Spanish). These *canonical forms* are spelled orthographically in the coding, with no additional mark up.
- **In-language alternative:** productions recognizable as pragmatically and semantically appropriate equivalent forms in Tseltal—but differing from the standard, expected item word form—were accepted if they could feasibly be the dominant way of referring to this concept in that child's family. These *in-language alternative forms* are spelled orthographically in the coding, and are enclosed in a single pair of parentheses.
- **Other-language alternative:** productions in Spanish or another language that were recognizable as pragmatically and semantically appropriate equivalent forms in Tseltal—but differing from the standard, expected item word form—were accepted if they could feasibly be the dominant way of referring to this concept in that child's family. These *other-language alternative forms* are spelled orthographically in the coding, and are enclosed in a double pair of parentheses. Note that many of the standard list items are shared directly with Spanish; those items are considered "canonical" Tseltal productions, since they represent expected borrowings.
- **Excluded:** productions that did not satisfy the research aim of identifying children's recognizable target wordforms for the items on the list. These most often included onomatopoeia as substitutes for target items (e.g., "moo" for "cow") but also included non-adult-like over-extensions (e.g., "car" for "taxi") and the production of an associated word in place of the target word (e.g., "bite" for "snake"). These *excluded forms* are spelled orthographically in the coding, and are enclosed in a single pair of square brackets. We note that, while the decision to exclude these responses may under-count some children's productive vocabularies (if, e.g., other researchers find onomatopoeic substitutes acceptable), they make up 0.66% (less than 1%) of the checklist responses we gathered.

Three other formatting decisions were made in conducting digital data entry that facilitate the use of this parenthesis-based coding and the identification of unique forms per list item:

- multi-word responses are separated by underscores (e.g., “ya_xban”)
- multi-alternative responses are ordered as follows: canonical > (in-language alternative) > ((other-language alternative)) > [excluded]
- recorded forms that *only* varied based on non-phonological, non-meaningful variation in continuant length were collapsed into a single form (e.g., for the sound a car engine makes: “rr”, “rrrr”, “rrr” were all converted to “rrrr”)

References Cited

Lüdecke, D. (2018). Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software, 3*(26), 772. https://doi.org/10.21105/joss.00772

License

The role of the environmental context in shaping teachers' linguistic input

Nicola Lester
Katherine E. Twomey
Anna Theakston
University of Manchester, UK

Abstract: While decades of research have demonstrated that the quality of linguistic input children receive from adults has significant effects on their language development, more recent work suggests that the quality of that input is affected by the environmental context in which it is delivered. In the current study, six teachers were audio-recorded teaching four- and five-year-old children similar content in either a museum, their classroom with museum resources, or their classroom with typical classroom resources. Quality of input was measured in terms of the proportion of decontextualised talk, *wh-* questions, rare words and multi-clausal sentences produced. Teachers produced a significantly higher proportion of decontextualised talk when teaching in the museum compared to teaching in the classroom with regular classroom resources. However, teachers used the highest proportion of rare words when teaching with museum resources in the classroom compared to the other two contexts. These data demonstrate that different learning contexts lend themselves to different aspects of high-quality input, with implications for children's language development.

Keywords: linguistic input; context; learning environment; museum; classroom

Corresponding author(s): Nicola Lester, School of Environment, Education and Development, University of Manchester, Oxford Road, Manchester, M13 9PL, UK.
Email: nicola.lester@manchester.ac.uk

ORCID ID(s): Nicola Lester: https://orcid.org/0000-0002-2909-734X; Katherine E. Twomey: https://orcid.org/0000-0002-5077-2741; Anna Theakston: https://orcid.org/0000-0002-9483-7893

Citation: Lester, N., Twomey, K. E., & Theakston, A. (2025). The role of environmental context in shaping teachers' linguistic input. *Language Development Research*, *5*(3), 276-301. http://doi.org/10.34842/ldr2025-923

Introduction

In recent decades, extensive research has been dedicated to uncovering the factors which may affect the rate at which children successfully acquire the necessary language skills to thrive. A critical factor is the linguistic input children receive, and its profound effect on children's early language development has been repeatedly demonstrated (e.g., Hart & Risley, 1995; Hoff-Ginsberg, 1985; Huttenlocher et al., 2002; Rowe, 2017). Importantly, many studies have demonstrated that quality, not just quantity, of linguistic input plays a key role (e.g. Rowe, 2012). Importantly, however, recent work has demonstrated that the environmental context linguistic input is received in can also affect language learning (Lester, Theakston & Twomey, 2023; Tamis-LeMonda et al., 2019). However, while the focus of research has largely been on parental language, a substantial amount of language input to school-age children comes from their teachers. Thus, in the present exploratory study, we examined whether there was a relationship between the quality of linguistic input children receive from their teachers and the context in which they receive it, with a particular focus on museum contexts and teaching resources. Below, we review existing evidence demonstrating the importance of language input for language development and discuss the effects of input quality (the richness and complexity of the language children hear) on early acquisition.

Linguistic Input from Caregivers

An extensive body of research has shown that language development in children is influenced by the linguistic input that they receive. There is now a consensus that a rich input, in terms of quantity, diversity, and complexity, promotes language development and leads to more rapid language acquisition (Hoff & Naigles, 2002; Rowe, 2008). Much previous research on the effects of linguistic input has focused on the sheer quantity of words (word tokens) heard by children and suggests this is positively correlated with vocabulary growth (Hart & Risley, 1995; Huttenlocher et al., 1991). However, recent years have seen a growing interest in the role of quality in linguistic input. Previous studies have defined input quality in a variety of ways, uncovering positive relationships between linguistic quality and children's language development across a range of language areas and stages of development (e.g. Blything et al., 2019; Huttenlocher et al., 2002; Rowe, 2012; Salo et al., 2016; Uccelli et al., 2019; Weizman & Snow, 2001).

For instance, there is accumulating evidence that the number of open-ended questions addressed to children is positively related to their language development (Hoff-Ginsberg, 1985; Rowe et al., 2017; Shatz et al., 1989). Open-ended questions, often characterised as *wh-* questions—that is, questions that begin with who, what, where, when, why or how—are thought to provoke more thoughtful and sophisticated responses compared to closed questions which require a simple yes or no answer (Ping, 2014). Moreover, when children attend to non-verbal environmental cues such as their caregiver's eye-gaze or potential referents of new words whilst simultaneously being exposed to *wh-* questions, they are given the opportunity to practice linking referents to objects. In particular, even in cases where children do not know the answer to a question, they are able to use both the question and non-verbal cues to determine the answer, thus developing their verbal reasoning skills (Rowe et al., 2017). Increased parental use of *wh-* questions has been positively related to children's auxiliary verb usage (Hoff-Ginsberg, 1985), vocabulary size (Blake et al., 2006; Rowe et al., 2017; Salo et al., 2016) and better verbal reasoning skills (Rowe et al., 2017). This growing body of evidence highlights the importance of *wh-* questions in linguistic input, and their role in developing and enhancing the language skills of the young language learner.

However, other research highlights the importance of alternative facets of input quality. In particular, an increasing number of studies show that parents' use of decontextualised talk positively enhances children's language acquisition (Gillam et al., 2012; Rowe, 2013; Rowe, 2012). Decontextualised talk is defined as language used to discuss absent objects or events occurring in the past or future (Snow, 1990), and often takes place when speakers are engaged in narrative, explanatory or pretend discussion (Snow et al., 2001). Decontextualised talk is considered more conceptually complex than contextualised talk, as children cannot rely on the visual context for comprehension, and it often includes more sophisticated vocabulary (Rowe, 2013). The proportion of decontextualised talk in linguistic input is positively related to children's later vocabulary, narrative comprehension and production (Beals, 1997; Katz, 2001; Reese et al., 2010; Snow et al., 2001), reading comprehension (Snow et al., 2001) and ability to provide formal definitions (Katz, 2001), skills that are essential to children's literacy development (Cunningham & Stanovich, 1997). Specifically, Rowe (2012) found that four-year-old children who were exposed to a higher proportion of parental decontextualised talk, particularly narrative utterances, had larger vocabularies one year later compared to those who heard less decontextualised talk. Rowe suggested that by exposing their children to narrative about topics that are removed from the present, parents challenge children to think abstractly and scaffold their child's ability to produce similar discourse, thus promoting their child's vocabulary development.

In addition to exploring the effects of decontextualised talk, Rowe (2012) also found that the proportion of sophisticated vocabulary (defined as words not commonly known by children aged 9-10 years) used by parents with their 30-month-old children was a stronger predictor of children's vocabulary size one year later than the quantity of talk that children were exposed to. Rowe argues that by this age, children have had substantial exposure to commonly used words and are ready to acquire more difficult and sophisticated vocabulary. Thus, sophisticated vocabulary is more beneficial to children of this age in terms of vocabulary size than sheer quantity of talk. Similarly, Weizman and Snow (2001) found that the proportion of sophisticated words heard by

5-year-old children during interactions with their mother's predicted variance in vocabulary skill during their first and third year of school (see also Beals, 1997).

Linguistic Input in Childcare or School Settings

Much of the existing research focuses on the quality of the linguistic input children receive from their primary caregivers who are often a child's initial source of input (although largely in WEIRD [Western, Educated, Industrialised, Rich and Democratic] cultures). However, as soon as children start full-time education, linguistic input from teachers constitutes a significant proportion of the language they hear: children spend an average of 32.5 hours a week in a typical UK classroom over the course of an academic school year (Department for Education, 2022). Consequently, researchers have begun to recognise that a significant proportion of a child's linguistic input may come from within their childcare setting or school, and that there is a consequent need to understand the nature of this input and its effects on language development. Studies focusing on the nature of the input children receive from teachers participating in normal teaching or classroom-based conversational activities have shown that when children receive high quality input from teachers (specifically, a higher proportion of multi-clausal sentences, decontextualised talk, and use of more varied word types), their communication skills, language development, and reading abilities accelerate (e.g., Bowers & Vasilyeva, 2011; Dickinson & Porche, 2011a; Huttenlocher et al., 2002). For instance, Huttenlocher et al. (2002) found that the quality of teachers' linguistic input to three- to four-year-old children was positively related to the children's syntactic development; when teachers used a higher proportion of multi-clausal sentences, children's comprehension of these structures increased over the school year. In addition, lexical richness – that is, the number of different word types used by teachers – predicted children's vocabulary growth (Bowers & Vasilyeva, 2011), and teachers' use of decontextualised talk has been found to have similar positive effects on children's vocabulary production and comprehension (De Temple & Snow, 2003; Mascareño et al., 2016). These effects have been found to last beyond early primary school (Burchinal et al., 2008; Dickinson & Porche, 2011a; Mashburn et al., 2009) and have also been observed in pre-school contexts where the language of education was different from the children's home language (Gámez, 2015, see also Bowers and Vasilyeva, 2011).

Linguistic Input and Context

So far, we have discussed evidence that the quality of the linguistic input provided by teachers in normal classroom interactions and during book reading activities can have a positive impact on children's language development. However, the linguistic input children receive, and thus the quality of that input, can be affected by the context in which it is provided. Again, the vast majority of research on how linguistic input differs between contexts focuses on the input provided by primary caregivers (e.g., Noble et al., 2018; Tamis-LeMonda et al., 2017; Tamis-LeMonda et al., 2019). However, there is a handful of studies that compare teacher input across contexts, to begin to address this gap in the literature (e.g., Chaparro-Moreno et al., 2022; Cote, 2001; Massey et al., 2008). For instance, differences in teachers' linguistic input have been found between different activities in the classroom. In a corpus analysis of 2,928 utterances in 453 conversations that 13 preschool children had with peers and teachers during a typical school week, Chaparro-Moreno and colleagues (2022) found that in activities that were mainly led by teachers such as circle time (whole class discussion) and reading, teachers used more decontextualised talk compared to activities such as centre time, where children move around more freely and have more autonomy over their interactions. Similarly, Farrow, Wasik and Hindman (2020) found that teachers used more complex sentences during book reading, compared to the 'morning message' or other small group activities. Further, Cote (2001) found different patterns in teachers' and children's use of sophisticated vocabulary in different learning contexts. Children used significantly more sophisticated vocabulary during circle time compared to book reading, mealtime, and free play. In contrast, teachers used significantly more sophisticated vocabulary during free play than at mealtime, and significantly less during book reading than mealtime (excluding the words from the book itself in the analysis).

Gest et al. (2006) explored how teachers' linguistic input differed across book reading, free play and mealtime contexts with three- and four-year-old children from 20 Head-Start classrooms in the USA. They found that teachers provided the highest rate of rich and complex child-directed talk (in terms of the variety of vocabulary, the proportion of extended, elaborated utterances, and the introduction of new concepts, ideas and information) during book reading, a finding they suggested was due to the new and varied language teachers used when asking and answering questions about a new book. In contrast, teachers provided the highest rate of pretend talk in the free play context, as they often assumed the role of play enhancer, encouraging pretend and imaginative play amongst children. Teachers used less rich and challenging talk, but more decontextualised talk at mealtimes compared to the book reading and free play contexts, which likely reflected the standard mealtime etiquette.

There is some evidence from more detailed studies of book reading in classrooms to suggest that these aspects of teacher input quality translate to children's vocabulary acquisition. For example, several studies show that providing explicit definitions of target words during book reading activities is positively related to pre-school and reception-aged children's receptive and expressive vocabulary growth (Beck & McKeown, 2007; Coyne et al., 2009; Gonzalez et al., 2010; Hadley et al., 2016; Wasik et al., 2016). There is also substantial evidence to suggest that incorporating questions into book-reading interventions is beneficial to word learning (Dickinson & Porche, 2011b; Dickinson & Smith, 1994; Sénéchal, 1997). For example, Sénéchal (1997) found that three- and four-year-old children's expressive and receptive vocabulary gains were greater when teachers incorporated 'what' or 'where' questions into their book-reading activities, compared to when teachers focused solely on producing the words contained in the book, and in a study with 14 teachers and their teaching assistants in their classrooms with their 4-year-old students, Massey and colleagues (2008) found that teachers used more cognitively challenging questions during shared book reading, compared to any other classroom activity. Cognitively challenging questions were defined as those that were conceptually focused (i.e., focused on non-present objects or past and future events) and included eliciting inferences or predictions, analysing information, and discussing vocabulary in this study.

Importantly, although much of the linguistic input from teachers is provided in the context of formally taught lessons in a classroom environment, this is not always the case (Barnes et al., 2020). Although this body of literature goes some way to uncovering how linguistic input from teachers differs between different classroom-based activities, to-date, few studies have explored whether delivering teaching outside of the classroom environment has any effect on the quality of the linguistic input children receive. In particular, museums are a popular destination for school trips and may offer enhanced learning experiences for children (Henderson & Atencio, 2007). Indeed, some researchers have suggested that information-rich museum environments provide an ideal learning context to promote language learning in young children (Henderson & Atencio, 2007; Kola-Olusanya, 2005; Rodriguez & Tamis-LeMonda, 2011). Henderson and Atencio (2007) proposed that museums' attractive exhibits and the presentation of open-ended questions promote conversation amongst visitors, and encourage children to actively explore their environment, thus creating language learning opportunities. One possibility is that museums are contexts that encourage caregiver/teacher language input that is high in quality, for example by promoting the use of rare words, *wh-* questions and decontextualised talk. It is therefore important to understand how different educational contexts, that is, different environments/locations and/or teaching resources and activities, may affect the linguistic input that teachers provide.

The Present Study

In the present study we aim to close this gap in the literature by exploring whether the quality of the linguistic input teachers provide differs when delivering the same learning activity (a) in a classroom with regular classroom resources (b) in classroom with museum resources, and (c) in a museum context with access to a wider variety of resources such as exhibits, information signs, and knowledgeable staff. In line with previous research, we chose to measure language quality in a variety of ways that have previously been found to enhance children's language development. Specifically, we asked whether the proportion of *wh-* questions, utterances containing decontextualised talk, multiclausal sentences and rare words used by teachers differed between contexts.

Although previous work suggests that there could be effects of context on linguistic input, the present study is exploratory in nature, aiming to highlight what these effects may be. Classroom-based teaching shows differences in the quality of input as a function of the learning activity (e.g., Chaparro-Moreno et al., 2022; Gest et al., 2006; Massey et al., 2008), though it is unclear whether this is affected by the types of classroom resources used in any given activity. Moreover, while classroom resources often include books which promote quality talk (e.g. Cameron-Faulkner & Noble, 2013; Wasik et al., 2016), museums provide enriched visual, unique learning contexts which could promote more decontextualised talk (e.g., Kola-Olusanya, 2005). Thus, for the current study, we do not make any directional hypotheses concerning which contexts might promote different types of high-quality input.

Method

Participants

Our sample consisted of six female teachers all of whom taught reception-aged children [4-5 years] at an inner-city primary school in the North West of England, UK, with a majority of pupils learning English as an additional language. Participants were recruited through a partnership between the school where they worked, a local museum and a university, and took part on a voluntary basis. Participants provided informed, written consent prior to taking part in the study. Participants were between the ages of 21-44 years. Participants had an average of 11.5 years' teaching experience (SD = 4.68). Five participants were monolingual English-speaking, and one participant was a bilingual English and Urdu speaker. English was the first language of all participants, and all teaching was delivered in English.

Procedure

Each session was delivered to a unique group of 11 to 14 children from the year group, with no child participating in more than one session. A single teacher led each session, although individual teachers delivered multiple sessions across different groups as indicated below. Due to Covid-19 and the associated restrictions, half of the sessions were taught prior to the Covid-19 outbreak with the 2019 Reception cohort of children, and half of the sessions were taught post-Covid-19 outbreak with the 2020 Reception cohort. Due to one failed recording, the final dataset consisted of 11 recorded sessions - three in the museum, three in the classroom with museum resources, and five in the classroom with regular classroom resources (see Table 1). For the two teachers who taught in both the museum context and the classroom with regular classroom resources context, one taught in the museum first, followed by the classroom with regular classroom resources, and the other taught in the classroom with regular classroom resources first, followed by the museum. For the three teachers who taught post-COVID-19 pandemic, two taught in the classroom with museum resources first, followed by the classroom with regular classroom resources, and the other teacher taught in the classroom with regular classroom resources first, followed by the classroom with museum resources.

Table 1. ***Number of sessions taught in each context by participant.***

Participant	Number of sessions taught in each context		
	Museum	Classroom with museum resources	Classroom with regular classroom resources
Teacher 1	1	0	0
Teacher 2	1	2	1
Teacher 3	1	0	1
Teacher 4	0	0	1
Teacher 5	0	1	1
Teacher 6	0	1	1

All teaching sessions pre-Covid-19 were taught over the course of three consecutive days; thus, the teachers who taught their sessions prior to the COVID-19 pandemic taught their second session either one day or two days after teaching their first session. All teaching sessions post-Covid-19 were taught over two consecutive days meaning teachers taught their second session the day after teaching their first. Teacher 2 taught four sessions in total as she had a class of children in the target year group both pre-Covid-19 and post-Covid-19. Teachers were unaware of the purpose of the study until they were debriefed at the end, and the lead researcher was present during all teaching sessions, though they were not directly involved with the session.

All teaching sessions were themed around the topic of insects, and the central goal of all sessions for all teachers was to teach children six novel vocabulary items (names of insects). The researcher (first author) instructed each teacher to spend approximately 20 minutes per session teaching these vocabulary items using the resources available in each context. While teachers were given flexibility in how they used the resources, they independently chose to use them in very similar ways across the three contexts. When teaching in the museum, teachers could use the insect displays in the exhibit: boxed specimens of each of the target insects that had been preserved for educational purposes, small plastic magnifying glasses, and a microscope. When teaching in the classroom using museum resources, teachers used the boxed specimens from the museum. When teaching in the classroom using standard classroom resources, teachers used coloured, laminated images of insects, and a book titled Everything Insects by Carrie Gleason (produced as part of the National Geographic for Kids range). The six participants collectively agreed prior to the beginning of the study that these resources reflected those that they would typically use in the classroom. Teachers were told they were free to use the resources in a way that reflected their individual teaching style, which ensured conditions were as natural as possible whilst controlling the materials used. Each taught session lasted approximately 20 minutes. Teachers were audio recorded using an Olympus DS-3500 digital voice recorder worn on a lanyard around their neck.

Transcription

All recordings were transcribed by the lead researcher verbatim in English using CLAN (MacWhinney, 2000). Only the teachers' speech was transcribed – any audible child speech was omitted. The flow of teacher speech was divided into utterances. We defined 'utterance' as having a single intonational contour within a single conversational turn and consisting of one or more syntactic units (e.g., constituents or clauses). An utterance was usually preceded and followed by a pause (Huttenlocher et al., 2010). A second coder transcribed a random 10% of the recordings. Agreement on exact wording and the breakdown of speech into utterances averaged 92% (range = 85% - 97%).

Analysis of speech

Transcripts were coded and analysed by the first author. No transcribed utterances were excluded from the analysis. Transcriptions were coded for quality of teacher talk. Specifically, in order to measure quality, transcriptions were coded for utterances containing *wh-* questions, decontextualised talk, multi-clausal sentences and rare word vocabulary (a measure of lexical richness), as defined in the coding scheme below. Finally, we performed between-setting comparisons for each measure.

Coding Scheme:

We created a coding scheme based on those previously used by Gest et al. (2006), Huttenlocher et al. (2002; 2010), Noble et al. (2018), Rowe (2012), and Snow et al. (2001). For our measures of *wh-* questions, decontextualised talk and multi-clausal sentences, we coded every utterance spoken by teachers as either containing each of the variables of interest (coded 1) or not (coded 0) to enable analyses to be conducted on the proportion of use of each measure as a function of the overall number of utterances produced. For our rare words measure, it was necessary to collapse the data across participants and teaching sessions to determine the total proportion of rare words used in each context (i.e. total rare words used per context / total words used per context). These measures were defined in the following ways:

***Wh*-Questions:** utterances transcribed with a question mark and having *wh*-question syntax in the main clause- who, what, where, why, when and how (e.g.

"What can you see?" "Where are its wings?").

Decontextualised Talk. talk about people, places, things, and events that that happened in the past or will happen in the future (Gest et al., 2006; Rowe, 2012) (e.g. "There was one of these in my garden once" "I was bitten by one of these on holiday"). Talk about events that took place just minutes before but were no longer taking place in the present moment were also treated as decontextualised talk (e.g. "We've just looked at the beetles."). This is consistent with other definitions in the literature, such that decontextualised talk can encompass either a spatial or a temporal detachment from immediate context (Snow et al., 2001). Spatially detached talk encompassed any references to things not immediately present. This includes discussions about people or places that are not physically there (Gest et al., 2006), interpretations of others' intentions or mental states (DeTemple, 2001), and abstract language use such as offering definitions and explanations of concepts that cannot be directly observed (Rowe, 2012).

Multi-Clausal Utterances. utterances containing more than one clause (e.g. "Have a look inside the box and then pass it on."). Additionally, utterances that contain more than one lexical verb were coded as multi-clausal (e.g. "she sat and listened carefully.").

Rare Words. Word tokens with a Log Zipf Frequency of three or below on the Subtlex database (van Heuven et al., 2014) were considered rare. The Subtlex-UK word frequencies are based on a corpus of 201.3 million words from 45,099 BBC broadcasts. There are separate measures for primary school children (the CBBC channel), which was the measure used for the present study; van Heuven et al. also present the word frequencies as Zipf-values (values 1-3 = low frequency words; 4-7 = high frequency words).

Reliability of Coding:

Across all transcripts, there was a total of 4,879 utterances. A second coder coded a random selection of 10% of the transcripts. Agreement when coding *wh-* questions was 98.2% with a Cohen's kappa value of 0.92. Agreement when coding for decontextualised talk was 93.3% with a Cohen's kappa value of 0.66, and agreement when coding for multi-clausal sentences was 93.5% with a Cohen's kappa value of 0.58. As only a subset of the data was coded for reliability, no changes were made to the coding where disagreements occurred. Rare words were coded automatically against the SUBTLEX database using (version 4.1.1; R Core Team, 2021).

Results

Across the three 20-minute recordings in the museum, there was a mean of 490 utterances per recording (range: 424-549). Specifically, there was a mean of 433 utterances across the five 20-minute recordings in the classroom with classroom resources (range: 372-475), and a mean of 438 utterances across the three recordings in the classroom with museum resources (range: 302-585). Analyses of input quality measures between contexts were carried out using the lme4 package (Bates, Maechler, Bolker, & Walker, 2021) to fit generalised linear mixed effects models for the proportion of utterances containing *wh-* questions, decontextualised talk and multiclausal sentences in R (R Development Core Team, 2021). Restricted maximum likelihood estimation was used for the reporting of generalised linear mixed model parameters. We tested if the inclusion of an additional term was justified using the likelihood ratio test for model comparisons (Pinheiro & Bates, 2000). The factor 'Context' was coded using treatment contrast (the default coding in R), with the reference level 'Classroom Resources'.

Table 2 reports the mean proportions and standard deviations of each measure of quality in each context. Each measure of linguistic quality occurs, on average, in only a small proportion of utterances (6-12%). We were interested in whether the quality of the linguistic input teachers used differed between the contexts in which they taught. First, the three binary outcome variables (*wh-* questions, decontextualised talk and multi-clausal sentences) were submitted to separate generalised linear mixed effects models with fixed effects of context (museum, classroom with classroom resources, classroom with museum resources), with a random intercept for participant. This random intercept was included to account for any individual differences amongst teachers given the fact that not all teachers taught in all contexts, and the fact that the combination of contexts in which they taught differed. Pairwise comparisons were then run using the Emmeans package with Bonferroni correction to determine exactly where the differences in linguistic input between each context lay (Lenth et al., 2023).

Table 2. ***Mean proportions and standard deviations by variable and context (and proportion of rare words).***

Context	*wh-* Questions		Decontextualised Talk		Multi-Clausal Sentences		Rare Words	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Museum	0.09	0.04	0.11	0.02	0.11	0.03	0.01	0.09
Museum Resources	0.09	0.04	0.10	0.04	0.10	0.05	0.05	0.21
Classroom Resources	0.09	0.03	0.07	0.02	0.07	0.01	0.02	0.15

The first analysis did not find a significant effect of context on the proportion of *wh*-questions teachers used. Compared to the reference level context (classroom resources), teachers were not significantly more likely to produce *wh*- questions in the museum context (β =0.11, *SE* = 0.14, *z* = 0.77, *p* = .44), or in the museum resources context (β = −0.22, *SE* = 0.12, *z* = –1.81, *p* = .07). The model explained a very small amount of variance, with a marginal R^2 of 0.005 and a conditional R^2 of 0.05, indicating that fixed effects alone explained little variance and random effects contributed only modestly to the overall variance. A chi-squared comparison showed that a null model containing only the random effect of participant, did not statistically differ from a model that contained the fixed effect of context ($\chi^2(2) = 4.51$, $p = .11$).

A second analysis indicated that context significantly influenced the likelihood of decontextualised talk. Compared to the classroom resources context, teachers were significantly more likely to use decontextualised talk in the museum context (β = 0.26, *SE* = 0.12, *z* = 2.20, *p* = .028), but less likely to use decontextualised talk in the museum resources context (β = −0.37, *SE* = 0.13, *z* = –2.83, *p* = .0047). Although the effect was statistically significant, the beta coefficient was small, indicating that the size of the effect was modest, and indeed, the mean values across contexts ranged from 6-12% indicating that decontextualised talk was relatively rare in all contexts. This suggests that while context has a measurable influence on the likelihood of decontextualised talk, the practical impact may be limited. Moreover, the model explained a small proportion of variance overall, with a marginal R^2 of 0.01 and a conditional R^2 of 0.02, indicating that fixed and random effects contributed modestly to explaining variation in decontextualised talk.

Bonferroni-adjusted pairwise comparisons revealed that decontextualised talk differed significantly between the museum and museum resources contexts ($p < .001$), while the difference between classroom resources and museum resources was also significant (*p* = .014). However, the difference between classroom resources and museum did not reach statistical significance after correction (*p* = .083).

Additionally, a GLMM revealed a significant effect of context on the proportion of multiclausal utterances in teacher's linguistic input. Specifically, teachers were significantly more likely to produce multiclausal utterances in the museum resources context compared to the classroom resources context (β = 0.49, *SE* = 0.15, *z* = 3.26, *p* = .001). However, the difference between the museum and classroom resources contexts was not significant (β = 0.04, *SE* = 0.16, *z* = 0.25, *p* = .80). Post Hoc comparisons indicated that the proportion of multiclausal utterances was highest in the museum resources context (*M* = –2.07, *SE* = 0.18), followed by the museum (*M* =−2.52, *SE* = 0.18), and lowest in the classroom resources (*M* = −2.56, *SE* = 0.16). Bonferroni-adjusted pairwise comparisons showed that the museum resources context had significantly more multiclausal utterances than the classroom resources context (*p* =.003).

However, on average only 7-11% of utterances across contexts were multiclausal, and while significant, the β values were again relatively small, suggesting that other factors may play a more substantial role in explaining variance in the proportion of multiclausal utterances. The difference between the museum and museum resources contexts approached significance ($p = .052$), while the classroom resources–museum contrast was not significant ($p = 1.00$). Overall, the model explained a modest proportion of variance, with a marginal R^2 of 0.01 and a conditional R^2 of 0.04, indicating that fixed and random effects together accounted for a small amount of the variability in multiclausal utterance production.

To examine the relation between context and the teachers' use of rare versus non-rare words in their linguistic input, a chi-square test of independence was performed. As this measure was taken at the level of a full teaching session rather than on an utterance-by-utterance basis (because each utterance contained multiple different words), it was necessary to collapse the data across participants and teaching sessions to determine the proportional use of rare words in each context. The relation between these variables was significant, $\chi^2(2, N = 6) = 221.8$, $p < .01$. Teachers used the highest proportion of rare words when teaching in the classroom with museum resources, and the lowest proportion of rare words when teaching in the museum.

It was necessary to pause data collection for the present study due to the Covid-19 pandemic, resulting in all data from the museum and half of the classroom resources data being collected prior to the pandemic, with the remainder data being collected post-pandemic. Thus, it was essential to check whether taking part in the study prior to or post the first UK government lockdown significantly affected the quality of teachers' linguistic input, as this could mean that any differences found between contexts reflected unexpected effects of the pandemic rather than effects of the contexts themselves; for example, teachers may have been conscious of the need to alter their input post-pandemic as children had missed in-person school time. We found that taking part in the study prior to or post Covid-19, measured through the addition of a 'Covid' variable (pre-Covid, post-Covid) to the models, had no significant effect on the proportion of utterances containing *wh-* questions ($\chi^2(1) = 0.52$, $p = .47$) or decontextualised talk ($\chi^2(1) = 0.05$, $p = .82$) produced by teachers. However, when teachers took part in the study pre-Covid, they produced significantly fewer multi-clausal utterances compared to taking part in the study post-Covid ($\chi^2(1) = 9.12$, $p = .003$). The addition of the Covid variable to the model containing the fixed effect of multi-clausal sentences meant that the beta value for the fixed effect of context on this variable in the model was no longer significant, suggesting the two variables account for some of the same variance.

General Discussion

The present study explored whether the context in which teachers taught children affected the quality of their linguistic input. For the purposes of this study, a change in context in an environmental sense referred to a change in either the physical environment, the teaching resources, or both (but not a change to the content of the teaching session). Quality of linguistic input was assessed according to the proportional use of utterances containing *wh-* questions, decontextualised talk, multi-clausal sentences, and the proportion of rare words used in each teaching session. Teachers taught groups of four- and five-year-old children either in the museum, in their regular classroom setting with regular classroom resources, or in their regular classroom with resources from the museum. Our findings show that some aspects of teachers' linguistic input were affected by teaching context, thus underscoring its potential importance when considering the promotion of language development in children. Specifically, the frequency of teachers' utterances containing decontextualised talk significantly increased when teaching in the museum compared to teaching in the classroom with museum resources and teaching in the classroom with regular classroom resources. The frequency of rare words used by teachers was significantly higher when teaching in the classroom with museum resources and was lowest when teaching in the museum. In addition, the use of multi-clausal sentences increased when teachers taught in the classroom with museum resources, compared to teaching in the museum, or teaching in the classroom with regular classroom resources, though this was no longer significant when participation in the study prior to or post-Covid-19 was taken into account. There were no significant contextual effects on teachers' proportional use of wh- questions. These findings both support and extend those of previous work, showing that the linguistic input provided by teachers differs across contexts, but demonstrate that these effects are not just present due to the task itself, but also the environmental context in which children are taught. Below, we discuss possible explanations for the differences in the quality of input in the different contexts.

Decontextualised Talk in the Museum

It could be that being in a novel museum environment compared to their normal classroom environment provided teachers with greater opportunity to discuss non-present events. Teachers were inclined to discuss previous museum trips with children and related the museum resources back to events that had taken place in school (e.g. "when we do our minibeast hunt have you ever found one?"). Teachers also utilised the substantial array of objects, information boards and interactive displays around the museum exhibit to prompt discussion about non-present events (e.g. "you might have seen some mosquitoes before when you've been on holiday."; "maybe when we go back to class this afternoon, we can have a go at drawing a hornet."). These observations echo the findings of Tamis-LeMonda and colleagues (2019) who found that parental level of decontextualised talk was greater when engaged in object-play activities compared to other tasks such as grooming, feeding and transition, whilst also extending it to teachers in an educational capacity (for converging evidence of an effect of context on both utterance complexity and amount of decontextualised talk, see Brinchmann et al., 2023).

Although the same core objects were used in the museum resources in the classroom condition, it could be that the novel environmental context of the museum coupled with the additional resources that it offers promoted teachers' decontextualised talk (e.g., "when I went on holiday I saw lots of mosquitos."). In contrast, because the objects themselves were novel and interesting in the museum resources condition compared to regular classroom resources, teachers focused their discussion on what the children could see in front of them and how they would describe the objects (e.g."what can you see?" "How many legs has it got?"), and less frequently diverted conversation towards non-present objects and events (e.g."maybe in the summer when we do a bug hunt we might find one"). Thus, novelty in learning environment promoting decontextualised talk in the museum, and novelty in the task reducing decontextualised talk in the classroom, may account for our finding that decontextualised talk was used significantly less often in the museum resources condition than in both the museum and classroom with classroom resources.

Importantly, this finding also provides some empirical support for the claims that museums make good language learning environments for young children (Henderson & Atencio, 2007; Kola-Olusanya, 2005; Rodriguez & Tamis-LeMonda, 2011). Specifically, the quality of linguistic input in terms of decontextualised talk was found to be greater in the museum context compared to the other two contexts, and importantly, the frequency of decontextualised talk in linguistic input is positively related to children's later vocabulary, narrative comprehension and production (Beals, 1997; Katz, 2001; Reese et al., 2010; Snow et al., 2001) and reading comprehension (Snow et al., 2001). Thus, museums may enhance language learning opportunities through the promotion of decontextualised talk.

Multi-Clausal Utterances

Like Gest and colleagues (2006), the present study also suggests that context affects the frequency of teachers' multi-clausal utterances. Teachers may have been more likely to use longer, elaborated utterances when using museum resources in the classroom, because the context allowed the task and discussion to be more in-depth and focused on the objects themselves (e.g. "so let's remember, this beetle is black and it's got six legs"; "it's definitely got lots of legs hasn't it?"). In contrast, being in the museum may have meant teachers paid additional attention to the novel environment. Previous research suggests that more structured contexts with focal tasks elicit higher quality linguistic input from caregivers (Tamis-LeMonda et al., 2017). Moreover, it could be that using regular resources in their normal classroom setting was not sufficient to promote long, elaborated discussion above and beyond using novel, more interesting resources. However, this finding should be interpreted with caution, as the effect of context was difficult to disentangle from the effects of the Covid-19 pandemic. Interestingly, teachers produced significantly fewer multi-clausal utterances overall before the Covid-19 pandemic began. It could be that on returning to school after lockdown, teachers were aware that children had missed out on a large amount of their education, and that they played an important role in helping children catch up. Thus, they talked to children using longer, more elaborate utterances. Further research would be required to clarify if this suggestion is correct.

Rare Words

Our findings demonstrate that the frequency of rare words used by teachers varies across contexts, a finding that has not yet been investigated in previous research. We found that teachers used the highest proportion of rare words when teaching with museum resources in the classroom. One possibility is that using novel, interesting resources in a familiar environment enabled them to focus their input more on the resources, thus incorporating more rare words. Although the resources used in the museum were the same, we observed that less time was spent discussing these compared to in the classroom, as children were also given time to explore the museum exhibit. It seems that because less time was spent on the structured activity discussing the boxed specimens, and more time was spent freely exploring the environment, teachers' linguistic input was interspersed with more periods of 'off-topic' conversation, such as managing children's excitement and guiding them to the right areas. Teachers also answered children's generic questions, for example about exploring other areas of the museum ("no we're just looking at bugs today") or what time the bus was coming to collect them from the museum ("I know we're going to have lunch soon") compared to the classroom when using museum resources, therefore allowing less opportunity to incorporate rare words.

Wh-Questions and Context

Although our contexts showed differences in some measures of linguistic quality, we found no differences for *wh*- questions. This could be due to the focal task of the taught sessions being very similar in all contexts, with the only difference being the resources used. For example, teachers would tend to ask similar *wh*- questions in all contexts regardless of the resources, often alluding to what the children could see in front of them, or how they would describe the target items.

Future Directions and Testable Hypotheses

Our findings highlight some variation in the quality of teachers' linguistic input across different teaching contexts, with some contexts appearing to promote different dimensions of quality compared to others. Specifically, we observed more decontextualised talk in the museum context, while classrooms using museum resources prompted more multi-clausal utterances and rare words. These findings offer a testable hypothesis for future work: that both physical context and teaching materials contribute, interactively or independently, to shaping teachers' linguistic input. Disentangling the relative contributions of these factors is important. For example, future studies could manipulate the materials used, while holding context constant, or vice versa, to determine whether particular features of the environment (e.g., novelty, authenticity, visual stimuli) or of the materials (e.g., conceptual richness, novelty, tangibility) are more influential in eliciting specific types of language from teachers. Additionally, given that we found different measures of quality were more prevalent in different contexts, this suggests that different features of the learning environment may selectively promote different aspects of quality in teachers' linguistic input. In light of this, a hypothesis that should be explored in future work is that there is no single "best" context for promoting high-quality input, but rather that diverse contexts may be suited differently to supporting particular linguistic goals. Future research should aim to identify which features of the learning environment most robustly promote which measures of input quality, and how these can be integrated into teacher training and curriculum design. This work will be critical for developing a more nuanced, theory-driven understanding of how teaching environments shape children's language learning opportunities.

Considerations and Conclusions

It is important to acknowledge that in the present study, the data were collected from a relatively small number of 20-minute teaching tasks per context, with only one set of resources being used in each context, and only a small number of teachers contributing data. Although we accounted for individual differences in teacher talk in our statistical models, and the number of utterances collected was high, it will be important to replicate these findings with a wider range of tasks and participants.

In addition, it is unclear to what extent the observed effects of context could change as a function of factors such as the time spent in the context or the familiarity of the context. For instance, longer, more frequent sessions in the museum could mean that teachers become more familiar with the resources and therefore more confident in designing tasks and talking around them, which could result in children remaining more on-task, potentially reducing the amount of decontextualized talk. Moreover, more frequent sessions in the museum could mean that the children are desensitised to the novelty of the environment, thus, teachers may spend less time managing behaviours that arise from distraction. Similarly, more frequent sessions in the classroom with different sets of museum resources could promote different kinds of talk from teachers once they were comfortable and confident using the resources. Overall, however, although data were only taken from a small set of teachers in the present study, findings are suggestive of the wider relations between learning context and linguistic input.

Since no existing research to our knowledge compares how environmental context affects linguistic input from teachers while controlling for the teaching activity, and no studies contrast museums specifically as a learning context with other educational settings, our findings, while exploratory and observational, make interesting and important novel contributions to the field. The present study is the first to provide empirical support for claims that museums could support children's language development by promoting higher quality linguistic input from teachers in terms of greater levels of decontextualised talk compared to in the classroom. Additionally, the present study demonstrates that teachers used a higher proportion of rare words when teaching in the classroom with museum resources compared to teaching in the museum or teaching in the classroom with regular classroom resources. These findings suggest that teaching in such environments may enrich children's language learning. Offering complex linguistic input in the form of decontextualised talk and the use of rare words can enhance vocabulary and abstract thinking, thus, providing children with opportunities to develop their language skills.

More broadly, the findings underpin the value of collaborative relationships between schools and museums by demonstrating the possible advantages of both class visits to museum contexts, but also of museums loaning resources to school settings. By integrating museum resources into classroom teaching and fostering immersive learning experiences in museums, educators can create richer, more varied language learning environments that benefit children's cognitive and linguistic development.

References

Barnes, E. M., Grifenhagen, J. F., & Dickinson, D. K. (2020). Mealtimes in Head Start pre-k classrooms: examining language-promoting opportunities in a hybrid space. *Journal of Child Language, 47*(2), 337-357. https://doi.org/10.1017/S0305000919000199

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beals, D. E. (1997). Sources of support for learning words in conversation: evidence from mealtimes. *Journal of Child Language, 24*(3), 673–694. https://doi.org/10.1017/S0305000997003267

Beals D. (2001). Eating and reading: links between family conversations with preschoolers and later language and literacy. In Dickinson, D. K., & Tabors, P. O.

(Eds.), *Beginning literacy with language: Young children learning at home and school* (pp. 75-79). Paul H. Brookes Publishing.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251–271. https://doi.org/10.1086/511706

Blake, J., Macdonald, S., Bayrami, L., Agosta, V., & Milian, A. (2006). Book reading styles in dual-parent and single-mother families. *British Journal of Educational Psychology, 76*(3), 501–515. https://doi.org/10.1348/000709905X49719

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology, 101*(2), 294–304. https://doi.org/10.1037/a0013844

Blything, L. P., Hardie, A., & Cain, K. (2019). Question asking during reading comprehension instruction: A corpus study of how question type influences the linguistic complexity of primary school students' responses. *Reading Research Quarterly*, rrq.279. https://doi.org/10.1002/rrq.279

Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics, 32*(1), 221–241. https://doi.org/10.1017/S0142716410000354

Brinchmann, E. I., Røe-Indregård, H., Karlsen, J., Schauber, S. K., & Hagtvet, B. E. (2023). The linguistic complexity of adult and child contextualized and decontextualized talk. *Child Development*, 94, 1368–1380. https://doi.org/10.1111/cdev.13932

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Developmental Science, 12*(3), 140–153. https://doi.org/10.1080/10888690802199418

Cameron-Faulkner, T., & Noble, C. (2013). A comparison of book text and child directed speech. *First Language, 33*(3), 268-279. https://doi.org/10.1177/0142723713487613

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula.* Brookline Books Publishing.

Chaparro-Moreno, L. J., Lin, T.-J., Justice, L. M., Mills, A. K., & Uanhoro, J. O. (2022). The influence of context on the abstraction level of Children's conversations in the preschool classroom. *Early Education and Development*, 34, 705–724. https://doi.org/10.1080/10409289.2022.2067429

Cote, L. R. (2001). Language opportunities during mealtimes in preschool classrooms. In Dickinson, D. K., & Tabors, P. O. (Eds.), *Beginning literacy with language: Young children learning at home and school* (pp. 205-221). Paul H. Brookes Publishing.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli Jr., R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal, 110*(1), 1–18. https://doi.org/10.1086/598840

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945. https://doi.org/10.1037/0012-1649.33.6.934

Dale, E., & Chall, J. S. (1977). *A formula for predicting readability*. Bureau of Educational Research.

Department of Education. (2022, March 28). *Length of the School Week: Minimum expectation*. GOV.UK. Retrieved July 31, 2022, from https://www.gov.uk/government/publications/length-of-the-school-week-minimum-expectation

De Temple, J. M., & Snow, C. E. (2003). Learning words from books. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 16–36). Lawrence Erlbaum. https://doi.org/10.4324/9781410607355

Dickinson, D. K. (2001a). Book reading in preschool classrooms: Is recommended practice common? In Dickinson, D. K., & Tabors, P. O. *Beginning literacy with language: Young children learning at home and school*. Paul H. Brookes Publishing.

Dickinson, D. K. (2001b). Large-group and free-play times: Conversational settings supporting language and literacy development? In Dickinson, D. K., & Tabors, P. O. *Beginning literacy with language: Young children learning at home and school*. Paul H. Brookes Publishing.

Dickinson, D. K., Nesbitt, K. T., Collins, M. F., Hadley, E. B., Newman, K., Rivera, B. L., Ilgez, H., Nicolopoulou, A., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Teaching for breadth and depth of vocabulary knowledge: Learning from explicit and implicit instruction and the storybook texts. *Early Childhood Research Quarterly, 47*, 341–356. https://doi.org/10.1016/j.ecresq.2018.07.012

Dickinson, D. K., & Porche, M. V. (2011a). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development, 82*(3), 870–886. https://doi.org/10.1111/j.1467-8624.2011.01576.x

Dickinson, D. K., & Porche, M. V. (2011b). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development, 82*(3), 870–886. https://doi.org/10.1111/j.1467-8624.2011.01576.x

Dickinson, D. K., & Smith, M. W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly, 29*(2), 104. https://doi.org/10.2307/747807

Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science, 25*(7), 1362–1370. https://doi.org/10.1177/0956797614533801

Gest, S. D., Holland-Coviello, R., Welsh, J. A., Eicher-Catt, D. L., & Gill, S. (2006). Language development subcontexts in head start classrooms: Distinctive patterns of teacher talk during free play, mealtime, and book reading. *Early Education & Development, 17*(2), 293–315. https://doi.org/10.1207/s15566935eed1702_5

Gillam, S. L., Gillam, R. B., & Reece, K. (2012). Language outcomes of contextualized and decontextualized language intervention: Results of an early efficacy study. *Language, Speech, and Hearing Services in Schools, 43*(3), 276–291. https://doi.org/10.1044/0161-1461(2011/11-0022)

Gonzalez, J. E., Pollard-Durodola, S., Simmons, D. C., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2010). Developing low-income preschoolers' social studies and science vocabulary knowledge through content-focused shared book reading. *Journal of Research on Educational Effectiveness, 4*(1), 25–52. https://doi.org/10.1080/19345747.2010.487927

Goodwin, A., Fein, D., & Naigles, L. (2015). The role of maternal input in the development of *wh* -question comprehension in autism and typical development. *Journal of Child Language, 42*(1), 32–63. https://doi.org/10.1017/S0305000913000524

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., & Nesbitt, K. T. (2016). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly, 51*(2), 181–198. https://doi.org/10.1002/rrq.130

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H. Brookes Publishing Co.

Henderson, T. Z., & Atencio, D. J. (2007). Integration of play, learning, and experience: What museums afford young visitors. *Early Childhood Education Journal, 35*(3), 245–251. https://doi.org/10.1007/s10643-007-0208-1

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*(2), 418-433. https://doi.org/10.1111/1467-8624.00415

Hoff-Ginsberg, E. (1985). Some contributions of mothers' speech to their children's syntactic growth. *Journal of Child Language, 12*(2), 367–385. https://doi.org/10.1017/S0305000900006486

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology, 45*(3), 337–374. https://doi.org/10.1016/S0010-0285(02)00500-5

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

Katz, J. R. (2001). Playing at home: The talk of pretend play. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and school* (pp. 53–73). Paul H. Brookes Publishing Co.

Kola-Olusanya *, A. (2005). Free-choice environmental education: understanding where children learn outside of school. *Environmental Education Research, 11*(3), 297–307. https://doi.org/10.1080/13504620500081152

Lenth, R. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.8.5) [R package]. https://CRAN.R-project.org/package=emmeans

Lester, N., Theakston, A., & Twomey, K. E. (2023). The role of the natural history museum in promoting word learning for young children. *Infant and Child Development, 32*(2), e2400. https://doi.org/10.1002/icd.2400

Mascareño, M., Snow, C. E., Deunk, M. I., & Bosker, R. J. (2016). Language complexity during read-alouds and kindergartners' vocabulary and symbolic understanding. *Journal of Applied Developmental Psychology, 44*, 39–51. https://doi.org/10.1016/j.appdev.2016.02.001

Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Development*, *80*(3), 686–702. https://doi.org/10.1111/j.1467-8624.2009.01291.x

Massey, S. L., Pence, K. L., Justice, L. M., & Bowles, R. P. (2008). Educators' Use of Cognitively Challenging Questions in Economically Disadvantaged Preschool Classroom Contexts. *Early Education and Development*, 19(2), 340–360. https://doi.org/10.1080/10409280801964119

McLeod, A. N., & McDade, H. L. (2011). Preschoolers' incidental learning of novel words during storybook reading. *Communication Disorders Quarterly*, *32*(4), 256–266. https://doi.org/10.1177/1525740109354777

Noble, C. H., Cameron-Faulkner, T., & Lieven, E. (2018). Keeping it simple: the grammatical properties of shared book reading. *Journal of Child Language*, *45*(3), 753–766. https://doi.org/10.1017/S0305000917000447

Perry, L. K., Prince, E. B., Valtierra, A. M., Rivero-Fernandez, C., Ullery, M. A., Katz, L. F., Laursen, B., & Messinger, D. S. (2018). A year in words: The dynamics and consequences of language experiences in an intervention classroom. *PLOS ONE*, *13*(7), e0199893. https://doi.org/10.1371/journal.pone.0199893

Ping, M. T. (2014). Group interactions in dialogic book reading activities as a language learning context in preschool. *Learning, Culture and Social Interaction*, *3*(2), 146-158. https://doi.org/10.1016/j.lcsi.2014.03.001

Pinheiro, J. C., & Bates, D. M. (2000). Mixed-effects models in S and S-PLUS. Springer-Verlag.

R Core Team (2021). *R: A language and environment for statistical computing* (Version 4.1.1). R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reese, E., Leyva, D., Sparks, A., & Grolnick, W. (2010). Maternal elaborative reminiscing increases low-income children's narrative skills relative to dialogic reading. *Early Education & Development*, *21*(3), 318–342. https://doi.org/10.1080/10409289.2010.481552

Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the home learning environment across the first 5 years: Associations with children's vocabulary and literacy skills at prekindergarten: Trajectories of the home learning environment. *Child Development*, *82*(4), 1058–1075. https://doi.org/10.1111/j.1467-8624.2011.01614.x

Rowe, M. (2013). Decontextualized language input and preschoolers' vocabulary development. *Seminars in Speech and Language, 34*(04), 260–266. https://doi.org/10.1055/s-0033-1353444

Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*(1), 185–205. https://doi.org/10.1017/s0305000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development: Child-directed speech and vocabulary. *Child Development, 83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Rowe, M. L., Leech, K. A., & Cabrera, N. (2017). Going beyond input quantity: *Wh*-questions matter for toddlers' language and cognitive development. *Cognitive Science, 41*, 162–179. https://doi.org/10.1111/cogs.12349

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences, 112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112

Salo, V. C., Rowe, M. L., Leech, K. A., & Cabrera, N. J. (2016). Low-income fathers' speech to toddlers during book reading versus toy play. *Journal of Child Language, 43*(6), 1385–1399. https://doi.org/10.1017/S0305000915000550

Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language, 24*(1), 123–138. https://doi.org/10.1017/S0305000996003005

Shatz, M., Hoff-Ginsberg, E., & Maciver, D. (1989). Induction and the acquisition of English auxiliaries: the effects of differentially enriched input. *Journal of Child Language, 16*(1), 121–140. https://doi.org/10.1017/S0305000900013477

Snow, C. E. (1990). The development of definitional skill. *Journal of Child Language, 17*(3), 697–710. https://doi.org/10.1017/S0305000900010953

Snow, C. E., Tabors, P. O., & Dickinson, D. K. (2001). Language development in the preschool years. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and school* (pp. 1-25). Paul H. Brookes Publishing Co.

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, *90*(6), 2135–2152. https://doi.org/10.1111/cdev.13089

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: language to infants in structured and naturalistic contexts. *Developmental Science*, *20*(6), e12456. https://doi.org/10.1111/desc.12456

Uccelli, P., Demir-Lira, Ö. E., Rowe, M. L., Levine, S., & Goldin-Meadow, S. (2019). Children's early decontextualized talk predicts academic language proficiency in midadolescence. *Child Development*, *90*(5), 1650–1663. https://doi.org/10.1111/cdev.13034

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*, 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Wasik, B. A., Hindman, A. H., & Snell, E. K. (2016). Book reading and vocabulary development: A systematic review. *Early Childhood Research Quarterly*, *37*, 39–57. https://doi.org/10.1016/j.ecresq.2016.04.003

Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265–279. https://doi.org/10.1037/0012-1649.37.2.265

Data, Code and Materials Availability Statement

The data supporting the findings reported in this paper are openly available from the Open Science Framework repository at: https://osf.io/zj4fe/?view_only=0bb630aed1674842b1d0b7c63d2ef798

Ethics Statement

This study was approved by the Psychology, Communication and Human Neuroscience Division Panel at the University of Manchester. Ref: 2019-8263-12545

Authorship and Contributorship Statement

All listed authors meet the *Language Development Research* journal criteria for authorship. Specifically, each author has made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; participated in drafting the manuscript or revising it critically for important intellectual content; provided final approval of the version to be published; agreed to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Individual author contributions are as follows. **Nicola Lester:** Conception and design of the study; data acquisition and interpretation; data analysis; manuscript drafting. **Katherine E. Twomey**: Conception and design of the study; statistical analysis; critical revision of the manuscript; final approval of the manuscript. **Anna Theakston**: conception and design of the study; critical revision of the manuscript; final approval of the manuscript.

Acknowledgements

We would like to thank the children, the teachers, and the support staff at our CASE partner school for their friendly cooperation. This work was supported by the ESRC International Centre for Language and Communicative Development (LuCiD) [AT & KT; ES/S007113/1] and an ESRC 1+3 CASE studentship [NL; 7995306]

License

Trajectories of early vocabulary growth in Hebrew-speaking toddlers: The role of comprehension

Hila Gendler-Shalev, Ph.D.
Department of Communication Sciences and Disorders
University of Haifa, Israel

Virginia A. Marchman, Ph.D.
Department of Psychology
Stanford University, USA

Esther Dromi, Ph.D.
Constatiner School of Education
Tel Aviv University, Israel

Abstract: The second year of life is characterized by rapid expressive vocabulary growth for most children, however, there is also substantial individual variation. Can variation in expressive vocabulary skill identified very early in the second year, prior to 16 months, provide useful signals about children's later development? Here, we examined growth trajectories in expressive vocabulary from 12-24 months of age in Hebrew-speaking children grouped according to their earliest expressive vocabulary level. Caregivers of 92 toddlers completed the Hebrew adaptation of the MacArthur-Bates Communicative Development Inventories: Words & Gestures three times, every few months. Children were classified at first administration (12-16 months) as either lower ($n = 21$, ≤ 25th percentile) or higher ($n = 71$, > 25th percentile) in expressive vocabulary. Trajectories of growth were significantly delayed in the lower group, compared to the higher group, but the shapes of trajectories were generally similar. Critically, children with initial weaker receptive skills (< 25th percentile) had significantly shallower growth trajectories than children with stronger comprehension skills. Moderate delays in receptive vocabulary (< 50th percentile) were not informative in predicting growth trajectories for children with lower initial expressive vocabulary scores. These results suggest that lower expressive and receptive skills defined as early as 16 months provide useful information about future expressive vocabulary growth. Theoretical and clinical implications are discussed.

Keywords: vocabulary growth, expressive and receptive vocabulary, MacArthur-Bates Communicative Development Inventories (CDI), language development, vocabulary growth trajectories, Hebrew

Corresponding author: Hila Gendler-Shalev, Department of Communication Sciences and Disorders University of Haifa, Abba Khoushy Ave 199, Haifa, 3498838, Israel. Email: hilags@welfare.haifa.ac.il

ORCID ID(s): Hila Gendler-Shalev: https://orcid.org/0000-0003-3624-2522; Virginia A. Marchman: https://orcid.org/0000-0001-7183-6743; Esther Dromi: https://orcid.org/0000-0003-0787-8396

Citation: Gendler-Shalev, H., Marchman, A.V., & Dromi, E. (2025). Trajectories of early vocabulary growth in Hebrew-speaking toddlers: The role of comprehension. *Language Development Research*, 5(3), 302–328. http://doi.org/10.34842/ldr2025-838

Introduction

In the second year of life, most children undergo substantial increases in the number of words that they can say (Bates et al., 1994; Fenson et al., 1994), going from producing just a few single words to amassing an expressive vocabulary of several hundred words[1] (Frank et al., 2021). However, there is also considerable variability across children in when and how quickly their expressive vocabulary grows. Some children begin to produce their first words around their first birthday and quickly expand their vocabularies over the next several months. Other children, often referred to as "late talkers" (LT) (Fisher, 2017; Rescorla & Dale, 2013), may not do so until months later and may experience delays in vocabulary development that persist throughout the preschool period.

But how stable are these individual differences over time? On the one hand, variation in a composite of early language skills identified late in the second year has been shown to be correlated with language and literacy skills in school age (Bornstein et al., 2016), suggesting persistence in individual differences in language skills across the first decade of life. However, few studies have explored whether variation in expressive language skills identified earlier, e.g., younger than 16 months, can meaningfully predict outcomes (e.g. Ukoumunne et al., 2012). On the other hand, even when assessed at 2 years, it is well-established that early expressive vocabulary skills alone are not a strong correlate of persistent language delays or developmental language disorder (DLD) (Rescorla, 2011). This suggests that some children with early delays experience "catch-up" in vocabulary, displaying more rapid growth over the same period of time than other children. A critical question is what additional factors might explain these individual differences? Several possibilities have been explored in the literature, including birth order, family history, and socioeconomic status (e.g., Fisher, 2017, Hammer et al, 2017). Here, we focused on one factor that has been shown to be particularly predictive of later outcomes, i.e., early comprehension abilities. We asked the extent to which receptive vocabulary scores moderated the longitudinal trajectories of expressive vocabulary growth across the second year of life in Hebrew-learning children who range in early expressive vocabulary skills.

Parent Reports of Early Vocabulary

Parent report questionnaires, such as the MacArthur-Bates Communicative Development Inventories (CDIs, Marchman et al., 2023) or the Language Development Survey (LDS, Rescorla, 1989), are commonly used tools for assessing children's vocabulary development, especially for toddlers under 3 years of age. The

[1] In line with the definition of a "word" in other language development studies (Dromi, 1987; Frank et al., 2021), we define a word as a linguistic unit that carries a consistent meaning understood by another person. The word can be a childish mispronunciation (e.g., 'all' for *ball,* or 'nana' for *banana*).

CDIs, developed originally for American English, have now been adapted to more than 60 languages (see https://mb-cdi-stanford.edu/adaptations.html) and have been demonstrated to provide reliable and cost-effective estimates of children's burgeoning language skills. Many of these instruments are now available online, easing the burden of administration and scoring required by traditional paper-and-pencil forms (deMayo et al., 2021). For example, the Hebrew adaptation of the Words & Gestures (WG) online form (HCDI-WG, Gendler-Shalev & Dromi, 2022) was developed and normed for Hebrew-speaking children 12 to 24 months of age. Via a vocabulary checklist, parents are asked to mark, from among a representative list of hundreds of possible items, the words that their child "understands" or "understands and says," yielding measures of receptive and expressive vocabulary size.

Based on data from thousands of children around the world from many language communities, studies using the CDIs have contributed substantial insights regarding both the universal features and the variation that characterizes early vocabulary development (Frank et al., 2021). For example, studies have explored the predictors of what words children are first likely to learn (Braginsky, Yurovsky, Marchman, & Frank, 2019; McDonough et al., 2011), and the sources of crosslinguistic differences that elucidate language-specific rather than universal characteristics of early language development (Bleses et al., 2008a; 2011). Importantly, several large-scale studies have provided lexical development norms that reveal the substantial individual differences that exist across children, regardless of the language they are learning. These norms also provide useful benchmark levels for the expected vocabulary size for children of a given age from different language communities (Bleses et al., 2008b; Hao et al., 2008; Kalashnikova et al., 2016), including Hebrew (Gendler-Shalev & Dromi, 2022).

Vocabulary Growth in Children Who Are Slow to Begin to Talk

One of the most common reasons that toddlers are referred for a clinical evaluation is late onset of recognizable words or a small expressive vocabulary size towards their second birthday (Whitehurst & Fischel, 1994). Toddlers younger than 3 years who produce only a few words, in the absence of a diagnosed developmental disability or hearing impairment, are often labeled as *late talkers* (LT) (Fisher, 2017). Most studies define LT at the age of about 24 months (Fisher, 2017; Rescorla, 2009), however, other studies do so earlier at the age of 18 months (Fernald & Marchman, 2012). While previous practice assumed that early delays in children younger than 2 years were not clinically relevant, there is now a growing shift towards early referral and intervention approaches, thereby, abandoning the "wait and see" approach (e.g., Edwards et al., 2021; Singleton, 2018).

Note that LT status, by itself, does not constitute a clinical diagnosis, as most LT children "catch up" to their peers by preschool. At the same time, some LT children do show continued delays and are diagnosed with DLD as they grow older (Rescorla, 2011). Moreover, as a group, even LT children who move into the typical range of expressive vocabulary distribution and therefore, who are not diagnosed with DLD, still show lower academic performance than children without a history of late talking (Bleses et al., 2016; Rescorla & Dale, 2013). As such, children identified as LT are considered to be at risk for exhibiting language and learning disorders later in development, thus making accurate early identification of LT children a clinical vital importance (Fisher, 2017). Few studies have asked whether early delays in the onset of expressive vocabulary can predict trajectories of vocabulary growth when assessed in children younger than 16 months, prior to the point when children are commonly labeled "late talkers" (e.g., Ukoumunne et al., 2012).

An important question is whether language learning mechanisms operate similarly in children who show initial delays in early expressive vocabulary compared to their peers who begin to talk earlier, or are delays indicative of atypical processes of vocabulary growth. For example, Jones and Brandt (2019) compared the characteristics of the words in the lexicons of English-speaking 18-month-old toddlers with smaller versus larger expressive vocabulary sizes. They found that toddlers with smaller vocabularies showed comparable vocabulary compositions to those of younger children, as reported by Braginsky et al. (2019), suggesting delayed but not atypical processes of vocabulary development. These findings align with a study demonstrating that children who were identified as LT and younger typically developing children with similar vocabulary sizes had similar vocabulary characteristics (Gendler-Shalev & Novogrodsky, 2024). However, Jiménez et al. (2021) found that English-speaking toddlers who were identified as LT showed a weaker noun bias compared to vocabulary-size matched typically-developing toddlers, suggesting atypical vocabulary growth. In the current longitudinal study, we compared the shapes of the trajectories in expressive vocabulary growth over the second year of life in Hebrew-speaking toddlers who have higher vs. lower expressive vocabulary skills when first inducted into the study (12-16 months). If children with lower and higher early expressive skills show similar shapes of vocabulary growth, this suggests continuity in mechanisms guiding development in children in both groups. However, if toddlers with early initial delays show patterns of vocabulary growth over time that are not parallel to those of children with initially higher early expressive skills, this could suggest different learning mechanisms are associated with delayed onset of expressive vocabulary.

Correlates of Early Vocabulary Delays

In addition to small vocabulary size, research, predominately in English-speaking children, has identified other differences between children with relatively larger vs. smaller early expressive vocabularies. As a group, late-talking 16–18-month-olds have been shown to demonstrate limited use of communicative gestures (Thal & Tobias,

1994) and limited range of symbolic play skills (Hall et al., 2013), as well as phonological delays and deviations (Carson et al., 2003). At older ages, children with early expressive vocabulary delays were also found to produce word combinations later than their typically developing peers (Dale et al., 2003) and upon reaching school-age, also showed delays in grammatical proficiency (Moyle et al., 2007; Rescorla, 2011).

Notably, early receptive abilities have been associated with delays in early expressive language skills. Laboratory experiments have shown that infants begin comprehending words around 6–9 months (Bergelson & Swingley, 2012) and after their first birthday, infants' word comprehension abilities typically improve notably (Bergelson, 2020; Meylan & Bergelson, 2022). Although most infants understand words a few months before they produce them, receptive and expressive vocabularies are strongly, although not perfectly, correlated (Frank et al., 2021). In a laboratory study assessing comprehension and production of novel object names (Gershkoff-Stowe & Hahn, 2013), children with larger expressive vocabularies, as measured by the CDI, tended to comprehend new words faster than those with smaller expressive vocabularies. Surprisingly, children with larger expressive vocabularies took more time to verbalize new words (which they comprehended earlier) compared to children with smaller expressive vocabularies. The authors suggested that this phenomenon could be attributed to the increase in semantic neighborhood density, aiding comprehension while possibly hindering lexical access during production. These results contribute to the idea that while comprehension and production are closely linked, they are enhanced by distinct underlying processes. Moreover, these findings suggest that receptive and expressive language skills undergo a complex interplay over the course of acquisition.

If children with early expressive delays also exhibit low receptive skills, this may signify additional risk for later expressive skills. Indeed, research has shown that some children with early vocabulary delays exhibit a large gap between the sizes of their receptive and expressive vocabularies (Rescorla, 2011), while others appear to have a small gap and experience delayed receptive as well as expressive vocabularies (Cheung et al., 2022; Verganti et al., 2024). Buschmann et al. (2008) reported that children classified as LT who also had smaller receptive vocabulary sizes tended to have lower nonverbal IQs (though still within age-typical ranges) than LT children with larger receptive vocabulary sizes. In contrast, LT children who had smaller expressive vocabulary sizes, but who did not also display smaller receptive vocabularies, did not. Similarly, Verganti et al. (2024) found that Italian-speaking children with delayed expressive and receptive vocabulary sizes produced fewer gestures than children with only delayed expressive vocabulary sizes. Finally, toddlers with lower expressive vocabularies at 18 months were slower and less accurate in comprehending language in real time, as measured by the looking-while-listening task, than toddlers without initial expressive delays (Fernald & Marchman, 2012). These results are consistent with the work of Thal et al. (2013), who compared the outcomes of English-speaking "late producers" (children with late expressive vocabulary only) with "late comprehenders" (children with both late receptive and expressive vocabulary) at age 6 years. These researchers found that "late comprehenders" were at higher risk for later language delays than "late producers", i.e., children with expressive delays who had receptive skills within the normal range.

At the same time, few studies have explored whether weaker early comprehension skills may potentially place expressively-delayed children at additional risk when those delays are assessed early in the 2nd year. Moreover, less is known about whether delays in early receptive skills are related to *trajectories* of expressive vocabulary growth longitudinally, over multiple time points during the 2nd year of life. From a clinical perspective, few studies have explored what level of concomitant receptive delays are likely to incur additional risk for children who show initial delays in expressive skills.

The Current Study

We investigated trajectories of Hebrew-speaking children's expressive vocabulary growth longitudinally at multiple administrations in the 2nd year of life. At the first administration (ranging in age from 12 to 16 months), we classified children as "higher" vs. "lower"[2] in expressive language skills based on their percentile score in expressive vocabulary (≤ 25th vs. > 25th), as reported on the HCDI-WG at the first administration. We also grouped children using two cut-off scores for receptive vocabulary percentile at that same administration (12-16 months). We first grouped children using a cutoff that was analogous to that used for expressive language (i.e., +/- 25th percentile). Toddlers with lower expressive vocabulary percentile scores who also have lower receptive vocabulary percentiles would fall in the lowest quartile in *both* expressive and receptive language skills. We then grouped children based on a more liberal cut-off for receptive vocabulary (i.e., +/- 50th percentile). Contrasting these two percentile cut-offs allows us to gain more insights into what profile of receptive skills shape trajectories of expressive language growth in children with initially higher vs. lower expressive vocabulary sizes at 12-16 months.

We conducted growth curve modeling over age to investigate the following questions:

(1) When defined at 12-16 months, is higher vs. lower expressive vocabulary a predictor of trajectory of growth in expressive vocabulary across the 2nd year of life?

[2]This group may be termed differently in the literature. Some researchers refer to this group as LT (e.g., Rescorla, 2011; Fisher, 2017). However, given the young age of our participants and the broad developmental variability at this stage, we opt for the more nuanced term "lower" to be consistent with the approach taken by Vehkavuori & Stolt (2019).

(2) Are trajectories of expressive vocabulary growth from 12-24 months similar in shape in children who have initially higher vs. lower expressive skills?

(3) Does level of early receptive vocabulary skill moderate trajectories of expressive vocabulary growth in children with initially higher vs. lower expressive vocabulary abilities? If so, are patterns similar when grouping children using the 25th vs. 50th percentile on receptive vocabulary?

Method

Participants

Participants were Hebrew-speaking caregivers of 92 toddlers aged 12-24 months. All parents reported that their child had not been diagnosed with a medical or developmental condition and that they had not been treated for an ear infection more than once in the last three months. Hebrew was the primary language for all children, with all children being exposed to Hebrew for at least ten hours a day. Finally, all parents reported that they were not worried about the rate of the child's development.

The families were from predominantly highly educated backgrounds, with 89% of primary caregivers in the sample holding a college degree, compared to 48% in the general population (Gendler-Shalev & Dromi, 2022). Using data from a larger study (Gendler-Shalev & Dromi, 2022), data were analyzed for all children in that sample whose caregivers completed a HCDI-WG for the first time when their child was 12-16 months and completed additional administrations at two subsequent time points prior to when the child was 24 months. A total of 3 additional families were considered for analyses but were later excluded because the child fell outside of the target age range (n = 2) and because it was later discovered that the child was exposed to a language other than Hebrew in the home (n = 1). Although the sample size in the current study may appear limited, it is consistent with prior studies in the field (i.e., Carson et al., 2003; Hadley & Holt, 2006; Hadley & Short, 2005).

Expressive Vocabulary Groups. Children were classified based on their expressive vocabulary scores at the first administration of the HCDI-WG according to the norms for Hebrew (Gendler-Shalev & Dromi, 2022) as either initially lower (n = 21, ≤ 25th percentile) or as higher expressive vocabulary scores (n = 71, > 25th percentile). While these children in the initially lower group could be conceptualized as "late talking", we do not use the term "late talker" here. Given the young age of our participants and the broad developmental variability at this stage, we opted for the more nuanced terms "lower" vs. "higher" in expressive vocabulary percentile score, relative to their age, similar to the approach taken by Vehkavuori & Stolt (2019). Note also that different studies have classified children as LT based on different criteria.

While some previous studies have used more stringent cut offs, for example, the 10th (e.g., Bleses et al., 2016), 15th (e.g., Thal et al., 2004), or 20th (Fernald & Marchman, 2012) percentiles, less stringent criteria of the 30th percentile have also been reported (Jones, 2003). Here, we opted for a cutoff of 25th percentile, which allowed us to differentiate children in the lowest quartile of expressive vocabulary skill from those in the top three quartiles, while still maintaining a sufficiently large sample in the lower expressive vocabulary group. Table 1 reports the number of children at enrollment by age, vocabulary group, and child sex.

Receptive Vocabulary Groups. In order to test the role of early receptive skill on growth trajectories, children were also classified into receptive vocabulary groups using two percentile score levels computed at the first administration of the HCDI-WG across the full sample. First, we grouped children into groups based on an analogous cut-off as that used for expressive vocabulary, i.e., in the bottom quartile vs. the top three quartiles for their age (lower receptive vocabulary: ≤ 25th percentile, n = 27; higher receptive vocabulary: > 25th percentile, n = 65). Second, we grouped children into receptive vocabulary groups using a more liberal definition, i.e., a 50th percentile cut-off. Here, children with lower receptive vocabulary percentiles would fall in the bottom vs. top half of the distribution for their age (lower: ≤ 50th percentile, n = 45, higher: > 50th percentile, n = 47).

Table 1. Number of participants by age in months at the first administration of the MB-HCDI and child sex in children identified as higher (*n* = 71) and lower (*n* = 21) in expressive vocabulary at 12-16 months and in the full sample (*n* = 92).

	Initial Expressive Vocabulary Group					
Age (mos)	Higher		Lower		Full Sample	
	Boys	Girls	Boys	Girls	Boys	Girls
12	7	9	1	1	8	10
13	11	6	0	0	11	6
14	6	10	0	1	6	11
15	10	5	6	5	16	10
16	6	1	4	3	10	4
Total	40	31	11	10	51	41

Procedure

Vocabulary data were collected via an online Hebrew adaptation of the MacArthur-Bates Communicative Development Inventory: Words & Gestures (HCDI-WG; Gendler-Shalev & Dromi, 2022). The HCDI-WG is a reliable and valid tool for evaluating early lexical development of Hebrew-speaking children from 12-24 months. The questionnaire consists of a vocabulary checklist with 428 words in 18 categories. Caregivers are asked to mark the words that the child "understands" or "understands and says." Total comprehension and production vocabulary scores were tabulated and percentiles were derived based on age in months and child sex following standard protocols.

Caregivers were initially approached via social media, web news sites, and radio and TV talk-shows to participate in a longitudinal study on child language development. After completing the first HCDI-WG, caregivers of children who were younger than 16 months were contacted approximately every four months via email and asked to complete two additional HCDI-WGs. Because the initial administration occurred at varying ages, the parent of one child may have completed the HCDI-WG when the child was, for example, 12, 16, and 20 months, and another when the child was, for example, 13, 17, and 21 months.

Analytic Plan

We first present descriptive statistics for raw vocabulary production scores as a function of child age and expressive vocabulary group (initially higher vs. lower at 12-16 months). We then present percentiles for vocabulary production and comprehension at each administration by group. To model trajectories of expressive vocabulary growth, we used generalized additive models for location, scale, and shape (GAMLSS). GAMLSS is a general regression framework for modeling fixed effect and mixed-level growth functions. Mixed-level growth functions are used to model longitudinal data with multiple administrations within participant over time or age. GAMLSS has advantages over other frameworks because it offers a flexible method for modeling trajectories that are non-linear and for capturing changes in variance over age. Models were fit using the GAMLSS function (Stasinopoulos et al., 2017) in the R statistical package (Version 4.0.3; R_Core_Team, 2020). Based on earlier work (Frank et al., 2021; Marchman et al., 2023), the growth trajectory of parent-reported expressive vocabulary scores over age is best described as a beta distribution, similar to a logistic function, that is limited by, but does not include, 0 and 1. Prior to conducting the models, raw vocabulary scores were converted to a proportion score and extreme values were set to 0.001 and .999. Following earlier studies, penalized B-spline smoothers (i.e., P-splines) were used to model the effect of age on the mean and variance in vocabulary. As in earlier studies with vocabulary data, the lambda values, or smoothing parameters, were set to a large number (10,000) to ensure sufficiently smooth growth over age without overfitting (Marchman et al., 2023).

Our main focus was to examine trajectories of expressive vocabulary growth over age as a function of expressive vocabulary group, i.e., in children initially identified as either lower or higher percentiles at first administration. Group and age of administration were fixed effects. Group was dummy coded with higher expressive vocabulary as the reference group. To capture the repeated nature of our data, participant was included as a random effect on the intercept. Significant main effects would indicate that the mean levels of expressive vocabulary growth differed in children with initially lower vs. higher expressive vocabulary percentiles. A significant age x group interaction would indicate that the shapes of the growth trajectories over age were different as a function of initial expressive vocabulary group.

A second goal was to examine whether growth trajectories for children with initially lower vs. higher expressive vocabularies were different as a function of early receptive vocabulary level. We conducted two parallel analyses testing the effects of two receptive vocabulary level groupings. In both sets of analyses, receptive vocabulary group was added to the model as a fixed effect, dummy coded with "higher" as the reference group. Participant was included as a random effect on the intercept. In both sets of models, a significant main effect of receptive vocabulary group would suggest that children's expressive vocabulary growth was related to their early receptive vocabulary level. A significant interaction of receptive vocabulary group by expressive language group would indicate that children with initially lower vs. higher expressive vocabularies who had lower vs. higher receptive vocabulary scores exhibited different mean levels of expressive vocabulary growth over age, i.e., receptive vocabulary level moderates expressive vocabulary development as a function of initial expressive vocabulary size.

Results

Descriptive Statistics

Table 2 presents mean age and raw expressive and receptive vocabulary scores at first administration for children in the higher vs. lower expressive vocabulary groups. As expected, children in the lower expressive vocabulary group had lower raw scores, on average, than the children in the higher group, $t(85) = 5.31$, $p < 0.001$, $d = .76$. Note, however, that raw scores for receptive vocabulary were comparable across expressive vocabulary groups, on average, $p = 0.71$, with comparable estimates of the variance, suggesting that there was similar variability in raw scores across both groups.

Table 2. Descriptive statistics of age and raw expressive and receptive vocabulary scores at the first administration for children classified with initially higher (*n* = 71) vs. lower (*n* = 21) expressive vocabulary scores at 12-16 months.

	Age (months)		Expressive Vocabulary (Raw Scores)	Receptive Vocabulary (Raw Scores)
Expressive Vocabulary Group	M (SD)	Range	M (SD)	M (SD)
Higher	13.7 (1.3)	12 - 16	30.4 (30.1)	133.9 (91.1)
Lower	15.0 (1.1)	12 - 16	10.1 (6.1)	125.6 (90.3)

Note: Initial expressive vocabulary group defined at first administration (12-16 months) as > 25th percentile (higher) vs. ≤ 25th percentile (lower). Expressive vocabulary ("understands and says") and receptive vocabulary ("understands and says" + "understands") raw scores were based on responses on the vocabulary checklist from the H-CDI (Gendler-Shalev & Dromi, 2022) (max = 428).

Table 3 presents descriptive statistics for expressive vocabulary percentile scores by group and administration. At Administration 1, all children in the higher group, by definition, had scores that were > 25th percentile, whereas all children in the lower group had scores ≤ 25th percentile. At subsequent administrations, expressive vocabulary percentile scores remained lower for the children in the initially lower group compared to those in the higher group, on average, but there was evidence of both developmental continuity and catch up: A total of 13 (of 21, 61.9%) children stayed ≤ 25th percentile at all three administrations, while the remainder moved into the normal range by either the second (*n* = 4, 19.05%) or the third (*n* = 4, 19.05%) administration. We can note that all 8 of the children in the initially lower expressive vocabulary group who demonstrated some recovery at administrations 2 and/or 3 had percentile scores > 11th percentile in expressive vocabulary at the first administration. The 13 children in the initially lower group who stayed delayed had percentile scores spanning from 1st to 25th. Interestingly, about two-thirds (48 of 71, 67.6%) of the children with initially higher scores at first administration had scores > 25th percentile at all three administrations, while the remainder (*n* = 23, 32.4%) had scores ≤ 25th percentile at one or both subsequent administrations.

Table 3 also shows that, at Administration 1, receptive vocabulary percentile scores were generally lower for children with initially lower expressive vocabulary scores than children with initially higher expressive vocabularies. However, note that scores spanned the full range (1 to 99) in each group. That is, some children with initially lower expressive percentile scores scored just as high in receptive vocabulary as children at the same age who had initially higher scores, even though their expressive vocabulary scores were lower. The same pattern held for each subsequent administration, i.e., receptive vocabulary percentiles spanned the full possible range for children in both groups.

Table 3. Descriptive statistics for expressive and receptive vocabulary percentile scores at each administration for children classified with higher (*n* = 71) or lower (*n* = 21) initial expressive vocabulary percentiles at first administration (12-16 months).

		Expressive Vocabulary (percentile)		Receptive Vocabulary (percentile)	
Expressive Vocabulary Group	Admin	M (SD)	Min-Max	M (SD)	Min-Max
Higher	1	56.3 (18.4)	28-99	52.9 (28.3)	4-99
	2	51.9 (30.5)	3-98	56.7 (27.5)	1-97
	3	60.3 (27.4)	5-99	61.8 (25.4)	4-99
Lower	1	15.1 (8.3)	1-25	36.9 (29.6)	1-95
	2	15.5 (16.2)	1-57	47.8 (28.9)	2-92
	3	24.7 (22.1)	1-86	44.9 (29.2)	3-96

Note: Group = Initial expressive vocabulary group defined at first administration (12-16 months) as > 25^{th} percentile (higher) vs. ≤ 25^{th} percentile (lower). Expressive and receptive vocabulary percentiles derived from the norming study of the HCDI (Gendler-Shalev & Dromi, 2022).

Modeling Production Vocabulary Growth

Expressive Vocabulary Group Status. We next sought to model vocabulary growth over age from 12 to 24 months as a function of initial expressive vocabulary group (higher vs. lower). In Table 4, as expected, Model 1 shows that vocabulary scores increased significantly over age, accounting for more than 60% of the variance in scores. Model 2 shows that adding the factor of initial vocabulary group significantly increased the overall fit of the model (LR test = 77.3, $p < 0.001$), adding approximately 9.6% variance to the overall model fit. As illustrated in Figure 1, children with higher scores at the initial time point were reported to know significantly more words across the period than children with lower initial scores. For example, at 18 months, model estimates indicated that children with initially lower expressive vocabulary scores produced about 46 words, on average, while children with initially higher expressive vocabulary scores were reported to produce more than twice that many, approximately 115 words. By 24 months, this group difference persisted, such that children with lower initial scores were estimated to produce about 220 words, on average, whereas children with higher initial scores were estimated to produce more than 100 more words, about 327 words, on average. Model 3 shows that adding the interaction term did not significantly increase the overall model fit (LR test = 0.31, p = .58), increasing the variance accounted for by less than 0.1%. Thus, Model 2 was the best fitting model, which suggests that, while offset in developmental time, the shapes of the vocabulary growth trajectories were not significantly different for toddlers with initially higher vs. lower expressive vocabulary scores.

Table 4. Model estimates (unstandardized B (SE)) and fit statistics for growth in expressive vocabulary over age by expressive vocabulary group (higher vs. lower initial scores).

	Model 1	Model 2	Model 3
Age	0.34 (0.02)***	0.37 (0.02)***	0.37 (0.02)***
Expressive Group	--	–1.12 (0.13)***	–0.58 (0.73)
Age x Expressive Group	--	--	–0.03 (0.04)
Deviance	–452.44	–529.79***	–530.00
AIC	–443.75	–519.04	–517.63
BIC	–428.03	–499.59	–495.23
R^2	60.8	70.4	70.4

Note: Expressive Group = Initial expressive vocabulary group defined at first administration (12-16 months) as > 25th percentile (higher) vs. ≤ 25th percentile (lower). HIgher = reference group. Model comparisons (Likelihood ratio (LR) tests) for Model 2 are in relation to Model 1 and Model 3 in relation to Model 2.

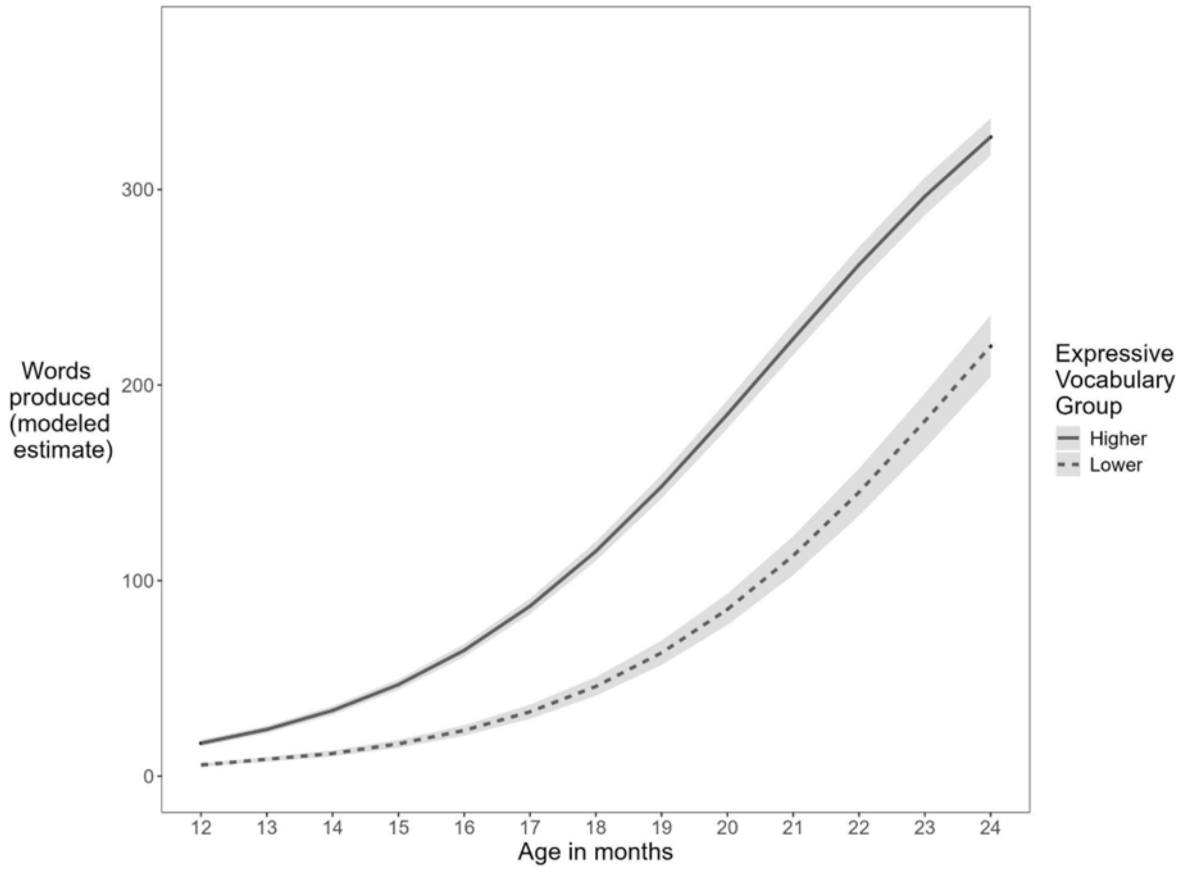

Figure 1. Predicted expressive vocabulary growth over age for children classified with higher (*n* = 71) and lower (*n* = 21) expressive vocabulary at first administration (12-16 months) based on the main effects only model (Model 2). Shaded area indicates +/- 1 SE.

Receptive Vocabulary. Next, we asked whether growth trajectories of expressive vocabulary were moderated by children's receptive vocabulary level also assessed at the first administration. We tested two percentile cut-offs to explore which comprehension level was most informative for predicting trajectories of expressive vocabulary growth in children with a range of initial expressive vocabulary scores.

25th Percentile Cutoff. Model 4 shows that adding receptive vocabulary group as a main effect to Model 2 significantly improved overall model fit (LR test = 16.23, $p < 0.001$), adding about 2% additional variance. Both initial expressive vocabulary group (lower vs. higher) and receptive vocabulary group (lower vs. higher) significantly made a unique contribution to mean expressive vocabulary levels over age. Model 5 adds two-way interaction terms (age x receptive level, and group x receptive level) but did not result in an increase in overall model fit (LR test = 4.34, p = .13). A final model adding the 3-way interaction also did not increase overall model fit (LR test = 4.5, p = .37). Thus, the best-fitting model with this receptive vocabulary level cut off is Model 4. As shown in Figure 2, both children with lower and higher initial expressive vocabulary scores who had receptive vocabulary levels > 25th percentile tended to have larger expressive vocabularies than children who understood fewer words, across the age period.

To further illustrate, at 18 months, the modeled expressive vocabulary size for toddlers with higher scores in both expressive and receptive vocabulary (>25th percentile) was about 124 words, about 35 words higher than for toddlers who had higher expressive scores, but with receptive scores ≤ 25th percentile (88 words). Similarly, the modeled expressive vocabulary size for toddlers with lower initial scores who also had ≤ 25th percentile receptive scores was about 36 words, almost 20 words fewer than for toddlers with lower initial scores who had higher (>25th percentile) receptive scores (54 words). These differences persisted until 24 months, such that toddlers with higher initial expressive scores but lower (≤ 25th percentile) receptive scores were predicted to produce about 300 words, about 35 fewer words than their higher-expressive peers with > 25th percentile receptive scores (336 words). Importantly, by 24 months, toddlers who had lower initial expressive scores and also had ≤ 25th percentile receptive scores were predicted to produce only 193 words, nearly 50 words fewer than their lower-expressive peers with higher receptive scores (240 words).

50th Percentile Cutoff. We now explore the role of receptive vocabulary with a more liberal cutoff, i.e., 50th percentile. Similar to the earlier analysis, Model 6 in Table 5 shows that adding receptive vocabulary group as a main effect significantly improved overall model fit compared to Model 2 (LR test = 22.56, p < 0.001), adding about 2% additional variance. Thus, across children with both higher and lower initial expressive vocabulary scores, children who were in the top half of the distribution in receptive vocabulary scores had higher expressive vocabularies than children in the lower half of the distribution. However, unlike with the 25th percentile cutoff, Model 7 shows that the addition of the two-way interactions significantly improved model fit (LR test = 9.98, p = 1.02). Specifically, +/-50th receptive level affected the expressive vocabulary growth of children differently in children in the higher vs. lower initial expressive vocabulary groups. As shown in Figure 3, a receptive group difference was evident for the children with higher initial vocabulary scores, but not for children with lower initial vocabulary scores. Follow-up analyses confirmed that for the children with higher initial expressive scores, receptive vocabulary level (+/- 50th) was a significant predictor of expressive vocabulary overall ($\boldsymbol{\beta}$ = –0.58, p < 0.001), with no interaction with age ($\boldsymbol{\beta}$ = –0.01, p = 0.86), suggesting that the effects were consistent across the age period. In contrast, receptive vocabulary level was not a significant predictor of expressive vocabulary for children with lower initial expressive vocabulary scores (***β*** = 0.24, *p* = 0.21) nor in interaction with age (***β*** = 0.02, *p* = 0.79). Thus, for children who were saying relatively fewer words between 12-16 months, knowing that their receptive skills were above the median did not provide additional information about their future growth in expressive vocabulary.

Table 5. Model estimates (unstandardized B (SE)) and fit statistics for growth in expressive vocabulary over age by initial expressive vocabulary group (higher vs. lower) and two cut-offs for higher vs. lower comprehension level (25th or 50th).

	Receptive Vocabulary Group			
	25th percentile cutoff		50th percentile cutoff	
	Model 4	Model 5	Model 6	Model 7
Age	0.37 (0.02)***	0.38 (0.02)***	0.37 (0.02)***	0.37 (0.02)***
Expressive Group	−1.05 (0.13)***	−0.90 (0.74)	−1.08 (0.13)***	−1.03 (0.74)
Receptive Group	−0.45 (0.11)***	0.35 (0.60)	−0.46 (0.10)***	−0.74 (0.53)
Age x Expressive Group	–	−0.02 (0.03)	–	−0.03 (0.04)
Age x Receptive Group	--	−0.05 (0.03)	–	0.01 (0.03)
Expressive x Receptive	--	0.39 (0.26)	–	0.78 (0.26)**
Deviance	−545.92***	−550.41	−552.24***	−561.76*
AIC	−533.53	−531.65	−539.85	−543.44
BIC	−511.10	−497.68	−517.41	−509.45
R^2	72.0	72.5	72.7	73.6

Note: Expressive Group = Initial expressive vocabulary group defined at first administration (12-16 months) as > 25th percentile (higher) vs. ≤ 25th percentile (lower). Higher = reference group. Receptive Group = Initial receptive vocabulary group defined at first administration using either > 25th or > 50th percentile (higher) vs. ≤ 25th or ≤ 50th percentile (lower). Model comparisons (Likelihood ratio (LR) tests) for Models 4 and 6 are in relation to Model 2 (Table 4), Model 5 is in relation to Model 4, and Model 7 in relation to Model 6.

To further illustrate, at 18 months, the modeled expressive vocabulary size for children with higher initial expressive vocabulary scores and with > 50th percentile initial receptive scores was about 137 words, about 48 words higher than for their higher-expressive peers with receptive scores ≤ 50th percentile (89 words). This difference persisted until 24 months, such that higher-expressive toddlers with > 50th percentile receptive scores were predicted to produce about 346 words, about 42 more words than their higher-expressive peers with ≤ 50th percentile receptive scores (304 words). However, the modeled expressive vocabulary size for toddlers with lower initial expressive vocabularies was not different for children with ≤ 50th percentile receptive scores at both ages (18 months: 47 words; 24 months: 212 words) vs. those with > 50th percentile receptive scores (18 months: 39 words; 24 months: 186 words).

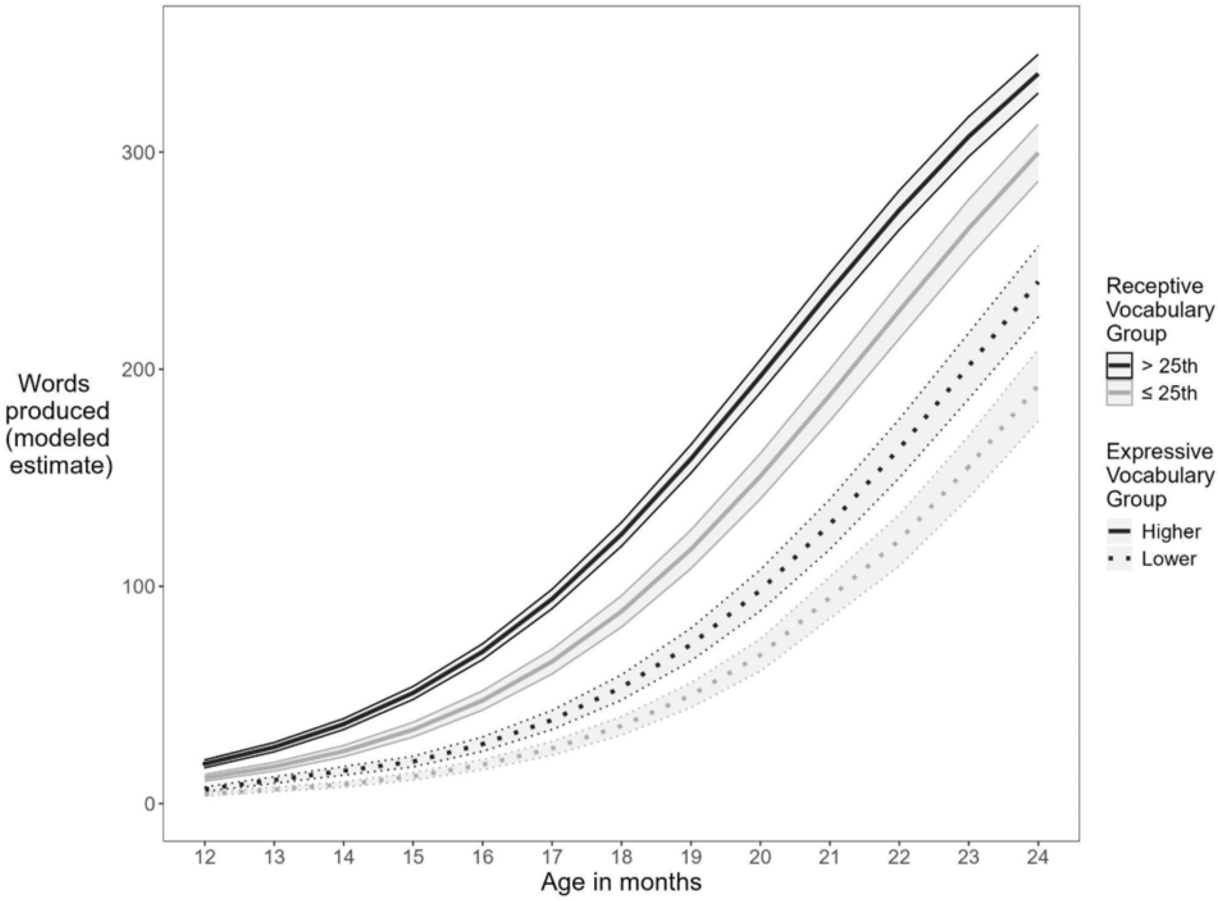

Figure 2. Predicted expressive vocabulary growth over age for children classified by initial expressive vocabulary size as higher (*n* = 71) vs. lower (*n* = 21) and as > 25th (higher) vs. ≤ 25th percentile (lower) in receptive vocabulary based on Model 4. Shaded area represents +/- 1 SE.

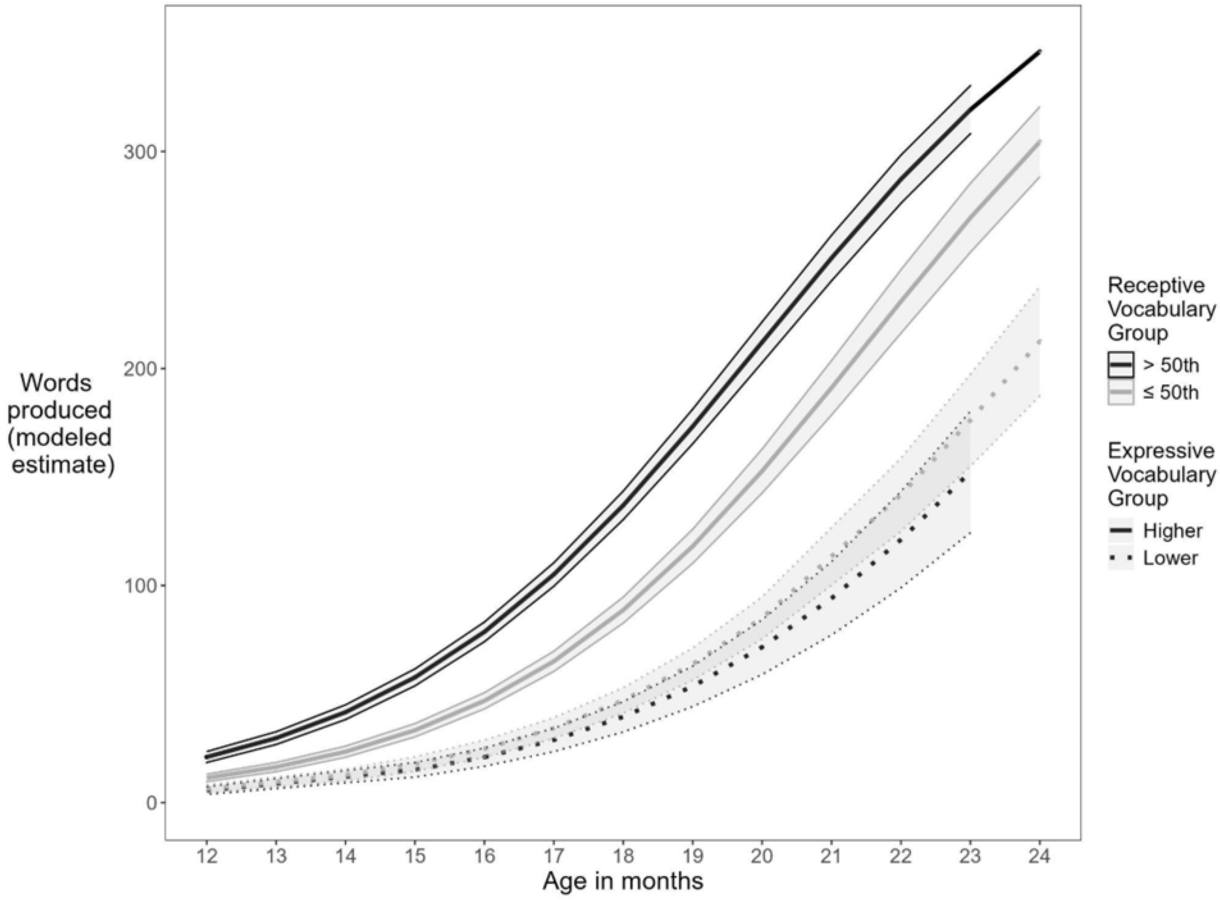

Figure 3. Predicted expressive vocabulary growth over age for children initially classified as higher (*n* = 71) and lower (*n* = 21) in expressive vocabulary and as > 50th (higher) vs. ≤ 50th (lower) in receptive vocabulary at first administration (12-16 months) based on Model 7. Shaded area represents +/- 1 SE.

In summary, across both initial expressive vocabulary groups, classifying children based on receptive vocabulary levels below/above the 25th percentile gave additional information about expressive vocabulary growth, such that children with higher receptive skills predicted consistently larger expressive vocabularies throughout development. This lack of a significant moderation must be interpreted with caution, and we invite future research to replicate these findings in larger and other language samples. In contrast, when a more liberal cut-off was used, information about receptive level was only informative for children with higher initial expressive vocabulary scores. Simply falling below the 50th percentile did not add additional predictive power beyond expressive vocabulary alone for children with initially lower expressive scores.

Discussion

This study used longitudinal growth modeling to investigate trajectories of expressive vocabulary growth from 12 to 24 months in toddlers learning Hebrew. There were three main group level findings. First, children who were ≤ 25th percentile in vocabulary development at first administration, i.e., lower initial expressive vocabulary, showed lower average vocabulary scores throughout the second year of life, compared to their peers with higher initial scores. Initial expressive vocabulary group status identified prior to 16 months was a significant predictor of vocabulary size over the second year. Thus, despite large variability across individuals (38% of the children who were classified in the lower group at the first administration moved into the normal range at the second or third administration), an early assay of a child's expressive vocabulary size relative to peers in the same age group can yield critical information regarding that child's trajectory of expressive vocabulary growth over the next several months.

Second, in spite of overall delays, initial expressive vocabulary group status did not significantly moderate the shape of the age-related trajectories of vocabulary growth. That is, our analyses suggested that toddlers across the full continuum of expressive vocabulary levels showed the same general pattern of growth, on average, over the period studied. This parallel growth trajectory suggests that toddlers with and without initial delays follow a similar developmental path, possibly reflecting similar mechanisms of learning in children with all ability levels. While it is possible that our study was not sufficiently powered to statistically detect a moderation, the interaction term contributed very little variance (< 1%) to the overall model and estimates appeared to be reasonably precise, relative to the other fixed effects of group and age. Future studies are necessary to further explore whether this pattern of findings replicates in larger samples and in other language communities. On the other hand, trajectories of expressive vocabulary growth were significantly developmentally delayed for the toddlers with initially lower expressive vocabulary scores on average, compared to their higher scoring peers. While it is unknown which (if any) of these children ultimately received a clinical diagnosis of language disorder, this study suggests some degree of developmental continuity in children at both the higher and lower-ends of the early expressive vocabulary spectrum.

We next asked whether skills involving comprehension may moderate expressive vocabulary growth trajectories differently for children at the higher- and lower-ends of the continuum of expressive vocabulary at 12-16 months. This question is particularly interesting given that receptive vocabulary scores were generally lower in the children with lower vs. higher expressive vocabulary groups, however, receptive vocabulary scores spanned the full possible range in both groups. We tested two groupings on receptive vocabulary percentile, one that was analogous to our grouping for expressive language (25th percentile) and another that was more lenient

(50th percentile). Thus, our third main finding was that grouping children < 25th percentile added information beyond initial expressive vocabulary group status in predicting expressive vocabulary growth until 24 months. Thus, for children in this age range, these results suggest that lower early receptive vocabulary skills are important to monitor and that receptive skills operate in tandem with expressive language abilities in promoting vocabulary growth (e.g., Bergelson et al., 2020; Fernald & Marchman, 2012). However, when the 50th percentile cut-off was used, information about receptive level was only informative in children with higher initial expressive vocabulary scores; simply falling below the median did not predict lower levels of expressive vocabulary growth for children who initially fell behind their peers in expressive vocabulary.

In general, this pattern of findings suggests that better early receptive abilities can provide a boost to expressive vocabulary development regardless of initial expressive vocabulary level, not just in children at the lowest-end of the continuum. At the same time, these results provide additional precision to the claim that low receptive skills may confer additional risk for those children with early expressive delays, by showing that only receptive levels in the bottom quartile were predictive of continued slowing of expressive vocabulary development. These results are consistent with those of earlier studies showing that children with concomitant receptive and expressive delays are likely to be more at risk for later language learning difficulties, compared to their peers with initial delays who do not also show comprehension delays (Buschmann et al., 2008; Thal et al., 2013). Our results expand this literature in two ways. First, they show that early receptive skills can be informative as early as 16 months. Second, that information is most useful for predicting expressive language growth for initially-delayed toddlers when it reflects relatively low levels of skill, rather than just moderate delays.

From a theoretical perspective, several insights emerged. First, evidence for continuity of growth rates throughout the second year of life indicates that early expressive skills could be an indicator of delayed foundational abilities, similar to that observed in children at older ages (e.g., Bornstein et al., 2016). Second, the generally comparable trajectories of lexical development in children with both higher and lower initial expressive vocabulary implies that linguistic skills follow a similar trajectory across the ability continuum, even when they manifest later in development. Third, although expressive and receptive vocabulary skills are interconnected, they are fortified by separate sets of underlying processes (Cheung et al., 2022). This suggests that comprehension skills play an important role in boosting expressive vocabulary development in children across the expressive language continuum (Bergelson et al., 2020; Fernald & Marchman, 2012).

These results have important clinical implications. First, continuities in early vocabulary development are observable even when assessed early in the second year of life. Monitoring of expressive vocabulary skills even early in development can provide useful information for children's development at least through the second year, when more comprehensive assessments are likely to be more feasible. Further, these results reinforce the significance of monitoring both receptive and expressive vocabulary skills as potential risk factors for difficulties in later language and literacy development. Moreover, the assessment of comprehension skills, as well as the enhancement of receptive vocabulary skills, should be considered as important targets in the early facilitation of vocabulary learning. While we cannot address this issue empirically with the data we have available here, higher scores in comprehension abilities among children with initial delays could potentially signify a chance for eventual "catching up" prior to the age of 3 years, when clinical evaluations are considered to be more robust. It would be useful for future studies to replicate our findings with children identified with delays early in development and following them into the 3rd year of life and beyond to further understand the processes underlying comprehension and production in relation to later language outcomes.

Limitations

While the longitudinal nature of our sample was a strength, the sample was relatively small and consisted of primarily higher-SES families, the majority coming from caregivers with college educations. Moreover, it is possible that our study was not sufficiently powered to statistically detect moderations. Future studies should explore whether these patterns of findings replicate in larger and more sociolinguistically-diverse samples and specifically test whether socioeconomic status moderates the observed patterns of developmental change. Second, data were only available until the children were 24 months. We do not know whether children with lower levels of initial expressive vocabulary as we have defined it here, i.e., < 25th percentile between 12-16 months, continued to show a similar degree of continuity after 2 years of age. It should be noted that there may have been other predictors or moderators that were not explored here, e.g., the presence or absence of a family history of developmental language disorders, cognitive maturation, or the use of communicative or symbolic gestures. Finally, the present analysis was based on quantitative data and represented group differences. Single case studies that allow a closer look on the interplay between production and comprehension at the early phases of lexical learning are desirable as a complement to the current analyses.

Conclusions

This study presented a unique look at growth trajectories in expressive vocabulary beginning early in the second year of life in Hebrew-speaking toddlers. Results indicated that low expressive and receptive vocabulary levels were each significant indicators of more protracted expressive vocabulary development. Similarities were observed in the general shapes of the growth trajectories in children identified with both higher and lower initial levels of expressive skill. Taken together, these results provide new evidence regarding the stability of early language skills and suggest parallel mechanisms guiding expressive language development in children who span the full continuum of ability levels. Moreover, these results add to the existing literature that early comprehension abilities buttress early productive skills early in development for children across the expressive ability spectrum. From a clinical perspective, these results suggest that monitoring of low levels of receptive, as well as expressive, skills as early as 16 months can provide useful information regarding trajectories of expressive vocabulary growth through 24 months of age.

References

Bates, E., Marchman, V. A., Thal, D., Fenson, L., Dale, P., Reznick, J. S., … Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*(1), 85–123. https://doi.org/10.1017/S0305000900008680

Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child Development Perspectives, 14*(3), 142-149. https://doi.org/10.1111/cdep.12373

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253-3258. https://doi.org/10.1073/pnas.1113380109

Bleses, D., Basbøll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive Processes, 26*(8), 1193–1231. https://doi.org/10.1080/01690965.2010.515107

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics, 37*(6), 1461–1476. https://doi.org/10.1017/S0142716416000060

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008a). Early vocabulary development in Danish and other languages: a CDI-based comparison. *Journal of Child Language, 35*(3), 619–650. https://doi.org/10.1017/S0305000908008714

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008b). The Danish Communicative Developmental Inventories: Validity and main developmental trends. *Journal of Child Language, 35*(3), 651–669. https://doi.org/10.1017/S0305000907008574

Bornstein, M. H., Hahn, C.-S., & Putnick, D. L. (2016). Stability of core language skill across the first decade of life in children at biological and social risk. *Journal of Child Psychology and Psychiatry, 57*(12), 1434–1443. https://doi.org/10.1111/jcpp.12632

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

Buschmann, A., Jooss, B., Rupp, A., Dockter, S., Blaschtikowitz, H., Heggen, I., & Pietz, J. (2008). Children with developmental language delay at 24 months of age: results of a diagnostic work-up. *Developmental Medicine & Child Neurology, 50*(3), 223-229. https://doi.org/10.1111/j.1469-8749.2008.02034.x

Carson, C. P., Klee, T., Carson, D. K., & Hime, L. K. (2003). Phonological profiles of 2-year-olds with delayed language development: Predicting clinical outcomes at age 3. *American Journal of Speech-Language Pathology, 12*(1), 28-39. https://doi.org/10.1044/1058-0360(2003/050)

Cheung, R. W., Hartley, C., & Monaghan, P. (2022). Receptive and expressive language ability differentially support symbolic understanding over time: Picture comprehension in late talking and typically developing children. *Journal of Experimental Child Psychology*, 214, 105305. https://doi.org/10.1016/j.jecp.2021.105305

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46*, 544–560. https://doi.org/10.1044/1092-4388(2003/044)

deMayo, B., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C., ... Marchman, V. A. (2021). Web-CDI: A system for online administration of the MacArthur-Bates Communicative Development Inventories. *Language Development Research, 1*(1), 55–98. https://doi.org/10.34842/kr8e-w591

Dromi, E. (1987). *Early lexical development*. Cambridge University Press.

Edwards, N. M., Kaiser, E., & Stapel-Wax, J. L. (2021). Investigating a "wait and see" mindset among pediatric health care providers missed opportunities to refer toddlers with autism spectrum disorder to part c early intervention. *Infants and Young Children, 34*(4), 284–305. https://doi.org/10.1097/IYC.0000000000000201

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child*

Development, 59(5), 1–173. https://doi.org/10.2307/1166093

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research, 60*(10), 2935–2948. https://doi.org/10.1044/2017_JSLHR-L-16-0310

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. MIT Press.

Gendler-Shalev, H., & Dromi, E. (2022). The Hebrew Web Communicative Development Inventory (MB-CDI): Lexical development growth curves. *Journal of Child Language, 49*(3), 486–502. https://doi.org/10.1017/S0305000921000179

Gendler-Shalev, H., & Novogrodsky, R. (2024). Word characteristics of late talkers' early lexicon. *First Language, 44*(5), 507-532. https://doi.org/10.1177/01427237241269568

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

Hadley, P. A., & Holt, J. K. (2006). Individual differences in the onset of tense marking: A growth-curve analysis. *Journal of Speech, Language, and Hearing Research, 49*, 984–1000. https://doi.org/10.1044/1092-4388(2006/071)

Hadley, P. A., & Short, H. (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(6), 1344-1362. https://doi.org/10.1044/1092-4388(2005/094

Hall, S., Rumney, L., Holler, J., & Kidd, E. (2013). Associations among play, gesture and early spoken language acquisition. *First Language, 33*(3), 294-312. https://doi:10.1177/0142723713487618

Hao, M., Shu, H., Xing, A., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods, 40*(3), 728–733. https://doi.org/10.3758/BRM.40.3.728

Jiménez, E., Haebig, E., & Hills, T. T. (2021). Identifying areas of overlap and distinction in early lexical profiles of children with autism spectrum disorder, late talkers, and typical talkers. *Journal of Autism and Developmental Disorders, 51*(9), 3109-3125. https://doi.org/10.1007/s10803-020-04772-1

Jones, S. S. (2003). Late talkers show no shape bias in a novel name extension task. *Developmental Science, 6*(5), 477-483. https://doi:10.1111/1467-7687.00304

Jones, S. D., & Brandt, S. (2019). Neighborhood density and word production in delayed and advanced learners. Journal of Speech, *Language, and Hearing Research, 62*(8), 2847-2854. https://doi.org/10.1044/2019_JSLHR-L-18-0468

Kalashnikova, M., Schwarz, I.-C., & Burnham, D. (2016). OZI: Australian English Communicative Development Inventory. *First Language, 36*(4), 407–427. https://doi.org/10.1177/0142723716648846

Marchman, V. A., Dale, P. S., & Fenson, L. (2023). *The MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual.* (V. A. Marchman, P. S. Dale, & L. Fenson, Eds.) (3rd Edition). Brookes Publishing Co.

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science, 14*(2), 181–189. https://doi.org/10.1111/j.1467-7687.2010.00968.x

Meylan, S. C., & Bergelson, E. (2022). Learning through processing: Toward an integrated approach to early word learning. *Annual Review of Linguistics, 8*, 77-99. https://doi.org/10.1146/annurev-linguistics-031220-011146

Moyle, M. J., Weismer, S. E., Evans, J. L., & Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research, 50*, 508–528. https://doi.org/10.1044/1092-4388(2007/035

R_Core_Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*, 587–599. https://doi.org/10.1044/jshd.5404.587

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research, 52*, 16–30. https://doi.org/10.1044/1092-4388(2008/07-

0171)

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*(2), 141–150. https://doi.org/10.1002/ddrr.1108

Rescorla, L., & Dale, P. S. (2013). *Late talkers: Language Development, Interventions, and Outcomes.* Brookes Publishing Co.

Singleton, N. C. (2018). Late Talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics, 65*(1), 13–29. https://doi.org/10.1016/j.pcl.2017.08.018

Stasinopoulos, M. D., Rigby, R. A., Heller, G. A., Voudouris, V., & De Bastiani, F. (2017). *Flexible smoothing using GAMLSS in R (1st ed.)*. CRC Press.

Thal, D. J., Marchman, V. A., & Tomblin, J. B. (2013). Late-talking toddlers: Characterization and prediction of continued delay. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 169–201). Paul H. Brookes Publishing Co.

Thal, D. J., Reilly, J., Seibert, L., Jeffries, R., & Fenson, J. (2004). Language development in children at risk for language impairment: Cross-population comparisons. *Brain and Language, 88*(2), 167-179. https://doi.org/10.1016/S0093-934X(03)00096-8

Thal, D., & Tobias, S. (1994). Relationships between language and gesture in normally developing and late-talking toddlers. *Journal of Speech & Hearing Research, 37*(1), 157-170. https://doi.org/10.1044/jshr.3701.15

Ukoumunne, O. C., Wake, M., Carlin, J., Bavin, E. L., Lum, J., Skeat, J., ... & Reilly, S. (2012). Profiles of language development in pre-school children: a longitudinal latent class analysis of data from the Early Language in Victoria Study. *Child: Care, Health and Development, 38*(3), 341-349. https://doi.org/10.1111/j.1365-2214.2011.01234.x

Vehkavuori, S. M., & Stolt, S. (2019). Early lexicon and language skills at 42 months. *Clinical Linguistics & Phonetics, 33*(9), 854-868. https://doi.org/10.1080/02699206.2019.1584721

Verganti, C., Suttora, C., Zuccarini, M., Aceti, A., Corvaglia, L., Bello, A., ... & Sansavini, A. (2024). Lexical skills and gesture use: A comparison between expressive and receptive/expressive late talkers. *Research in Developmental Disabilities, 148*, 104711. https://doi.org/10.1016/j.ridd.2024.104711

Whitehurst, G. J., & Fischel, J. E. (1994). Practitioner Review: Early developmental language delay: What, if anything, should the clinician do about it? *Journal of Child Psychology and Psychiatry, 35*(4), 613–648. https://doi.org/10.1111/j.1469-7610.1994.tb01210.x

Data, Code and Materials Availability

The Hebrew webCDI is available at https://webcdi.org/. The data and code to generate these analyses are available at https://github.com/vmarchman/Vocab-Hebrew

Ethics Statement

Ethics approval was obtained from the ethics committee of the University of Tel-Aviv. All participants gave informed written consent before taking part in the study.

Authorship and Contributorship Statement

All authors conceived of the study. **Hila Gendler-Shalev** and **Virginia A. Marchman** designed the study. **Hila Gendler-Shalev** and **Esther Dromi** collected the data. **Virginia A. Marchman** analyzed the data. **Hila Gendler-Shalev** and **Virginia A. Marchman** wrote the first draft of the manuscript. All authors revised the manuscript and approved the final version. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgments

We would like to thank the families who participated in our study. We would also like to thank the editor and reviewers, as well as Alvin Tan, George Kachergis, and Michael Frank for their valuable assistance with statistical analysis. Their insightful comments and suggestions significantly improved the quality of this manuscript.