A Corpus and Behavioral Analysis of the Quality of Young Children's STEM Books

Hilary E. Miller, Lucy M. Cronin-Golomb, and Patricia J. Bauer

Abstract

Promoting early STEM knowledge helps to prepare children for formal schooling. Shared book reading may promote early STEM knowledge. This research examined the quality of available STEM books in children's environments and investigated how such books influenced children's learning in shared book reading contexts. In Study 1, we used both meaning-based human coding and computerized latent semantic analysis to categorize books based on the extent to which they provided support for encoding and demand for active recall. We found similarity in the ratings using the two approaches. Most books fell into categories characterized by low Support and Demand. In Study 2, we found that 4- to 5-year-olds learned more STEM facts when books were high in Support and/or Demand, although few books fell into those categories. This research highlights the importance that textual features of books play in promoting early STEM knowledge during shared book reading.

Keywords: Shared Book Reading; STEM learning; Textual Analysis

Informal STEM Learning

Individual differences in children's early knowledge are pervasive across STEM (science, technology, engineering, and math) domains, impacting school readiness and later academic achievement (Duncan et al., 2007; Verdine et al., 2017). Early experiences in informal settings can facilitate children's knowledge acquisition and interest in STEM domains (e.g., Haden, 2010). Shared book reading (caregiver reading to child) of STEM books is an informal setting where children can acquire such early knowledge. Yet, it is unknown whether STEM books are structured in ways that align with cognitive principles of learning and memory and how children learn from books varying in their alignment with such principles. The current research bridges this gap by conducting: 1) corpus analyses of preschool-aged children's STEM books to measure book alignment with cognitive principles; and 2) an experimental investigation of how variations in alignment impact preschool-aged children's STEM learning in shared book reading contexts.

Cognitive Principles of Learning

Research on learning and memory provide insights into how books' textual features can align with cognitive principles to support semantic knowledge acquisition, including STEM knowledge. To align with cognitive principles, books should include textual elements that: (1) provide support for encoding factual information to compensate for children's low working memory (Cowan, 2014); and (2) present demand for children's active engagement to accommodate their underdeveloped deliberate memory strategies (Courage & Cowan, 2009). Acquiring facts from books can be taxing on working memory, outpacing resources (van den Broek, 2010). Adults have high working memory, which helps them process facts (Baddeley, 1992). They also use active strategies, such as re-reading, to help deal with book content that outpaces working memory (van den Broek & Helder, 2017). In school, children are taught memory strategies (Armbruster et al., 2003). However, young children do not use such strategies spontaneously (Coffman et al., 2008). Instead, children likely depend more heavily on external support from book structures and adult readers' scaffolding.

Books can support semantic knowledge without requiring active memory strategies. Readers will recall more from texts when they have support for encoding the text and when their memory of earlier text is easily reactivated (van den Broek & Helder, 2017). Books can support encoding by elaborating on facts and by making causal connections between elements. Both children and adults recall more information when there are more causal connections between elements, such as when one event motivates another (Lynch et al., 2008; Trabasso & van den Broek, 1985).

Additionally, books can be demanding of deliberate memory strategies, without children spontaneously using them. They can include questions or interactive prompts that encourage active engagement. Research on testing effects show that having students actively process information through pretest or retrieval questions enhances learning (e.g., Karpicke & Grimaldi, 2012; Pressley, et al., 1990).

In sum, books' textual features can align with cognitive principles by being supportive of encoding and demanding of active memory processing. To be maximally valuable for learning, STEM books should align with these principles.

Shared Book Reading

Shared book reading is an everyday practice in homes (Bus et al., 1995). Decades of research show that the quality of caregiver extra-textual talk during shared book reading affects young children's developing language and literary skills (e.g., Bus et al., 1995). Most research on shared book reading focused on extra-textual talk (talk beyond the text) and on reading of narrative storybooks.

Much less shared book reading research focused on expository (factual) books or on how books' textual features affect learning. Those that focused on reading expository books showed that caregivers had more extra-textual talk and used more complex language when reading such books compared to narratives (e.g., Price et al., 2009). Similarly to narratives, caregivers' extra-textual talk when reading expository books predicts children's language and literary skills (Robertson & Reese, 2017). Expository books fulfill similar objectives as narrations in promoting language and literary skills; however, they have added potential to be valuable in promoting knowledge acquisition, especially STEM knowledge.

Despite the importance of expository shared book reading, there are open questions regarding how young children acquire semantic information from book text. Most of the research on learning from books in early childhood focuses on how caregivers' talk supports learning (e.g., Haden et al., 1996). Unlike research with older children and adults, there has been little consideration for how the textual features of the books themselves affect the quality of learning. We extend past research in two ways. First, we conducted a corpus analysis to investigate whether young children's STEM books provide: a) Support for encoding semantic information by being coherent and elaborating on facts; and b) Demand for active retrieval by posing questions and including interactive prompts. Second, we test whether and how these textual features of Support and Demand affect children's recall of STEM facts in the context of shared book reading.

Study 1: Corpus Analyses

We investigated whether STEM books targeted towards preschool-aged children provide supportive and demanding contexts for acquiring STEM knowledge consistent with cognitive principles. STEM books are widely available and have potential to effectively facilitate knowledge acquisition. Despite their potential, it is unknown whether these books are designed in ways that promote STEM knowledge acquisition.

In Study 1, we conducted corpus analyses of readily available preschool-aged children's STEM books. We took two different approaches to conducting the analysis: human coding and computerized latent semantic analysis (LSA). The human coding provided a sensitive meaning-based measure for coding books, whereas the LSA provided an automated assessment of linguistic features expected to predict comprehensibility (McNamara et al., 2014) that may also capture elements of the human coding. We compared the methods, as due to the LSA's automated nature, it may have long-term benefits of being an easy method for categorizing books based on quality of promoting STEM knowledge.

For the human coding, raters evaluated the extent to which books provided support for encoding by being coherent and elaborating on facts and provided Demand for active factual recall. For the LSA, we used Coh-Metrix software (McNamara et al., 2014). Coh-Metrix was trained on a large corpus of texts for K-12 grade students. Using this corpus, it assigned values to word vectors in high dimension semantic space based on frequency of co-occurrence (e.g., "nail" and "saw" share semantic similarity). We used LSA to identify overlap in semantic space among words in each STEM book.

In sum, Study 1 had two primary goals: 1) to investigate how children's books cluster into categories based on levels of Support and Demand; and 2) whether we can use LSA to categorize books similarly to the human coding metric of Support. We predicted that the books would vary based on dimensions of Support and Demand. We also expected that the human coding and LSA would share similarities to the human ratings of Support. However, there will likely be cases when there is high semantic overlap but the sentences do not reach the level of criterion for Support in the human coding. This would occur when text is on the same topic but provides new facts on the topic rather than elaborates on them. Bauer and Larkina (2017) showed that young children have difficulty integrating novel facts when there is high semantic similarity among distractor items. Thus, high semantic similarity may not equate to optimal learning. Also, the human coding characterizes books as High in Support and Demand based on their facts; while the Coh-Metrix LSA does not distinguish between facts and non-facts (e.g., narration).

Method

Corpus The corpus included 52 STEM books designed for preschool-aged children. It included books on the following STEM topics: animals, human body, nature, weather, geography, space, and physical science. The books were selected to capture opportunities for STEM learning prior to formal education. We created our own corpus as there was no known available corpus. Books were selected if they met the following criterion: (1) listed on children's bestseller book lists and/or currently sold at bookstores; (2) available at public libraries; and (3) normed for preschool-aged children, and (4) less than 15,000 characters due to limits of the online Coh-Metrix software (McNamara et al., 2014). To create the corpus, a research assistant transcribed each sentence in the book and a different research assistant checked them for accuracy. Sentences were not included in analysis if they were related to an activity to be conducted outside of the book (e.g., instructions for an experiment) or were section headings. Two books were excluded for being outliers on our human coding measures of Support and Demand (more than 2 Standard deviations above the mean on both measures).

Human Coding Human coders rated the levels of Support and Demand for each book. Each book received a score for

Table 1: Example of Human Coding

Sentence	Support	Demand	Topic	Function
Whose food is this, crawling up a tiny hill?	1	1	Anteater	Pre-tested
These are an anteater's ants.	0	0	Anteater	Introduced
It licks up lots and lots of ants.	0	0	Anteater	Extended
An anteater can eat up to 30,000 insects in one day.	1	0	Anteater	Elaborated
Whose food is this, deep inside a tube-shaped flower?	1	1	Hummingbird	Pre-tested
This is a hummingbird's nectar.	0	0	Hummingbird	Extended

Note. **This is an excerpt from the book *Whose Food is This?* by Nancy Kelly Ann**

Support and for Demand based on the number of sentences of each type per topic. Table 1 presents coding examples. Each book was coded for the number of topics, which were defined as sentences on the same subject. Typically, a topic contained multiple sentences. However, books that were low in coherence contained topics with only a few sentences.

Next, the books were coded for the number of Support and Demand sentences. A sentence was coded as containing Support if it: 1) elaborated on facts by providing further details, examples, or definitions; 2) made analogies or antonyms to facts; 3) pre-tested information later presented; or 4) made connections to earlier facts. Sentences were not coded as providing support if they introduced or extended facts by presenting new conceptual information or non-factual information. A sentence was coded Demand if it was a question or interactive prompt on the books' facts.

Two research assistants served as coders and each rated about half the books. About every fourth book was coded by both for reliability. Seventeen books were double coded (33%). The first author corrected discrepancies. For Support coding, there was 89% agreement with Cohen's Kappa at .78. For Demand coding, there was 99% agreement with Cohen's Kappa at .98. For the number of topics, we conducted a correlation which was $r = .96$ and the average absolute value difference in disagreement between topic count was .35.

Latent Semantic Analysis We used two measures provided by Coh-Metrix for the latent semantic analysis: LSA similarity between adjacent sentences ($LSASS_1$) and LSA given-new (LSAGN). $LSASS_1$ measures the mean overlap in semantic similarity between words in adjacent sentences. The LSAGN calculates a mean for the proportion of how much information in the current sentence was given in earlier sections of the text versus how much is new. The $LSASS_1$ provides indexes of the coherence of the text. The LSAGN is likely most similar to the human coding measures of Support as it provides indexes of when information is coreferential, sharing similarity to earlier presented information.

Results

We first report results from the human coding following by the Coh-Metrix. Descriptive statistics for our coding measures are presented in Table 2 for all the books and for each cluster based on human coding and Coh-Metrix LSA. As shown, across all books, there was great variability in the number of Support and Demand sentences per topic.

Our human cluster analysis categorized books based on measures of Support and Demand per Topic. We normalized the data using z-scores and ran a k-means clustering analysis. This method groups data such that the sum of squares between a predetermined number of clusters and the cluster center is minimized (Hartigan & Wong, 1979). We determined the number of clusters by observing dendrogram, silhouette, and scree plots. As seen in Table 2 (left panel), the data clustered into three categories characterized by 1) Low Support and Demand; High Support and Low Demand; and High Support and High Demand. The between group sum of squares was 73.18 and accounted for 72% of total sum of squares. Most books clustered as Low Support and Demand. The results suggest that young children's STEM books tend not to align highly with learning principles.

In terms of LSA analyses, we analyzed the alignment between the Support human coding variable and the Coh-Metrix coding. Using Pearson's correlations, we found that the human ratings of Support by Topic significantly correlated with $LSASS_1$ ($r_{40} = .36$, $p < .019$) and LSAGN ($r_{40} = .46$, $p = .002$). As expected, the LSAGN was more highly associated with human ratings of Support. We then ran a cluster analysis replacing the human coding of Support with the LSAGN. Descriptive statistics are presented in Table 2 (right panel). Coh-Metrix does not have an equivalent rating of Demand, and thus we used the human coding for that variable in the analysis. We found that 3 clusters were a good fit for the data and the clusters were similarly characterized in groups based on high/low Support and Demand as in the human coding. The between group sum of squares was 75.05 and accounted for 74% of total sum of squares. Thirty-nine (75%) fell into the same cluster categories. The central difference was that more books categorized as Low Support and Demand in the human ratings were categorized into the High Support Low Demand group based on LSAGN.

Discussion

Study 1 investigated whether and how books cluster into categories based on their alignment with learning principles and whether a computerized LSA coding captures similar

Table 2: Human Coding and Coh-Metrix Cluster Analysis Means and Standard Deviations

	All	Human Coding Clusters			Coh-Metrix LSA Clusters		
Variables		LSLD	HSLD	HSHD	LSLD	HSLD	HSHD
Support by Topic	3.46 (3.16)	1.50 (1.09)	6.58 (3.00)	6.66 (2.69)	1.28 (1.31)	4.53 (3.23)	6.66 (2.69)
Demand by Topic	0.65 (0.95)	0.26 (0.33)	0.55 (0.47)	2.96 (0.72)	0.24 (0.28)	0.45 (0.45)	2.96 (0.72)
LSA Given New	0.37 (0.06)	0.35 (0.06)	0.40 (0.05)	0.39 (0.05)	0.31 (0.03)	0.41 (0.04)	0.39 (0.05)
LSA Overlap Adjacent	0.26 (0.11)	0.24 (0.11)	0.31 (0.08)	0.29 (0.09)	0.17 (0.05)	0.34 (0.08)	0.29 (0.09)
Count of Books	52	32	14	6	21	25	6

Note. **Table presents means with standard deviations in parentheses. LS = Low Support; LD = Low Demand; HS = High Support; HD = High Demand. The variables used in the cluster analysis were: Support by Topic and Demand by Topic for the human coding; and LSA Given New and Demand by Topic for the Coh-Metrix LSA.**

variance as a human meaning-based coding. We found that in a representative sample of readily-available STEM books designed for preschool-aged children, the majority of books fell into categories characterized by low Support for encoding and low Demand for active retrieval. The LSA coding shared similar overlap to the human coding of Support. This was especially evident in the LSAGN coding, which quantified the extent to which information in new sentences was coreferential with the previous text. However, when comparing cluster analyses between the human and Coh-Metrix coding, a portion of books received high ratings based on LSAGN but low ratings based on human coding of Support. This likely occurred as there can be similarity in words captured by the LSA that is not considered Supportive of learning and the LSA was not sensitive to differences in factual versus non-factual information.

Study 2: Behavioral Analysis

Study 2 investigated how high and low levels of Support and Demand textual features influence STEM learning during shared book reading. Study 1 quantified variations in books' alignment with known cognitive principles of learning. However, these principles have primarily been developed through investigations of adults and have not been evaluated in contexts of either shared book reading or of learning STEM facts. In Study 2, caregivers read four STEM books that were crossed on high/low levels of Support and Demand. The high/low levels were based on high/low Support and Demand ratings from the human and Coh-Metrix clusters from Study 1. We investigated whether there were differences based on books' textual features in terms of how many facts children recalled.

Method

Participants There were 35 children (*M* age = 4.81 years, range = 4.15 to 5.57, 20 Females). Each child participated alongside their caregiver (34 females). Based on parental self-report, the sample was Asian (3%), Black or African American (6%), White or Caucasian (77%), and mixed race (14%); 9% identified as Hispanic or Latino. Caregivers gave informed consent for themselves and their child. An additional seven dyads participated but were excluded: not completing second session (2), experimenter error (1), having difficulty understanding the tasks (1) and having prior familiarity with the books (3). Participants were recruited through a university subject pool of families interested in research participation. Children received a small toy and families were given a $10 gift card for participation.

Materials We used four books: *Biggest, Strongest, Fastest* by Steven Jenkins, *Whose Food is This?* by Nancy Kelly Ann, *What Lives in a Shell* by Kathleen Weidner Zoehfeld, and *Bugs are Insects* by Anne Rockwell. These books are normed for preschoolers and reflect common STEM books sold to that age group. The books are of similar word length, are on similar STEM topics of animal facts, and vary in their Support and Demand levels. Table 3 presents descriptive statistics of the books. All books fit into the 3 categories clustered in Study 1, except for the Low Support and High Demand book. We selected this book to have balanced conditions. Its Demand per topic level was a bit lower than the books clustered as High Demand. There were many Demand questions; however, because the book was Low Support, it had many topics with fewer questions per topic. We ensured that the books were novel to the participants through a questionnaire assessing if the children had prior exposure to the book (Participant Section details exclusions).

To assess STEM learning, we designed open-ended questions related to the books' STEM facts. There were six questions per book. Half were global and related to multiple facts and the main theme of each book (e.g., "why do animals need a shell?") The other half were local and related to specific facts (e.g., "What food do hummingbirds eat"). The questions were randomized with three orders per book.

Procedure Each dyad participated in two sessions. Both sessions began with the caregiver reading two of the four books. The book assignment was counterbalanced within and across sessions, with the constraint that per session participants received one high and one low Support book. Before reading, caregivers were instructed to read the books to their children as they would normally read to them. They were told that after reading, their children would be tested on information from the books. During the shared book reading period, dyads sat together on a couch and were offered snacks. They were left alone during this period. We undertook these procedures to increase participants' comfort in this setting. The shared book reading period lasted approximately 20-30 minutes.

After reading, the experimenter tested children on the book fact recall. Children first participated in a free recall task (results not reported in proceeding). Then in the open-ended recall test, they were tested on questions related to the first book read followed by the second. After the test questions in Session 1 only, children participated in the Comprehension-Knowledge subtest of the Woodcock-Johnson-III (WJ-III). The experimenter marked children's answers online, which were later checked offline by a research assistant.

During testing, the caregiver sat in a space separated by a partition. They completed questionnaires while wearing

Table 3: Support and Demand Levels Across Books

Support/ Demand	Book Name	Word Count	Topics	Support	Demand	LSAGN	$LSASS_1$
LSLD	Biggest	641	15	1.73	0.00	0.353	0.285
LSHD	Food	650	9	2.56	1.22	0.374	0.23
HSLD	Shell	650[b]	6	6.50	0.50	0.429	0.397
HSHD	Bugs	666	6	8.00	2.00	0.47	0.407

Note. LS = Low Support; LD = Low Demand; HS = High Support; HD = High Demand, LSAGN = Latent Semantic Analysis Given New, $LSASS_1$ = Latent Semantic Analysis Adjacent Sentences Overlap. Book names abbreviated. [B] This book was cut by 8 pages for equivalent word count across books.

noise-canceling headphones. The headphones prevented caregivers from hearing questions asked to the child, as to avoid influence on the second session. The full session (shared book reading and testing) was video- and audio-taped.

Results

Study 2 examined whether textual features of Support and Demand influence preschool-aged children's STEM fact recall during shared book reading interactions. We first conducted preliminary analyses to examine if Age or Session affected children's open-ended recall performance. Using Pearson correlations, we found no significant correlation between Age and Proportion Correct, r (33) = -.02, p = .905. Using two-tailed paired samples t-test, we found no significant difference between Session 1 (M = .44, SD = .24) and Session 2 (M = .46, SD = .22) on Proportion Correct, t(34) = -.21, p = .832. We removed these variables from further analyses.

Table 4: Results of ANCOVA Predicting Open-Ended Recall Accuracy

	Open-Ended Recall Accuracy			
Predictors	Sum of Squares	F	p	partial η^2
Support	2.29	44.24	<.001	.620
Demand	0.66	8.20	.008	.320
QT	0.00	0.02	.886	.002
WJ-III	5.32	30.51	<.001	.791
Support x Demand	0.50	7.07	.012	.264
Support x QT	0.89	18.63	<.001	.387
Demand x QT	0.01	0.21	.648	.010
Support x Demand x QT	0.00	0.19	.663	.006

Note. QT is abbreviated for Question Type. WJ-III is abbreviated for Woodcock-Johnson-III. Degrees of freedom are 1,30 for all variables except for covariate of WJ-III which is 1,29.

Next, we investigated how textual features of Support and Demand influenced children's recall of STEM facts. We used ANCOVA analyses to predict children's proportion correct on the recall questions, with the predictor variables of Support (high, low), Demand (high, low), Question Type (global, local). Children's standardized WJ-III Comprehension-Knowledge score was used as a covariate. Results are presented in Table 4 and Figure 1.

We found significant main effects of Support and Demand, significant interactions of Support x Demand, and Support x Question Type, and a significant covariate effect of WJ-III score. Children performed better in the High than Low Support condition and in the High than Low Demand condition. Children with higher verbal comprehension had higher fact recall. To follow-up on the interaction between Support x Demand, we conducted two-tailed paired samples t-tests on the effect of Demand separately for High and Low Support books. For High Support books, we found no significant difference between High and Low Demand, t(34) = .24, p =.810. For Low Support books, we found significant effects of Demand, such that children performed better on the High compared to Low Demand books, t(34) = 3.58, p =.001.

To follow-up on the interaction between Support x Question Type, we conducted two-tailed paired samples t-tests on the effect of Support for each Question Type. On the Global questions, participants performed better on the High Support compared to Low Support books, t(34) = 7.78, p<.001. On the Local Questions, participants also performed better on the High Support compared to Low Support books, t(34) = 2.09, p =.044, but the size of the effect was smaller.

Discussion

In Study 2, we found that Support and Demand textual features influenced young children's recall of STEM facts during shared book reading. Children recalled more from the books when there was Support for encoding, irrespective of levels of Demand. However, when Support was low, children still benefited from the books when Demand was high.

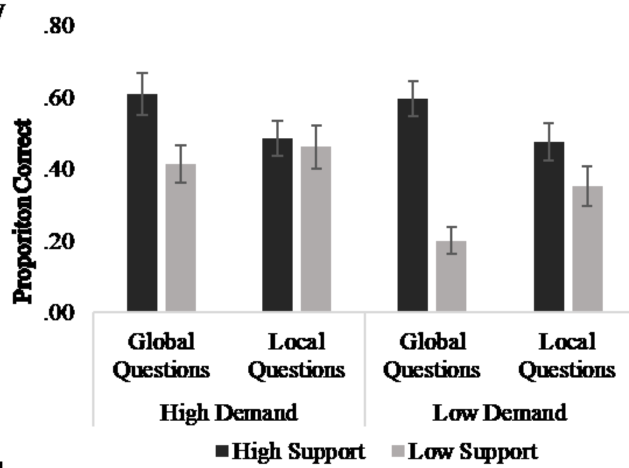

Figure 1: Proportion Correct Based on Support, Demand, and Question Type. Error bars are plotted +/- 1 SE of the mean.

We also found interactions with Question Type such that children recalled more facts related to the global themes of the book when the books had high levels of Support compared to low, likely due to those books being more coherent. Children also recalled more local information in the High Support books; yet, this effect was markedly smaller. This may result from the Local questions not being highly elaborated on in either the High or Low Support Books.

Overall, these findings suggest that the quality of the books themselves matter for children's learning during shared book reading of STEM books. In future directions, we will investigate whether the caregivers' extra-textual talk interacts with the books' textual features to scaffold children's learning.

General Discussion

The current research identified a new mechanism for promoting STEM knowledge early in childhood prior to formal schooling. In Study 1, we showed that STEM books readily available to preschool-aged children tend not to align with cognitive learning principles in that they provide limited support for encoding and limited demand for active recall. We arrived at this conclusion by identifying how books cluster into categories based on Support and Demand using human ratings and computerized LSA. In Study 2, we found that children learn little from books that are low in Support and in Demand during shared book reading, despite the high prevalence of such books in children's environments. However, we identified that children benefit from the high presence of Supportive and Demanding textual elements within books. Thus, to promote early STEM knowledge, it is critical to increase children's exposure to such high-quality books.

In the future, it will be important to extend this research in several directions. First, there was 25% disagreement among the human coding and Coh-Metrix LSA analysis. It will be important to understand whether this discrepancy has differential predictability in children's learning outcomes. Can computerized methods such as Coh-Metrix be used to identify books that are likely facilitative of children's factual STEM learning? Additionally, future research should identify the long-term effects of reading high quality STEM books on children's learning. How long do the enhanced effects of providing Support and Demand during shared book reading on learning last? Does increasing children's exposure to high quality STEM books increase their general STEM learning beyond the context of the book? Lastly, these learning principles are not specific to STEM learning and it will be important to see how Supportive and Demanding textual features affect learning in other domains of children's books.

Overall, in focusing on textual features, this research highlighted the importance of considering the quality that book content has in shaping STEM knowledge acquisition. Both research on shared book reading and research investigating STEM learning in settings such as museums focused primarily on shared interactions (e.g., Haden, 2010; Price et al., 2009). However, learning also depends on the cultural tools available within our environment. Research that fully considers learning mechanisms needs to consider both the cultural tools available (e.g., quality of STEM books) and how such tools are used in learning (e.g., shared book reading).

Acknowledgments

Thank you to the families who participated in this research, as well as the research assistants who aided in coding and data entry. Funding for this study was provided by the NSF Fellowship 1911639 awarded to Hilary E. Miller.

References

Armbruster, B. B., Lehr, F., & Osborn, J. (2003). *Put reading first: The research building blocks for teaching children to read : kindergarten through grade 3*. National Institute for Literacy, National Institute of Child Health and Human Development, U.S. Dept. of Education.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Bauer, P. J., & Larkina, M. (2017). Realizing relevance: The influence of domain-specific information on generation of new knowledge through integration in 4- to 8-year-old children. *Child Development*, *88*(1), 247–262. https://doi.org/10.1111/cdev.12584

Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, *65*(1), 1–21. https://doi.org/10.3102/00346543065001001

Coffman, J. L., Ornstein, P. A., McCall, L. E., & Curran, P. J. (2008). Linking teachers' memory-relevant language and the development of children's memory skills. *Developmental Psychology*, *44*(6), 1640–1654. https://doi.org/10.1037/a0013859

Courage, M., & Cowan, N. (Eds.). (2009). *The development of memory in infancy and childhood* (2nd ed). Psychology Press.

Cowan, N. (2014). Working Memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Haden, C. A. (2010). Talking about science in museums. *Child Development Perspectives*, *4*(1), 62–67. https://doi.org/10.1111/j.1750-8606.2009.00119.x

Haden, C. A., Reese, E., & Fivush, R. (1996). Mothers' extratextual comments during storybook reading: Stylistic differences over time and across texts. *Discourse Processes*, *21*(2), 135–169. https://doi.org/10.1080/01638539609544953

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means clustering algorithm. *Applied Statistics*, *28*(1), 100. https://doi.org/10.2307/2346830

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24*(3), 401–418. https://doi.org/10.1007/s10648-012-9202-2

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and Its relation to other early reading skills. *Reading Psychology*, *29*(4), 327–365. https://doi.org/10.1080/02702710802165416

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology*, *15*(1), 27–35.

https://doi.org/10.1016/0361-476X(90)90003-J
Price, L. H., Kleeck, A., & Huberty, C. J. (2009). Talk during book sharing between parents and preschool children: A comparison between storybook and expository book conditions. *Reading Research Quarterly*, *44*(2), 171–194. https://doi.org/10.1598/RRQ.44.2.4
Robertson, S.-J. L., & Reese, E. (2017). The very hungry caterpillar turned into a butterfly: Children's and parents' enjoyment of different book genres. *Journal of Early Childhood Literacy*, *17*(1), 3–25. https://doi.org/10.1177/1468798415598354
Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X
van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, *328*(5977), 453–456. https://doi.org/10.1126/science.1182594
van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, *54*(5–6), 360–372. https://doi.org/10.1080/0163853X.2017.1306677
Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). Links between spatial and math skills across the preschool years. *Monographs of the Society for Research in Child Development*, *82*(1), 71–80. https://doi.org/10.1111/mono.12283