JOURNAL OF COGNITION AND DEVELOPMENT
2025, VOL. 26, NO. 1, 90–110
https://doi.org/10.1080/15248372.2024.2395515

Routledge
Taylor & Francis Group



# Information Sources for Word Meaning in Children's Picture Books: A Human Simulation Paradigm Study

Amanda Mankovich, Sadie MacDonald, Brianna Kinnie, Sara C. Johnson, and Sumarga H. Suanda

University of Connecticut

**ABSTRACT**

Picture book reading is widely regarded as an activity that promotes multiple aspects of children's language acquisition, including their vocabulary development. Historically, researchers interested in what underlies the link between picture book reading and vocabulary development have investigated a suite of caregiver behaviors during picture book reading that might support learning. More recently, with the advent of modern text analytic tools, researchers have begun to look at the properties of the picture books themselves and have revealed many ways in which the language in picture book text may support learning. The current study adds to this growing literature but focuses not only on how the language in picture books might support word learning but also on how its visual content supports learning. In an adaptation of the Human Simulation Paradigm applied to commonly read picture books to English-learning children, the current study involved adults ($N = 96$). Results showed that a word's linguistic and visual contexts in picture books are both valuable sources of information about that word's meaning, and that their relative contributions may vary as a function of the book's target audience and the word in question. Moreover, the combined visuo-linguistic context may be an especially valuable source of information for word learning. The implications of this work for the mechanisms by which picture books contribute to word learning are discussed.

Picture books have long been recognized as a valuable tool for young children's vocabulary development (Dickinson et al., 2012; Flack et al., 2018; Noble et al., 2019). A growing number of studies suggest that one candidate explanation for their prowess in cultivating learning is the rich language in their text (for review, see Nation et al., 2022). That is, several investigations of picture book texts have revealed that the language children are exposed to in picture books is both distinct from and richer than the language children are exposed to in everyday speech (e.g., see Cameron-Faulkner & Noble, 2013; Hsiao et al., 2023; Montag et al., 2015). For example, compared to child-directed speech, picture books expose children to a more diverse set of words (Montag et al., 2015), to words that are more morphologically advanced (Dawson et al., 2021), to sentence constructions that are more complex (Cameron-Faulkner & Noble, 2013), and to syntactic structures that are rare in child-directed speech (Hsiao et al., 2023; Montag, 2019). Importantly, many of the ways in

**CONTACT** Sumarga H. Suanda s.suanda@uconn.edu Department of Psychological Sciences, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020

Supplemental data for this article can be accessed online at https://doi.org/10.1080/15248372.2024.2395515

which picture book language is distinct from child-directed speech have been independently identified as accelerators of vocabulary acquisition (Hoff & Naigles, 2002; Rowe, 2012; Weizman & Snow, 2001), lending credence to the idea that a key pathway by which picture books and picture book reading support early word learning is the rich linguistic input it stimulates.

How exactly though does picture books' rich linguistic input facilitate the process of learning new word meanings? One possibility is that it does so through an indirect route. That is, picture books may expose children to new words, but it is their caregivers and teachers who then infuse those words with meaning. Indeed, studies have shown that vocabulary acquisition from shared book reading increases when caregivers and teachers provide explanations for new words (Barnes et al., 2017; Coyne et al., 2004; Elley, 1989), when they encourage children to actively process the text and illustrations (e.g., by asking them questions, see Ard & Beverly, 2004; Blewitt et al., 2009; Sénéchal et al., 1995), and when they recast new words using different sentence structures that help reiterate the meanings of those words (Demir-Lira et al., 2019; Pemberton & Watkins, 1987). A different line of research has also shown that caregivers and teachers may support word learning from picture books through non-verbal means (Murphy, 1978; Rohlfing et al., 2015; Zhang & Yu, 2022). For example, Zhang and Yu (2022) recently used head-mounted eye-tracking to investigate toddlers' moment-by-moment visual attention during shared book reading with their caregivers. They found that toddlers' attention was more likely to be directed to the referents of words if caregivers pointed to those referents as they labeled them. Thus, whether the exact behaviors are verbal, non-verbal, or a mixture of the two, there is good reason to believe in the indirect route from picture books to word learning via caregiver behavior. The current study seeks evidence for an alternative and more direct route by which picture books could support the acquisition of new word meanings: the linguistic and visual contexts of the picture books themselves.

## How picture book linguistic and visual contexts could support word Learning

A long list of computational and experimental analyses has revealed that much of a word's meaning can be gleaned from the linguistic contexts that surround it (see Lenci, 2018). In a classic study, Miller and Charles (1991) found that the more similar two words were in meaning (e.g., "car" and "automobile"), the more similar they were in sentential contexts (i.e., the sentences in which those words appeared; see also Charles, 2000). More modern computational analyses of large amounts of text have largely supported Miller and Charles's finding that there is a strong link between a word's meaning and its linguistic contexts (see Jones et al., 2017; Lupyan & Lewis, 2019). Interestingly, several recent findings have shown that one way in which picture book text is unique is the rich linguistic contexts that envelope their words. For example, in one recent comparison of picture book text to child-directed speech, Montag (2019) found that picture book text contains more sentences with the passive construction and more sentences containing relative clauses (see also Cameron-Faulkner & Noble, 2013; Hsiao et al., 2023). Although these data are more frequently discussed in terms of their contributions to children's grammatical development, many studies have revealed how rich syntactic contexts illuminate word meaning (for review, see Fisher et al., 2010, Naigles & Swensen, 2007). Thus, one possible direct path for how picture books support new word learning is not simply that picture books introduce children to

new words (Dawson et al., 2021; Mesmer, 2016; Montag et al., 2015), but also that picture books provide the appropriate linguistic contexts that help children interpret the meanings of those words (see also Robbins & Ehri, 1994; Sénéchal et al., 1995).

A different path by which picture books might provide children with direct access to a word's meaning is through the link between a word and its non-linguistic, visual contexts. This may be possible because picture books are multimodal forms of art that convey stories both textually and pictorially. Although the relationships between picture book text and illustrations have long been of interest within the literary analysis of children's picture books (e.g., see Nikolajeva & Scott, 2001; Nodelman, 1988; Sipe, 1998), this relationship has received much less attention within the literature on word learning via picture books (for some notable exceptions, see Dyer et al., 2000; Flack & Horst, 2018; Kümmerling-Meibauer & Meibauer, 2011). Interestingly, there has, however, been an increased interest in the visual environment for children's word learning outside of the picture book learning literature (Smith et al., 2018; Yu et al., 2021). That is, due in large part to technological advancements like head-mounted cameras and head-mounted eye-trackers (see Franchak & Yu, 2022 for a recent review), several studies have revealed that the visual features of referents, as viewed from the toddler learner's perspective, are important contributors to children's word learning (Clerkin & Smith, 2022; Slone et al., 2019; Yu & Smith, 2012). Thus, the extent to which picture book illustrations also depict a word's referent in a visually clear and transparent way could possibly support children's linking of words to their meanings.

Finally, picture books might offer direct access to a word's meaning through the combination of the word's linguistic and visual contexts. Interestingly, a number of children's literature scholars have argued that the relationship between a picture book's text and its illustrations is often one of interdependence, whereby critical narrative information relies on the interaction between text and illustration (Arizpe & Styles, 2003; Martinez & Harmon, 2012; Nikolajeva & Scott, 2001). For example, in one quantitative analysis of commonly-read picture books targeting young readers, Martinez and Harmon (2012) found that nearly a quarter of plot elements could only be deduced from a combination of pictorial and textual information (see also Wagner, 2013). One intriguing implication of this is the possibility that jointly processing the linguistic and visual contexts around a word may yield a superadditive effect. That is, joint processing might involve not only collecting the independent contributions of the linguistic and visual contexts but also taking in additional information that emerges from the interaction between the two contexts.

## The human simulation paradigm as a tool for assessing how different contexts support word learning

The broad goal of the current study is to assess the validity of these direct paths through which picture books support learning word meanings. That is, the current study examines the extent to which the sentential contexts around words and the illustrations that co-occur with words in picture books are informative of those word's meanings. To achieve this, we employ the Human Simulation Paradigm (HSP), a paradigm originally designed to assess the relative and combined contributions of different sources of information for word meaning from child-directed speech (Cartmill et al., 2013; Gillette et al., 1999; Snedeker & Gleitman, 2004). In most HSP studies, researchers

begin with audio-visual recordings of toddlers interacting with their caregivers. Researchers then identify within those recordings the moments in which caregivers uttered target words of interest (e.g., "ball," "toy," "throw," "play," etc.). To assess the informativity for word learning of the surrounding non-linguistic context, the researchers present naïve adult observers with shortened muted video vignettes depicting those utterance events. At the precise moments that caregivers uttered the words of interest within those vignettes, the adult observers hear a beep and must then guess the word uttered. To assess the informativity of the linguistic contexts surrounding words, the observers have to identify words solely from their linguistic contexts. For example, in one condition in Gillette et al. (1999) foundational HSP study, observers would be presented with caregiver utterances in which the target word of interest was replaced with a nonsense word (e.g., "Can you GORP Markie on the phone?" if the caregivers had said "Can you call Markie on the phone?"). The observers' task in this condition was to guess what the nonsense word represented.

The logic of the HSP is that the degree to which naïve observers can guess word identity from their non-linguistic and/or linguistic contexts speaks to the informativity of those contexts for identifying and acquiring word meaning. The HSP has been widely used to characterize the informativity of different contexts for different types of words (e.g., nouns, verbs; see Gillette et al., 1999; Kako, 2005; Snedeker & Gleitman, 2004), as well as to characterize the informativity of different contexts across languages (e.g., Fitch et al., 2021). Importantly, in one HSP study, Cartmill et al. (2013) found a correlation between the informativity of children's learning experiences (specifically their non-linguistic contexts) as assessed via the HSP and those children's later vocabulary development, suggesting that the HSP may measure not only the information that is available in the learning environment for children to use but also the information that children actually use for learning.

Considering that the current study's goal is to investigate the contributions of picture book text and illustrations to the acquisition of word meanings, it is worth noting that a number of previous HSP studies of child-directed speech have shown that different sources of information appear to play different roles for different types of words (see Gillette et al., 1999; Piccin & Waxman, 2007; Snedeker & Gleitman, 2004). Perhaps, the most well documented of these differences is that between nouns and verbs, where Gillette et al. (1999) showed that non-linguistic contexts are more informative for learning noun meanings and that linguistic contexts appear especially informative for learning verb meanings (see also Piccin & Waxman, 2007; Snedeker & Gleitman, 2004). An important follow-up to Gillette and colleagues' finding is that the division of labor between non-linguistic and linguistic sources of information does not simply fall along grammatical class lines. For example, Kako (2005) showed that the value of non-linguistic contexts for learning noun meanings may be restricted to nouns that denote object names. That is, Kako (2005) found that whereas non-linguistic information (i.e., the observational contexts in which nouns occurred) aided the identification of nouns that denote basic-level object categories (e.g., "ball," "hat;" what Kako referred to as *BLOC* nouns), which information was not particularly helpful for identifying nouns that did not denote basic-level object categories (e.g., "music," "tail;" or *non-BLOC* nouns). Just like Gillette and colleagues had found for verbs, Kako found that learning the meanings of non-BLOC nouns appeared to require access to their linguistic contexts (Kako, 2005).

## Current study

In the current study, we adapt the HSP to investigate the independent and joint informativity of the linguistic and visual contexts for word learning within picture books commonly read to English-learning children in North America. Similar to the HSP studies of child-directed speech described above, the informativity of the linguistic contexts of picture books was assessed by asking naïve English-speaking adult observers in the US and the UK to identify target words from the text that surrounds them. Informativity of the visual contexts, on the other hand, was assessed by asking observers to identify the same words solely from their pages' illustrations. The goal of this study was to address two research questions. First, what is the relative and combined informativity of the linguistic and visual contexts for word learning in commonly read picture books? Second, does this informativity change as a function of the picture book's target audience and as a function of the target word? Thus, the pages studied here include pages from both picture books targeting younger and older English-learning children. With regard to word type, although the current study focuses exclusively on nouns, following Kako (2005), it includes both BLOC nouns (e.g., "ball") and non-BLOC nouns (e.g., "day"). This variation allows us to explore whether the key sources of information for learning to differ across words in picture books, as they appear to do in child-directed speech (e.g., Gillette et al., 1999; Kako, 2005; Piccin & Waxman, 2007; Snedeker & Gleitman, 2004).

## Methods

### *Participants*

Participants were 96 adults 18–35 years of age (62 Females, 29 Males, 2 Nonbinary, 3 Other/Unknown). Participants were either US- or UK-based enrollees in the Prolific online participant pool (www.prolific.co) and received $5 for participation ($n = 74$) or were students enrolled in an introductory psychology course at a large public university in the northeastern United States and received course credit for participation ($n = 22$). The participants self-identified as Asian ($n = 18$), Black ($n = 4$), White ($n = 71$), or more than one race ($n = 3$); six participants self-identified as Hispanic or Latino or Spanish Origin. Although all instructions were in written English, participants were not required to be native English speakers (37 participants reported speaking at least one additional language).

### *Materials and stimuli*

#### *Target nouns*

The stimuli for this study were picture book vignettes which were derived from pages in children's picture books where one of forty target nouns appeared in the page's text. The target nouns were chosen to mirror the types of nouns used in previous HSP studies of child-directed speech (see Gillette et al., 1999). As such, these nouns are frequent in speech directed to English-learning children (e.g., "bird," "cow," "day;" for the full list, see Appendix Table A), are typically acquired by English-learning children by 30 months of age (Frank et al., 2017), and vary in kind, with some referring to basic-level object categories

(BLOC[1]; e.g., "dog," "bag," or "duck;" see (Kako, 2005) and others referring to non-basic-level object categories (non-BLOC; e.g., "boy," "eye," or "day").[2]

### Source materials

The picture books used to create the vignettes were initially purchased to serve as props in semi-naturalistic observational studies of parent–toddler interactions. Most books were purchased due to their appearance on both classic (e.g., *Time* magazine's "100 Best Children's Books of All Time") and current bestseller lists (e.g., Barnes & Noble top sellers; as of 2019) of picture books targeting English-learning children. The books were either books that targeted infants and younger toddlers (age ranges within 0–5 years; "First Books") or books that targeted older toddlers and younger children (age range within 3–8 years; "Early Books").[3] A detailed list of the 69 unique books (34 First Books and 35 Early Books) used in the present study are listed as part of the Supplementary Materials for this paper hosted on OSF (https://osf.io/utv35/).

### Vignette selection and creation

For each of the 40 target nouns, four instances from four different picture books were selected. Although the four instances shared the same root morpheme, they may have differed in grammatical number (e.g., "bird" vs. "birds") or in diminutive form (e.g., "piggy" vs. "pig"). Two instances were chosen from First Books, and two instances were chosen from Early Books. Thus, in total, there were 160 unique noun events. For each event, three separate vignettes were created for the three experimental conditions: *Image* condition, *Text* condition, and *Combo* condition. For the Image condition, vignettes were created by scanning the picture book pages (see Figure 1(a)) and then editing the scenes (using the picture-formatting tools within *Microsoft PowerPoint*) so as to remove all of the text on the page (see Figure 1(b)). For the Text condition, vignettes were created by displaying all of the text from the page, with the target noun removed and replaced with an underline spanning five characters (see Figure 1(c)). In cases where the target noun appeared multiple times on the same page (14.38% of the vignettes), each instance was replaced with an underline. Finally, for the Combo condition, vignettes were created by appending the Text condition vignette to the Image condition vignette (see Figure 1(d)).

## Design & procedure

The 160 vignettes in each condition were divided into four experimental lists of 40 vignettes, one vignette per target noun. These lists (i.e., which noun events made up each list) were identical across the three conditions. Two of the four experimental lists were of vignettes from First Books and two of the four lists were of vignettes from Early Books. On average, any given book contributed 2.25 vignettes (ranging from one to six vignettes). No vignettes derived from the same book were of the same exact scene. Both the condition and

[1]Following Kako (2005), nouns were classified as BLOCs if they: (1) name whole objects, (2) name situationally-independent kinds, and (3) name objects similar in shape.

[2]The choice to focus on the BLOC vs. non-BLOC distinction over other distinctions (e.g., concreteness) reflects previous findings demonstrating that the BLOC-non-BLOC distinction better captures word identification trends in the Human Simulation Paradigm (see Kako, 2005).

[3]Target age range classifications were based on age ranges printed on the books and on Amazon and Barnes & Noble recommendations.

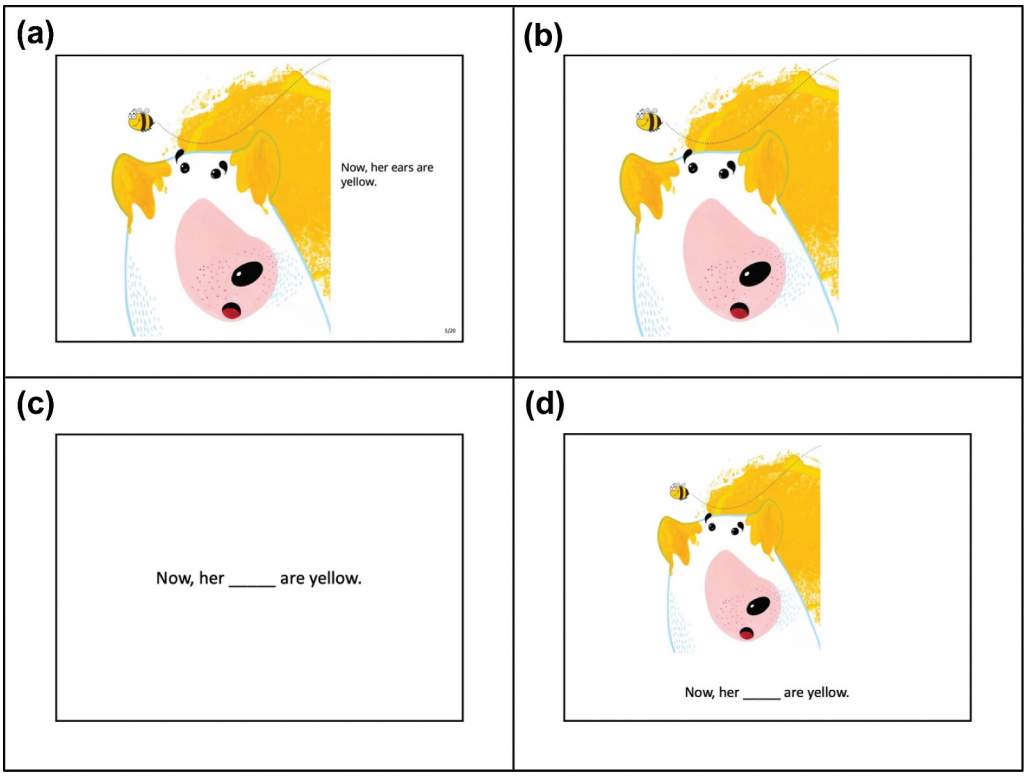


**Figure 1.** Illustrative stimulus for each condition. *Note*. Due to copyright issues, the actual stimuli used in the current study are unable to be depicted. This scene, which comes from "Colours" (Priyanka et al., 2015) is released under CC BY 4.0 license, is not a stimulus that was actually used and is thus for illustrative purposes only (depicted are illustrative vignettes for the noun "ear"): (a) Raw-scanned page taken from the picture book; (b) Doctored illustration with all text removed (*Image* condition); (c) All of the page's text without the mystery noun (*Text* condition); (d) Doctored illustration paired with the page's text without the mystery noun (*Combo* condition).

experimental list were between-subjects variables. Thus, the 96 participants were randomly assigned to and completed only one of the three conditions and one of the four experimental lists.

This study was built and hosted on the Gorilla Experiment Builder platform (www.gorilla.sc; Anwyl-Irvine et al., 2020). Participants accessed the study via an onward link from Prolific (for paid participants) or Sona Systems (for course credit participants). Prior to the study, participants consented to the study and filled out a brief demographic and language history questionnaire. Next, to acquaint participants with the task, participants completed four practice trials. Each practice trial started with the display of an Image, Text, or Combo vignette (depending on condition). After 3 seconds, a prompt appeared below the vignette asking participants to "Type in the noun you think fits best on this page" (Image condition) or "Type in the noun you think best fills in the blank" (Text and Combo condition). Upon completing each practice trial, participants were shown the correct noun alongside their guess. The practice vignette nouns were not one of the 40 target

nouns used in the experiment, and the practice vignettes were not derived from any of the 69 books used in the experiment. Immediately following the practice trials, participants completed the 40 experimental trials. The experimental trials were identical in structure to the practice trials with the exception that participants were not provided with the correct answer. The entire session lasted approximately 20–30 minutes in duration. Video screencasts depicting the study can be viewed on the paper's OSF link (https://osf.io/utv35/).

### Data scoring

Two trained coders independently evaluated participant guesses based on the scoring criteria used in previous HSP experiments (see Gillette et al., 1999).[4] Guesses were coded as correct if: (1) they were an exact match to the target noun, (2) they shared the same root morpheme as the target noun but differed in grammatical number (e.g., "cars" vs. "car"), (3) they shared the same root morpheme as the target noun but included (or dropped) a diminutive suffix (e.g., "doggy" vs. "dog"), (4) they were the clipped (or unclipped) form of the target noun (e.g., "phone" vs. "telephone"), or (5) they were a single-letter typo that did not form an alternative English word (e.g., "scjool" instead of the target noun "school"). Guesses were coded as incorrect if they did not share the target noun's root form in any of the ways just outlined, or if they included a modifying word before or after the target noun (e.g., "sandcastle" instead of the target noun "sand"). Note that guesses that likely referred to the same object as the target noun but differed in the root morpheme (e.g., "puppy" instead of "dog") were scored as incorrect. In the rare case that participants offered multiple answers where one of the answers was correct, we scored those cases as incorrect.[5] The two trained coders agreed on 99.43% of the trials (Cohen's Kappa = 0.99). All disagreements were resolved by a third trained coder.

### Data analysis

Given the categorical nature of these data (i.e., most participant guesses were scored as either correct or incorrect), standard continuous analyses like *t*-tests and ANOVAs were inappropriate (Jaeger, 2008). Thus, for all statistical analyses, we employed mixed-effects logistic regression models (i.e., Generalized Linear Mixed Models, or GLMM, for binomially distributed outcomes; Baayen et al., 2008). All analyses were executed in R version 4.3.1 (R Core Team, 2020), utilizing the lme4 package version 1.1–34 (Bates et al., 2020). Although we report some statistics in support of the key model statistics in the text body, the full specifications and parameter estimates of each model are described in Appendix B. Where means and standard deviations are reported, these reflect the mean of subject-level proportions and the standard deviations around those subject-level means. The raw data that support the findings of this study are openly available on the study's OSF site (https://osf.io/utv35/).

[4]See Gillette et al. (1999) for a discussion of the challenges associated with alternative HSP scoring methods. Although such methods might yield different absolute estimates of informativity, there is no reason to believe such methods would yield different trends across conditions.

[5]The total percentage of responses that included multiple answers (regardless of whether some were correct) was less than 2%.

## Results

### *The informativity of the linguistic and visual context in picture books*

Our first analyses assessed and compared the individual contributions of a word's visual and linguistic context for word identification from picture books. To do this, we first reduced our larger trial-level dataset to only data in the Image and Text conditions. We then asked whether adding the fixed effect of condition (Image vs. Text) into a GLMM built to predict participants' trial-level accuracy would improve model fit above and beyond a baseline model that consisted solely of random effects (see Appendix B for full details on the random effects structures and for model summary outputs). Results revealed that adding condition to the model did indeed improve model fit, $\chi^2(1) = 58.28$, $p < 0.001$, highlighting how participants in the Text condition ($M = .35$, $SD = .10$) performed reliably better than participants in the Image condition ($M = .15$, $SD = .05$; see Figure 2(a)).

We next asked whether this text advantage varied as a function of target audience and word type. To do this, we built upon the model above (which only included the main effect of the condition), by first adding fixed effects for the Target Audience (i.e., First vs. Early books) and Word Type (i.e., BLOC vs. non-BLOC), and then by adding interaction terms (e.g., the interaction between Condition and Target Audience, etc.). We kept additions to the model only when they significantly improved the model fit (indexed by likelihood ratio tests using by $\chi^2$). Results revealed a significant main effect of Word Type, $\chi^2(1) = 13.01$, $p < 0.001$, highlighting better performance for BLOC words ($M = .33$, $SD = .13$) than for non-BLOC words ($M = .20$, $SD = .19$), but no significant main effect of Target Audience, $p = 0.86$. Additionally, analyses revealed significant two-way interactions between Condition and Word Type, $\chi^2(1) = 209.70$, $p < 0.001$, and between Condition and Target Audience, $\chi^2(1) = 8.54$, $p < 0.01$. The Condition-by-Word Type interaction reflects the fact that the greater informativity of the linguistic over visual contexts was largely driven by the non-BLOC words (BLOC Words: $M_{text} = .31$, $SD_{text} = .14$, $M_{image} = .35$, $SD_{image} = .12$; non-BLOC Words: $M_{text} = .37$, $SD_{text} = .11$, $M_{image} = .04$, $SD_{image} = .04$; see Figure 2(b)). The Condition-by-Target Audience, on the other hand, reflects the fact that although there is greater informativity in the Text over Image conditions regardless of target audience (First Books: $M_{text} = .32$, $SD_{text} = .09$, $M_{image} = .18$, $SD_{image} = .04$; Early Books: $M_{text} = .37$, $SD_{text} = .11$, $M_{image} = .13$, $SD_{image} = .05$), the difference appears greater in the Early Books over First Books (see Figure 2a). Finally,

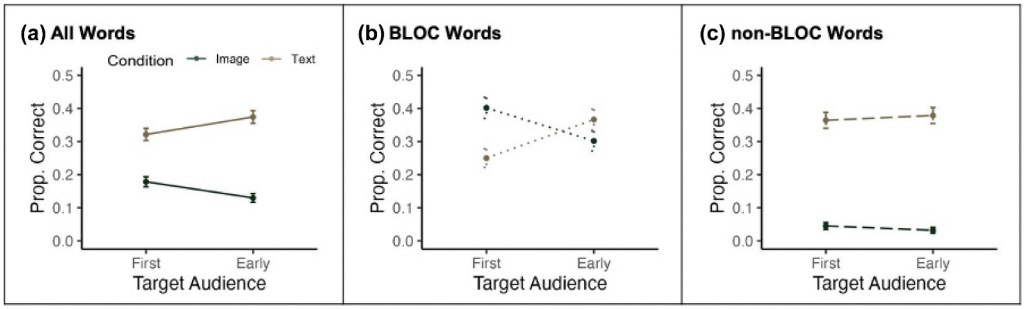


**Figure 2.** Performance across image and text conditions as a function of word type and target audience.
*Note*. Mean proportion correct by target audience in the Image and Text conditions for (a) All words; (b) BLOC words; (c) Non-BLOC words. Error bars represent the standard errors of participant-level means.

although a visual inspection of the trends (see Figure 2b,c) appear to suggest a three-way interaction between Condition, Target Audience, and Word Type, the model did not reveal a statistically significant three-way interaction, $p = 0.25$.

### The informativity of the combined linguistic and visual contexts in picture Books

We next examined how the linguistic and visual contexts *jointly* inform word meaning. We first explored this issue by examining how participants' ability to identify word meaning improved when they received a combination of linguistic and visual contexts relative to when they received each context in isolation. Following the same procedures described above, we constructed one GLMM to assess the difference between performance in the Combo condition to performance in the Text condition and one GLMM to assess the difference between performance in the Combo condition to performance in the Image condition.

In the model comparing the Combo and Text conditions, adding a fixed effect of Condition (Combo vs. Text) significantly improved model fit, $\chi^2(1) = 52.21$, $p < 0.001$, highlighting that the participants in the Combo condition ($M = .58$, $SD = .11$) were better at identifying words than participants in the Text condition. Both the interaction terms of Condition-by-Target Audience, $\chi^2(1) = 6.36$, $p < 0.05$, and Condition-by-Word Type, $\chi^2(1) = 56.32$, $p < 0.001$, also improved the model fit above and beyond a model with only the fixed effect of the condition. The Condition-by-Target Audience interaction suggests that the Combo condition's advantage over the Text condition was stronger for First Books ($M_{Combo} = .62$, $SD_{Combo} = .07$; $M_{Text} = .32$, $SD_{Text} = .09$) than Early Books ($M_{Combo} = .54$, $SD_{Combo} = .13$; $M_{Text} = .37$, $SD_{Text} = .11$; see Figure 3(a)). The Condition-by-Word Type interaction, on the other hand, highlights that while the Combo condition's advantage over the Text condition was evident for both BLOC ($M_{Combo} = .70$, $SD_{Combo} = .13$; $M_{Text} = .31$, $SD_{Text} = .14$; see Figure 3(b)) and non-BLOC words ($M_{Combo} = .50$, $SD_{Combo} = .12$; $M_{Text} = .37$, $SD_{Text} = .11$; see Figure 3(c)), it was more prominent for BLOC words. The three-way interaction between Condition, Target Audience, and Word Type did not improve model fit, $p = 0.39$.

In a comparable model comparing the Combo and Image conditions, adding a fixed effect of Condition (Combo vs. Image) significantly improved model fit, $\chi^2(1) = 134.47$, $p < 0.001$, suggesting that Combo condition participants were also better at identifying words than Image condition participants (Figure 3(d)). One interaction term, between Condition and Word Type, also significantly improved model fit, $\chi^2(1) = 67.83$, $p < 0.001$. This interaction highlights that the difference in performance between Combo and Image conditions was greater for non-BLOC words ($M_{Combo} = .50$, $SD_{Combo} = .12$; $M_{Text} = .04$, $SD_{Text} = .03$) than BLOC words ($M_{Combo} = .70$, $SD_{Combo} = .13$; $M_{Text} = .35$, $SD_{Text} = .12$; see Figure 3(e,f). Neither the two-way interaction between Condition and Target Audience, $p = 0.61$, nor the three-way interaction between Condition, Target Audience, and Word Type, $p = 0.64$, improved model fit.

### A superadditive effect from the combined linguistic and visual contexts?

In addition to comparing the Combo condition performance to the *individual* performances of the Image and Text conditions, we also compared the performance of the Combo condition to the *summed* performance of the two single-modality conditions. The

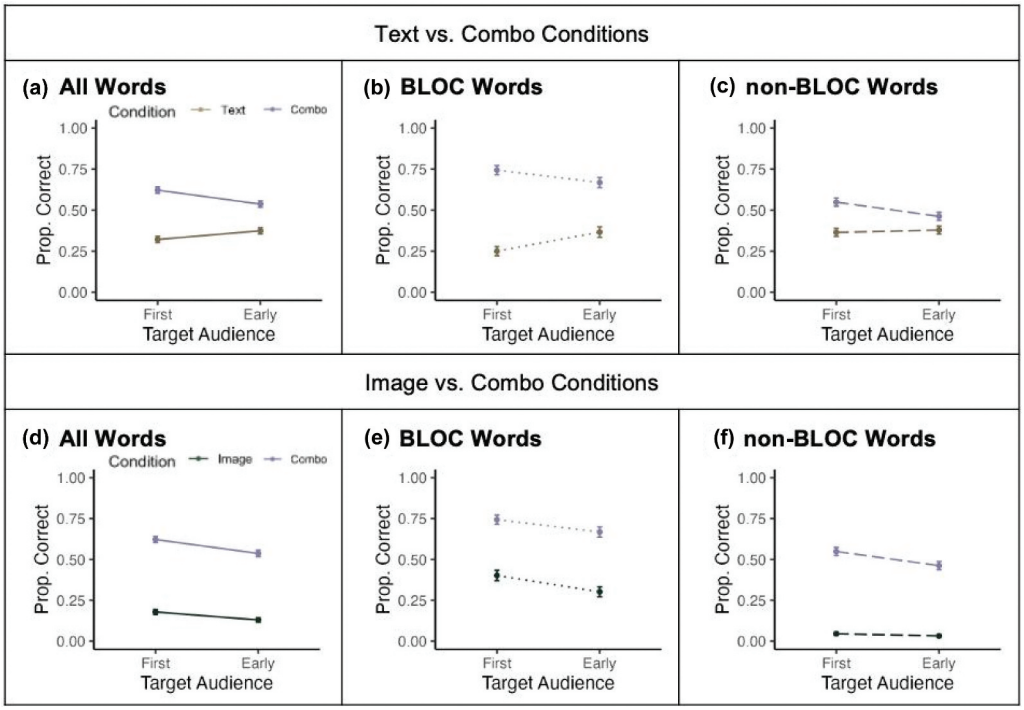


**Figure 3.** Pairwise comparisons between performance in the text and Combo Conditions and between the performance in the image and Combo Conditions. *Note*. Mean proportion correct responses by target audience for (a-c) All words, BLOC words, and non-BLOC words in the Combo vs. Text conditions; and for (d-f) All words, BLOC words, and non-BLOC words in the Combo vs. Image conditions; Error bars represent the standard errors of participant-level means.

goal of this analysis was to ask whether the holistic information that emerges from the combined linguistic and visual contexts is more informative than the summed information from each individual context. To do this analysis, we first created chimeric participants by matching participants from the Image condition with participants from the Text condition (from here on, we will refer to these chimeric participants as "summed participants" or participants in the "Summed condition"). For each individual trial (i.e., scene), we scored the summed participants' responses as follows. If either the Image participant or Text participant that made up the summed participant identified the word correctly, the summed participant's response was scored as correct. If neither the Image participant nor the Text participant identified the word correctly, the summed participant's response was scored as incorrect. Utilizing similar modeling building techniques as described above, we then asked whether participants in the Combo condition were more likely to identify word meanings than the participants in this newly constructed Summed condition.

Figure 4 depicts the performance of the Combo and Summed participants. As Figure 4a illustrates, participants in the Combo condition ($M = .58$, $SD = .11$) were more likely to identify word meanings than participants in the Summed condition ($M = .46$, $SD = .09$). This visual interpretation was supported in a GLMM whereby adding a fixed effect of Condition (Summed vs. Combo) significantly improved model fit, $\chi^2(1) = 22.62$, $p < 0.001$. Following a similar iterative model technique as described above, the interaction term

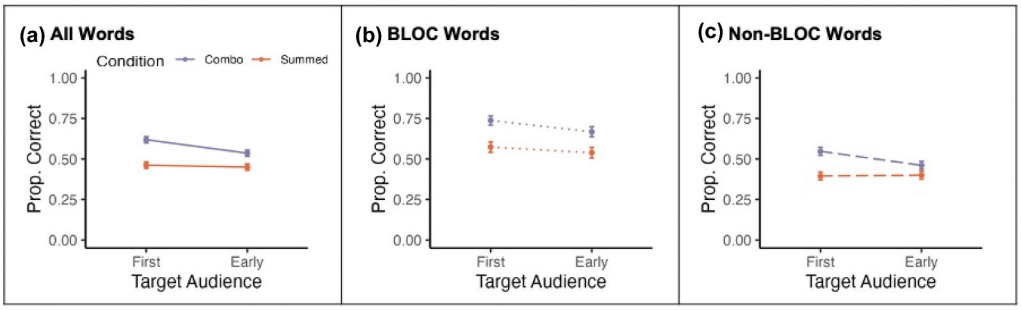


**Figure 4.** Performance across combo and summed conditions as a function of word type and target audience. *Note*. Mean proportion correct responses across the Combo and Summed conditions for (a) All words; (b) BLOC words; and (c) Non-BLOC words. The error bars indicate the standard errors of participant-level means.

between Condition (Summed vs. Combo) and Target Audience (First vs. Early) marginally improved model fit, $\chi^2(1) = 3.27$, $p = 0.07$. This interaction term illustrates a trend whereby the advantage of the Combo condition over Summed condition was stronger in First Books ($M_{combo} = .62$, $SD_{combo} = .07$; $M_{summed} = .46$, $SD_{summed} = 0.10$) over Early Books ($M_{combo} = .54$, $SD_{combo} = .13$; $M_{summed} = .45$, $SD_{summed} = .10$). Neither the two-way interaction between Condition and Word Type, $p = 0.22$, nor the three-way interaction between Condition, Target Audience, and Word Type see Figure 4(b,c), $p = 0.13$, improved model fit.[6]

## General discussion

A rich body of correlational and experimental research has revealed the great value that shared picture book reading has for children's early language development (for reviews see Dickinson et al., 2012; Flack et al., 2018; Noble et al., 2019). The current study builds on a growing body of recent work (for reviews see Nation et al., 2022, Strouse et al., 2018) that has sought to characterize the features of children's picture books as a window into understanding why they are an invaluable tool for boosting language development. By analyzing how the text and illustrations in picture books commonly read to English-learning children facilitate the identification of meaning of early-learned English nouns, the current study yielded three key findings. First, the text that surrounds nouns in picture books is a particularly powerful source of information for acquiring the meanings of those nouns, and one that is generally more potent than the illustrations that co-occur with those nouns. Second, this broad-stroke advantage of picture book text over picture book illustrations is qualified by interactions with both the book's target audience and the type of noun, suggesting a more complex picture of where the information lies for learning noun meanings from picture books. Finally, the combined signal from a picture book's text and its illustrations is not only more informative than the signal from each individual source, it also appears more informative than the sum of those two sources. This suggests that an important source of a word's meaning in the current corpus of picture books may also come from the emergent information that arises when text and illustrations are processed

[6]Permutation analyses were conducted to test the robustness of the Combo vs. Summed condition comparisons and are reported in Appendix C.

simultaneously. Together, these findings highlight some candidate pathways by which picture books support early word learning and offer some insights into future research that can further deepen our understanding of children's picture books and their role in learning.

### *Linguistic context as a source of information for word meaning in picture books*

The current finding that the linguistic context in this corpus of picture books is a rich information source for acquiring word meanings is consistent with several recent analyses documenting that picture book texts offer a rich linguistic experience, and that the linguistic environment picture book reading cultivates may be responsible for its prowess in promoting word learning (Cameron-Faulkner & Noble, 2013; Dawson et al., 2021; Hsiao et al., 2023; Mesmer, 2016; Montag, 2019; Montag et al., 2015). The current study's contribution to this growing body of evidence is that it reveals one way by which that rich linguistic environment could support the identification of individual word meanings. That is, whereas previous analyses have revealed in many ways that picture book text is both linguistically rich and unique (see Nation et al., 2022 for review), the precise ways by which that text would translate to word learning were not examined. After all, there are many ways that a linguistically rich and unique experience could lead to learning. Some of these ways are rather indirect in nature, where picture book text may expose children to new word forms, but the process of linking those forms to meaning happens either outside of the picture book reading context or through verbal and non-verbal support from caregivers (e.g., see Demir-Lira et al., 2019). Although the current study did not assess children's word learning directly (an issue we return to in the *Limitations and Future Directions* below), it provides proof of concept that picture book text may not only expose children to a rich set of words but it also provides the linguistic context to learn those words' meanings. Thus, just as the linguistic context that envelopes a new word in a book's passage can yield learning for older children and adults (see Joseph & Nation, 2018; Swanborn & Glopper, 1999), so too might the linguistic context in picture books for younger learners (see also Robbins & Ehri, 1994). Importantly, the capacity to identify and learn new word meanings via their linguistic context is well within the repertoire of young word learners (for reviews, see Bloom, 1994; Naigles & Swensen, 2007; for evidence specific to noun learning, see ; Fisher et al., 2010).

Although the current work highlights that the linguistic contexts surrounding words in this corpus of picture books are informative about their word meanings, the precise aspect of the linguistic contexts that spotlight their meanings is unclear. A likely possibility is that these linguistic context effects are driven by the meanings of the individual words in that context. For example, correct identification of the words "chair" and "door" in sentences like " . . . he sat in his _____ . . . " and " . . . she turned out the light and closed the _____. . ." is likely driven by understanding the meanings of the verbs "sit" and "close." The linguistic context effects may also; however, be driven by the syntactic characteristics of different kinds of nouns (see Bloom, 1994; Samuelson & Smith, 1999; Srinivasan & Barner, 2016). For example, nouns that denote countable things (e.g., book, chair, dog) can both take the plural form and be modified by numerals. In contrast, nouns that denote uncountable substances (e.g., sand) can do neither. Thus, the syntactic frames in which the current target nouns are embedded may also signal something about the noun's meaning. Future work that distills the wholistic linguistic contexts examined in the current study into their component parts

(for examples of this approach, see Gillette et al., 1999; Snedeker & Gleitman, 2004) may be able to further isolate where picture book's linguistic informativity lie and thus further delineate how picture book text supports word learning.

### *More than words: the contributions of picture book illustrations to word meaning identification*

Although the findings from the current study undoubtedly reinforce the message that picture books can offer a rich linguistic input (e.g., Nation et al., 2022), they also highlight that a picture book's text is not the only feature that might explain its potency for supporting word learning. For basic-level object categories (e.g., "book," "chair," "dog"), the co-occurring visual illustrations appear equally informative for identifying word meaning and possibly more informative in picture books geared toward the youngest listeners and readers. In fact, the observed informativity of the visual illustrations for BLOC nouns in the current corpus of First Books (40%) is quite a bit higher than the informativity of the extralinguistic contexts observed in studies of child-directed speech (22%; see Cartmill et al., 2013). These data complement recent experimental studies demonstrating the importance of picture book illustrations for children's learning of novel object names (see Flack & Horst, 2018; Strouse et al., 2018). Moreover, the findings that source informativity varies as a function of noun type (i.e., BLOC nouns vs. non-BLOC nouns) are consistent with previous HSP studies of child-directed speech revealing that different sources of information (i.e., observational, linguistic) support the acquisition of different types of words (e.g., see Gillette et al., 1999; Kako, 2005). Interestingly, in one of the few analyses of how the illustrations in commonly read picture books to English-learning children may support their acquisition of word meanings, Dyer et al. (2000) found rather anemic levels of informativity in those illustrations. Importantly, however, the words of focus in Dyer and colleagues' work were mental state terms (e.g., "feel," "happy," "know," "think"), a class of words whose meanings are notoriously opaque from their observational contexts (e.g., see Snedeker & Gleitman, 2004). Building upon the recent and growing body of non-picture book research on the importance of children's visuo-referential world for learning nouns that denote object names (Clerkin & Smith, 2022; Slone et al., 2019; Yu & Smith, 2012), the current study highlights that the illustrations in picture books should not be neglected as a candidate explanation for their role in learning and development (see also Kuwabara et al., 2020).

The current findings are also revealing of the ways in which the visual information in picture books could work in conjunction with its text to constrain word learning. Thus, the key to understanding how picture books support learning may lie not only in the individual modalities but also in the ways those two modalities connect. That is, when the current study's participants processed the linguistic and visual contexts together, they were not only better at identifying meanings than participants who processed the linguistic or visual contexts in isolation but also better than the *summed performance* of those participants. Exactly what underlies these superadditive effects is unclear. One possibility is that these findings are an epiphenomenon of the fact that picture book authors and illustrators, in an effort to tell their stories in a limited space, will frequently convey complementary, rather than redundant, information, in the text and illustrations (see Nikolajeva & Scott, 2001; Nodelman, 1988; Sipe, 1998). Thus, having access to both sources of information allows

learners to extract the deeper and true meaning of the scenes and stories in picture books, which are not transparent from any single source. It is extracting that deeper meaning that allows learners to process each source of information in ways they would not otherwise. A different possibility is that these findings are a by-product of a more general principle of communication (Tomasello, 2001; Yurovsky, 2018) and semantic knowledge (Elman, 2004; Lupyan & Lewis, 2019) that is not unique to the picture book context. For example, a number of cognitive and developmental scholars have argued that rather than words as mapping onto meanings, words, along with other non-linguistic processes (e.g., sensory, perception), cue meaning (e.g., Elman, 2004). One consequence of this perspective, is that the nature of linguistic information depends heavily on their non-linguistic contexts, and vice versa (for a recent discussion, see Lupyan & Lewis, 2019). Applying this perspective to the current study, the sentential contexts and scenes that co-occur with words in picture books are mutually constraining in precisely the same way that linguistic and non-linguistic experiences mutually constrain in everyday conversations. This mutually constraining process, which has been noted by several literary scholars of children's picture books (e.g., see Nodelman, 1988), may be especially important for young children whose linguistic capabilities are still developing.

### *Developmental tuning of sources of meaning in picture books*

Many of the current findings revealing the contributions of visual context for word meaning identification appear to be more prominent in books targeting the youngest of audiences. For example, the advantage of the linguistic over visual contexts is smaller in books for younger learners. Additionally, the advantage of processing the combined linguistic and visual contexts over processing only the linguistic context diminishes with the increased target audience age. Finally, the superadditive effects of linguistic and visual contexts also appear to diminish with increased target audience age. Interestingly, developmental studies of early word learning to have revealed that whereas infants' and toddlers' learning to appear to rely more on visuo-perceptual properties of referents, older toddlers' and young children's learning appears to rely more on social and linguistic cues (e.g., Hollich et al., 2000). It is highly unlikely that picture book authors are motivated by creating the ideal word learning environments for young children. Instead, the confluence of developmental trends observed in the current study likely reflects the increased sophistication of the narratives in picture books geared toward older audiences and the greater reliance on language to convey those narratives (e.g., Martinez & Harmon, 2012). Nonetheless, what the current findings highlight is that a by-product of those motivations is a learning environment that is tuned to the word learning requirements of the audience.

### *Limitations and future directions*

A few limitations to the current work are worth noting. First, some of the trends observed in the current study may have been impacted by certain methodological decisions. For example, the current study measured the informativity of illustrations and text within a single page. Given that narratives are often depicted across multiple pages, and that children are exposed to sequences of pages, the current work may underestimate the actual level of informativity of picture book text and/or illustrations. Additionally, the current

study represents only a sampling of the nouns present in these books, only a sampling of the books to which English-learning children are exposed (see Hudson Kam & Matthewson, 2017), and focuses exclusively on picture books popular among English-learning audiences in North America. Thus, future work that expands the current study to a wider range of nouns, books, cultures, and languages, would be invaluable to test the generalizability and robustness of the current findings. Second, the current study's approach to measuring contextual informativity was to follow prior studies and use adults as "simulations" of the word learning process (e.g., see Gillette et al., 1999). Although prior studies have revealed that children's performance in the HSP mimics that of adults (e.g., Piccin & Waxman, 2007) and that informativity estimates derived from adult ratings are correlated with children's learning outcomes (Cartmill et al., 2013), it is still important to explicitly test whether these findings extend to developmental populations. One possibility, for example, is that due to the linguistic maturity of our adult participants, we may have overestimated the overall linguistic advantage observed in the current study.

Two final and broad limitations of the current work are worth mentioning. First, this study follows other recent research that has focused on characterizing the features of picture books as a window into the learning experience that picture book reading and listening offers (for review, see Nation et al., 2022). It is important to acknowledge that the information in the books themselves is only one component of that experience, and that the adult reader, as well as the child listener, also shape that experience (e.g., Fletcher & Reese, 2005; Mol et al., 2009). Thus, the most relevant learning experience is , ultimately, some blend of the book, the adult, and the child (for discussion, see Mesmer, 2016; Montag, 2019; Read et al., 2023). Future work that combines the current analyses of picture book features with analyses of the real-time dynamics of parent–child interactions will be helpful not only in revealing more concretely whether and how the current study misestimates informativity but also in revealing how parents augment, complement, and supplement the informativity offered directly by the picture books themselves. Second, the current study was focused on the potential benefits of picture book illustrations for children's word learning. It is worth noting that other research suggests that book illustrations can have detrimental effects on other kinds of tasks, including learning to read (e.g., Willows, 1978), information integration (Pike et al., 2010), and story comprehension (Eng et al., 2020). Although many of the detrimental effects have not gone undisputed (e.g., see Greenhoot et al., 2014; Kennedy & Cariveau, 2024), they are an important reminder that picture books serve a multitude of functions and are deployed in a multitude of contexts in children's lives (see Breitfeld et al., 2021; Grøver et al., 2023). As a result, the ways in which picture books and their illustrations influence learning may well depend on the function and context in question.

## Conclusions

Picture books communicate memorable stories through a tailored set of words and illustrations. Understanding why picture book reading and listening are reliably linked to optimal language outcomes requires an understanding of how their words and illustrations independently and jointly inform word meaning. By adapting a paradigm originally designed to understand the contributions of linguistic and extralinguistic input for word learning from child-directed speech and applying it to picture books commonly read to English-learning

children, the current study represents one step toward characterizing the informativity of words and illustrations for learning within picture books. Because vocabulary development, in general, involves accumulating information across multiple sources of information, the hope is that by understanding the ways in which learners assemble meaning from the different sources of information within picture books, we will also shed light on the general principles that apply to word learning more broadly.

## Acknowledgments

We thank Audra Logan, Zoe Green, Katharine Ustorf, and Nicholas Willett for assistance with this project, the Communication and Development Laboratory, especially Miranda Long and Nina Schoener, for discussion and feedback on this project, and Gerry Altmann for discussion and feedback on the manuscript.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was supported, in part, by the James S. McDonnell Foundation [JSMF 220020549] and the National Institutes of Health [R00-HD082358].

## References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Ard, L. M., & Beverly, B. L. (2004). Preschool word learning during joint book reading: Effect of adult questions and comments. *Communication Disorders Quarterly*, *26*(1), 17–28. https://doi.org/10.1177/15257401040260010101

Arizpe, E., & Styles, M. (2003). *Children reading pictures: Interpreting visual texts*. Routledge Falmer.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barnes, E. M., Dickinson, D. K., & Grifenhagen, J. F. (2017). The role of teachers' comments during book reading in children's vocabulary growth. *Journal of Educational Research*, *110*(5), 515–527. https://doi.org/10.1080/00220671.2015.1134422

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., & Green, P. (2020). *Package 'lme4' (Version 1.1-26) [Computer software]*. Comprehensive R Archive Network. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, *101*(2), 294–304. https://doi.org/10.1037/a0013844

Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua*, *92*, 297–329. https://doi.org/10.1016/0024-3841(94)90345-X

Breitfeld, E., Potter, C. E., & Lew-Williams, C. (2021). Children simultaneously learn multiple dimensions of information during shared book reading. *Journal of Cognition and Development*, *22*(5), 744–766. https://doi.org/10.1080/15248372.2021.1939353

Cameron-Faulkner, T., & Noble, C. (2013). A comparison of book text and child directed speech. *First Language*, *33*(3), 268–279. https://doi.org/10.1177/0142723713487613

Cartmill, E. A., Armstrong, B. F., III, Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Charles, W. G. (2000). Contextual correlates of meaning. *Applied Psycholinguistics*, *21*(4), 505–524. https://doi.org/10.1017/S0142716400004057

Clerkin, E. M., & Smith, L. B. (2022). Real-world statistics at two timescales and a mechanism for infant learning of object names. *Proceedings of the National Academy of Sciences*, *119*(18), Article e2123239119. https://doi.org/10.1073/pnas.2123239119

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, *12*(3), 145–162. https://doi.org/10.1207/s15327035ex1203_3

Dawson, N., Hsiao, Y., Tan, A. W. M., Banerji, N., & Nation, K. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research*, *1*(1), 9–53. https://doi.org/10.34842/5we1-yk94

Demir-Lira, E., Applebaum, L. R., Goldin-Meadow, S., & Levine, S. C. (2019). Parents' early book reading to children: Relation to children's later language and literacy outcomes controlling for other parent language input. *Developmental Science*, *22*(3), e12764. https://doi.org/10.1111/desc.12764

Dickinson, D. K., Griffith, J. A., Golinkoff, R. M., & Hirsh-Pasek, K. (2012). How reading books fosters language development around the world. *Child Development Research*, *2012*, 1–15. https://doi.org/10.1155/2012/602807

Dyer, J. R., Shatz, M., & Wellman, H. M. (2000). Young children's storybooks as a source of mental state information. *Cognitive Development*, *15*(1), 17–37. https://doi.org/10.1016/S0885-2014(00)00017-4

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 174–187. https://doi.org/10.2307/747863

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306. https://doi.org/10.1016/j.tics.2004.05.003

Eng, C. M., Godwin, K. E., & Fisher, A. V. (2020). Keep it simple: Streamlining book illustrations improves attention and comprehension in beginning readers. *Science of Learning*, *5*, Article 14. https://doi.org/10.1038/s41539-020-00073-5

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews Cognitive Science*, *1*(2), 143–149. https://doi.org/10.1002/wcs.17

Fitch, A., Arunachalam, S., & Lieberman, A. M. (2021). Mapping word to world in ASL: Evidence from a human simulation paradigm. *Cognitive Science*, *45*(12), e13061. https://doi.org/10.1111/cogs.13061

Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology*, *54*(7), 1334–1346. https://doi.org/10.1037/dev0000512

Flack, Z. M., & Horst, J. S. (2018). Two sides to every story: Children learn words better from one storybook page at a time. *Infant and Child Development*, *27*(1), Article e2047. https://doi.org/10.1002/icd.2047

Fletcher, K. L., & Reese, E. (2005). Picture book reading with young children: A conceptual framework. *Developmental Review*, *25*(1), 64–103. https://doi.org/10.1016/j.dr.2004.08.009

Franchak, J. M., & Yu, C. (2022). Beyond screen time: Using head-mounted eye tracking to study natural behavior. *Advances in Child Development and Behavior*, *62*, 61–91. https://doi.org/10.1016/bs.acdb.2021.11.001

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Greenhoot, A. F., Beyer, A. M., & Curtis, J. (2014). More than pretty pictures? How illustrations affect parent-child story reading and children's story recall. *Frontiers in Psychology*, *5*, Article 738. https://doi.org/10.3389/fpsyg.2014.00738

Grøver, V., Snow, C. E., Evans, L., & Strømme, H. (2023). Overlooked advantages of interactive book reading in early childhood? A systematic review and research agenda. *Acta Psychologica*, *239*, Article 103997. https://doi.org/10.1016/j.actpsy.2023.103997

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

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., Hennon, E., Rocroi, C., & Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, i–135. http://www.jstor.org/stable/3181533

Hsiao, Y., Dawson, N. J., Banerji, N., & Nation, K. (2023). The nature and frequency of relative clauses in the language children hear and the language children read: A developmental cross-corpus analysis of English complex grammar. *Journal of Child Language*, *50*(3), 555–580. https://doi.org/10.1017/S0305000921000957

Hudson Kam, C. L., & Matthewson, L. (2017). Introducing the Infant Bookreading Database (IBDb). *Journal of Child Language*, *44*(6), 1289–1308. https://doi.org/10.1017/S0305000916000490

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. *Psychology of Learning and Motivation*, *67*, 239–283. https://doi.org/10.1016/bs.plm.2017.03.008

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Kako, E. (2005). Information sources for noun learning. *Cognitive Science*, *29*(2), 223–260. https://doi.org/10.1207/s15516709cog0000_11

Kennedy, T., & Cariveau, T. (2024). Picture-text compounds in early reading: A descriptive review. *Behavior & Social Issues*, *33*, 322–338. https://doi.org/10.1007/s42822-023-00139-0

Kümmerling-Meibauer, B., & Meibauer, J. (2011). Early-concept books: Acquiring nominal and verbal concepts. In B. Kümmerling-Meibauer (Ed.), *Emergent literacy: Children's books from 0 to 3* (pp. 91–114). John Benjamins Publishing Company.

Kuwabara, M., Alonso, J., & Ayala, D. (2020). Cultural differences in visual contents in picture books. *Frontiers in Psychology*, *11*, 304. https://doi.org/10.3389/fpsyg.2020.00304

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *34*(10), 1319–1337. https://doi.org/10.1080/23273798.2017.1404114

Martinez, M., & Harmon, J. M. (2012). Picture/Text relationships: An investigation of literary elements in picturebooks. *Literacy Research and Instruction*, *51*(4), 323–343. https://doi.org/10.1080/19388071.2012.695856

Mesmer, H. A. E. (2016). Text matters: Exploring the lexical reservoirs of books in preschool rooms. *Early Childhood Research Quarterly*, *34*, 67–77. https://doi.org/10.1016/j.ecresq.2015.09.001

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1–28. https://doi.org/10.1080/01690969108406936

Mol, S. E., Bus, A. G., & De Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*(2), 979–1007. https://doi.org/10.3102/0034654309332561

Montag, J. L. (2019). Differences in sentence complexity in the text of children's picture books and child-directed speech. *First Language*, *39*(5), 527–546. https://doi.org/10.1177/0142723719849996

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science*, *26*(9), 1489–1496. https://doi.org/10.1177/0956797615594361

Murphy, C. M. (1978). Pointing in the context of a shared activity. *Child Development*, 371–380. https://doi.org/10.2307/1128700

Naigles, L. R., & Swensen, L. D. (2007). Syntactic supports for word learning. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 212–231). Blackwell Publishing. https://doi.org/10.1002/9780470757833.ch11

Nation, K., Dawson, N. J., & Hsiao, Y. (2022). Book language and its implications for children's language, literacy, and development. *Current Directions in Psychological Science*, *31*(4), 375–380. https://doi.org/10.1177/09637214221103264

Nikolajeva, M., & Scott, C. (2001). *How picturebooks work*. Garland.

Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J. (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review*, *28*, Article 100290. https://doi.org/10.1016/j.edurev.2019.100290

Nodelman, P. (1988). *Words about pictures: The narrative art of children's picture books*. University of Georgia Press.

Pemberton, E. F., & Watkins, R. V. (1987). Language facilitation through stories: Recasting and modelling. *First Language*, *7*(19), 79–89. https://doi.org/10.1177/014272378700701905

Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, *3*(4), 295–323. https://doi.org/10.1080/15475440701377535

Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology*, *105*(3), 243–255. https://doi.org/10.1016/j.jecp.2009.10.006

Priyanka, Guhathakurta, A., & Berger, A. (2015). *Colours*. Room to Read.

R Core Team. (2020). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing. http://www.R-project.org/

Read, K., Rabinowitz, S., & Harrison, H. (2023). It's the talk that counts: A review of how the extra-textual talk of caregivers during shared book reading with young children has been categorized and measured. *Journal of Early Childhood Literacy*. https://doi.org/10.1177/14687984231202968

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, *86*(1), 54–64. https://doi.org/10.1037/0022-0663.86.1.54

Rohlfing, K., Grimminger, A., & Nachtigäller, K. (2015). Gesturing in joint bookreading. In B. Kümmerling-Meibauer, J. Meibauer, K. Nachtigäller, & K. J. Rohlfing (Eds.), *Learning from picturebooks* (pp. 99–116). Routledge.

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

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

Sénéchal, M., Thomas, E., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, *87*(2), 218. https://doi.org/10.1037/0022-0663.87.2.218

Sipe, L. R. (1998). How picture books work: A semiotically framed theory of text-picture relationships. *Children Literature in Education*, *29*, 97–108. https://doi.org/10.1023/A:1022459009182

Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-generated variability in object images predicts vocabulary growth. *Developmental Science*, *22*(6), Article e12816. https://doi.org/10.1111/desc.12816

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Snedeker, J., & Gleitman, L. R. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 257–293). MIT Press.

Srinivasan, M., & Barner, D. (2016). Encoding individuals in language using syntax, words, and pragmatic inference. *Wiley Interdisciplinary Reviews Cognitive Science*, *7*(5), 341–353. https://doi.org/10.1002/wcs.1396

Strouse, G. A., Nyhout, A., & Ganea, P. A. (2018). The role of book features in young children's transfer of information from picture books to real-world contexts. *Frontiers in Psychology*, *9*, Article 50. https://doi.org/10.3389/fpsyg.2018.00050

Swanborn, M. S. L., & Glopper, K. D. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*(3), 261–285.

Tomasello, M. (2001). Could we please lose the mapping metaphor, please? *Behavioral and Brain Sciences*, *24*(6), 1119–1120. https://doi.org/10.1017/S0140525X01390131

Wagner, L. (2013). By the numbers: A quantitative content analysis of children's picturebooks. *Frontiers in Psychology*, *4*, Article 850. https://doi.org/10.3389/fpsyg.2013.00850

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

Willows, D. M. (1978). A picture is not always worth a thousand words: Pictures as distractors in reading. *Journal of Educational Psychology*, *70*(2), 255–262. https://doi.org/10.1037/0022-0663.70.2.255

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

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, *118* (52), Article e2107019118. https://doi.org/10.1073/pnas.2107019118

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73–79. https://doi.org/10.1016/j.newideapsych.2017.09.001

Zhang, Y., & Yu, C. (2022). Examining real-time attention dynamics in parent-infant picture book reading. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society* (pp. 1367–1374). Cognitive Science Society, Toronto, Canada.