

ARTICLE

The timescales of word learning in children with language delays: In-the-moment mapping, retention, and generalization

Sarah C. KUCKER^{1*} and Erin SEIDLER²

¹Oklahoma State University, USA

²University of Wisconsin Oshkosh, USA

*Corresponding author. Sarah C. Kucker, Oklahoma State University, 116 Psychology Building, Stillwater, OK 74078. Email: sarah.kucker@okstate.edu

(Received 26 August 2020; revised 17 July 2021; accepted 18 November 2021)

Abstract

Learning new words and, subsequently, a lexicon, is a time-extended process requiring encoding of word-referent pairs, retention of that information, and generalization to other exemplars of the category. Some children, however, fail in one or more of these processes resulting in language delays. The present study examines the abilities of children who vary in vocabulary size (including both children with normal language (NL) and late talking (LT) children) across multiple timescales/processes – known and novel word mapping, novel word retention, and novel noun generalization. Results indicate that children with lower language skills suffer from deficits in quick in-the-moment mapping of known words compared to their NL peers, but age and vocabulary size rather than normative vocabulary ranking or NL/LT status better predicts performance on retention and generalization processes. Implications for understanding language development as a holistic process with multiple interacting variables are discussed.

Keywords: late talkers; word learning; referent selection; generalization

Introduction

Word learning is a core part of early development, but requires multiple, interactive and related cognitive processes to be successful. To learn just a single new word a child must engage in fast in-the-moment mapping and encoding of a word and its referent, strengthen those initial word-referent associations such that they can be retained over time, and be able to generalize this information to new examples in future moments. Each process builds on each other and has cascading effects on the others – successful mapping lays a foundation for more robust retention, and robust retention strengthens the network that supports generalization (Kucker, McMurray & Samuelson, 2015; McMurray, Horst & Samuelson, 2012). Thus, in order to successfully build their vocabulary, a child needs to integrate each of these processes relatively well. Success across each process (mapping,

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

retention, and generalization) predicts future lexical growth. Conversely, failure at one or more points can lead to delays in development, most notably recognized as being a late talker.

Late-talking (LT) children are those who represent the lowest end of a normative vocabulary distribution, scoring lower in expressive vocabulary for their age (MacRoy-Higgins, Shafer, Fahey & Kaden, 2016). For the purposes of this study we used a slightly more liberal 25th percentile as the cut-off to capture heterogeneity within the group (see also Jones, 2003). Late-talking is not attributed to cognitive or developmental disorders, genetic abnormalities, neuromuscular disabilities, or hearing disorders (Rescorla, 2011). By definition, LT children exhibit slower expressive language growth and have smaller normative vocabularies than their peers with normal language (NL). However, less is known about why LT children may struggle to gain a robust vocabulary and what impact their small vocabularies have across the timescales of word learning.

Timescales of word learning

Initial in-the-moment processes of word learning

Communication starts with real-time, in-the-moment processing. In order for a child to comply when her father says, “go get your cup,” the child must, in the span of mere seconds, parse *cup* from the rest of the speech stream, interpret it as referring to the plastic blue hollow container on the counter, and ignore the silver metal spoon next to it, and the white liquid in the milk carton in the fridge. This in-the-moment behavior draws on a child’s lexicon, requiring recollection of prior knowledge and, critically, an application of such knowledge in real time. By around one year of age, children can do this, correctly attending to a known target when presented with its auditory label (Oviatt, 1980; Woodward, Markman & Fitzsimmons, 1994); a process which becomes efficient between 17- and 24-months (Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998) and is seen both in the home and lab (Garrison, Baudet, Breitfeld, Aberman & Bergelson, 2020).

A similar in-the-moment process supports communication and learning about unknown, novel words as well, such as when her father says, “let’s play with your new *slinky*.” Children as young as 16-months-old can reliably map novel words (*slinky*) to their unknown referents (silver metal coil) by either looking or reaching toward the target (Mervis & Bertrand, 1994). This behavior is thought to be driven by multiple word learning constraints which operate in the span of seconds as the child disambiguates a target from the foil items. For instance, Mutual Exclusivity (Markman & Wachtel, 1988) proposes that children avoid a second label for an already known item, instead attributing any novel words to novel items (the Novel-Name Nameless-Category principle; Mervis & Bertrand, 1994). Children can also use the principle of contrast (Clark, 1990), disjunctive syllogism (Halberda, 2003), or even social pragmatic cues such as joint attention (Tomasello & Akhtar, 1995) to, upon hearing a word, select the correct referent from an array and avoid foils.

Decades of work have demonstrated that this initial, in-the-moment behavior is driven by both lexical and perceptual biases, suggesting that children with poorer lexical or perceptual processing skills may struggle with this phase of word learning. In fact, younger children (e.g., 18-month-olds) and those with lower vocabularies for their age have poorer disambiguation skills, selecting a novel or known target from an array at or near chance levels (Bion, Borovsky & Fernald, 2013; see also Kucker, McMurray & Samuelson, 2018). Similar patterns are seen with LT children. Work measuring visual

attention during novel word mapping (MacRoy-Higgin & Montemenaro, 2015) and millisecond level eye-tracking during referent selection (Ellis, Borovsky, Elman & Evans, 2015) reveals real-time processing differences for LT vs. NL children. However, the process of (quickly) identifying and selecting the referent (traditionally referred to as “fast-mapping”) is not equivalent to learning and does not mean a word has integrated robustly into the lexicon (Horst & Samuelson, 2008). This initial in-the-moment period is followed by a slower process of strengthening word-referent links such that the information can be recalled at a later time (known as “slow mapping”, or here referred to as “Retention”; see Carey, 2010; Horst & Samuelson, 2008; Kucker et al., 2015; Swingley, 2010). Though fast and slow mapping were at the forefront of research on NL children for decades, less work has explored the processes in LT children. In fact, the majority of “fast-mapping” work with LT children does not directly test children’s accuracy at known word selection, nor explicitly measure mapping, mutual exclusivity, disambiguation, or encoding abilities for novel words. Instead, canonical “fast-mapping” work in LT children has examined RETENTION or recollection of a new word-referent pair after an extensive (albeit “fast”) training of the name via ostensive naming or repeated exposures (e.g., Ellis Weismer, Venker, Evans & Moyle, 2013).

This distinction in prior work is important because mapping and retention are related but distinct processes (Kucker et al., 2015): being good at mapping does not guarantee success at retention (Horst & Samuelson, 2008). For example, 18-month-old children typically are very good at selecting a novel target when asked, but also demonstrate a robust novelty bias when asked to select a known item, choosing a novel foil item instead. Despite good novel referent selection, these children largely fail to select known referents when there is a novel distractor and fail to retain novel word-referent pairs after a short break. However, there is also evidence that in-the-moment mapping ability may impact retention. For instance, some 18-month-old children who, for an unknown reason, lack a novelty bias (selecting the known referents correctly in-the-moment) subsequently show above-chance retention (Kucker et al., 2018). Similarly, Bion et al. (2013) found a positive correlation between referent selection accuracy and retention in a looking-based task. There might then be possible downstream effects of mapping behaviors on later retention. Moreover, other work suggests LT children are also subject to the same cascading effects. For instance, real-time processes like attention are correlated with later retention in LT children (MacRoy-Higgins & Montemarano, 2015; see also Tenenbaum et al., 2014). However, we also know that the impact of LT children’s in-the-moment behaviors may be unique from that of their NL peers – LT children who are more efficient at known word recognition at 18-months (look to a target faster) are more likely to catch-up to their NL peers a year later (Fernald & Marchman, 2012). That is, early in-the-moment behavior and mapping of words and referents are distinct from slower learning and retention processes, but have cascading effects on later vocabulary growth. This in-the-moment behavior and mapping of words and referents will hence be referred to as in-the-moment mapping.

Slower associative processes of retention in word learning

In addition to in-the-moment mapping, a child must also retain the new word-referent pair. Retention can be boosted with a richer initial encoding/mapping context that includes repetition or additional cues. With multiple exposures to the word-referent pair, children as young as 16-months retain (Mervis & Bertrand, 1994). Likewise,

24-month-old children retain when given additional familiarization time with the target items (Kucker & Samuelson, 2012). By 30-months, normal language children demonstrate retention with as little as one prior exposure to the word-referent pair (Bion et al., 2013). This suggests that over development, the lexicon strengthens such that it can support both old and new words well. Children's ability to retain newly learned words, especially as they near 24-30-months, then gives insight into their potential for further growth.

However, not all children demonstrate the same ability to retain word-referent pairs. Ellis Weismer et al. (2013) found that *KNOWN* word *COMPREHENSION* (i.e., in-the-moment selection of a known target) did not differentiate LT and NL children, but performance on *NOVEL* word *retention* was significantly lower for LT children. Similarly, after an extended training session, LT two-year-olds retain fewer new words than their NL peers (MacRoy-Higgins & Montemaranò, 2015; MacRoy-Higgins et al., 2013). On the other hand, Ellis et al. (2015) found that in-the-moment looking and fixation behaviors of 18-month-old children more reliably differentiated LT from NL children than their subsequent retention/learning accuracy. That is, LT children may not be behind their peers on all aspects of word learning, but which processes are impacted is less clear. Taken together, this again demonstrates the importance of analyzing the process of word learning across timescales, especially in late talking children.

Generalization and the refinement of learning

Critically, learning a word does not end with retention. Children must also learn that most words do not map to just a single referent in the world, but instead represent a category of items. As a child is exposed to each new example of a referent, not only is their lexical network becoming refined and chances of retention increasing, but they are also building statistics about what a referent can be and which properties are relevant for generalizing the word to other exemplars. In the case of nouns, these relevant properties are often aspects of the item's physical shape – e.g., the majority of cups have a similar round, hollow shape and most things called “ball” are spherical but can vary on color or material. As a child's vocabulary (and the proportion of nouns in particular) increases, so does attention to shape (Samuelson & Smith, 1999). As such, by the time children have 150-250 words in their productive vocabulary (roughly 20-months-old), they have experienced enough regularities in language to reliably generalize labels to other novel items with similar shape, an ability called the shape bias (Landau, Smith & Jones, 1988). This bias helps children learn how to learn, such that the next encounter with a word is more efficient. In this way, the shape bias is a critical mechanism for predicting future vocabulary growth (Perry, Samuelson, Malloy & Schiffer, 2010; Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002).

Some children, however, do not acquire a shape-based vocabulary and do not show a robust shape bias. A handful of studies have found that LT children lag behind their peers in both the structure of nouns in their vocabulary (Colunga & Sims, 2017) and novel word extensions/generalizations (Jones, 2003). When given a novel word for a novel item and asked to find similar items that match in a single feature (e.g., same shape but different color and material, or same material but different color and shape), NL children choose the shape-match, but as a group, LT children were less likely to do so (Jones, 2003; Perry & Kucker, 2019). Furthermore, LT children have difficulty recognizing shape caricatures of known items (Jones & Smith, 2005), suggesting weaker knowledge for overall shape

features. These same deficits in generalization likely persist as children develop – older preschool children diagnosed with Specific Language Impairment also fail to demonstrate a shape bias on the standard novel noun generalization (NNG) task (Collisson, Grela, Spaulding, Rueckl & Magnuson, 2015). Overall, the ability to generalize is a critical skill for advancing one's vocabulary and deficits in this area may help explain why some children fail to demonstrate significant growth over time.

Current Study

While each process of word learning is intimately related to the next, being good at one process does not necessarily mean success at the others. However, most prior studies examine only one process in isolation and few have combined various measures of both known and novel word processing in the same children. Furthermore, recent work with children with autism spectrum disorder suggests that a holistic approach examining multiple timescales and processes can lend itself to a more thorough picture of language delay (Hartley, Bird & Monaghan, 2019, 2020). The first goal of the current study is to do so. First, in-the-moment known word comprehension was examined in two different referent selection contexts. Then, in-the-moment novel word-referent mapping was tested via a referent selection task. Extended learning processes were then examined – novel word retention from two tasks were measured and children's ability to generalize novel words via a Novel Noun Generalization (NNG) task was assessed.

Importantly, tasks were administered through a within-subject design across a wide age range of children who fell along the entire normal bell curve for productive vocabulary. That is, both children who would traditionally be classified as “late talkers” (LT) as well as those defined as having “normal language” (NL) participated. The second goal of the current study is to examine this vast heterogeneity in children's language abilities. To do so, all children were first analyzed as a large group with age and vocabulary percentile as predictors. A second analysis looked at the younger subset of children (those under 30-months) more in-depth and included vocabulary size and vocabulary structure as a predictor for performance on each task. A secondary analysis using traditional dichotomous groups (LT vs. NL) as the independent variable replicated most findings and allows for comparison to prior work. The results also highlight the variability of vocabulary within LT children, reinforcing the utility of using continuous measures (e.g., vocabulary) to classify children with potential delays.

Methods

Participants

A total of 114 children (61 female) between 17- and 40-months ($M = 26;30$, $SD = 6;23$) from an urban metropolitan area participated. See [Table 1](#) for demographic information. All children were reported to be monolingual English speakers with no major developmental delays (e.g., no autism, William's syndrome, Down's syndrome). However, in order to ensure a wide distribution of language abilities, children were allowed to participate if parents had concerns about their child's language or if the child was enrolled in a birth-to-three program provided they had no diagnosed developmental disorders. A vocabulary percentile was calculated for each child by comparing each child's total productive vocabulary size to the national norms (Fenson, Dale, Reznick, Bates, Thal,

Table 1. Demographic information for children in the full sample and NL/LT sub-groups

	Overall Sample	Age-Matched Group		Vocab-Matched Group	
		NL	LT	NL	LT
<i>Age (Mo;Days)</i>	26;30 (6;23) [17;4-40;30]	25;25 (6;16) [17;5-40;14]	25;7 (6;14) [17;4-40;12]	25;12 (7;1) [17;11-35;21]	30;6 (6;21) [19;18-40;12]
<i>Child Vocab*</i>					
<i>MCDI-WS</i>	277.48 (216.30) [3-672]	345.73 (201.87) [33-654]	58.67 (47.65) [3-193]	186.70 (151.21) [33-434]	185.10 (153.35) [30-446]
<i>MCDI-III</i>	62.75 (20.13) [23-96]	87.25 (7.85) [77-96]	50.2 (10.76) [39-63]	56.8 (12.64) [39-68]	53.8 (13.03) [34-64]
<i>Parent Education</i>	5.74 (1.23) [1-8]	5.87 (1.45) [1-7.5]	5.57 (1.16) [3.5-8.0]	5.97 (.399) [5-6.5]	5.33 (1.19) [3-7]
<i>Parent Income</i>	7.86 (1.68) [1-9]	8.12 (1.77) [1-9]	7.52 (1.92) [1-9]	8.57 (.756) [7-9]	7.07 (2.20) [1-9]
<i>Race</i>					
<i>White</i>	<i>n</i> = 95	<i>n</i> = 23	<i>n</i> = 21	<i>n</i> = 12	<i>n</i> = 11
<i>Black</i>	<i>n</i> = 6	<i>n</i> = 1	<i>n</i> = 2	<i>n</i> = 1	<i>n</i> = 2
<i>Mixed/NR</i>	<i>n</i> = 11	<i>n</i> = 2	<i>n</i> = 3	<i>n</i> = 2	<i>n</i> = 2
<i>Hispanic</i>	<i>n</i> = 19	<i>n</i> = 5	<i>n</i> = 4	<i>n</i> = 2	<i>n</i> = 5

Note: Means shown with standard deviation in parentheses and range in brackets. **N* for the overall MCDI-WS sample was 86, MCDI-III was 28. *N* for the Age-matched MCDI-WS in NL and LT were respectively 22 and 21, MCDI-III was 4 and 5. *N* for the Vocab-matched MCDI-WS and MCDI-III was 10 and 5 for both groups. Education was rank ordered with 1 as less than 7th grade and 8 as Doctoral degree; a score of 6 indicated a 4-year college degree. Income was also rank ordered with a 1 as "less than \$10,000" and 9 as "more than \$100,000".

Pethick, Tomasello, Mervis & Stiles, 1994). This number was used to identify LT children (those <25th percentile). The 25th percentile was used to have a broad sample of children at the lower end of the range, similar to other recent studies (Colunga & Sims, 2017; Perry & Kucker, 2019) emphasizing the heterogeneity of children's language skills. An additional 45 children were dropped for not completing all four tasks ($n = 20$), exposure of more than 50% of another language ($n = 21$), or parents not completing the vocabulary form ($n = 4$). The demographic and vocabulary distribution of dropped participants did not differ significantly from those kept.

In order to draw comparisons to prior work using specific groups of LT vs. NL, multiple subsets were created from the full sample. First, an LT group that fell below the 25th percentile for productive vocabulary and a NL group above the 25th percentile were selected and matched on gender and age within 2 weeks. This Age-Matched subset of 52 children (26 NL) were equal on parent education and income. A separate Vocabulary-Matched subset of 30 children (15 LT) were matched on vocabulary (within 12 words on the MCDI-WS and 5 words on the MCDI-III), see Table 1.

Stimuli

Each task consisted of a selection of known and/or novel objects. Known items (Figure 1a) were selected from a pool of items commonly known by the majority of 17-month-old children (Fenson et al., 1994). Novel items were unknown objects that children under 40-months-old are not likely familiar with (see Figures 1b and 1c). Novel Noun Generalization items (Figure 1d) consisted of four sets of novel items; each set consisted of one exemplar object, two items that matched the exemplar in shape (but differed in color and material), and two material-matching items (differing on shape and color). Parents confirmed their child knew the names of the known objects and was unfamiliar with all novel items. Items were replaced as necessary. Novel words were simple words that were phonologically legal in English and had no known referents.

Procedure

The child was seated across a table from the experimenter in a booster seat next to their parents or on their parents' lap. Parents completed an age-appropriate version of the MacArthur-Bates Communicative Development Inventory (MCDI) during the session (MCDI: Words and Sentences [MCDI-WS] for those 30-months and under, and MCDI-III for those over 30-months; Dale & Fenson, 1996; Fenson et al., 1994) and were instructed to avoid interacting with their child, offering minimal encouragement only if needed. Each child completed a series of four separate tasks (Figure 2), taking breaks as needed between tasks. Each of the tasks paralleled the procedures used in prior work and tap into different timescales of word learning. In order to ensure that even the youngest, lowest vocabulary children could complete the tasks, receptive accuracy was the main dependent variable of interest in each task.

Known Word Comprehension. Known word comprehension was designed to test children's in-the-moment mapping of known words by testing children's ability to recognize prototypical items by name in the lab context. These trials also serve a secondary purpose of familiarizing children with the testing procedures. The task mimicked that used in prior comprehension tests (e.g., looking-while-listening) but used real items and a live experimenter for both logistic reasons and to capture a more



Figure 1. Stimuli. Known items used in known word comprehension and referent selection (a), novel items in Referent Selection (b), novel items used in Direct Naming (c), and Novel Noun Generalization items (d).

ecologically valid learning experience that parallels with that used in the referent selection tasks below.

This phase began with the experimenter arranging three name-known items equidistant apart in a horizontal row on a white tray out of sight of the child. Then, while maintaining eye contact with the child, the experimenter placed the tray within sight, but out of reach. After three seconds, the experimenter prompted the child to retrieve one item by name (e.g., “Can you get the shoe?”) and pushed the tray forward. The child was re-prompted up to three times if necessary and then praised and/or corrected as needed. Children’s initial answer (prior to any praise or correction) as well as their final responses (after praise/correction) were noted. Each object was the target only once. Target location and trial order were randomized.

Referent Selection and Retention. Referent selection and retention consisted of two phases tapping into the in-the-moment selection of known and novel referents (mapping), as well as retention of novel word-referent links after a five-minute break. This is similar to prior referent selection (fast-mapping) paradigms (Horst & Samuelson, 2008). Importantly, these trials tap children’s ability to use their prior knowledge (or lack of) in-the-moment to disambiguate a familiar from a novel word and item. Retention trials followed to test if the new words were encoded robustly enough to be recalled from memory.




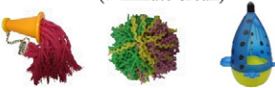




	Example Trial(s)	Example Prompt	# of trials
In-the-moment mapping (known & novel)	<i>Known Word Comprehension</i>		
		<i>Can you get the cup?</i>	3-6
	<i>Referent Selection (RS)</i>		
	<i>Known RS</i> 	<i>Can you get the shoe?</i>	4-8
<i>Novel RS</i> 	<i>Can you get the cheem?</i>	4-8	
Retention (novel)	<i>(5-minute break)</i>		
	<i>RS Retention</i> 	<i>Can you get the cheem?</i>	2-4
	<i>Direct Naming</i>		
	<i>Training</i> 	<i>This is the neem.</i> <i>This is the lorp.</i>	
<i>Retention</i> 	<i>Can you get the neem?</i>	4	
Generalization	<i>Novel Noun Generalization (NNG)</i>		
	<i>Training</i> 	<i>This is my zup.</i>	16
<i>Generalization</i> 	<i>Can you get your zup?</i>		

Figure 2. Schematic of the procedure. During training periods, items were held up by the experimenter one at a time. During the remainder of the trials, items were presented as real 3D items in a row on a white tray

Initial referent selection consisted of eight to sixteen trials¹, each consisting of two randomly selected known items from the prior Known Comprehension task and a single randomly selected novel item from the pool. On half the trials, children were asked to select a known item by name in the same manner as the Known Comprehension trials (“Can you get the shoe?”); these are referred to as the Known RS trials and represent another approach to testing children’s in-the-moment known word comprehension. On the alternating trials, children were asked to select a novel item by name (“Can you get the

¹Due to a change in procedure, 16 children received only 8 RS trials (4 novel and 4 known). There were no differences in performance for children receiving 8 vs. 16 trials and, thus, all are collapsed together.

cheem?"); these Novel RS trials tap children's in-the-moment novel word mapping. Across trials, novel items were present only once and object locations and trial order were randomized. Children were re-prompted up to three times if needed, but without correction or praise.

A five-minute coloring break immediately followed the referent selection trials and preceded retention. A single warm-up trial (mimicking Known Comprehension) followed the break to re-engage children; four to eight retention trials immediately followed. These retention trials critically tested children's memory for newly mapped words. On each retention trial, children were presented with three novel items – two of which had previously been targets on Novel RS trials (and thus, presumably had a novel label), and one of which was a novel foil item on Known RS trials. Targets and item locations were randomized and did not repeat across trials; however, the novel targets from the first half of the retention trials served as foils on the second half of the retention trials. There was no difference on performance between the first half and second half of the trials, however, so all trials were analyzed together.

Direct Naming and Retention. Direct Naming parallels classic novel word learning paradigms in which a child is ostensibly taught the name for a new item, then immediately tested on their recollection of its label (e.g., Ellis Weismer et al., 2013). While the direct, ostensive naming/training period includes mapping of words to objects, unlike referent selection, the training here presented items in isolation, without competitors or foils, and with repetition to make mapping easy. Thus, the critical test of children's retention of these newly learned words is (mostly) independent of their disambiguation abilities.

This task began with a training period in which children were presented with one novel item at a time (see Figure 1c) and taught its label. Items were held up by the experimenter and named once, then given to the child for the child to explore, during which it was named three more times. After roughly 10 seconds, the experimenter retrieved the item, named it once more while holding it up, and then removed it from sight. The process was repeated for each of the four stimuli in a random order. After all four items were trained, children were immediately tested on their retention for the novel labels². Items were presented in pairs on a white tray equidistant apart. While maintaining eye contact with the child, the tray was placed on the table out of reach of the child. Children were then asked to retrieve an item by name ("Can you get the *neem?*") and the tray was pushed within reach of the child. Children were re-prompted up to three times if needed, but no correction or praise were given. Order of trials were counterbalanced.

Novel Noun Generalization (NNG). The final task tapped generalization skills and used the canonical Novel Noun Generalization procedure by Smith and colleagues (Landau et al., 1988; Smith et al., 2002). A key aspect of this task is children's ability to generalize a novel label to other members of the same category. Importantly, exemplars were present for reference at all times, so children's ability to respond required less emphasis on mapping or retention abilities. The key question is if children demonstrate a shape bias, applying a new word to other items with similar shapes.

The NNG task began with two warm-up trials, followed by the test trials. On each warm-up trial, the experimenter placed the target item (e.g., a fabric flower) and the foil

²The break time here was less than that in the RS trials for two reasons. First, prior work using direct or ostensive naming paradigms rarely have a long break (see Ellis Weismer et al., 2013) and second, pilot studies suggested that children had difficulty returning for the rest of the tasks when a second break was implemented. Thus, tasks were set-up to mirror prior work as closely as possible.

item (e.g., wood block) on a white tray equidistant apart and out of reach of the child. The experimenter then held up an item identical to the target (e.g., flower), naming it three times in a neutral syntax (e.g., “This is my flower”). The child was then asked “Can you get your flower?” and the tray was pushed forward for the child to make a selection. As above, children were re-prompted up to three times if needed and corrected and/or praised for their selections during warm-up. Sixteen test trials with novel items immediately followed warm-up and were nearly identical, but without correction or praise. On each trial, three items were present – a novel exemplar, a novel item matching the exemplar in shape, and a novel item matching the exemplar in material (see [Figure 1d](#)). The experimenter placed the shape-matching test item and the material-matching test item on a white tray equidistant apart, out of reach of the child. The experimenter then named the exemplar item with a novel name three times in a syntactically neutral context (e.g., “This is my zup”). Children were then asked “Can you get your zup?” and the tray was pushed forward. Children were re-prompted up to three times if needed but no praise was given. There were four test trials for each stimuli set with each test item appearing twice, but paired with a different competitor and in a different location on each trial. Order of sets and trials were randomized.

Coding

Coders blind to the hypotheses and talker status indicated children’s final selections offline via video recordings. Data from 50 random subjects (44%) were re-coded for reliability purposes. Inter-coder agreement was 97.3%. Discrepancies were settled via independent judgments by a third blind coder.

Analysis

While the sample of participants included both children who met the criteria for LT and children with NL, the vocabulary abilities of the participants represented the entire range of the normative continuum. We aimed to capture that variability in the analysis by first using a continuous measure of vocabulary percentile as the primary predictor. This dimensional approach avoids the assumptions of an arbitrary cut-off for a delay and has the potential to better capture individual variability (see JLSHR special issue on statistical approaches to language, 2019; Perry & Kucker, 2019). In order to compare to prior work, a second set of traditional group-level analyses comparing LT and NL children were also conducted.

First, binomial linear mixed models were used to examine the effect of a child’s vocabulary abilities and age on performance in each task. Each task was analyzed separately with a linear mixed model of children’s trial by trial performance (correct or not). Vocabulary percentile (centered) and age (in days, centered) were included as fixed effects and the maximum random effect structure justifiable was used (Barr, Levy, Scheepers & Tily, 2013). In order to examine how vocabulary size (not just normative percentile) directly impacts performance, a total vocabulary score was calculated for all children who completed the longer, more detailed MCDI-WS (n=86). A shape residual score was also calculated for these children in order to obtain a measure of vocabulary structure. This shape residual score represents the proportion of the child’s vocabulary dedicated to shape-based words, without being skewed by their overall vocabulary size (Perry, Axelsson & Horst, 2016; Perry & Samuelson, 2011; Slone & Sandhofer, 2017).

Shape vocabulary is known to be critical for generalization performance (Samuelson, 2002) and prior work suggests it may differ between LT and NL children (Colunga & Sims, 2017; Perry & Kucker, 2019). Additional mixed model regression analyses were then run on this MCDI sub-set with fixed effects of vocabulary size and shape residual score and the maximal justified random effects structure. Vocabulary percentile and age were not included in these additional models because both are highly collinear with vocabulary size.

Second, in order to draw comparisons to prior work, we also used between groups *t*-tests to compare LT vs. NL performance on each of the tasks. These subsets of LT and NL children were matched on age and gender, selected from the larger sample.

Finally, an important part of language development is understanding how the different processes of learning impact each other. Thus, a final analysis compared across tasks to examine if performance in one task predicted a child's performance in a different task. Overall correlations between the tasks were run and then separate generalized linear mixed models examined trial-by-trial performance on each task as predicted by average ability on the other tasks. Each model included vocabulary percentile and performance on each other task as fixed effects, a co-variate of age, and the maximal justified random effect structure. Data files and R scripts for all analyses are available via The Open Science Framework (OSF) https://osf.io/qf73r/?view_only=bbcbdb54c9cc4f6e806322a4e2183ff8.

Results

In-the-moment mapping

Known Word Comprehension

The known word comprehension task served two purposes – to test children's ability to use known words in-the-moment and familiarize the child with the testing procedure. Thus, children were ultimately corrected and/or praised on these trials, and responses were coded for both their initial response (prior to any praise or correction) and their final response (after praise and/or correction). Initial responses represent a child's ability to use their lexical knowledge in-the-moment; final responses represent a child's ability to ultimately learn the procedure and respond accurately after feedback. Thus, initial responses are the primary variable of interest here, but both initial and final are analyzed.

The best fitting model for initial responses included only a random intercept for subject. There were significant positive correlations of both age, $\beta = 1.01$, $z = 5.70$, $p < .0001$, and vocabulary percentile, $\beta = .51$, $z = 3.16$, $p = .002$ (see Figure 3, red circles, top panels). In order to further explore the impact of specific vocabulary size and structure on performance, children who completed the MCDI-WS were entered into a second model that included a fixed factor of total vocabulary size and shape residual score. The best fitting model included a random effect of subject. This vocabulary model revealed a significant effect of vocabulary size, $\beta = .92$, $z = 5.24$, $p < .0001$, but not specific shape vocabulary (shape residual), $\beta = -.02$, $z = -.01$, $p = .99$ (Figure 3, red circles, bottom panels). Notably, the age-matched categorical analysis also revealed a significant difference between LT and NL children, $t(50) = 2.31$, $p = .025$ (see Figure 7, after NNG results), with NL children (*M* accuracy = .80, *SD* = .20) out-performing LT (*M* accuracy .62, *SD* = .36). The vocabulary-matched comparison did not reveal a significant difference between groups (NL: *M* = .70, *SD* = .26; LT: *M* = .79, *SD* = .26), $t(28) = -.77$, $p = .45$.

After children were corrected and praised, their final selections were analyzed (Figure 3, blue triangles); the best fitting model included only a random effect of subject.

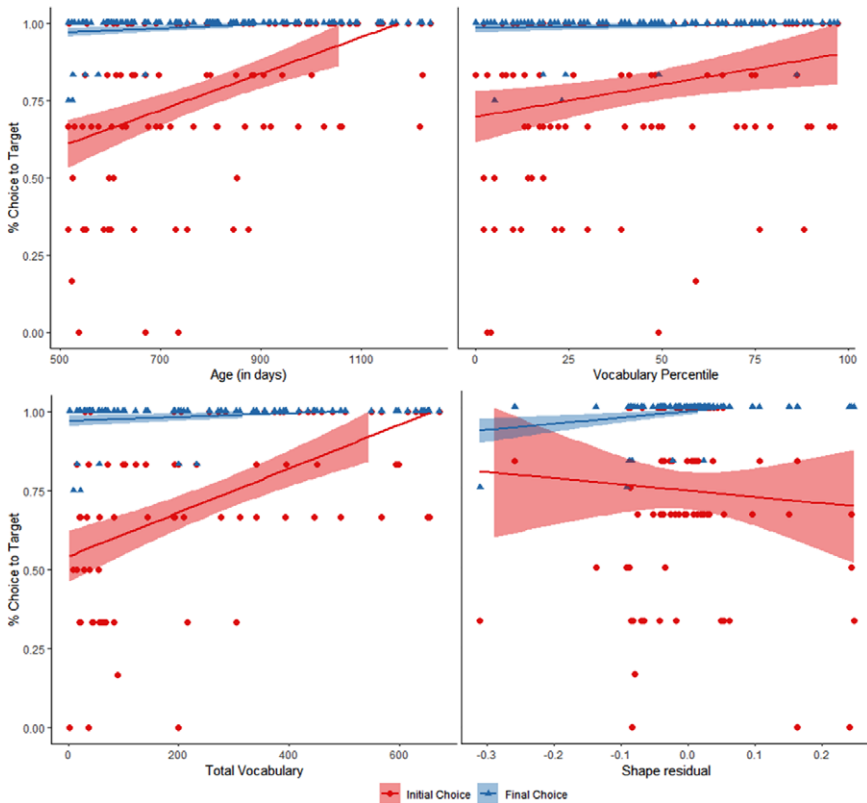


Figure 3. Performance on Known Comprehension trials. All children's performance included in the top two panels; children with the MCIDI-WS included in the bottom panels. Note: lines represent linear regressions for visualization purposes only.

Final selections were significantly predicted by a child's age, $\beta = 3.70$, $z = 2.18$, $p = .03$, but not vocabulary percentile, $\beta = .26$, $z = .58$, $p = .56$. The second model with total vocabulary size and shape residual scores included a random effect of subject and revealed no effect of vocabulary size, $\beta = 1.29$, $z = 1.51$, $p = .13$ and a marginal effect of shape residual, $\beta = 6.94$, $z = 1.95$, $p = .05$. Neither the analysis with age-matched groups, $t(43.56) = .28$, $p = .78$ (NL: $M = .99$, $SD = .02$; LT: $M = .99$, $SD = .02$), nor the vocabulary-matched groups, $t(14) = -1.00$, $p = .33$ (NL: $M = .99$, $SD = .03$; LT: $M = 1.00$, $SD = 0$), show an effect of talker status. However, nearly all children were at ceiling on the final choices here with low variability, thus the lack of group differences is not surprising.

Performance on Known Comprehension offers two conclusions. First, as seen in initial performance, both vocabulary and age matter for in-the-moment known word processing. Though nearly all children were well above chance overall, there were clear differences in the robustness of this ability dependent on vocabulary and age. On the surface, even LT children can operate in-the-moment to respond accurately to a known label as demonstrated by their above-chance performance. However, a child's ability to do so seamlessly appears to be influenced by their specific vocabulary size and age – LT children

(and those with lower vocabularies or who were younger in general) performed much more poorly than their NL peers (and those with higher vocabularies or whom were older). A second conclusion is that the near-ceiling performance on children's final choices demonstrates that children understood the procedures, giving faith that any poor performance in the subsequent tasks is unlikely to be due to confusion over basic methods used.

Referent Selection

Known Referent Selection. Known Referent Selection (Known RS) presents children with the same familiar, name-known items presented in the Known Comprehension phase, but now in the context of a novel foil distractor item. Thus, Known RS represents a critical test of the robustness of a child's word knowledge against their novelty biases (see Kucker et al., 2018). Again, multiple linear mixed models were run. The best fitting model included random intercepts for both subject and item. Vocabulary percentile marginally predicted a child's Known RS performance, $\beta = .36$, $z = 1.69$, $p = .09$, while age significantly predicted performance, $\beta = 1.32$, $z = 5.67$, $p < .0001$ (Figure 4, red circles, top panels). A second vocabulary model included a random effect of subject and item and revealed a significant effect of vocabulary size, $\beta = 1.29$, $z = 5.27$, $p < .0001$, but no effect of shape residual, $\beta = -3.86$, $z = -1.41$, $p = .16$ (Figure 4, red circles, bottom panels). Furthermore, age-matched groups, $t(50) = 1.85$, $p = .07$, but not vocabulary-matched groups, $t(28) = -1.23$, $p = .23$, showed marginally significant differences between LT and NL. Within the age-matched groups, the LT children selected the known target at levels no different than chance (.33) ($M = .39$, $SD = .30$), $t(25) = .77$, $p = .45$, whereas the NL children selected the known target above chance ($M = .57$, $SD = .32$), $t(25) = 3.48$, $p = .002$ (Figure 7).

Overall, there was a substantial drop in performance from the just-prior Known Comprehension trials for both LT and NL children, with LT children now at chance. Unlike the Known Comprehension trials, however, here novel foil items were present, which requires the child to ignore novel distractors while mapping the known word and referent. Known RS is thus an arguably more difficult test of children's in-the-moment use of their lexicon and provides a measure of a child's novelty bias – how often do they select the novel foil item instead of the familiar target? To test this, additional models were run to examine the correlation between choices to the NOVEL item with vocabulary and age. The mixed model of children's novel choices included only a random intercept of subject and revealed no effect of vocabulary $\beta = -.30$, $z = -1.53$, $p = .13$, but a significant effect of age, $\beta = -1.11$, $z = -5.39$, $p < .0001$. A second vocabulary regression included a random effect of subject and found a significant main effect of vocabulary size, $\beta = -1.08$, $z = -5.07$, $p < .001$, but not shape residual, $\beta = 2.09$, $z = .91$, $p = .36$. The age-matched groups also show a marginal effect of talker status, $t(50) = -1.66$, $p = .104$ (NL: $M = .40$, $SD = .33$; LT: $M = .55$, $SD = .35$), but the vocabulary-matched groups do not, $t(28) = 1.29$, $p = .21$. In fact, in the age-matched group, the LT children were selecting the novel foil distractor the majority of the time, significantly above chance (33%), $t(25) = 3.25$, $p = .003$, whereas the NL group was at chance for selecting the novel distractor, $t(25) = .99$, $p = .33$. That is, LT children's chance performance on Known RS appears due to an overriding bias toward novelty, similar to that seen in 18-month-old word learners (Kucker et al., 2018).

Novel Referent Selection. The Novel RS trials are set up identically to Known RS with two known items and a single novel item, but here, children are asked to select the novel

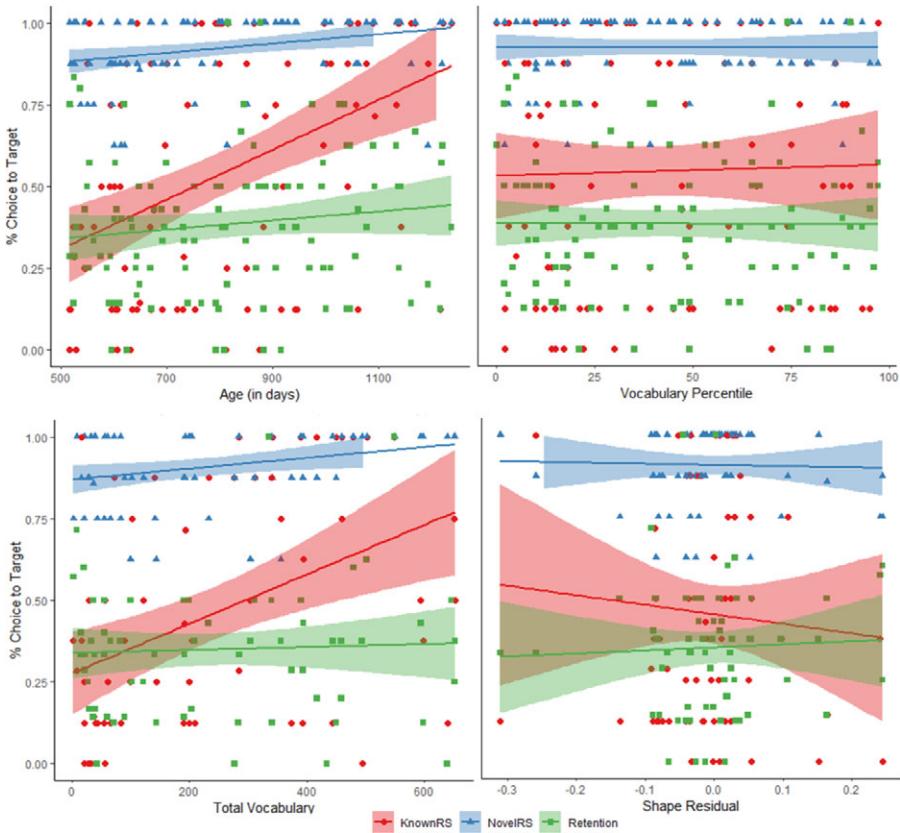


Figure 4. Performance on Referent Selection and Retention. All children's performance included in the top two panels; children with the MCDI-WS included in the bottom panels. Note: lines represent linear regressions for visualization purposes only.

item by name. This provides a measure of initial novel word mapping or disambiguation ability. The best fitting model included a random effect of subject and found a significant effect of age, $\beta = .48$, $z = 2.79$, $p = .005$, but no effect of vocabulary percentile, $\beta = .05$, $z = .30$, $p = .76$ (Figure 4, blue triangles, top panels). The vocabulary model included only a random effect of subject and found a significant effect of vocabulary size, $\beta = .40$, $z = 2.13$, $p = .03$, but not shape residual, $\beta = .18$, $z = .11$, $p = .91$ (Figure 4, blue triangles, bottom panels). Neither the age-matched, $t(50) = .58$, $p = .56$, nor the vocabulary-matched, $t(28) = -.38$, $p = .71$, groups found significant differences between LT and NL. Overall, all children were well above chance, near ceiling (age group NL: $M = .92$, $SD = .16$; LT: $M = .90$, $SD = .21$, see Figure 7). Given the novelty biases seen on the Known RS trials, selection of the novel item when it is the target, even by the LT children, is not terribly surprising. However, the overall high performance combined with the relatively poor performance on Known RS (especially for LT children) should lend caution to interpreting a child's ability to learn new words on these Novel RS trials given their indiscriminate ability to always select the unknown item regardless of what label is given. Testing learning is the goal of the Retention trials.

Retention

Referent Selection Retention (RS Retention)

After a short break, children were tested on their retention for the words exposed during the RS trials. A linear mixed model for RS retention was run with fixed effects of vocabulary percentile and age. The best fitting model included both subject and item as random effects. There was a marginally significant effect of age on performance, $\beta = .16$, $z = 1.81$, $p = .07$, but no effect of vocabulary, $\beta = .01$, $z = .08$, $p = .94$ (Figure 4, green squares, top panels). The vocabulary model included both subject and item as random effects and showed no effect of vocabulary, $\beta = .13$, $z = 1.17$, $p = .24$, nor shape residual, $\beta = .64$, $z = .50$, $p = .62$ (Figure 4, green squares, bottom panels). The categorical analysis revealed no significant differences between LT and NL for either age-matched, $t(50) = .87$, $p = .39$ or vocabulary-matched, $t(28) = .24$, $p = .81$ groups.³ However, NL children in the age-matched groups selected the target 41% of the time, significantly above chance (33%), $t(25) = 2.14$, $p = .04$, whereas the LT children did not, selecting it only 36% of the time, $t(25) = .78$, $p = .44$ (Figure 7, in next section). This is consistent with prior work that, for most children under three years old, retention is difficult (Horst & Samuelson, 2008), but as children get older, their ability to recall new word-referent pairs increases (see also Bion et al., 2013). Notably, however, prior work with LT children has found mixed results for retention in this group and LT children here (matched to NL on age) were not above chance, suggesting a possible area of weakness.

Direct Naming Retention

For the Direct Naming task, children's ability to retain the novel word-referent pairs after a repetitive, ostensive, naming period was tested. While also a test of retention like RS Retention, this task eliminates the competition during learning and minimizes the in-the-moment processing that is required. Thus, it is arguably a more direct test of retention because it is less dependent on disambiguation abilities. It also parallels some of the more traditional "fast-mapping" tasks in which children are ostensibly taught a label and immediately tested on that knowledge. The best fitting model included both subject and item and revealed a marginal main effect of vocabulary percentile, $\beta = .20$, $z = 1.89$, $p = .06$, but not age, $\beta = .11$, $z = 1.06$, $p = .29$ (Figure 5, top panels). The vocabulary model included only a random effect of subject and found a significant effect of vocabulary size, $\beta = .24$, $z = 2.10$, $p = .04$, but not shape vocabulary, $\beta = -.25$, $z = -.19$, $p = .85$ (Figure 5, bottom panels). The categorical analysis was not significant for the age-matched, $t(50) = 1.60$, $p = .11$, or vocabulary-matched, $t(28) = .35$, $p = .73$, groups. However, as with RS Retention, the age-matched NL group retained the newly taught words above chance (50%), whereas the LT group did not – NL: ($M=.62$, $SD = .40$), $t(25)=2.37$, $p = .026$; LT: ($M=.50$, $SD = .41$), $t(25)=0.0$, $p = 1.0$, Figure 7 (end of NNG section). Though there were

³In the current analysis, only items for which the child accurately mapped during the prior Novel RS trials were included. This is consistent with prior work and ensures the child had an initial word-referent pair to retain. If all retention trials are included instead (model includes both subject and target intercepts), the results are the same with only an effect of age, $\beta = .20$, $z = 2.39$, $p = .017$, and no vocabulary percentile, $\beta = .04$, $z = .44$, $p = .66$, in the main regression. The vocabulary model is also similar – a marginal effect of vocabulary size, $\beta = .18$, $z = 1.78$, $p = .076$, and no effect of shape residual, $\beta = .40$, $z = .35$, $p = .73$. There are no differences with either the age-matched, $t(50) = 1.07$, $p = .29$, or vocabulary-matched, $t(28) = .65$, $p = .52$ groups. Age-matched NL children are still marginally above chance, $t(25)=1.93$, $p=.06$, whereas LT children are not, $t(25) = .52$, $p=.60$.

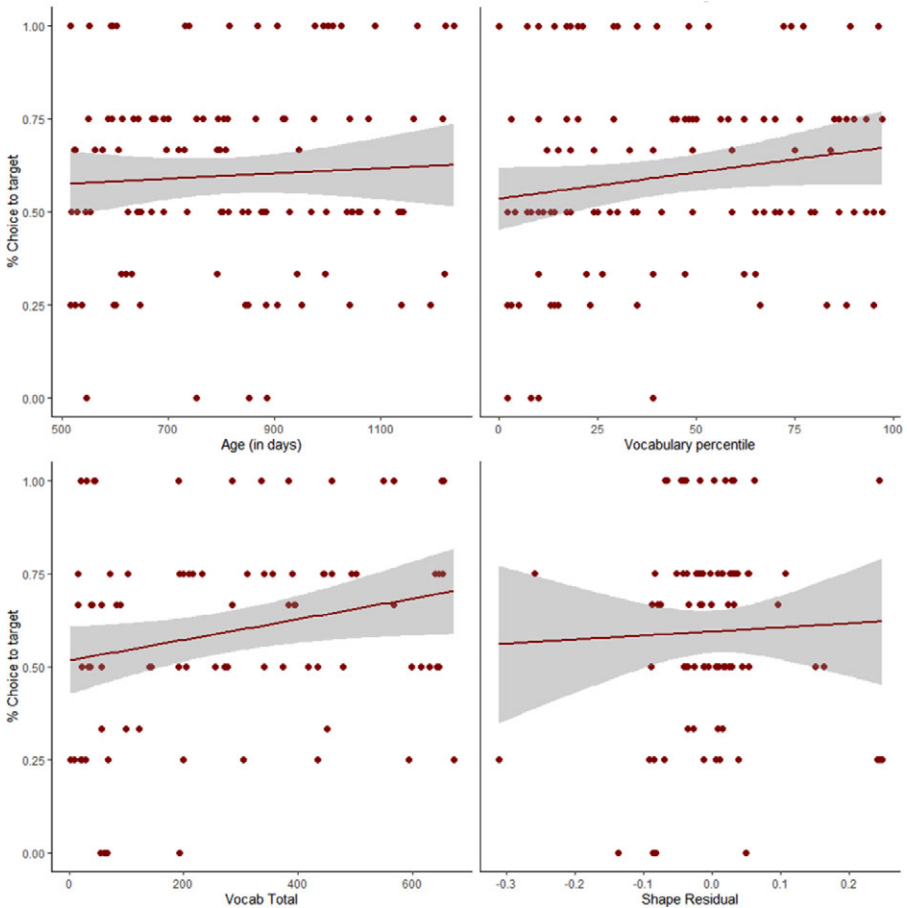


Figure 5. Performance on Direct Naming and Retention. All children's performance included in the top two panels; children with the MCDI-WS included in the bottom panels. Note: lines represent linear regressions for visualization purposes only.

not group-level differences or consistent vocabulary effects across both retention tasks (neither RS or here), the chance performance of the LT children suggests that overall lexical knowledge may impact retention abilities in subtle ways, but further work may be needed to tease this apart.

Generalization

Novel Noun Generalization

Once a child makes an association between a novel word and its referent, another critical step is the ability to generalize those novel labels to other exemplars of the same category – for solid nouns, this is demonstrated via a shape bias (Smith et al., 2002). The linear mixed model used predicted children's selection of the shape-matching item and included

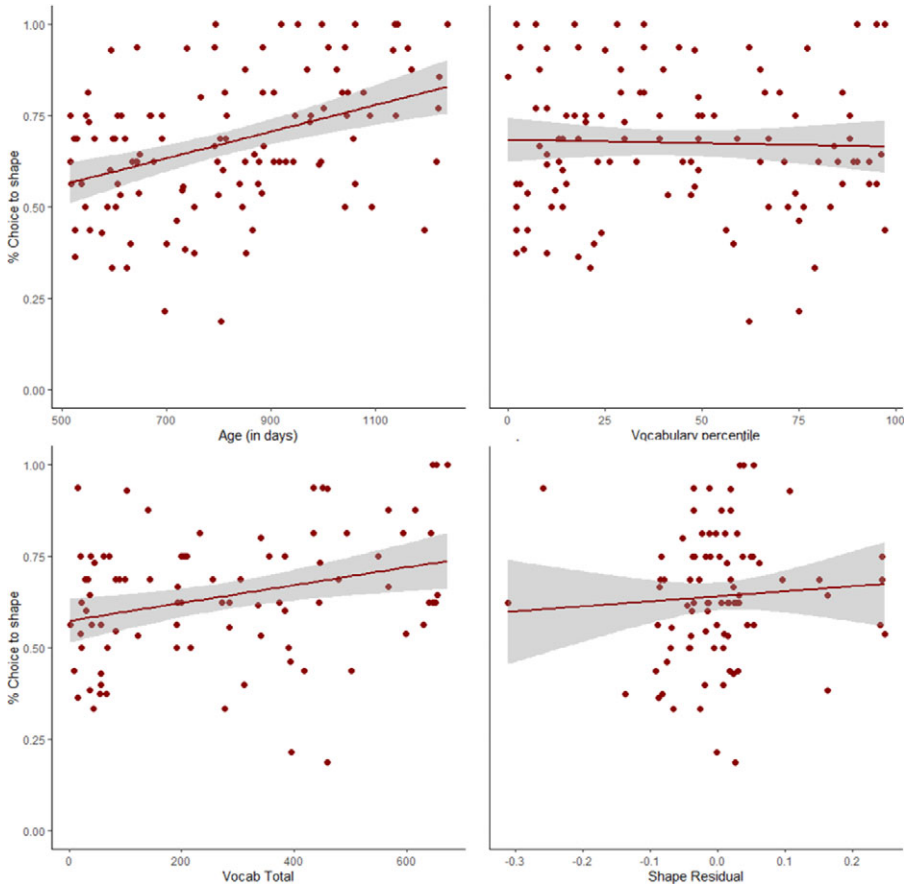


Figure 6. Performance on Novel Noun Generalization. All children’s performance included in the top two panels; children with the MCDI-WS included in the bottom panels. Note: lines represent linear regressions for visualization purposes only.

random effects of both subject and item. The model revealed a main effect of age, $\beta = .40$, $z = 4.61$, $p < .001$, but not vocabulary percentile, $\beta = .02$, $z = .23$, $p = .82$ (Figure 6, top panels). The vocabulary model included a random effect of both subject and item, revealing a significant effect of vocabulary size, $\beta = .25$, $z = 2.91$, $p = .003$, but no effect of shape vocabulary, $\beta = .54$, $z = .54$, $p = .59$ (Figure 6, bottom panels). Similar to prior work (Colunga & Sims, 2017), there were no differences in the age-matched NL and LT groups, $t(50) = .93$, $p = .36$, nor in the vocabulary-matched groups, $t(28) = .97$, $p = .34$. However, the majority of children were above chance (50%), showing a shape bias in both the age-matched NL group ($M = .68$, $SD = .45$), $t(25) = 5.40$, $p < .001$, and the LT group ($M = .63$, $SD = .45$), $t(25) = 3.60$, $p = .001$ (see Figure 7). However, like prior work by Perry and Kucker (2019), a child’s status as a late talker may not tell the entire story – for children under 30-months, the size of their vocabulary is predictive of performance, not their category label as LT. And, we know from prior work that the structure of one’s vocabulary is key for showing a shape bias. In fact, evidence suggests that LT children’s

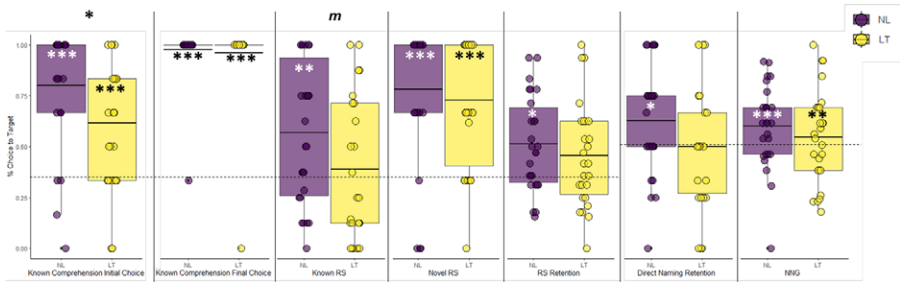


Figure 7. Performance on each task for age-matched NL children vs. LT children. Dashed line represents chance. Asterisk within bars represent difference from chance; above bars represent difference between NL and LT groups on that task. *** $p < .001$ ** $p < .01$, * $p < .05$, ^m $p < .10$

noun structure is unique and that uniqueness matters for generalization (Beckage, Smith & Hills, 2011; Colunga & Sims, 2017; Perry & Kucker, 2019). Though shape residual was not significant in the full model, exploratory analysis showed an interaction of the LT group and shape residual such that LT children with fewer shape-based nouns in their vocabulary than what is expected are less likely to show a shape bias, much like prior work.

Across tasks

Word learning is not a quick, instant process. Instead, it unfolds over multiple, inter-related timescales. How those phases are related is critical. Thus, two final sets of analyses compared performance across each of the tasks. First, bivariate correlations between performance on each task were run. These correlations give an overview of relationships between processes, but are limited to the group-level data. Thus, a second analysis incorporating within-subject repeated measures was conducted. Here, a generalized mixed model predicting an individual's performance on a given task from their average performance on each of the other tasks were run. To examine if language delay predicts different patterns of relationships, binomial mixed models were run for each task and talker group (LT vs. NL) individually. These models included the maximal random effects structure justified.

Correlations between average performance in each task are presented in Table 2. Much like the by-task analyses, a child's age significantly correlates with their performance in nearly every task, except those tapping retention (RS Retention and Direct Naming Retention). Vocabulary percentile is also significantly positively associated with Initial Known Comprehension and marginally with Direct Naming Retention. Importantly, there are also strong correlations between tasks, and across phases (in-the-moment mapping, retention, and generalization). All tasks tapping in-the-moment known selection are related – Initial Known Comprehension correlated with Final Known Comprehension and Known RS. However, the retention tasks do not correlate, suggesting as we hypothesized, that they either tap into slightly different elements of retention or that the differences in prior mapping (learning/selection) exposures indeed have different cascading effects. There are also relationships across timescales – Initial Known Comprehension positively correlated with NNG, Known RS correlated with Direct Naming

Table 2. Correlations between vocabulary percentile, age, and task performance

	1	2	3	4	5	6	7	8
1. Vocabulary Percentile	--							
2. Age	-.15	--						
3. Known Comprehension Initial	.24**	.46***	--					
4. Known Comprehension Final	.09	.30**	.29**	--				
5. Known RS	.09	.48***	.57***	.14	--			
6. Novel RS	-.03	.24*	.10	-.04	.05	--		
7. RS Retention	-.00	.13	.02	-.09	.13*	.03	--	
8. Direct Naming Retention	.16 ^m	.05	.16 ^m	.08	.19*	.19*	-.03	--
9. NNG shape	-.03	.40***	.32***	.14	.36***	.07	.05	.17 ^m

*** $p < .001$ ** $p < .01$ * $p < .05$ ^m $p < .1$

Retention as well as with NNG, Novel RS is associated with Direct Naming, and Direct Naming and NNG are positively correlated.⁴

To further probe these correlations, repeated measures regressions were run. As before, in each of these models, children's trial-by-trial performance on a given task was the outcome variable. Fixed effects included average performance (centered) on each of the other tasks as well as covariates of vocabulary percentile (centered) and age (in days, centered). Final Known Comprehension was not included as a predictor due to the ceiling effect seen by most children and its less central goal of orienting children to the task procedure. The best fitting models all included random intercepts of subject and/or item. See Table 3 for the full data table of each model. There was some evidence of a relationship between the tasks tapping in-the-moment mapping; Known RS significantly predicted Initial and Final Known Comprehension performance, and Initial Known Comprehension performance predicted Known RS ability. Novel RS and Direct Naming also predicted each other. Overall, however, few other predictors or models revealed significant effects, suggesting that while overall performance across tasks are correlated in the bivariate correlations, the variability makes relationships within a specific child harder to detect.

A critical final question is if this pattern of relationships across tasks/processes is different for NL vs. LT children. Thus, a final set of follow-up models for the age-matched NL and LT groups were run (see Table 3, right columns). As with the full sample, predictions between tasks was largely non-significant. However, there were a few notable differences between NL and LT patterns. For instance, NNG predicts Initial Known Comprehension performance for NL, but only predicts Final Known Comprehension in

⁴Nearly identical patterns of results are found when partial correlations accounting for age and vocabulary are run instead.

Table 3. Predicting performance on each task.

Outcome	Predictor	Full sample			NL sample			LT sample		
		β	z	p	β	z	p	β	z	p
<i>Initial Known Comprehension</i>	Vocab Percentile	.33	2.21	.03*						
	Age	.54	3.00	.003**						
	Known RS	1.97	4.14	<.001***	5.16	2.24	.03*	2.10	2.73	.01**
	Novel RS	1.10	.86	.39	.85	.21	.84	.95	.57	.57
	RS Ret	-.17	-.24	.81	-1.12	-.38	.70	-.94	-.92	.36
	Direct Naming	.13	.22	.83	-.22	-.11	.91	.56	.76	.45
	NNG	.91	1.04	.30	-8.99	-1.71	.09^m	2.21	1.46	.14
<i>Final Known Comprehension</i>	Vocab Percentile	.26	2.08	.04*						
	Age	.52	3.46	<.001***						
	Known RS	1.53	3.88	<.001***	2.10	2.73	.006**	5.16	2.24	.03*
	Novel RS	.68	.66	.51	.95	.57	.57	.85	.21	.84
	RS Ret	-.29	-.49	.62	-.94	-.92	.36	-1.11	-.38	.70
	Direct Naming	.20	.43	.67	.56	.76	.45	-.22	-.11	.91
	NNG	.83	1.14	.25	2.21	1.46	.14	-8.99	-1.71	.09^m
<i>Known RS</i>	Vocab Percentile	.01	.06	.95						
	Age	.67	2.73	.006**						
	Initial Known Comp	3.70	4.04	<.001***	2.03	1.70	.09^m	3.10	2.24	.03*
	Novel RS	-1.88	-1.06	.29	3.06	1.02	.31	.49	.18	.86
	RS Ret.	1.35	1.42	.16	4.93	2.24	.03*	-.06	-.04	.97
	Direct Naming	1.16	1.47	.14	2.28	1.63	.10^m	-1.09	-.90	.37
	NNG	1.66	1.42	.16	1.01	.45	.65	5.81	2.54	.01**

Table 3. (Continued)

Outcome	Predictor	Full sample			NL sample			LT sample		
		β	<i>z</i>	<i>p</i>	β	<i>z</i>	<i>p</i>	β	<i>z</i>	<i>p</i>
<i>Novel RS</i>	Vocab Percentile	.002	.014	.99						
	Age	.55	2.56	.01**						
	Initial Known Comp	.21	.27	.78	-.28	-.22	.82	.43	.37	.72
	Known RS	-.55	-.99	.32	.79	.62	.54	.38	.38	.71
	RS Ret.	.20	.26	.79	-2.39	-1.22	.22	.24	.19	.85
	Direct Naming	1.35	2.05	.04*	2.21	1.53	.13	.05	.05	.96
	NNG	-.11	-.12	.91	-1.55	-.57	.57	.02	.01	.99
<i>RS Retention</i>	Vocab Percentile	.03	.35	.72						
	Age	.15	1.27	.21						
	Initial Known Comp	-.48	-1.02	.31	-.19	-.27	.78	-.26	-.28	.78
	Known RS	.40	1.28	.20	1.12	1.84	.07^m	.40	.50	.62
	Novel RS	.22	.24	.81	-1.33	-.83	.41	1.18	.68	.49
	Direct Naming	-.22	-.60	.55	.21	.28	.78	-1.79	-2.37	.02*
	NNG	.11	.20	.84	-.28	-.28	.78	-.58	-.39	.69
<i>Direct Naming Retention</i>	Vocab Percentile	.17	1.52	.13						
	Age	-.09	-.69	.49						
	Initial Known Comp	-.05	-.10	.92	-.02	-.02	.98	.70	.66	.51
	Known RS	.60	1.59	.11	1.20	1.52	.13	-.21	-.23	.82
	Novel RS	1.82	1.90	.06^m	2.61	1.46	.15	.42	.21	.84
	RS Ret	-.33	-.66	.51	.51	.36	.72	-2.37	-2.00	.05*
	NNG	.82	1.30	.19	.75	.53	.59	1.08	.64	.52

Table 3. (Continued)

Outcome	Predictor	Full sample			NL sample			LT sample		
		β	<i>z</i>	<i>p</i>	β	<i>z</i>	<i>p</i>	β	<i>z</i>	<i>p</i>
NNG	Vocab Percentile	-.06	-.73	.47						
	Age	.27	2.60	.009***						
	Initial Known Comp	.33	.82	.41	-.65	-1.11	.27	.69	1.21	.23
	Known RS	.45	1.57	.12	.42	.72	.47	1.17	2.43	.02*
	Novel RS	-.25	-.34	.73	-.36	-.27	.79	-.05	-.05	.96
	RS Ret	-.04	-.12	.91	-.46	-.48	.63	.12	.19	.85
	Direct Naming	.50	1.52	.13	.27	.42	.67	.60	1.31	.19

****p*<.001

***p*<.01

**p*<.05

[†]*p*<.1

the LT sample. This reinforces that robust known vocabulary (as tested via Initial Known Comprehension performance) may support a shape bias and potentially on-track vocabulary growth. Likewise, Known RS performance is predicted by Initial Known Comprehension ability for both groups – suggesting that both tap the ability to use known words in-the-moment. However, other predictors of Known RS diverge across the groups – RS Retention predicts Known RS in the NL group, whereas NNG predicts Known RS in the LT group. In the LT group, there is some evidence that the two tasks tapping retention correlate – RS Retention and Direct Naming predict each other in the LT group, but not the NL group. Given the large amount of null results, the evidence here is merely preliminary, but suggests LT and NL children may differ in the types of processes that support word learning and cascades for future learning.

Discussion

The current study aimed to systematically measure the multi-faceted nature of word learning across timescales (in-the-moment mapping, retention, and generalization) in a wide range of children, including those with language delays. There were two key conclusions. First, the results reinforce that across multiple timescales both age and vocabulary abilities contribute to children's word learning. Second, children with smaller vocabularies, including LT children, demonstrate the biggest deficits in tasks that tap in-the-moment processing of known words, with some evidence for additional cascading deficits in retention. Taken together, the results suggest that approaches to studying word learning that cross multiple timescales may offer a holistic view for capturing word learning potential in both NL and LT children.

The importance of age and vocabulary across timescales

Replicating prior work (Bion et al., 2013; Samuelson, Schutte & Horst, 2009), child's age was a significant predictor of their language ability. What is expanded upon here is that age positively predicted across multiple different timescales/processes of the word learning process – significantly influenced known word identification (in Initial Known Comprehension, Final Known Comprehension, and Known RS), novel word selection and disambiguation (in Novel RS), some elements of novel word retention (in the RS Retention task), and generalization (in NNG). Moreover, while the children here represent a wide age range from as young as 17-months to as old as 40-months, there are still large changes in abilities even across a few months within that span. Age thus continues to be a critical variable to consider when comparing children's word learning abilities at *any* point in the word learning process.

A second consistent finding is that timescales of word learning are all influenced in some way by vocabulary, normative percentile rank, as well as specific vocabulary size and structure. Children with lower vocabulary sizes, those lower on normative vocabulary, and those classified as LT children perform poorer on in-the-moment known word comprehension tasks (Initial Known Comprehension, Known RS) and some elements of retention (Direct Naming). Total vocabulary size (but not percentile rank or LT status) predicts in-the-moment Novel RS, and generalization (NNG). This emphasizes the critical role that the vocabulary knowledge a child brings to the task plays in their ability to use that knowledge and acquire new words. Importantly, it suggests that deficits in word learning seen in LT children with smaller vocabularies may stem from their weak

lexical foundation, which prevents them from acting in-the-moment as accurately as their NL peers. Though, in part, unsurprisingly, this lays a foundation for future work more closely tracking cascading effects across processes.

Finally, the results lend initial support that word learning processes were likely related – overall performance on some tasks correlate with performance in another – namely, some in-the-moment processes (Initial Known Comprehension, Known RS) correlate with some retention processes (Direct Naming) and generalization (NNG). The within-child analyses comparing performance for an individual child to their performance in other tasks were largely non-significant, however. Though prior work has demonstrated relationships between in-the-moment mapping accuracy and retention in the same child (Bion et al., 2013; Kucker et al., 2018), the lack of relationship here may be due to the high variability even within children and limited number of trials in each task – e.g., to minimize the chance of fatigue, Direct Naming included only four new words/test trials. Nonetheless, when correlations were separately examined for LT vs. NL children, there was some indication that different relationships may exist for each group. The LT group's in-the-moment known word selection (Known RS) was predicted by their generalization ability (and vice versa), whereas NL children's Known RS was correlated with both RS Retention and NNG. The fact that Known RS and NNG are related in both LT and NL groups reinforces that the strength of current vocabulary (demonstrated by the ability to use that vocabulary in-the-moment) supports stronger lexical foundations, which supports generalization (see also Samuelson, 2002). The current results also found that LT children's RS Retention was predicted by Direct Naming and vice versa, which was not the case for NL children. This tentatively suggests that LT children may be using similar processes for both retention tasks, whereas for NL children, their retention is more likely supported by cascading effects of strong in-the-moment processes (given that RS Retention was predicted by Known RS in NL children). Taken together, these findings demonstrate that discrepancies in word learning abilities across children may lie in poor in-the-moment known word processing, and because any individual word does not exist in isolation (it is part of a larger lexical network), this can cascade to influence retention and generalization and, thus, possibly vocabulary growth more broadly.

Implications for LT children

The results suggest that as a group “late talkers” (and those with smaller vocabularies in general) may struggle most prominently with tasks that tap their (limited) vocabulary, as demonstrated by their significantly lower performance on Initial Known Comprehension and Known RS. On the surface, this is unsurprising because we know that, though a caregiver may report that a child says a word, it does not necessarily mean that the child has robust knowledge of that word. That fact is underscored here – caregivers reported that all children knew the known items used, but children still varied quite remarkably in their ability to bring such knowledge to bear in the lab setting. In fact, recent laboratory tests recognize this variability, showing that while, overall, children do recognize “known” items in the lab, there is a bias for specific familiar exemplars (Garrison et al., 2020), suggesting that representation of known items may still be weak in some ways. What the current work adds is that overall vocabulary size predicts the use of any individual word in a given moment. The smaller the vocabulary, the weaker the lexical network and, thus, the harder it is for children to use their vocabulary even for well-known words. This is particularly true for LT children here and consistent with prior work with LT children

(Ellis et al., 2015) and younger children with smaller vocabularies (Kucker et al., 2018). This also means that LT children may not be able to use the same mechanisms for learning new words as their NL peers. Because many of the known word learning constraints require a focus on vocabulary (e.g., mutual exclusivity, contrast), a lack of robust knowledge can impair use of these mechanisms. In fact, LT children here showed a robust novelty bias on Known RS trials, suggesting that novelty (not vocabulary) may be driving their in-the-moment behavior and selections, even on the Novel RS trials. Nevertheless, teasing the role of weak lexical knowledge apart from novelty remains an open question (see Mather, 2013).

In NL children, good in-the-moment processing tends to correlate with better retention, both in looking-based measures (Bion et al., 2013) and real item reaching measures (Kucker et al., 2018). However, this does not seem to be the case for the LT children here. While there are generally strong correlations across tasks at the group level, LT children showed stronger correlations between the in-the-moment tasks (Initial Known Comprehension and Known RS) and between the retention tasks (RS Retention and Direct Naming Retention) than they did across timescales. NL children's retention (in RS Retention at least) was predicted by their RS performance, however. This is consistent with prior work showing differences in known word processing between LT and NL children, but not overall retention (Ellis et al., 2015). Work with children with autism spectrum disorder also found delays in in-the-moment processing, but less so with retention and generalization (Hartley et al., 2019, 2020). Nonetheless, the processes have reciprocal effects – a smaller vocabulary may be a result of that child's inability to retain and integrate words into their lexicon in the first place.

The final task here tested children's generalization abilities. Overall, children showed a stronger shape bias as they got older, and there were no large differences between NL and LT children. This contrasts with prior work (Jones, 2003; Perry & Kucker, 2019) which found deficits in generalization for LT children. However, the children here spanned a wider age range than that in prior work and the majority of children were already demonstrating a shape bias. As discussed previously, the structure of one's vocabulary has increasingly been found to be a key predictor for generalization, especially in LT children (Beckage et al., 2011; Colunga & Sims, 2017; Perry & Kucker, 2019). Indeed, exploratory analyses found an interaction of shape vocabulary and the LT group on NNG performance here, replicating prior work (Perry & Kucker, 2019). Further exploration of vocabulary structure and deeper analysis of heterogeneity within the vocabulary abilities of LT children will be critical for future work.

Future steps and conclusions

The current study takes a first step toward examining variability in word learning across multiple timescales. It offers both a foundation highlighting the complex nature of word learning as well as incentive for future work further examining these processes. For one, LT children in the current sample were identified based on caregiver report of PRODUCTIVE vocabulary size, yet the tasks used here were COMPREHENSION based and did not require children to verbalize their response. As such, the current methods serve as a validity test that reported deficits in production extend to at least some observable comprehension deficits. Comprehension, however, does also require slightly different processes – robust semantic representations of the referents are required, for instance, but only weak phonological representations are needed as the labels are provided for the child. Thus,

differences between comprehension and production ought to be further examined, especially as they relate to potential phonological delays in LT children.

While the current study offers only a first step to assessing cascading effects, it does suggest that a lack of an ability to use current vocabulary in-the-moment may have cascading effects on other word learning processes. Critically, this means that targeting (and improving) in-the-moment processing of LT children may be fundamental for improving their language trajectories, and potentially have implications for interventions and the assessment of children with other language delays (e.g., Developmental Language Disorder). Long-term growth and longitudinal analyses are needed to fully quantify these effects, however.

Taking a more dynamic, pathways-based approach, in which different timescales as well as different individual abilities are included, is critical for a holistic picture of word learning. The results here suggest that fast and slow processes cannot be fully separated – performance on one task correlates with performance on others – but that the mechanisms driving behavior and therefore learning may shift as children grow their vocabularies. Such variability calls for more multi-dimensional approaches that vary the weight of in-the-moment context-dependent processing vs. memory processes vs. generalization abilities. To fully understand both normative development and also the nature of delays in LT children, multiple underlying word learning processes ought to be measured and considered.

Acknowledgments. Funding was provided by a Callier Center postdoctoral fellowship at the University of Texas Dallas to the first author. The authors would like to thank the former members of the Oshkosh Development Lab at the University of Wisconsin Oshkosh and the current members of the Language, Learning, and Development Lab at Oklahoma State for help with coding.

References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (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>
- Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE*, *6*(5), e19348. <https://doi.org/10.1371/journal.pone.0019348>
- Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*, *126*(1), 39–53. <https://doi.org/10.1016/j.cognition.2012.08.008>
- Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, *6*(3), 184–205. <https://doi.org/10.1080/15475441.2010.484379>
- Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*(02), 417. <https://doi.org/10.1017/S0305000900013842>
- Collisong, B. A., Grela, B., Spaulding, T., Rueckl, J. G., & Magnuson, J. S. (2015). Individual differences in the shape bias in preschool children with specific language impairment and typical language development: Theoretical and clinical implications. *Developmental Science*, *18*(3), 373–388. <https://doi.org/10.1111/desc.12219>
- Colunga, E., & Sims, C. E. (2017). Not only size matters: Early-talker and late-talker vocabularies support different word-learning biases in babies and networks. *Cognitive Science*, *41*, 73–95. <https://doi.org/10.1111/cogs.12409>
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 125–127. <https://doi.org/10.3758/BF03203646>
- 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. E., Evans, J. L., & Moyle, M. J. (2013). Fast mapping in late-talking toddlers. *Applied Psycholinguistics*, *34*(01), 69–89. <https://doi.org/10.1017/S0142716411000610>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), i. <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: Lexical Processing and vocabulary growth. *Child Development*, *83*(1), 203–222. <https://doi.org/10.1111/j.1467-8624.2011.01692.x>
- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. <https://doi.org/10.1111/1467-9280.00044>
- Garrison, H., Baudet, G., Breifeld, E., Aberman, A., & Bergelson, E. (2020). Familiarity plays a small role in noun comprehension at 12–18 months. *Infancy*, *25*(4), 458–477. <https://doi.org/10.1111/inf.12333>
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–34. doi: 10.1016/S0010-0277(02)00186-5. PMID: 12499109.
- Hartley, C., Bird, L.-A., & Monaghan, P. (2019). Investigating the relationship between fast mapping, retention, and generalisation of words in children with autism spectrum disorder and typical development. *Cognition*, *187*, 126–138. <https://doi.org/10.1016/j.cognition.2019.03.001>
- 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*, 104265. <https://doi.org/10.1016/j.cognition.2020.104265>
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. <https://doi.org/10.1080/15250000701795598>
- Jones, S. S. (2003). Late talkers show no shape bias in a novel name extension task. *Developmental Science*, *6*(5), 477–483.
- Jones, S. S., & Smith, L. B. (2005). Object name learning and object perception: A deficit in late talkers. *Journal of Child Language*, *32*(1), 223–240. <https://doi.org/10.1017/S0305000904006646>
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, *9*(2), 74–78. <https://doi.org/10.1111/cdep.12110>
- Kucker, S. C., McMurray, B., & Samuelson, L. K. (2018). Too much of a good thing: How novelty biases and vocabulary influence known and novel referent selection in 18-month-old children and associative learning models. *Cognitive Science*, *42*, 463–493. <https://doi.org/10.1111/cogs.12610>
- Kucker, S. C., & Samuelson, L. K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*, *17*(3), 295–323. <https://doi.org/10.1111/j.1532-7078.2011.00081.x>
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. [https://doi.org/10.1016/0885-2014\(88\)90014-7](https://doi.org/10.1016/0885-2014(88)90014-7)
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, *20*(2), 121–157. [https://doi.org/10.1016/0010-0285\(88\)90017-5](https://doi.org/10.1016/0010-0285(88)90017-5)
- MacRoy-Higgins, M., Shafer, V. L., Fahey, K. J., & Kaden, E. R. (2016). Vocabulary of toddlers who are late talkers. *Journal of Early Intervention*, *38*(2), 118–129. <https://doi.org/10.1177/1053815116637620>
- MacRoy-Higgins, M., & Montemarano, E. A. (2016). 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/neighborhood density on lexical learning in late talkers. *International Journal of Language & Communication Disorders*, *48*(2), 188–199.
- Mather, E. (2013). Novelty, attention, and challenges for developmental psychology. *Frontiers in Psychology*, *4*(491), 1–4. doi:10.3389/fpsyg.2013.00491
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. <https://doi.org/10.1037/a0029872>
- Mervis, C. B., & Bertrand, J. (1994). Acquisition of the Novel Name-Nameless Category (N3C) Principle. *Child Development*, *65*(6), 1646. <https://doi.org/10.2307/1131285>

- Oviatt, S. L. (1980). The emerging ability to comprehend language: An experimental approach. *Child Development*, 51(1), 97. <https://doi.org/10.2307/1129595>
- Perry, L. K., Axelsson, E. L., & Horst, J. S. (2016). Learning what to remember: Vocabulary knowledge and children's memory for object names and features. *Infant and Child Development*, 25(4), 247–258. <https://doi.org/10.1002/icd.1933>
- Perry, L. K., & Kucker, S. C. (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
- Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00345>
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, 21(12), 1894–1902. <https://doi.org/10.1177/0956797610389189>
- 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>
- Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, 38(6), 1016–1037. <https://doi.org/10.1037/0012-1649.38.6.1016>
- Samuelson, L. K., Schutte, A. R., & Horst, J. S. (2009). The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalization. *Cognition*, 110(3), 322–345. <https://doi.org/10.1016/j.cognition.2008.10.017>
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1–33. [https://doi.org/10.1016/S0010-0277\(99\)00034-7](https://doi.org/10.1016/S0010-0277(99)00034-7)
- Slone, L. K., & Sandhofer, C. M. (2017). Consider the category: The effect of spacing depends on individual learning histories. *Journal of Experimental Child Psychology*, 159, 34–49. <https://doi.org/10.1016/j.jecp.2017.01.010>
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1), 13–19. <https://doi.org/10.1111/1467-9280.00403>
- Swingle, D. (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development*, 6(3), 179–183. <https://doi.org/10.1080/15475441.2010.484412>
- Tenenbaum, E. J., Amso, D., Abar, B., & Sheinkopf, S. J. (2014). Attention and word learning in autistic, language delayed, and typically developing children. *Frontiers in Psychology*, 5, 490. <https://doi.org/10.3389/fpsyg.2014.00490>
- Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, 10(2), 201–224. [https://doi.org/10.1016/0885-2014\(95\)90009-8](https://doi.org/10.1016/0885-2014(95)90009-8)
- Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30(4), 553–566.

Cite this article: Kucker, S.C., & Seidler, E. (2023). The timescales of word learning in children with language delays: In-the-moment mapping, retention, and generalization. *Journal of Child Language* 50, 245–273, <https://doi.org/10.1017/S0305000921000817>