

# Book Language and Its Implications for Children's Language, Literacy, and Development

Kate Nation<sup>id</sup>, Nicola J. Dawson, and Yaling Hsiao

Department of Experimental Psychology, University of Oxford

Current Directions in Psychological Science

2022, Vol. 31(4) 375–380

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09637214221103264

www.psychologicalscience.org/CDPS



## Abstract

The onset of literacy marks a significant change in children's development. Written language is more complex than everyday conversation, and even books targeted at preschoolers contain more varied words and more complex syntax than child-directed speech does. We review the nature and content of children's book language, focusing on recent large-scale corpus analyses that systematically compared written and spoken language. We argue that exposure to book language provides opportunities for learning words and syntactic constructions that are only rarely encountered in speech and that, in turn, this rich experience drives further developments in language and literacy. Moreover, we speculate that the range, variety, depth, and sophistication of book language provide key input that promotes children's social and emotional development. Becoming literate changes things, and researchers need to better understand how and why reading experience shapes people's minds and becomes associated with a range of skills and abilities across the life span.

## Keywords

book language, reading, development, language, literacy

It is impossible to imagine the world without the written word. Written words fill people's daily lives, and even within the confines of laboratory experiments in psychological science, the visual word is probably the most ubiquitous stimulus. Written language is predicated on spoken language both in human evolutionary history (spoken language is biologically primary, whereas writing systems are cultural inventions) and in individuals' developmental history (children's progress in learning to read and write rests intimately on their underlying language prowess).

Although language is complex, it contains structure via a multitude of probabilistic relationships between multiple levels of correlated features (e.g., Seidenberg & MacDonald, 2018). Several theoretical accounts treat language development as the product of general learning mechanisms operating across this rich input (e.g., Ramscar, 2021). In this view, the transition to adultlike language comprehension and production is a gradual one, reflecting the accumulation of experience with sounds, words, sentences, and their contexts over a number of years.

The onset of literacy marks a significant change in children's language experience. Most obvious is the

acquisition of orthography—visual symbols that allow spoken language to be transported to and from the page. With instruction and practice, children learn how orthography connects with sounds and meaning, and the reading system comes to embody the distributional properties of their writing system (Castles et al., 2018). As Frith (1998) noted, “Language is never the same again” (p. 1011). The acquisition of an orthographic code changes how spoken language is processed, as evidenced by systematic differences between literate and nonliterate people across a range of nonreading tasks, including speech perception and production, and verbal working memory (for a review, see Dehaene et al., 2015).

Understandably enough, how children learn the orthographic code has been a central focus of research (Treiman, 2020). However, written language is different from spoken language in important ways (e.g., Biber, 1988), and learning to read provides children with access

## Corresponding Author:

Kate Nation, Department of Experimental Psychology, University of Oxford

Email: kate.nation@psy.ox.ac.uk

to these differences. What are the implications of this for children's language and literacy development?

## Writing and Speaking, Reading and Listening

Speakers and listeners work together to communicate in the moment, whereas written language is traditionally solitary and remote. Incomplete and ambiguous utterances are common in conversations but rarely trouble listeners for long. In the absence of a shared situation and shared cues such as facial expression, intonation, and gesture, written language has a difficult job to do—it has to work hard so that the intended meaning of the writer can be re-created in the mind of the reader. Writing is characterized by greater precision and increased syntactic complexity relative to speech (Roland et al., 2007). There are also differences in lexical richness (Hayes, 1998), as people use more varied words and more sophisticated vocabulary when writing than when speaking. Although factors such as formality and genre influence patterns of language use within each modality (and, as social media demonstrate, both written and spoken language continue to adapt and evolve), written language has greater linguistic variety at its disposal (Biber, 1988). Writers can choose from its repertoire to communicate nuance and complexity effectively.

As children progress through the education system, reading becomes an increasingly important vehicle for learning. To benefit from this, children need to become fluent in the language of the book. This framing resonates with the concept of academic language (e.g., Phillips Galloway et al., 2020). As the name suggests, research on academic language tends to be considered in the context of formal education extending through the middle- and high-school years. However, opportunities to learn about book language begin much earlier.

## Early Exposure to Book Language

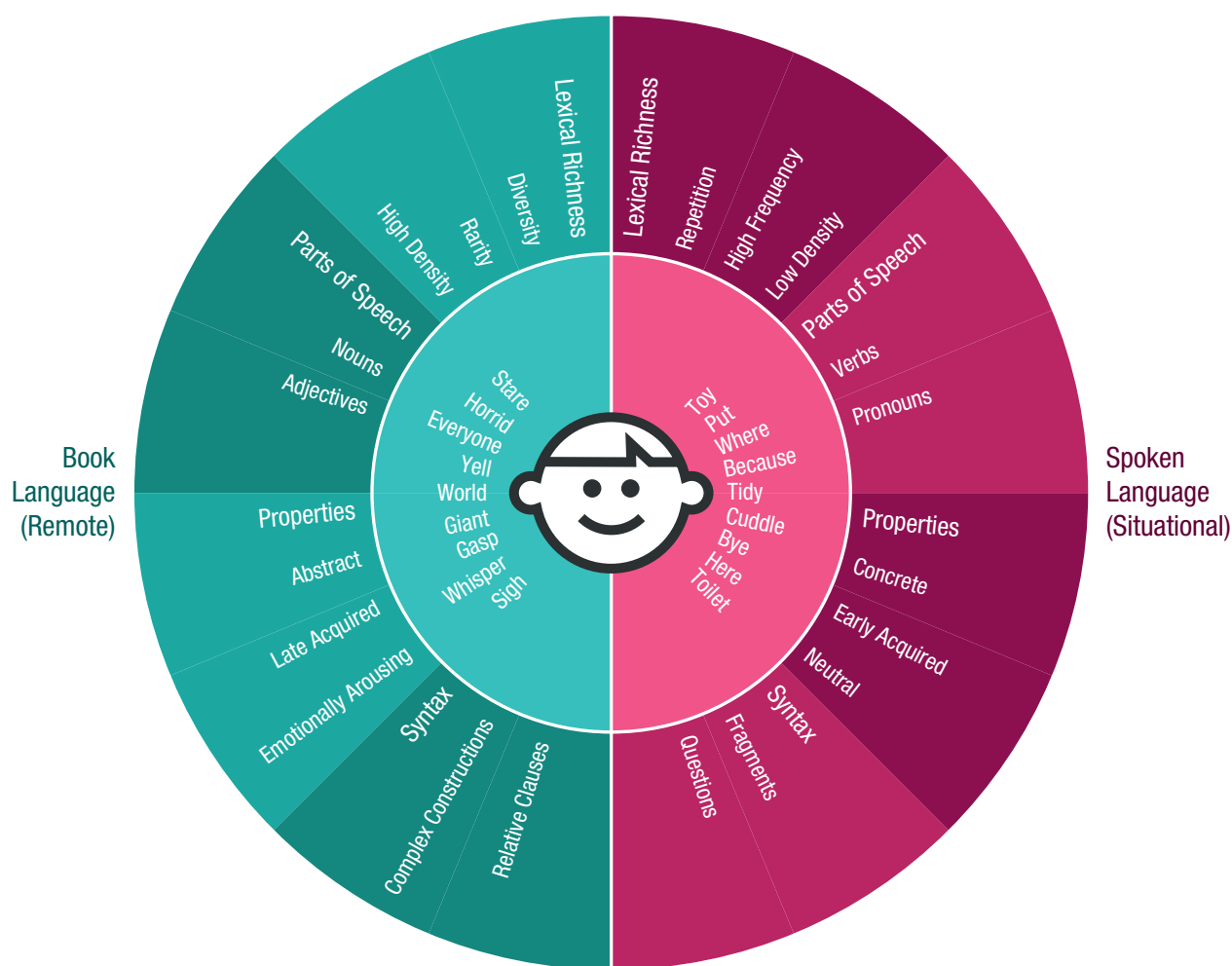
Montag et al. (2015) examined the vocabulary content of picture books targeted at preschoolers and found that books have greater lexical density and diversity than child-directed speech does; that is, they contain more words and more unique words. We (Dawson, Hsiao, et al., 2021) replicated these differences in density and diversity and found “book words” to be more sophisticated than words common in child-directed speech. Book words were more commonly nouns and adjectives, and they tended to be longer and more morphologically complex; they were also more abstract, acquired later in development, and more emotionally arousing. Listening to book language therefore provides exposure to vocabulary that is quantitatively and

qualitatively different from that experienced via day-to-day conversation.

A similar conclusion holds for syntactic complexity. Cameron-Faulkner and Noble (2013) analyzed 20 picture books written for 2-year-olds and found that they contained more complex constructions than child-directed speech. Focusing on relative clauses (clauses connected to the main part of a sentence by a pronoun such as “that,” “who,” or “which”; e.g., “She loved the garden *that she used to tend*”), we compared child-directed speech, picture books targeted at preschoolers for shared reading, and reading books for older children's independent reading. Relative clauses were rare in speech relative to both samples of book language (Hsiao et al., 2022; see also Montag, 2019).

The findings from these studies, summarized in Figure 1, show that learning about book language can begin before children can read themselves. Children who hear written language in the context of shared reading will experience words and syntactic structures that are systematically different from those they experience in speech. Given that language and literacy build from distributional information in the input, children who experience less book language will likely be at a learning and achievement disadvantage. Although the importance of shared reading has been long recognized (Noble et al., 2019), theoretical accounts of its contribution have emphasized factors such as dialogic exchange and talk about or around the text. Demir-Lira et al. (2019) found that parents used a broader range of vocabulary and more complex sentence structures during shared book-reading activities with 1- and 2-year-olds than they did in other interaction contexts, a finding consistent with book language itself being central to the importance of shared reading. Part of this effect was driven by the quality of the language parents used when talking around the text (e.g., providing descriptions of pictures), but it was mainly due to the complexity and diversity of the text itself.

Many factors are likely to be relevant in research on the complex relationship between shared reading and children's language and literacy outcomes, including parental education level and family risk for language and literacy difficulties. It is also important to address issues of causality. Parents do not just create an environment for their child—genes are also at play, as language and literacy are heritable traits (Hart et al., 2021). This means that the shared-reading environment parents provide might reflect their own capabilities, passed to their children via biology as well as culture. The fact that differences between people in language proficiency are associated with genetic differences does not undermine the need to understand the nature of experience, however. The hypothesis that exposure to particular patterns of language brings about language



**Fig. 1.** Infographic summarizing key differences in the nature of language input children experience via books and spoken language. The outer ring contrasts features and properties for the two registers, and example words are shown in the inner ring (for data and annotated examples, see Dawson, Hsaio, et al., 2021; Hsiao et al., 2022).

development, and that book experience provides a particular means for this, should be testable experimentally by directly manipulating exposure and relating this to patterns of learning. Such research has potential to inform intervention as well as theory. To date, attempts to improve language and literacy outcomes via general shared-reading interventions have generated only small effects (for a meta-analysis, see Noble et al., 2019). Greater gains might arise if specific aspects of book language were targeted in a focused manner (e.g., Dawson, Brockbank, & Nation, 2021).

### Reading Experience and Variation in Print Exposure

As children become independent readers, their opportunities to learn from books increase, and by adulthood, skilled readers have accumulated vast knowledge of written words. Reading experience is typically quantified via

proxy measures, such as the number of authors' names a person can recognize or the number of books in the home. Substantial evidence indicates a close association between print exposure and a range of outcomes, including individual differences in reading and vocabulary across the life span (Mol & Bus, 2011). This is not surprising given that text is lexically richer than speech and is therefore the primary supplier of new words, once children can read. Print exposure has beneficial effects beyond those for vocabulary, however. It also correlates with how well adults deal with spoken language in tasks tapping sentence comprehension and grammaticality judgment (e.g., Dąbrowska, 2018; Favier & Huetigg, 2021).

In line with usage-based accounts, these findings support the idea that reading experience shapes language development and leaves a legacy that is evident in how well adults deal with language. Questions remain—not least about specificity and the type of

experience that text provides, as distinct from language experience more generally. The importance of text is supported by findings that the contextual history of individual words in written language (derived by tracking usage across a large corpus of children's books) is associated with how well children process the same words in laboratory tasks, such as those requiring decisions about lexicality or meaning (Hsiao et al., 2020).

Arnold et al. (2019) reported a correlation between children's print exposure and their comprehension of ambiguous pronouns: More avid readers were more likely to show the adult bias of linking a pronoun with a grammatical subject (e.g., linking "he" with "Puppy" in "Panda Bear is having lunch with Puppy. He wants a pizza slice"). Arnold et al. speculated that this relationship is driven by exposure to literate language: In the absence of social cues such as the speaker's gaze, written language needs to convey who did what to whom, and by attending to this while reading, children gradually develop adultlike patterns in pronoun comprehension.

Focusing on relative clauses, Montag and MacDonald (2015) found that children and adults who read more (as captured by measures of print exposure) were more likely to use passive relative clauses (e.g., "the book *that was carried by the woman*") in their own speech, presumably a reflection of the observation that passive relative clauses are more common in text than conversation. Building on this, we (Hsiao et al., 2022) separated the content of a large corpus of books written for 5- to 16-year-olds according to the targeted age groups and found that as the targeted reading age increased, so too did the frequency of relative clauses. Furthermore, and as in speech, different types of relative-clause structures in children's books co-occurred in predictable ways with specific lexical properties. More work is needed to relate input statistics from large and developmentally sensitive written-language corpora to children's performance on a range of tasks. Nevertheless, it is worth noting that distributional patterns we identified in book language align with the ease with which adults process lexical-syntactic combinations in laboratory tasks. Directly manipulating children's exposure to specific forms in book language and then tracking consequences for language processing will be particularly valuable in helping research in this area move beyond correlational evidence.

## Beyond Language and Literacy

Although differences between written and spoken language have been recognized for a long time, it is only more recently that these differences have been systematically described and quantified for young children's

language experience (Fig. 1). The cumulative effect of reduced exposure to book language is likely to be large. Logan et al. (2019) estimated that by the time children are 5 years old, those who have been read to five times a week will have experienced an additional 1.4 million words, compared with children who have not been read to. And it is not just the number of words that matters; also important are the nature of the words (Dawson, Hsiao, et al., 2021) and the type of syntactic structures in which they appear, relative to speech (Hsiao et al., 2022; Montag, 2019). Reading experience shapes learning of the orthographic code (Castles et al., 2018), but it also has a wider influence in that text provides a rich substrate for learning words, sentence structures, and discourse patterns that are rare in speech. In turn, this knowledge is available to support growth in reading comprehension and to be used by children in their own writing.

Taking a closer look at the content of children's books also points to more distal consequences of variation in experience with book language. In adult corpora, fiction contains more complex emotion words (e.g., "despair," "relief," "irritation," "pride") than both nonfiction books and everyday speech, and adults who read more are more adept at recognizing complex emotions (Schwering et al., 2021). Written language may therefore provide unique linguistic input that propels the formation of emotional categories over time, driven by the need for readers to be able to derive emotional situations from text. Our finding that even books written for preschoolers contain more emotion words than child-directed speech does (Dawson, Hsiao, et al., 2021) once again suggests that opportunities for learning start early. This observation also invites speculation regarding outcomes for children who do not experience much book language, or those who are less able to access it fully because of language-learning difficulties. In a longitudinal study, Griffiths et al. (2020) found that language skills at age 5 to 6 years predicted emotion recognition at age 10 to 12 years, and that children with language disorders were poor at recognizing facial and vocal emotion cues. Causal relations cannot be inferred from these data, but language has been posited as a "critical ingredient" in the perception and experience of emotion (e.g., Lindquist, 2017). Perhaps the wide-ranging, varied, deep, nuanced, and sophisticated language needed to serve this function is the language that is most representative of the book.

Similar arguments can be made about syntax. Children's grasp of theory of mind (the ability to attribute mental states to oneself and others) is intimately linked with language development and, in particular, the acquisition of complex syntax (for a recent overview, see Kaltefleiter et al., 2021). This makes sense. English sentences that express another person's mental state



are complex, as they contain an embedded element: In “Mary thinks *that the sky is green*,” the subordinate element (in italics) is false, but the overall sentence may be true. Explicit training with this type of sentence structure improves children’s performance on theory-of-mind tasks (Hale & Tager-Flusberg, 2003), which suggests a relationship that is at least partially causal. Although day-to-day exposure to such utterances is grounded in social interaction and conversation, we note again the utility of books not only for providing situations that naturally invite such conversational exchange, but also for providing greater exposure to the rich linguistic forms themselves. Dyer et al. (2000) concluded that the text of books for 3- to 6-year-olds is laden with mental-state language: emotional, cognitive, and evaluative content words used in complex sentences with verbs such as “feel,” “think,” “want,” and “know.” Over time, this rich input provides opportunity to build and communicate knowledge about the psychological world; 9- to 12-year-olds who use more mental-state terms in their own narratives have stronger language skills (Hamilton et al., 2021), and they also read more fiction. These findings complement the various associations between fiction reading, emotion processing, empathy and theory-of-mind processing reported in the adult literature (e.g., Oatley, 2016; Schwering et al., 2021). Discussion has tended to focus on the role of fiction in promoting imagination and emotional experience. Plausibly, book language itself may play a more direct role, and although patterns of causality remain to be established and are likely to be complex, it is clear that books provide access to situations and characters beyond the everyday, and that the language needed to communicate these is different from everyday language.

## Concluding Remarks

Exposure to book language provides opportunities to experience words and sentences that are rarely encountered in conversations. These systematic differences start early in life and are evident in the books children hear in the context of shared reading. From infancy onward, children have opportunity to learn from this input. Such learning establishes the foundations for more advanced language development as well as literacy, and establishes the distributional properties that will become inherent in their adult language systems. Becoming literate changes things, and researchers need to better understand how and why reading experience shapes people’s minds and becomes associated with a range of skills and abilities across the life span. Progress toward this goal will come from further consideration of the nature of book language itself; how it builds and changes over time as children develop; how it is used

by caregivers, teachers, and children; and how reading experience varies across individuals.

## Recommended Reading

- Dawson, N., Hsiao, Y., Tan, A. W. M., Banerji, N., & Nation, K. (2021). (See References). Provides an in-depth analysis of the lexical content of books targeted at preschool children, comparing it with child-directed speech.
- Montag, J. L., & MacDonald, M. C. (2015). (See References). Details the syntactic complexity of children’s book language relative to child-directed speech, focusing on different types of relative clauses.
- Seidenberg, M. S., & MacDonald, M. C. (2018). (See References). Reviews language and literacy development from the perspective of statistical learning and considers how variability in the input (i.e., language and reading experience) influences language development and language processing.

## Transparency

*Action Editor:* Robert L. Goldstone

*Editor:* Robert L. Goldstone

*Declaration of Conflicting Interests*

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

*Funding*

Preparation of this manuscript was supported by a grant from the Nuffield Foundation (EDO/43392) to K. Nation and a British Academy Post-Doctoral Fellowship (PF2/180013) to Y. Hsiao.

## ORCID iD

Kate Nation  <https://orcid.org/0000-0001-5048-6107>

## References

- Arnold, J. E., Castro-Schilo, L., Zerkle, S., & Rao, L. (2019). Print exposure predicts pronoun comprehension strategies in children. *Journal of Child Language*, 46(5), 863–893. <https://doi.org/10.1017/S0305000919000102>
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- 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>
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51. <https://doi.org/10.1177/1529100618772271>
- Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, 178, 222–235. <https://doi.org/10.1016/j.cognition.2018.05.018>
- Dawson, N., Brockbank, S. E., & Nation, K. (2021). *Features of lexical richness in storybooks on children’s narrative retellings and word knowledge* [Stage 1 Registered Report]. OSF. <https://osf.io/eqdtt>

- 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>
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience, 16*(4), 234–244. <https://doi.org/10.1038/nrn3924>
- 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), Article e12764. <https://doi.org/10.1111/desc.12764>
- 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](https://doi.org/10.1016/S0885-2014(00)00017-4)
- Favier, S., & Huettig, F. (2021). Long-term written language experience affects grammaticality judgements and usage but not priming of spoken sentences. *Quarterly Journal of Experimental Psychology, 74*(8), 1378–1395. <https://doi.org/10.1177/17470218211005228>
- Frith, U. (1998). Literally changing the brain. *Brain, 121*(6), 1011–1012. <https://doi.org/10.1093/brain/121.6.1011>
- Griffiths, S., Goh, S. K. Y., Norbury, C. F., & the SCALES team. (2020). Early language competence, but not general cognitive ability, predicts children's recognition of emotion from facial and vocal cues. *PeerJ, 8*, Article e9118. <https://doi.org/10.7717/peerj.9118>
- Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*(3), 346–359. <https://doi.org/10.1111/1467-7687.00289>
- Hamilton, L. G., O'Halloran, I., & Cutting, N. (2021). Individual differences in narrative production in late childhood: Associations with age and fiction reading experience. *First Language, 41*(2), 179–199. <https://doi.org/10.1177/0142723720946553>
- Hart, S. A., Little, C., & van Bergen, E. (2021). Nurture might be nature: Cautionary tales and proposed solutions. *npj Science of Learning, 6*, Article 2. <https://doi.org/10.1038/s41539-020-00079-z>
- Hayes, D. P. (1998). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language, 27*(5), 572–585. [https://doi.org/10.1016/0749-596X\(88\)90027-7](https://doi.org/10.1016/0749-596X(88)90027-7)
- Hsiao, Y., Bird, M., Norris, H., Pagán, A., & Nation, K. (2020). The influence of item-level contextual history on lexical and semantic judgments by children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(12), 2367–2383. <https://doi.org/10.1037/xlm0000795>
- Hsiao, Y., Dawson, N. J., Banerji, N., & Nation, K. (2022). 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. Advance online publication*. <https://doi.org/10.1017/S0305000921000957>
- Kaltefleiter, L. J., Sodian, B., Kristen-Antonow, S., Wiesmann, C. G., & Schuwerk, T. (2021). Does syntax play a role in Theory of Mind development before the age of 3 years? *Infant Behavior and Development, 64*, Article 101575. <https://doi.org/10.1016/j.infbeh.2021.101575>
- Lindquist, K. A. (2017). The role of language in emotion: Existing evidence and future directions. *Current Opinion in Psychology, 17*, 135–139. <https://doi.org/10.1016/j.copsyc.2017.07.006>
- Logan, J. A. R., Justice, L. M., Yumus, M., & Chaparro-Moreno, L. J. (2019). When children are not read to at home: The million word gap. *Journal of Developmental & Behavioral Pediatrics, 40*(5), 383–386. <https://doi.org/10.1097/DBP.0000000000000657>
- Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*(2), 267–296. <https://doi.org/10.1037/a0021890>
- 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>
- Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General, 144*(2), 447–468. <https://doi.org/10.1037/xge0000054>
- 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>
- Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in Cognitive Sciences, 20*(8), 618–628. <https://doi.org/10.1016/j.tics.2016.06.002>
- Phillips Galloway, E., McClain, J. B., & Uccelli, P. (2020). Broadening the lens on the science of reading: A multifaceted perspective on the role of academic language in text understanding. *Reading Research Quarterly, 55*(S1), S331–S345. <https://doi.org/10.1002/rrq.359>
- Ramscar, M. (2021). How children learn to communicate discriminatively. *Journal of Child Language, 48*(5), 984–1022. <https://doi.org/10.1017/S0305000921000544>
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language, 57*(3), 348–379. <https://doi.org/10.1016/j.jml.2007.03.002>
- Schwering, S. C., Ghaffari-Nikou, N. M., Zhao, F., Niedenthal, P. M., & MacDonald, M. C. (2021). Exploring the relationship between fiction reading and emotion recognition. *Affective Science, 2*(2), 178–186. <https://doi.org/10.1007/s42761-021-00034-0>
- Seidenberg, M. S., & MacDonald, M. C. (2018). The impact of language experience on language and reading: A statistical learning approach. *Topics in Language Disorders, 38*(1), 66–83. <https://doi.org/10.1097/TLD.0000000000000144>
- Treiman, R. (2020). Learning to write words. *Current Directions in Psychological Science, 29*(5), 521–526. <https://doi.org/10.1177/0963721420951585>