ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Cross-Situational Statistical Word Learning in Late Language Emergence: An Online Study

Elizabeth Schoen Simmons,[a] Olivia Cayward,[b] and Rhea Paul[a]

[a] Department of Communication Disorders, Sacred Heart University, Fairfield, CT [b] Department of Speech-Language Pathology, Sacred Heart University, Fairfield, CT

ARTICLE INFO

Article History:
Received September 20, 2024
Revision received November 22, 2024
Accepted January 2, 2025

Editor-in-Chief: Julie A. Washington
Editor: James Booth

https://doi.org/10.1044/2025_JSLHR-24-00670

ABSTRACT

Purpose: Cross-situational statistical learning is one mechanism by which typically developing toddlers map words to referents. Yet, this type of statistical learning has been found less efficient in children with developmental language disorder (DLD). The purpose of this article is to evaluate cross-situational statistical learning in very young children with language delay, late talkers (LTs), compared to typically talking toddlers. We predict that LTs will show inefficiency in cross-situational statistical word learning similar to older children with DLD.
Method: LT (*n* = 15, 18–34 months) and typical talker (TT; *n* = 15, 18–35 months) groups matched on chronological age and sex completed a cross-situational statistical learning task in which they were trained on six novel word–referent pairs and then tested on these word–referent associations. The experiment was completed on the participant's home computer, and gaze was recorded for the duration of the experiment. Mixed-effects models were used to evaluate group differences in time spent looking at labeled referents as a measure of learning.
Results: The LT group spent an equal proportion of time looking at the named targets and the unnamed distractors when tested, suggesting minimal learning had occurred. The TT group, in contrast, spent a significantly greater proportion of time looking at the targets when labeled, indicating more established word–referent links.
Conclusions: These findings suggest that LTs, like older children with DLD, are less efficient at leveraging cross-situational statistical learning opportunities that may, in addition to other factors, contribute to their slow expressive vocabulary development.

Statistical learning is an implicit mechanism that represents one critical component, out of many, required for vocabulary acquisition during infancy and toddlerhood. As a form of pattern detection that supports learning of rules from input without conscious awareness or explicit instruction, statistical learning has been shown to facilitate early word learning processes, including the segmentation of the speech stream for identifying word boundaries (Saffran et al., 1996) and linking words with their referents (Yu & Smith, 2007). Indeed, the majority of toddlers will acquire a spoken vocabulary without explicit instruction and with great ease despite the seemingly daunting computational requirements of the task over the first few years of life.

There is a heterogeneous subset of toddlers under 3 years of age who fail to meet typical expressive language milestones in absence of any overt sensory or cognitive impairments (Fisher, 2017; Perry & Kucker, 2019). These children are described as having late language emergence and are commonly referred to as late talkers (LTs). LTs are generally defined as toddlers, between 18 and 35 months of age, who are at the low end of the productive vocabulary distribution (i.e., producing fewer than 50 words by their second birthday and/or not yet producing two-word combinations; Paul, 1991; Rescorla, 1989). They are at increased risk of persistent language impairment, known as developmental language disorder (DLD; Matte-Landry et al., 2020). Much of the research on LTs has

Correspondence to Elizabeth Schoen Simmons: simmonse3@sacredheart.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

focused on describing their spoken vocabularies with less work directed toward understanding the mechanisms that may result in delayed spoken language onset.

Statistical Word Learning in Typical Toddlers

Toddlers learn words in perceptually chaotic environments where they must make assumptions about spoken word–referent pairs to link the words they hear and the objects present within their environment. One way in which they do this is through tracking co-occurrences of words and their referents across many learning opportunities and generating hypotheses about which spoken words map to which referents (Yu & Smith, 2007). This tracking of co-occurrences is a form of statistical learning (Aslin, 2017) that has been demonstrated by infants as young as 12 months (Smith & Yu, 2008).

Infants' and toddlers' ability to implicitly detect patterns within auditory and visual input contributes to the foundation for acquiring a lexicon. The landmark study by Saffran et al. (1996) demonstrated that infants as young as 8 months can identify word boundaries in a synthesized speech stream after only a 2-min presentation of syllable strings using only the transitional probabilities between syllables. Segmentation of the continuous acoustic signal is one of the first steps in word learning. Once infants can reliably carve up the speech stream into individual word forms, they must start making connections between the words and the referents within their environment. Cross-situational statistical word learning is one such approach that infants use to make these connections.

The traditional cross-situational statistical word learning paradigm for toddlers is a passive task that includes a set of training trials where the word–object mappings are learned and then a set of test trials are presented to measure learning (Smith & Yu, 2008, 2013; Yu & Smith, 2011). Eye movements are recorded for the duration of the task and used to evaluate preference for items during the experiment. For training trials, participants are simultaneously presented with two novel objects on opposite sides of a screen, followed by the presentation of two novel spoken words. Within each trial, word–object mappings are ambiguous, but across trials when a recent word is presented with a new object, that word–object pair can be eliminated as a match. Thus, over training, the accumulation of co-occurrence statistics provides sufficient evidence as to which word is mapped to which object. The test phase uses a procedure much like that of preferential looking paradigms. Participants are presented with a single word and two potential referents, the target and a distractor, if participants have learned the word–object mapping, they spend more time looking at the target, labeled items and significantly less time looking at the distractor, unlabeled items. This paradigm is unique as it does not simply measure whether or not an infant can learn word–object associations, but rather it measures whether these associations can be learned implicitly by tracking the co-occurrence of words and objects presented in an ambiguous fashion over the course of many trials, potentially simulating real-life learning (Roembke et al., 2023). It is generally agreed that a constraint of successful cross-situational statistical word learning is attention and that the toddler must attend to the spoken words and referents in order for learning to occur (Aslin, 2017; Smith & Yu, 2008, 2013).

Cross-situational statistical word learning has been demonstrated by infants (Escudero et al., 2016; Smith & Yu, 2008, 2013; Yu & Smith, 2011), children (Benitez & Li, 2023; Hartley et al., 2020; Suanda et al., 2014), and adults (Monaghan & Mattock, 2012; Roembke & McMurray, 2016). Learner strategies are generally divided into two approaches: (a) aggregation of probabilistic data, or implicit strategies that align with the principles of statistical learning, and (b) propose-but-verify, a more explicit strategy where hypotheses are created about word–object links. The data aggregation strategy typified by Smith and Yu (2008) and Yu and Smith (2007) suggests that cross-situational statistical word learning is simply the tracking of co-occurrences of the word–object pairs and then aggregating this co-occurrence data over time. Conversely, the propose-but-verify strategy posits that learners generate a hypothesis about which word goes with which object, and if that hypothesis is not supported by subsequent data, the learner starts the process again with a new prediction (Medina et al., 2011; Roembke et al., 2023; Trueswell et al., 2013; Yurovsky & Frank, 2015). Passive, laboratory-based paradigms, designed for infants and toddlers, as the ones described above (Smith & Yu, 2008, 2013; Yu & Smith, 2011), are generally thought to measure implicit data aggregation strategies that more closely align with statistical learning (McGregor et al., 2022).

A handful of studies demonstrate that infants are able to link novel word forms and their visual referents during cross-situational statistical learning paradigms (Escudero et al., 2016; Smith & Yu, 2008, 2013; Yu & Smith, 2011). For example, Smith and Yu (2008) presented 12- and 14-month-old toddlers with six novel word–referent pairs using the cross-situational statistical learning paradigm. For this task, participants were presented with two spoken novel words and two novel shapes. The trial on its own was ambiguous as the participant did not know which word mapped to which object. Over the course of many trials, participants learned to associate each word–object pair by extracting the co-occurrence data. Through the tracking of co-occurrences, both age groups were able to learn at least a subset of the spoken word–referent pairs as measured by more time spent looking at the labeled object compared to the unlabeled object during the test phase. Escudero et al. (2016) replicated the findings above with older infants 17 and 20 months of age. They also extended the findings by modifying the paradigm to test whether infants could use cross-situational statistical learning opportunities to acquire phonetically similar novel word (e.g., *bon-ton*)–object pairs. Indeed, in this perceptually more complex test, infants across both age groups linked at least some of the novel word–object pairings between lexical forms that differed by a single phoneme.

Of course during word learning in natural settings, very young children have access to redundant environmental cues above and beyond the statistical probabilities provided by language patterns. For example, nonlinguistic cues such as gaze (Baldwin et al., 1996; Bannard & Tomasello, 2012; Tenenbuam et al., 2014) and gestures (Briganti & Cohen, 2011; Cheung et al., 2021; Grassmann & Tomasello, 2010) assist toddlers' mapping of words to referents in their environment. These cues, along with the statistical patterns of spoken language, provide redundant support to help toddlers make word–referent links within their environment.

Word Learning in LTs

LTs are generally described as having limited expressive vocabularies without any obvious cause such as overt neurodevelopmental disorders (e.g., autism or developmental delay in nonlinguistic domains) or sensory differences (Paul, 1991). The outcomes observed for LTs are heterogeneous, and a subset are referred as *late bloomers* (approximately 17%) whose delays are seemingly transient and eventually close the expressive language achievement gap with their typically developing (TD) peers, at least when tested on static, standardized measures (Chilosi et al., 2019; Rescorla, 2009; Rescorla & Turner, 2015). LTs who fail to catch up by age 5 of years will eventually meet criteria for DLD (approximately 15%; Collison et al., 2016; Singleton, 2018), and even with those LTs who seemingly achieve typical language functioning, a majority (approximately 68%) will retain subclinical language weaknesses (e.g., score in the average range yet significantly lower than socioeconomically matched typical peers on standardized assessment; Rescorla, 2009). Research on LTs has historically focused on characterizing their spoken vocabularies and other symbolic behaviors related to language (e.g., gestures, play skills). Yet, mounting evidence highlights that less efficient word learning (Cheung et al., 2022; Ellis-Weismer et al., 2013; MacRoy-Higgins & Dalton, 2015; MacRoy-Higgins & Montemarano, 2015) may be part of a broader syndrome beyond spoken vocabulary.

Studies of fast mapping in LTs suggest they learn fewer word–object relationships than their TD counterparts when taught novel noun–object pairs (Asadi et al., 2019; Ellis-Weismer et al., 2013; MacRoy-Higgins & Dalton, 2015; Rujas et al., 2019) and novel verb–action pairs (Asadi et al., 2019; Rujas et al., 2019). Fast-mapping paradigms have some variants but generally use a set of procedures in which a toddler is presented with a known object and a novel object and then hear the word for the novel object. After only a single exposure, many toddlers are above chance at identifying the novel object when named (Carey & Bartlett, 1978; Heibeck & Markman, 1987). An issue within fast-mapping paradigms is that they fail to mirror the type of input received by infants and toddlers when learning language. In one longitudinal study of language development, fast-mapping performance accounted for less than 10% of the variance in spoken language development in toddlers, whereas nonword repetition was the best predictor accounting for 36% of variance (Stokes & Klee, 2009).

Another method for testing toddler word learning is through use of naturalistic, play-based paradigms with repeated exposure to the words and objects with the idea that enriched input reduces referential ambiguity (Yu & Smith, 2011). MacRoy-Higgins et al. (2013) and MacRoy-Higgins and Montemarano (2015) used 10 play-based sessions over the course of 5 weeks to teach 12 novel word–object pairs to a group of LTs and age-matched typical controls. Participants were exposed to the objects and heard each word 50 times over the course of training. Comprehension probes posttraining using a four-alternative forced-choice task (*point to the toov*) showed that both participant groups comprehended at least some of the words, suggesting learning, but the LTs recognized significantly fewer word–object pairs than their typical peers posttraining. Moreover, they found that the LTs spent significantly less time visually attending to the objects during training (MacRoy-Higgins & Montemarano, 2015). Taken together, these studies of fast-mapping and more naturalistic procedures provide hints of word learning differences between LTs and TD toddlers.

Statistical Learning in DLD

There are no known studies that measure statistical learning in LTs, but they may be at risk for poor pattern detection, as seen in both preschool and school-age children with DLD. Impoverished auditory statistical learning has been demonstrated in school-age children with DLD (Evans et al., 2009; Haebig et al., 2017; Lukács et al., 2021) using a task similar to that of Saffran et al. (1996). Specifically, children with DLD are less able than language typical peers to identify word boundaries based on transitional probabilities when exposed to an artificial language.

In addition to demonstrating difficulties using statistical learning to parse the speech stream, children with DLD may also be less efficient at using cross-situational statistical word learning opportunities to link a word with its referent. Ahufinger et al. (2021) presented children between 6 and 12 years old, with and without DLD, eight novel word–object pairs using a cross-situational statistical learning design. They used a classic four-alternative forced-choice task and eye tracking to measure learning during the test phase. Results of the forced-choice task revealed that the DLD group made fewer word–object mappings compared to their TD language counterparts. Moreover, even for the items correctly identified, the DLD participants spent more time looking at the competitor objects compared to TD participants suggesting that the distractors competed more strongly for recognition in the DLD group, potentially pointing to less robust word–object mappings even for learned pairs. McGregor et al. (2022) replicated and extended these findings in 7-year-olds with DLD. In this study, children with DLD were presented with cycles or blocks of input and then tested after each cycle in order to evaluate the role of input amount on pattern extraction. Similar to findings of Ahufinger et al., these children with DLD made fewer word–object mappings after one cycle and were less accurate across all cycles compared to their TD peers.

To our knowledge, only one study has tested cross-situational statistical learning with preschoolers who met criteria for late talking at age 2 years (Cheung et al., 2022). These former LTs, sometimes referred to as *late bloomers,* scored in the average range on measures of expressive and receptive vocabulary at age 4 years. Findings from this study showed that the former LTs did not differ from TD peers in accuracy of selecting the target object from a foil when labeled. Interestingly, their retention for these words after 5 min post–initial testing was significantly poorer than their peers. Thus, it would seem that former LTs may be able to use cross-situational probabilities for word–object mappings, but these associations may fail to consolidate in memory or they may require more input for consolidation. It is important to note that although 50%–75% of LTs perform within the typical range on standardized tests by 4 years of age, not all do (Paul & Roth, 2011); thus, these findings may not be representative of known heterogeneity observed within LTs.

The Current Study

There are no known studies of cross-situational statistical word learning in late-talking toddlers; yet, there is emerging evidence in the literature that impaired word learning may be part of the broader syndrome expression of this condition (Cheung et al., 2022; Ellis-Weismer et al., 2013; MacRoy-Higgins & Dalton, 2015; MacRoy-Higgins & Montemarano, 2015). Thus, this study aims to begin to fill the gaps in the word learning literature in LTs. Specifically, we ask whether LTs use cross-situational statistical learning opportunities to map a small set of word–referent relationships.

LTs are a heterogeneous group of toddlers (Fisher, 2017; Perry & Kucker, 2019) as are their linguistic trajectories as they reach school age and beyond. While some indeed close the language gap and catch up to typical peers as measured by standardized language assessments; nonetheless, they are at increased risk for DLD (Collison et al., 2016; Matte-Landry et al., 2020; Rescorla, 2009; Singleton, 2018), and even for those who do not receive a formal diagnosis, the majority retain subclinical language weaknesses (Rescorla, 2009). Children with DLD are less efficient at using statistical cues to identify word boundaries during artificial language learning tasks (Evans et al., 2009; Haebig et al., 2017; Lukács et al., 2021). More specific to the current study, preschool children (McGregor et al., 2022) and school-age children (Ahufinger et al., 2021) with DLD are less efficient at mapping word–referent pairs during cross-situational statistical word learning opportunities.

Given that the majority of LTs will present with language differences at school age (Collison et al., 2016; Matte-Landry et al., 2020; Rescorla, 2009; Singleton, 2018), whether a formal diagnosis of DLD or subthreshold weaknesses, we predict that they will be less able to leverage the co-occurrence statistics, as are older children with DLD, resulting in less efficient word–object mappings compared to TD toddlers.

For our primary analysis, we look at word–object mappings collapsed across all pairs, similar to other studies of toddler cross-situational word learning (Escudero et al., 2016; Smith & Yu, 2008, 2013; Suanda et al., 2014). We predict that the LT group will spend a similar amount of time looking at the labeled referents and unlabeled distractor items during the test phase, as observed in older children with DLD—a pattern that does not indicate learning. For the typical toddlers, we predict they will spend significantly more time looking at the target labeled items compared to the unlabeled distractor items during the test phase, thus demonstrating learning.

Next, we will move to our exploratory analyses, which are more tentative given the small sample size. First, we will evaluate pair-level word–referent mappings. We predict here that the LTs will acquire fewer word–referent pairs compared to their TD peers, again as reported in older children with DLD (Ahufinger et al., 2021; McGregor et al., 2022). Second, since visual attention has been implicated in cross-situational statistical word learning, we evaluated whether the LTs and typical toddlers spent similar amounts of time looking at the referents during the learning phase as this could easily be measured in this task. We predicted that LTs may spend less time looking at the objects similar to the findings observed by MacRoy-Higgins and Montemaro (2015) during naturalistic word learning sessions. Auditory attention is much more complex to quantify in our experiment; thus, we were unable to measure this variable within our paradigm.

Method

Participants

All procedures involving human subjects were approved by the institutional review board at Sacred Heart University (Protocol No. 201027A) prior to recruitment and data collection. Participants' parent or legal guardian provided verbal consent for their toddler to participate before data were collected and families were compensated in the form of a $20 giftcard for their time.

Two groups of toddlers from monolingual English-speaking households, with no reported history of vision or hearing differences, were recruited from Children Helping Science (CHS; formally known as *Lookit*; Scott & Schulz, 2017). The two groups of toddlers included an LT group and a typical talker (TT) group. Inclusion criteria for each group are described below. Parents completed two questionnaires related to their toddler's language and communication development. The MacArthur–Bates Communicative Development Inventories, Words and Sentences (MB-CDI:WS; Fenson et al., 2007) is a norm-referenced parent report measure of toddler-spoken vocabulary. The form has 668 items organized into semantic and syntactic categories, and parents are asked to indicate which words their toddler produces regularly. The web-based version of the MB-CDI:WS, via an online administration portal, was used to measure the expressive vocabularies of each participant and collect information on hearing and vision status. Expressive vocabulary results from the MB-CDI:WS were used for group assignment (either LT or TT group). In addition, each parent completed the Modified Checklist for Autism in Toddlers–Revised (M-CHAT-R; Robins et al., 2009), which measures the presence of autism symptoms and provides a cutoff for "high risk."

This project was conducted entirely online using Massachusetts Institute of Technology's platform CHS/Lookit (Scott & Schulz, 2017). Our original target sample size was 30 late-talking toddlers (n = 30) and 30 TTs (n = 30). Despite the flexibility offered via online data collection, including ease of recruitment, we also encountered several barriers to data acquisition, which left us with a smaller sample size (N = 30). This is discussed further in the Limitations section of this article. See Figure 1 for flowchart of participant enrollment and attrition.

LTs

The LT group (n = 15) consisted of toddlers ages 18–34 months, with delayed expressive language as indicated by the MB-CDI:WS (Fenson et al., 2007) below the 16th percentile. Age range and MB-CDI:WS cutoff were selected using those widely reported in the literature (Collison et al., 2016; Curtis et al., 2023; Ellis et al., 2015; Horvath et al., 2019; MacRoy-Higgins & Montemarano, 2015; MacRoy-Higgins et al., 2013). The sample included more males than females, consistent with gender ratios for late talking (Zubrick et al., 2007). None of the LT participants met cutoff for high risk of autism on the M-CHAT-R (Robins et al., 2009).

TTs

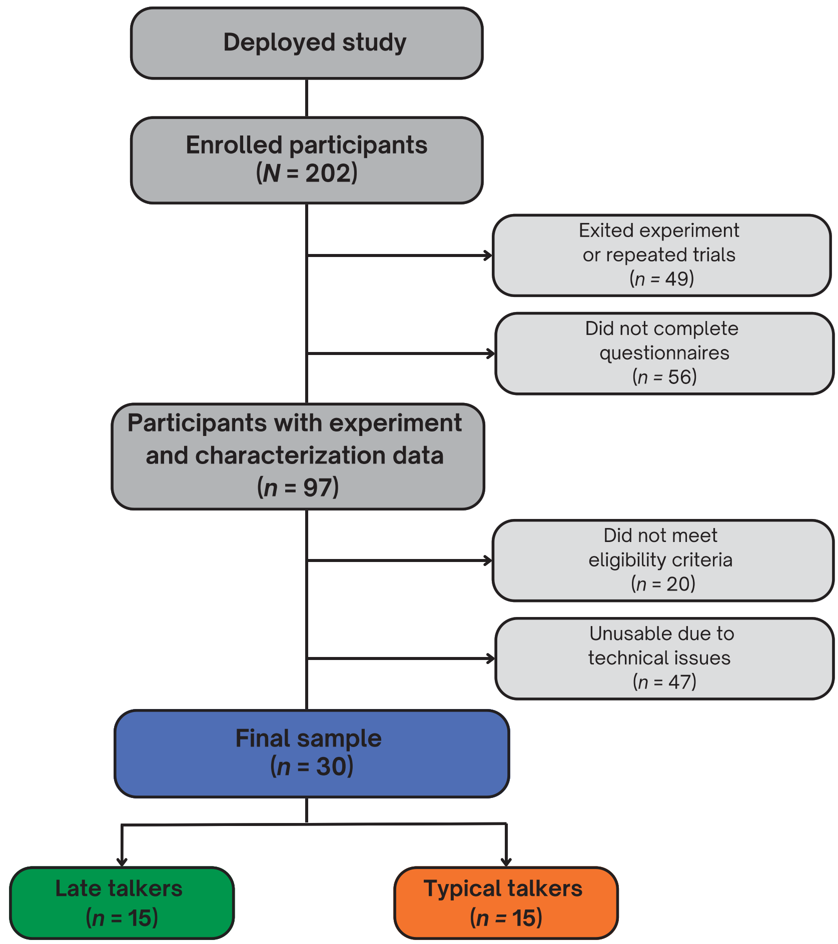
TD toddlers served as a comparison group. The TT group (n = 15) consisted of toddlers, ages 18–35 months with MB-CDI:WS expressive vocabulary ≥ 30th percentile and below cutoff for high risk of autism on the M-CHAT-R (Robins et al., 2009). The 30th percentile was selected to ensure none of our TTs were close to the threshold for late talking (Sosa & Stoel-Gammon, 2012). This group was matched on chronological age, $t(28) = -1.37$, $p = .20$; reported sex, $\chi^2(1) = 1.22$, $p = .27$; race, $\chi^2(1) = 0.24$, $p = .62$; and primary caregiver education, $\chi^2(3) = 1.53$, $p = .61$, to the LT group as measured by either independent t tests (chronological age) or χ^2 analyses (sex, race, caregiver education).

Independent t tests revealed significant differences between the percentile ranks, $t(28) = -13.88$, $p < .001$, on the MB-CDI:WS and spoken vocabulary sizes, $t(28) = -7.47$ $p < .001$, were present between groups, as expected. It is important to note that the sample consisted primarily of highly educated, White families. Implications for this narrow sample are included in the Discussion section. Table 1 provides an overview of participant characteristics.

Stimuli

Materials for the experiment were similar to those of Smith and Yu (2008) and included six novel word–referent pairs. Auditory stimuli consisted of six bisyllabic, phonotactically American English legal novel words (*bosa, gasser, manu, colat, kati, regli*) with a trochaic stress pattern and two carrier phrases (*look at the, point to the*) recorded by a female, native speaker using child-directed

Figure 1. Flowchart of participant enrollment and attrition.



speech. Both the novel words and carrier phrases were controlled for duration and intensity. The mean duration of the spoken novel words was 665 ms (*SD* = 9 ms), and the mean duration for the carrier phrases was 859 ms (*SD* = 37 ms). The visual stimuli included six bold, uniquely colored novel objects that were standardized for size and luminance and placed on a white background (see Figure 2).

Procedure

Study Platform

The study was completed online using the participant's home computer and webcam through Massachusetts Institute of Technology's CHS/Lookit online data collection platform for developmental studies (Scott & Shultz, 2017). To participate in the study, families entered the study webpage and were provided with a brief description of the procedures. Next, they advanced to the consent page where verbal consent was obtained from the parent in the form of video recording. Finally, the parent was guided through a short series of setup instructions to ensure their webcam was working properly, that their toddler was within the video frame, and to test the volume of their computer speakers. Parents were also given the opportunity to preview the experiment using a different set of stimuli, so they had a clear set of expectations of the study procedures. They could pause the experiment by pressing the "spacebar" and they could end the experiment at any time by pressing the "X" on their keyboard.

Cross-Situational Statistical Word Learning Experiment

Participants completed a version of Smith and Yu's (2008) and Yu and Smith's (2013) cross-situational statistical word learning paradigm, which consisted of training on the six novel word–referent pairs and then a set of test trials to measure learning. Participant's gaze was recorded throughout the duration of the experiment. Toddlers were seated on their parent's lap or in a highchair in front of

Table 1. Participant characteristics by the talker group.

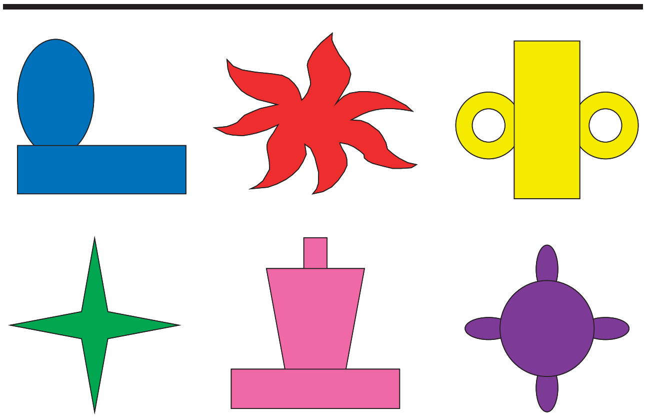
Variable	Group		t or χ^2	p values from t test or χ^2
	Typical talker (n = 15)	Late talker (n = 15)		
Mean chronological age in months (*SD*)	29.00 (5.07)	26.40 (5.33)	−1.37	.20
Percent male	53%	67%	1.22	.27
Race			0.24	.62
White	13	12		
More than one race	2	3		
Mean MB-CDI:WS percentile (*SD*)	71.20 (19.03)	9.47 (4.52)	−13.88	< .001
Mean no. of words produced on MB-CDI:WS (*SD*)	558 (204)	127 (90)	−7.47	< .001
Primary caregiver education			1.53	.61
High school	0	1		
Some college	1	2		
College degree	5	3		
Professional degree	9	9		

Note. Numbers in parentheses are standard deviations. MB-CDI:WS = MacArthur–Bates Communicative Development Inventories, Words and Sentences (Fenson et al., 2007).

their computer. Parents were instructed to close their eyes for the duration of the experiment as to not influence the toddler's looking. They were also instructed not to speak or point to the screen during the experiment.

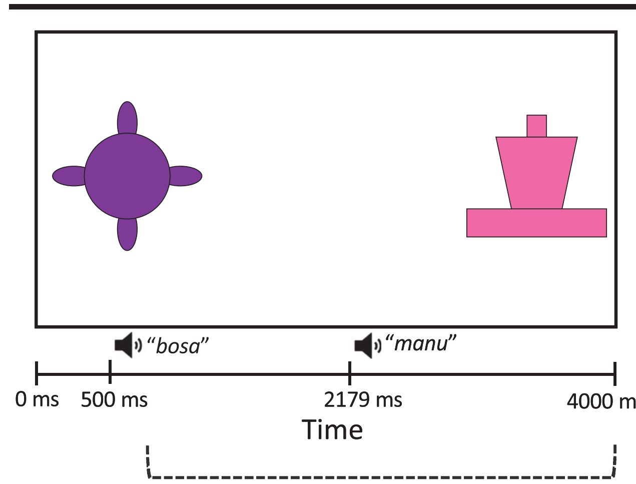
Training. For training trials, participants were simultaneously presented with two novel referents on opposite sides of the computer screen, either left or right, for a total of 4,000 ms. Each training trial started with a 500-ms preview window of the two referents. Next, two novel words were spoken with a 1,000-ms pause between the presentations of the two words. See Figure 3 for trial structure. Within each trial, word–referent mappings were ambiguous, but across trials when a recent word was presented with a new object, that word–referent pair could be eliminated as a match. Thus, over training, the accumulation of co-occurrence statistics should provide sufficient evidence as to which word is mapped to which referent. A total of 30 training trials with each of the six word–referent pairs were presented 10 times. To help maintain the toddlers' interest during the task, the first two trials were preceded by a 2,000-ms attention grabber (e.g., giggling Elmo that loomed on the screen), and eight additional 2,000-ms attention grabbers were randomly presented between the remaining trials.

Figure 2. Six novel objects used in the cross-situational learning experiment.



Testing. Immediately following training, participants completed 12 test trials each lasting 8,000 ms. For each test trial, participants were presented with two of the same six referents from training, but only one word was spoken (target word). After a 500-ms preview window of the two referents, participants heard either the carrier phrase

Figure 3. Example of stimuli presentation for the training phase.



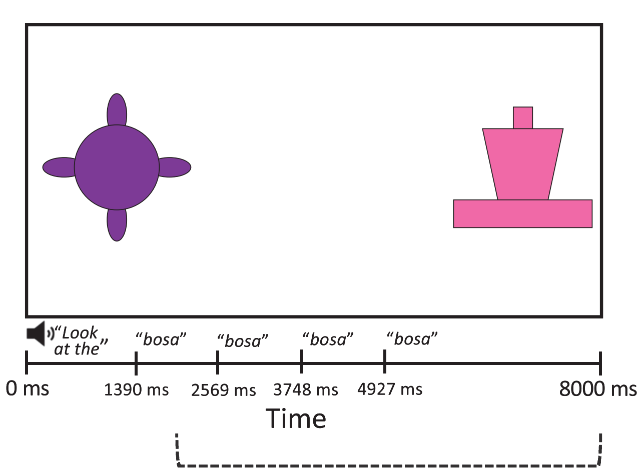
“Look at the” or “Point to the,” followed by a 500-ms pause, and then the target word was spoken four times with a 500-ms pause between target word presentations. Thus, for a test trial, participants would see two referents, the target referent and the distractor, on opposites sides of the screen and hear “Look at the bosa, bosa, bosa, bosa.” Each word was tested two times. Participants were presented with a 2,000-ms attention grabber between each test trial to maintain their interest. See Figure 4 for test trial structure. The training lasted approximately 3 min, and the test phase lasted approximately 2 min, with the entire experiment taking about 5 min to complete.

Design. To minimize potential bias, toddlers were randomly assigned to one of two unique lists of training trials. Each list had a different mapping of the words and the referents, unique combinations of word–referent pairs, and a unique presentation order (Smith & Yu, 2008). Left- and right-side presentations of referents were pseudorandomized with the constraint that half the target words presented first were shown on the left side of the screen and half of the target words were presented second and on the right side of the screen. There were no consecutive presentations of word–referents pairs. Participants also received one of two unique lists of test trials with randomized target–distractor pairs. Half of the targets appeared on the left side of the screen, and half the targets appeared on the right side of the screen. The distractor for each test trial was randomly selected and appeared twice as the distractor across testing trials. Each target word was presented with each carrier phrase once.

Gaze Coding

To evaluate looking behavior during training and test trials, participant’s gaze was coded by a research assistant, blind to participant group assignment and trial information, frame-by-frame for all trials, including training and test phases, using ELAN software (Version, 6.4; Sloetjes & Wittenburg, 2008). Looks were categorized based on direction of gaze to objects on the screen, either left or right. Ambiguous looks or looks away from the screen (e.g., looking at parent) were not coded. A random 20% of the sample (n = 6) was recoded by a second coder with agreement on direction of gaze (left or right) for frames 95.7%.

Figure 4. Example of stimuli presentation for the test phase.



Data Analysis Plan

A window of analysis (WoA) was identified for training and test trials. The start time of the WoA began 367 ms post–first word onset as this is the approximate time required for infants and toddlers to launch a language-modulated eye movement response (Bergelson & Aslin, 2017; Orena et al., 2022; Swingley, 2012). The end of the trial was selected for the end of the WoA. LTs are generally slower at lexical processing during similar preferential looking tasks (Curtis et al., 2023; Ellis et al., 2015; Fernald & Marchman, 2012); thus, we wanted to ensure the LTs had ample time to process the spoken words.

For training, the WoA began at 367 ms post–first word onset and ended at the end of the trial (WoA for training trials: 867–4,000 ms). For test trials, the WoA started post–target word onset through the end of the trial. Since test trials included a carrier phrase and a 500-ms pause prior to the first presentation of the target word, the WoA was adjusted so that it started after the carrier phrase +500 ms pause plus 367 ms post–target word onset (*Look* trials WoA: 1,757–8,000 ms; *Point* trials: WoA: 1,693–8,000 ms). WoAs are visualized by dashed lines in Figures 3 and 4.

The proportion of time spent looking at the referents during training and test phases was calculated for each WoA. For the training trials, the proportion of time spent looking at the referents was operationalized as time spent looking at each referent post–first word onset divided by the trial duration. For test trials, the proportion of time spent looking at the target and distractor referents was calculated using the following formula: time spent looking at either the target or distractor divided by time spent looking at both the target and distractor combined. Trials in which the participants did not look at both referents were eliminated from analyses. Independent t tests revealed no significant differences between groups on the number of trials included in analysis for either the training, $t(28) = 0.81$, $p = .40$, or the test, $t(28) = -0.13$, $p = .95$, phases.

Linear mixed-effects models were used as our primary set of analyses to compare the proportion of time looking at referents during training and test trials between the LT and TT groups. For the training phase, the dependent variable was "proportion of looking time" and the independent variable was "group." For the test phase, the dependent variable was "proportion of looking time" and the independent variables were "looking location" (target vs. distractor) and "group," along with the interaction between "looking location" and "group." Participant was used as the random effect (intercepts) for both models. A maximally complex random effects structure (Barr et al., 2013), including participant slopes and word type (intercepts and slopes), could not be achieved as models failed to converge.

Given our a priori exploratory prediction that the LTs would map fewer word–referent pairs than their TT peers, paired *t* tests were used to explore differences in looking at the target and distractor for each group by word–referent pair. Critically, a word was considered *mapped* if the toddler spent significantly more time looking at the target image when labeled (compared to the distractor) during the test phase. A Shapiro–Wilk normality test was conducted prior to running the paired *t* tests to ensure data were normally distributed given the small sample size. Holm correction was used for multiple comparisons.

Analyses were conducted in R (Version 1.1.463; R Core Team, 2020) using the *lme4* package (Version 1.1-21; Bates et al., 2015) for multilevel modeling. The *lmerTest* package (Version 3.10; Kuznetsova et al., 2017) was employed to evaluate main effects, and the *emmeans* package (Version 1.4.6; Lenth et al., 2020) was used to explore planned comparisons. The *rstatix* package (Version 0.7.0; Kassambara, 2021) was used to complete the paired *t* tests.

Results

For test trials, there was no main effect of group (b = .05, SE = .03, t = 1.55, p = .12) or looking location (b = −.03, SE = .03, t = −0.41, p = .68), but there was a significant interaction between group and looking location (b = −.10, SE = .04, t = −2.19, p = .03). Holm-adjusted post hoc comparisons revealed that the TT group spent significantly more time looking at the target objects when labeled, compared to the distractor objects, suggesting that the TTs had made a connection between the label and referent for at least some of the pairs (b = .11, SE = .03, t = 3.58, p = .002). This effect was not present for the LT group. Rather, they spent a similar proportion of time looking at the target and distractor objects when presented with the target object label (b = .01, SE = .03, t = 0.41, p = .68; see Figure 5).

Next, we explored whether the LTs showed evidence of mapping any of the word–referent pairs. Visual inspection of the data suggested that both the LT and TT groups did map a subset of the words as measured by a greater proportion of the time looking at the target image compared to the distractor (see Figure 6). Paired *t* tests using Holm correction for multiple comparisons were conducted by group and word to evaluate specific word–referent pairs acquired for each group. After correction for multiple comparisons, the LTs showed a trend for a

Figure 5. Proportion of time looking at target and distractor objects by group during the test trials.

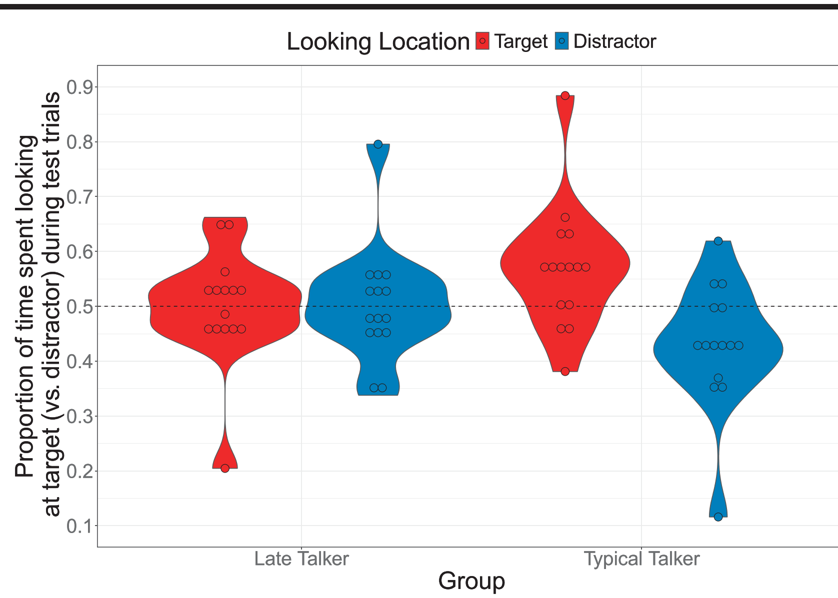
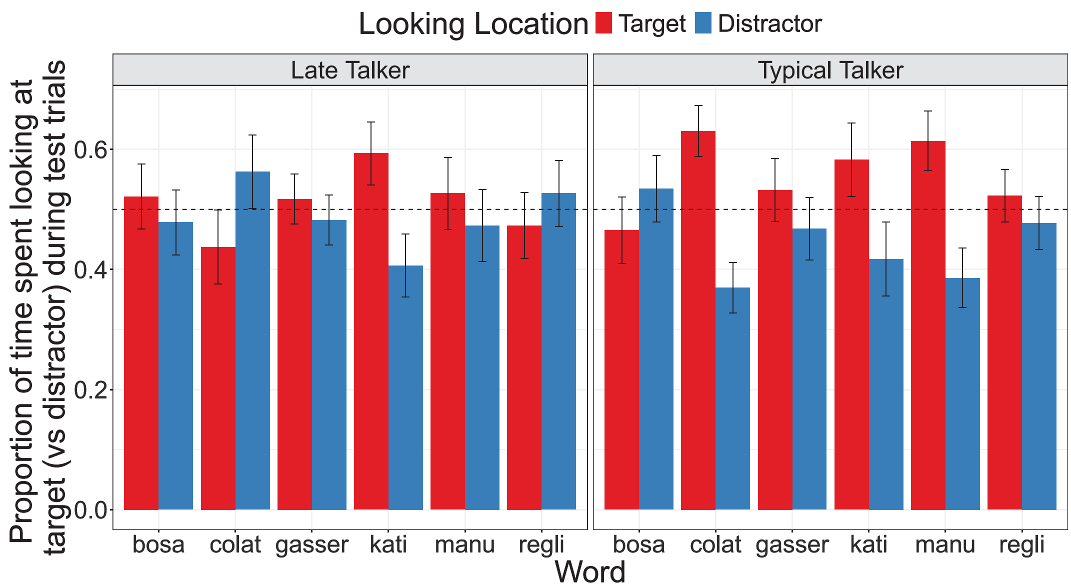


Figure 6. Proportion of time looking at target and distractor objects by group and word during the test trials.



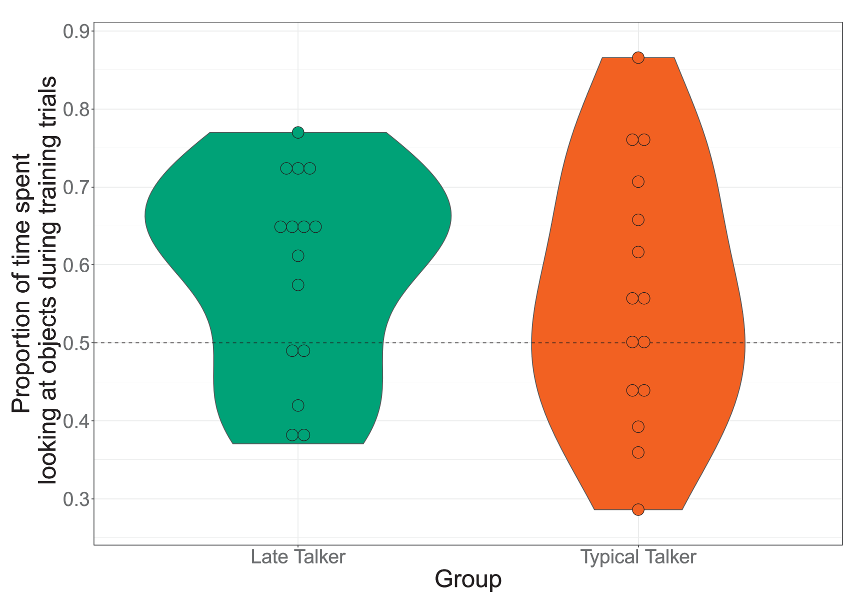
preference for one out of the six word–referent pairs, *kati,* which was marginally significant (p = .07) with a moderate effect size (d = 0.60). The TT group showed a preference for one out of the six pairs including *colat* (p = .008, d = 1.07, large effect size) and a trend toward significance with a moderate effect size for two more (*kati* [p = .06, d = 0.73, moderate effect size] and *manu* [p = .07, d = 0.60, moderate effect size]).

Lastly, during training, there was no main effect of group, indicating that both the LT and TT groups spent similar proportions of time looking at the objects (b = −.01, SE = .04, t = −0.28, p = .78; see Figure 7).

Discussion

The purpose of this article was to compare TD and late-talking toddlers' application of statistical learning to support word–object mappings and acquire a small set of novel words in a cross-situational word learning paradigm.

Figure 7. Proportion of time looking at novel objects by group during training trials.



Given LTs general difficulty learning word–object mappings during fast-mapping opportunities, coupled with the known difficulties older children with DLD show in cross-situational statistical word learning, it was predicted that the LT group would perform worse than their peers as evidenced by more limited linking of the novel word–object connections.

Indeed, the TD toddler group did appear to map at least some of the word–referent pairs as evidenced by spending significantly more time looking at the labeled objects during test compared to the distractor objects. Our small group of LTs did not show this pattern. Rather, they spent a similar proportion of time looking at both the target and distractor items during the test phase. At the participant level, learners and nonlearners emerged in both groups reflecting heterogeneity across the sample. It is possible the individual differences observed in our language delayed group may help predict which LTs achieve typical language functioning and which may have more persistent language problems. A closer inspection of the individual word–referent pairs suggest that the LTs mapped only a single pair, while the TTs linked about half the words to their intended referents (three out of six word–referent pairs). However, looking time differences for two out of these three pairs were only marginally significant, perhaps due the small sample size. Application of Bayesian statistics may be useful if null effects are found in future studies.

A known constraint on statistical learning is attention (see Aslin, 2017, for a review). It is possible that our LTs may not have learned the words due to reduced visual attention to the objects when labeled. Eye movements relative to duration of fixations and number of gaze shifts have been shown to differentiate between strong and weak statistical word learning (Yu & Smith, 2011). While our protocol does not allow this level of granularity relative to eye movements, we did find that groups did not differ on the proportion of time spent looking at all objects combined during the training phase. This suggests that, on a macrolevel, the LTs are allocating similar amounts of visual attention to the objects during training. Our findings are in contrast to MacRoy-Higgins and Montemaro (2015), who showed, during naturalistic word training, that LTs spent less time looking at novel objects when labeled compared to peers. Our task was computer-based with fewer distractor items, whereas the naturalistic task was play-based with a great number of foils. It is possible that the use of a computer-based task plus fewer on-screen distractors helped organize the LTs' visual attention during the experiment, although this is only speculation at this point and would have to be empirically tested.

Even though LTs and TTs demonstrated similar amounts of visual attention to objects during training, LTs did not acquire as many of the novel word–referent mappings. This suggests, of course, that visual attention alone is not responsible for learning. Auditory attention could not be measured during this experiment. Indeed, weaker sustained auditory attention has been implicated in preschool children with language disorder (Spaulding et al., 2008) yet intact in older, elementary school–age students with DLD (Victorino & Schwartz, 2015); thus, it is difficult to know if auditory attention may be driving our effects. To our knowledge, auditory attention has not been directly evaluated in LTs. It is possible that although the LTs were visually attending to the images on the screen, they were not attending in the same way as their TT peers to the auditory input. The *Goldilocks effect* has been described to drive auditory (Kidd et al., 2014) and visual attention (Kidd et al., 2012) in toddlers, such that they seek information that is within their zone of proximal development; that is, input that has the appropriate amount of both novelty and predictability. It is possible that while our visual stimuli were adequate to maintain our LTs' attention (*Goldilocks effect*), the processing of our auditory stimuli was more challenging, thus attention to the words was not as robust and negatively impacted learning. Again, this would need to be unpacked further by empirical testing.

There is emerging evidence to suggest that LTs may rely heavily on phonological features of words, including neighborhood density (ND), when adding words to their spoken lexicons (Simmons & Paul, in press, 2024), whereas TTs are less reliant on this supportive cue. It is posited that when a new word form is encountered, its phonological neighbors already stored in memory become activated. These previously consolidated word forms then act as templates to support the encoding and storage of the new word forms. Thus, forms from dense phonological neighborhoods facilitate word learning, whereas words from sparse neighborhoods are more challenging to learn and store. While we did not manipulate the phonological composition of the novel words used in this study, a post hoc analysis of their phonological features revealed potentially interesting patterns. We calculated the phonological ND for each of the novel spoken words from this study, using procedures outlined in Simmons and Paul (2024). The ND density metric reflects the number of phonological neighbors, with a larger number reflecting more neighbors and a lower number reflecting fewer neighbors. Both groups appeared to benefit from word forms that arise from phonologically dense neighborhoods. That is, both the LTs and TTs linked the word–referent pair with the highest ND (*kati;* ND = 19). Yet, the TTs, also appeared to map the word–referent pair with the lowest ND (*colat*; ND = 1), although this effect was marginally significant. The other word form, *manu*, again only marginally significant for the TTs, did not follow this ND pattern (ND = 8). It is possible that our LTs were able to acquire the *kati* word–referent association due to the scaffolding provided by its dense phonological neighborhood. Careful manipulation relative to the phonological composition of the words presented in these paradigms could provide empirical evidence of this hypothesis in future studies.

Clinical Implications

There are some language-based interventions that have successfully used principles derived from statistical learning accounts to increase spoken vocabulary size in LTs and preschoolers with language disorder (Alt et al., 2014, 2020; Fey et al., 2016; Plante et al., 2014). Alt et al.'s (2014) Vocabulary Acquisition and Usage for Late Talkers (VAULT) program, for example, provides significantly higher than typical levels of input dosage (e.g., 45–90 opportunities per word/30-min sessions) for a small set of target words across many training sessions spread over the course of months. It provides variability of both linguistic and contextual input, which yields opportunities for the toddler to track co-occurrences of the target word and referent across varying linguistic frames (e.g., target used in different sentential contexts) and contextual environments (e.g., different referent exemplars are used for each target word). These intervention studies, along with studies of fast mapping (Asadi et al., 2019; Ellis-Weismer et al., 2013; MacRoy-Higgins & Dalton, 2015; Rujas et al., 2019) and our previous work evaluating the role of word frequency on vocabulary acquisition in a large database analysis of LTs (Simmons & Paul, 2024) converge on the interpretation that LTs need higher doses of input to establish word–referent pairings. These higher doses can be achieved through a variety of implicit training approaches such as milieu teaching (Fey et al., 2016), focused language stimulation (Ellis-Weismer et al., 2017), enhanced conversational recasting (Plante et al., 2014), and VAULT (Alt et al., 2014), as well as through explicit approaches such as elicited imitation (Eisenberg, 2014).

Limitations

Our study has some limitations that are important to acknowledge. The sample size was small and likely underpowered, as suggested by our marginally significant findings of the by-word analyses. Additionally, like much of the work in child language research, participants were primarily White and from well-educated families. Replication of this work with a larger, more diverse sample is critical to our understanding of language learning in LTs. While use of an online data collection platform is novel and affords enhanced opportunities to collect data, it introduces variability that cannot easily be controlled. For example, we do not know if all parents kept their eyes closed during the procedures as instructed, as many were out of frame as participants were seated in a highchair as instructed. We cannot be certain if noncaregivers were present during testing as participants sat close to the camera and occluded much of the background. We also cannot rule out other developmental conditions that may be present in our LT group. While parents reported no other developmental concerns; because we did not complete a full developmental assessment, it is possible that some of our LTs may have autism or global developmental delays, although we attempted to mitigate this issue by use of the autism screener.

While we enrolled a large pool of participants (N = 202), our final sample was small largely due to technological and coding difficulties associated with online, home-based studies, as highlighted in Figure 1. For example, some of the participants ended the experiment before completion or paused the experiment, which resulted in the code repeating trials. These participants had to be excluded. Internet connectivity was also problematic, particularly with slow and/or unstable connections in participants' homes, resulting in unusable data. Finally, the ambient lighting was critical for gaze coding, and if the room was too bright or too dark, coding could not be accurately completed. In addition, some parents did not submit the questionnaires after finishing the experiment, so those participants were excluded. These unforeseen challenges left us with a limited set of usable data. Interestingly, gaze coding in preferential looking paradigms, such as those used in this study, has been shown to be less sensitive to data loss compared to automated, precision eye tracking and generally captures similar types of information (Venker et al., 2020). However, an extension of this study in a well-controlled laboratory environment with an eye tracker may yield additional insights into cross-situational statistical word learning in LTs that our home-based, gaze-coding procedures failed to capture, for example, in the fine-grained time course of processing of the novel words.

Another limitation, common to other studies of statistical learning reviewed here, is that although significant preferences for some intended word–referent pairings were found, these preferences were not absolute, as might be expected if the word–referent mappings were fully learned, and this was true for both LTs and TTs. Even in the cases where toddlers matched words and their referents correctly, their proportion of looking at the target was around .55 (see Figure 5). While these findings reflect some emerging connections being formed between the referents and their labels, the pairings do not yet appear to be stable and robust. For this reason, we have not described our results as "learning" but instead have followed the model used by Kapa and Mettler (2024) in referring to them in more tentative terms (e.g., word–referent mapping, word–referent linking). Further research is needed to determine in both LTs and TTs how and when the boundary between fast mapping or incipient associations and stable, consolidated learning is crossed.

This observation leads us to recognize another limitation of the study. We were not able to investigate more granular questions regarding the dosage necessary for learning word–referent pairings at any level by LTs. Empirical studies are needed to determine optimal dosage of input and therapy intensity. While, in general, it appears that more exposure is needed for language learning by LTs than by typical toddlers, the dosage needed to obtain positive outcomes at maximum efficiency, without providing more than is strictly necessary, is not yet known. Studies that provide varying amounts of cross-situational learning exposures in blocks until the LTs can relate words and referents at high levels of accuracy would help to identify the optimal level of input for this group. Such studies might be, for example, similar to the design used by McGregor et al. (2022) in their investigation of 7-year-olds with DLD. This research would be extremely valuable in planning intervention for this population.

Conclusions

The aim of this study was to compare cross-situational statistical word learning in late-talking and typical-talking toddlers. Our results show that, even in this small-group comparison, LTs were less able to link words with referents, given cross-situational statistical opportunities, with results similar to those reported in older children with DLD. These findings highlight the need for further exploration of word learning in LTs, including both mechanisms that underpin word learning, such as auditory attention, as well as the dosage levels required by LTs to make stable word–referent associations for vocabulary development.

Data Availability Statement

The data used for analyses in this manuscript are not publicly available due to institutional review board restrictions. Requests for the data set can be sent via e-mail to Elizabeth Schoen Simmons at simmonse3@sacredheart.edu.

Acknowledgments

This article was written using support from the National Institute on Deafness and Other Communication Disorders under Award R15DC020290 (PI: Elizabeth Schoen Simmons) and from Sacred Heart University's Research and Creative Grants (PI: Elizabeth Schoen Simmons). The authors would like to thank Children Helping Science, the families that participated in this project, and the research assistants who contributed to acquisition and/or coding of the data including Nicole Khrimian, Julia Linhares, Olivia Plaia, and Allison Roberge. The authors would also like to acknowledge Richard Aslin, PhD (Yale University) for his feedback on various iterations of this project.

References

Ahufinger, N., Guerra, E., Ferinu, L., Andreu Barrachina, L., & Sanz Torrent, M. (2021). Cross-situational statistical learning in children with developmental language disorder. *Language, Cognition and Neuroscience, 36*(9), 1180–1200. https://doi.org/10.1080/23273798.2021.1922723

Alt, M., Mettler, H. M., Erikson, J. A., Figueroa, C. R., Etters-Thomas, S. E., Arizmendi, G. D., & Oglivie, T. (2020). Exploring input parameters in an expressive vocabulary treatment with late talkers. *Journal of Speech, Language, and Hearing Research, 63*(1), 216–233. https://doi.org/10.1044/2019_JSLHR-19-00219

Alt, M., Meyers, C., Ogilvie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically based word learning intervention for late-talking toddlers. *Journal of Communication Disorders, 52,* 207–220. https://doi.org/10.1016/j.jcomdis.2014.07.002

Asadi, M., Zarifian, T., Kazemi, M. D., & Ghaedamini Harouni, G. (2019). Fast-mapping in Persian-speaking late-talking toddlers. *Early Child Development and Care, 191*(5), 726–740. https://doi.org/10.1080/03004430.2019.1645136

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Cognitive Science, 8*(1–2). https://doi.org/10.1002/wcs.1373

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word–object relations. *Child Development, 67*(6), 3135–3153. https://doi.org/10.1111/j.1467-8624.1996.tb01906.x

Bannard, C., & Tomasello, M. (2012). Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds? *PLOS ONE, 7*(11), Article e49881. https://doi.org/10.1371/journal.pone.0049881

Barr, D., Levy, R., Scheeprers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Benitez, V. L., & Li, Y. (2023). Cross-situational word learning in children and adults: The case of lexical overlap. *Language Learning and Development, 20*(3), 195–218. https://doi.org/10.1080/15475441.2023.2256713

Bergelson, E., & Aslin, R. (2017). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development, 13*(4), 481–501. https://doi.org/10.1080/15475441.2017.1324308

Briganti, A. M., & Cohen, L. B. (2011). Examining the role of social cues in early word learning. *Infant Behavior and Development, 34*(1), 211–214. https://doi.org/10.1016/j.infbeh.2010.12.012

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports in Child Language Development, 15,* 17–29.

Cheung, R. W., Hartley, C., & Monaghan, P. (2021). Caregivers use gesture contingently to support word learning. *Developmental Science, 24*(4), Article e13098. https://doi.org/10.1111/desc.13098

Cheung, R. W., Hartley, C., & Monaghan, P. (2022). Multiple mechanisms of word learning in late-talking children: A longitudinal study. *Journal of Speech, Language, and Hearing Research, 65*(8), 2978–2995. https://doi.org/10.1044/2022_JSLHR-21-00610

Chilosi, A. M., Pfanner, L., Pecini, C., Salvadorini, R., Casalini, C., Brizzolara, D., & Cipriani, P. (2019). Which linguistic measures distinguish transient from persistent language problems in late talkers from 2 to 4 years? A study on Italian speaking children. *Research in Developmental Disabilities, 89,* 59–68. https://doi.org/10.1016/j.ridd.2019.03.005

Collisson, B. A., Graham, S. A., Preston, J. L., Rose, M. S., McDonald, S., & Tough, S. (2016). Risk and protective factors for late talking: An epidemiologic investigation. *Journal of Pediatrics*, 172, 168–174e1. https://doi.org/10.1016/j.jpeds.2016.02.020

Curtis, P. R., Estabrook, R., Roberts, M. Y., & Weisleder, A. (2023). Specificity of phonological representations in U.S. English-speaking late talkers and typical talkers. *Infancy, 28*(4), 771–792. https://doi.org/10.1111/infa.12536

Eisenberg, S. (2014). What works in therapy: Further thoughts on improving clinical practice for children with language disorders. *Language, Speech, and Hearing Services in Schools, 45*(2), 117–126. https://doi.org/10.1044/2014_LSHSS-14-0021

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers? *Journal of Communication Disorders, 58,* 143–157. https://doi.org/10.1016/j.jcomdis.2015.06.011

Ellis-Weismer, S., Venker, C., Evans, J. L., & Moyle, M. J. (2013). Fast mapping in late-talking toddlers. *Applied Psycholinguistics, 34*(1), 69–89. https://doi.org/10.1017/S0142716411000610

Ellis-Weismer, S., Venker, C, & Robertson, S. (2017). Focused stimulation approach to language intervention. In R. McCauley, M. Fey, & R. Gillam (Eds.), *Treatment of language disorders in children* (2nd ed., pp. 175–202). Brookes.

Escudero, P., Mulak, K., & Vlach, H. (2016). Infants encode phonetic detail during cross-situational word learning. *Frontiers in Psychology, 7,* Article 1419. https://doi.org/10.3389/fpsyg.2016.01419

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*(2), 321–335. https://doi.org/10.1044/1092-4388(2011/10-0298)

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur–Bates Communicative Development Inventories*. Brookes.

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

Fey, M., Warren, S., Bredin-Oja, S., & Yoder, P. (2016). Responsivity education/prelinguistic milieu teaching. In R. McCauley, M. Fey, & R. Gillam (Eds.), *Treatment of language disorders in children* (2nd ed., pp. 57–86). Brookes.

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

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science, 13*(1), 252–263. https://doi.org/10.1111/j.1467-7687.2009.00871.x

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry, 58*(11), 1251–1263. https://doi.org/10.1111/jcpp.12734

Hartley, C., Bird, L. A., & Monaghan, P. (2020). Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition, 200,* Article 104265. https://doi.org/10.1016/j.cognition.2020.104265

Heibeck, T., & Markman, E. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58*(4), 1021–1034. https://doi.org/10.2307/1130543

Horvath, S., Rescorla, L., & Arunachalam, S. (2019). The syntactic and semantic features of two-year-olds' verb vocabularies: A comparison of typically developing children and late talkers. *Journal of Child Language, 46*(3), 409–432. https://doi.org/10.1017/S0305000918000508

Kapa, L. L., & Mettler, H. M. (2024). Statistical learning among preschoolers with and without developmental language disorder: Examining effects of language status, age, and prior learning. *Journal of Speech, Language, and Hearing Research, 67*(9), 3081–3093. https://doi.org/10.1044/2024_JSLHR-23-00602

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests* (Version 0.7.0) [R package]. https://CRAN.R-project.org/package=rstatix

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLOS ONE, 7*(5), Article e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development, 85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTestPackage: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.4.8) [R package]. https://CRAN.R-project.org/package=emmeans

Lukács, Á., Lukics, K. S., & Dobó, D. (2021). Online statistical learning in developmental language disorder. *Frontiers in Human Neuroscience, 15,* Article 715818. https://doi.org/10.3389/fnhum.2021.715818

Macroy-Higgins, M., & Dalton, K. P. (2015). The influence of phonotactic probability on nonword repetition and fast mapping in 3-year-olds with a history of expressive language delay. *Journal of Speech, Language, and Hearing Research, 58*(6), 1773–1779. https://doi.org/10.1044/2015_JSLHR-L-15-0079

MacRoy-Higgins, M., & Montemarano, E. (2015). Attention and word learning in toddlers who are late talkers. *Journal of Child Language, 43*(5), 1020–1037. https://doi.org/10.1017/S0305000915000379

MacRoy-Higgins, M., Schwartz, R. G., Shafer, V. L., & Marton, K. (2013). Influence of phonotactic probability/neighbourhood density on lexical learning in late talkers. *International Journal of Language & Communication Disorders, 48*(2), 188–199. https://doi.org/10.1111/j.1460-6984.2012.00198.x

Matte-Landry, A., Boivin, M., Tanguay-Garneau, L., Mimeau, C., Brendgen, M., Vitaro, F., Tremblay, R., & Dionne, G. (2020). Children with persistent versus transient early language delay: Language, academic, and psychosocial outcomes in elementary school. *Journal of Speech, Language, and Hearing Research, 63*(11), 3760–3774. https://doi.org/10.1044/2020_JSLHR-20-00230

McGregor, K. K., Smolak, E., Jones, M., Oleson, J., Eden, N., Arbisi-Kelm, T., & Pomper, R. (2022). What children with developmental language disorder teach us about cross-situational word learning. *Cognitive Science, 46*(2), Article e13094. https://doi.org/10.1111/cogs.13094

Medina, T., Snedeker, J., Trueswell, J., & Gleitman, L. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America, 108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

Monaghan,P., & Mattock, K. (2012). Integrating constraints for learning word–referent mappings. *Cognition, 123*(1), 133–143. https://doi.org/10.1016/j.cognition.2011.12.010

Orena, A. J., Mader, A. S., & Werker, J. F. (2022). Learning to recognize unfamiliar voices: An online study with 12- and 24-month-olds. *Frontiers in Psychology, 13,* Article 874411. https://doi.org/10.3389/fpsyg.2022.874411

Paul, R. (1991). Profiles of toddlers with slow expressive language development. *Topics in Language Disorders, 11*(4), 1–13. https://doi.org/10.1097/00011363-199111040-00003

Paul, R., & Roth, F. (2011). Characterizing and predicting outcomes of communication delays in infants and toddlers: Implications for clinical practice. *Language, Speech, and Hearing Services in Schools, 42*(3), 331–340. https://doi.org/10.1044/0161-1461(2010/09-0067)

Perry, L. & Kucker, S. (2019). The heterogeneity of word learning biases in late-talking children. *Journal of Speech, Language, and Hearing Research, 62*(3), 554–563. https://doi.org/10.1044/2019_JSLHR-L-ASTM-18-0234

Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., Lieser, A. M., & Burton, B. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology, 23*(4), 530–545. https://doi.org/10.1044/2014_AJSLP-13-0038

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rescorla, L. (1989). The language development survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*(4), 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*(1), 16–30. https://doi.org/10.1044/1092-4388(2008/07-0171)

Rescorla, L., & Turner, H. L. (2015). Morphology and syntax in late talkers at age 5. *Journal of Speech, Language, and Hearing Research, 58*(2), 434–444. https://doi.org/10.1044/2015_JSLHR-L-14-0042

Robins, D. L., Fein, D., & Barton, M. (2009). *Modified Checklist for Autism in Toddlers–Revised, With Follow-Up (M-CHAT-R/F)*. https://www.mchatscreen.com/

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language, 87,* 105–127. https://doi.org/10.1016/j.jml.2015.09.005

Roembke, T. C., Simonetti, M. E., Koch, I., & Philipp, A. M. (2023). What have we learned from 15 years of research on cross-situational word learning? A focused review. *Frontiers in Psychology, 14,* Article 1175272. https://doi.org/10.3389/fpsyg.2023.1175272

Rujas, I., Casla, M., Mariscal, S., Lázaro López-Villaseñor, M., & Murillo Sanz, E. (2019). Effects of grammatical category and morphology on fast mapping in typically developing and late talking toddlers. *First Language, 39*(3), 249–280. https://doi.org/10.1177/0142723719828258

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Scott, K., & Schultz, 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

Simmons, E. S., & Paul, R. (in press). The role of word form in lexical selection of late talkers. *Journal of Speech, Language, and Hearing Research*.

Simmons, E. S., & Paul, R. (2024). Are late talkers just late?: Neighborhood density and word frequency properties of late talkers' spoken vocabularies. *Journal of Speech, Language, and Hearing Research, 67*(10), 3794–3802. https://doi.org/10.1044/2024_JSLHR-23-00769

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

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category—ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*. http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf [PDF]

Smith, L., & Yu, C. (2008). Infants rapidly learn word–referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Smith, L., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word–referent learning in infants. *Language Learning and Development, 9*(1), 25–49. https://doi.org/10.1080/15475441.2012.707104

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

Spaulding, T. J., Plante, E., & Vance, R. (2008). Sustained selective attention skills of preschool children with specific language impairment: Evidence for separate attentional capacities. *Journal of Speech, Language, and Hearing Research, 51*(1), 16–34. https://doi.org/10.1044/1092-4388(2008/002)

Stokes, S. F., & Klee, T. (2009). Factors that influence vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry, 50*(4), 498–505. https://doi.org/10.1111/j.1469-7610.2008.01991.x

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology, 126,* 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Swingley, D. (2012). The looking-while-listening procedure. In E. Hoff (Ed.), *Research methods in child language: A practical guide* (pp. 29–42). Blackwell. https://doi.org/10.1002/9781444344035.ch3

Tenenbaum, E., Amso, D., Abar, B., & Sheinkopf, S. (2014). Attention and word learning in autistic, language delayed and typically developing children. *Frontiers in Psychology, 5,* Article 490. https://doi.org/10.3389/fpsyg.2014.00490

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

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

Victorino, K. R., & Schwartz, R. G. (2015). Control of auditory attention in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 58*(4), 1245–1257. https://doi.org/10.1044/2015_JSLHR-L-14-0181

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14*(2), 165–180. https://doi.org/10.1111/j.1467-7687.2010.00958.x

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition, 145,* 53–62. https://doi.org/10.1016/j.cognition.2015.07.013

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50*(6), 1562–1592. https://doi.org/10.1044/1092-4388(2007/106)