Picturebooks Increase the Frequency and Diversity of Emotion Vocabulary in Children's Language Environments: Modeling Potential Benefits to Emotional Literacy, with Pedagogical Resources

Clarence Green & He Sun

Published online: 04 Nov 2024.

Submit your article to this journal ☞

Article views: 828

View related articles ☞

View Crossmark data ☞

Routledge
Taylor & Francis Group

Picturebooks Increase the Frequency and Diversity of Emotion Vocabulary in Children's Language Environments: Modeling Potential Benefits to Emotional Literacy, with Pedagogical Resources

Clarence Green [a] and He Sun [b]

[a]Faculty of Education, University of Hong Kong; [b]Centre for Research in Child Language, National Institute of Education, Nanyang Technological University

ABSTRACT

Research Findings: This study examines the frequency and diversity of emotion vocabulary in children's picturebooks, as well as its deviation from typical spoken language environments. We analyzed a corpus of 2146 transcripts of individuals reading picturebooks aloud on social media platforms, using an emotion-specific vocabulary list, and two emotion-related word lists: the Valence Aware Dictionary for Sentiment Reasoning (VADER) List, and the NRC Word-Emotion Lexicon Intensity Lists. Models of daily input assess how regular book reading impacts the diversity and quantity of emotion vocabulary in children's language environments. The findings reveal that emotion vocabulary in children's picturebooks is more diverse than found in typical samples of child-directed speech and that regular reading substantially increases exposure to emotion vocabulary in the language environment of children beyond their input from child-directed speech alone. The study underscores the significant role of regular picturebook reading in enriching children's quantity and quality of emotion vocabulary exposure and the opportunities this may provide for emotional literacy development. *Practice or Policy:* The findings are supplemented with emotion vocabulary wordlists produced by the study and provided here (https://tinyurl.com/4pvcb8b4). These constitute new resources for supporting children's emotion vocabulary acquisition during early education and development.

Emotions are states (e.g., *anger*, *joy*, *sadness*) experienced by individuals and perceived in others that are fundamental to daily life (Satpute & Lindquist, 2021). Children develop their emotional competence, also known as "emotional literacy" in educational research (Harper, 2016), over time, and early childhood research has long been interested in understanding this development and how it might be facilitated (Grosse & Streubel, 2024; Weimer et al., 2012). Such research is important because children's emotional literacy predicts positive social, mental health, and academic outcomes (Bruchinal et al., 2020; Sturrock & Freed, 2023). Internationally, public interest in emotional literacy has been increasing, as evidenced in educational policy trends toward stipulating specific emotional learning outcomes within legislated curricula (Barton et al., 2019).

Emotional literacy involves socio-cultural, cognitive, and linguistic learning (Sharp, 2014). Children learn emotion concepts, the causes of emotional states, self-regulation strategies,

CONTACT Clarence Green cgreen@hku.hk Faculty of Education, University of Hong Kong, Room 660, Meng Wah Complex, Pok Fulam, Hong Kong; He Sun he.sun@nie.edu.sg Centre for Research in Child Language, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616.

emotion-associated physiological reactions and cognitive processes, and emotion vocabulary (Streubel et al., 2020; Weimer et al., 2012). Emotion vocabulary, the focus of the current paper, encompasses two subtypes of words: emotion-specific vocabulary, referring to words that label specific emotions (e.g., *happy, sad, excited, angry* etc.) (Ogren & Sandhofer, 2021), and emotion-related vocabulary, referring to emotionally laden words that invoke or collocate with emotional states (e.g., *friendship, marvelous, scary* etc.) (Sabater et al., 2023). Emotion vocabulary is critical for children's capacity to construct and categorize their own and others' emotional experiences (Barrett, 2017; Ogren & Sandhofer, 2021). It also facilitates learning from others, enabling children to affectively align in peer interactions (Morris et al., 2013). In general, children who know more emotion vocabulary and know such words well (i.e., with greater depth) exhibit better emotional literacy outcomes (Grosse & Struebel, 2024; Sturrock & Freed, 2023).

However, emotion vocabulary is challenging to learn and teach due to its abstractness and complexity (Grosse et al., 2021). The current study investigates children's picturebooks as a resource for developing emotion-specific and emotion-related vocabulary. Picturebooks are a unique text type that integrate words and images and are typically read-aloud to children in home and school literacy environments (Crawford et al., 2024). Since children's vocabulary development is highly correlated with their linguistic input (Ogren & Sandhofer, 2021) and recent research has shown that picturebooks increase abstract and complex vocabulary in children's language environments (Dawson et al., 2021; Green et al., 2024; Montag et al., 2018; Nation et al., 2022), this study is motivated to explore how picturebooks may enhance children's input with regard to emotion vocabulary. Dawson et al. (2021) recently found that picturebooks contain vocabulary that is more emotionally intense than models of child-directed speech, and picturebooks typically feature narratives containing emotional situations and opportunities for emotion talk (Chan et al., 2023; Garces-Bacsal, 2022; Sun, 2019; Ng & Sun, 2022). However, no study has yet contributed to researchers and educators a technical analysis of exactly which emotion-specific and emotion-related words occur in picturebooks nor modeled how children's language environments might change with respect to emotion vocabulary input from regular picturebook read-alouds.

This paper makes main contributions. The first is to present a large-scale study quantifying the emotion vocabulary in picturebooks. Models are developed to represent how regular reading (1–3 books per day over a year) alters the daily vocabulary input for a hypothetical child over-and-above child-directed speech. The study shows which emotion vocabulary occurs in picturebooks, how frequently, and evaluates how regular reading enriches the quantity and quality (i.e., diversity) of emotion vocabulary in a child's input. Emotion vocabulary is represented by a set of age-appropriate emotion-specific words (ages 4–6;11) derived from Baron-Cohen et al. (2010) and emotion-related words associated with eight basic emotions developed in emotion and opinion mining (Plutchik, 1991; Mohammad, 2017).

The second contribution is pedagogical. Through the analysis of emotion vocabulary in picturebooks, how often it occurs and which are "keywords" (i.e., those statistically associated with books compared to a child's spoken language environment), three vocabulary resources are developed for use by educators and, potentially, parents in the home literacy domain (Patel et al., 2024). The resources may assist in vocabulary instruction, including the identification of "Tier 2" vocabulary (Beck et al., 2013). The tier 2 vocabulary framework is widely recommended for instructional vocabulary selection in k-12 internationally, which defines optimal instructional targets as those that are frequent in a child's language environment, associated with print, important to comprehension, and that contribute to increases in lexical diversity beyond what might be learned through exposure to oral language. The study describes how teachers can use the resources to support the development of emotion vocabulary.

Emotion Vocabulary Development

The more diverse vocabulary a child has for emotions, the greater their ability to understand a wider range of emotional states (Grosse et al., 2021). Children initially learn broad valence-based emotion vocabulary related to themselves, such as *happy, sad* as young as 2 years old (Widen & Nelson, 2022), and by 3;5-4;0, consistently apply emotion words to others, including imaginary narrative characters. Their emotion-specific vocabulary begins to double in size approximately every 2 years until age 12 (Baron-Cohen et al., 2010; Grosse & Streubel, 2024).

In theories of emotion such as Conceptual Act Theory (CAT) (Barrett, 2017), vocabulary facilitates the construction of emotions (Grosse & Streubel, 2024). CAT proposes that emotions are constructed based on prior learning experiences; a product of stored representations of previously encountered sensory input and accumulated knowledge. Emotion vocabulary development provides stable labels for these accumulated experiences. For example, the emotion-specific word *"anger"* labels a range of prior experiences for the child that involved the emotion, some quite different from others, yet unified once the word is learned. Vocabulary becomes the "statistical regularity" (Gendron & Barrett, 2018, p. 103) holding emotion concepts together across diverse instances, functioning as the "glue for emotion concept knowledge, binding concepts to embodied experiences and in turn shaping the ongoing processing of sensory information ... to create emotional experiences and perceptions" (Lindquist et al., 2015, p. 2).

An important distinction is made in the following study between emotion-specific and emotion-related vocabulary. Emotion-specific words are defined as those used to express emotion states, such as *"anger"* or *"trust"* (Baron-Cohen et al., 2010). However, emotion-specific vocabulary is not learned in isolation but in the context of other words associated with the emotion (Morris et al., 2013), which we refer to as emotion-related words, often also called emotionally laden words (Sabater et al., 2023). For example, the word *"anger"* expresses the emotion, and words that co-occur in the same language environment often include *"fight," "yelling," "angry," "upset"* (Mohammad, 2017). Children typically acquire a new word, including emotion words, after repeated encounters in the context of related words that allow meaning to be inferred (Hiebert, 2019). There is therefore an important role for both emotion-specific and emotion-related words in a child's emotion vocabulary development.

The importance of emotion vocabulary development may extend beyond emotion learning itself. Borghi (2020) suggests that since emotion words are among the first vocabulary acquired by children without concrete referents, they may constitute the foundation from which children learn abstract vocabulary more generally. Most academic vocabulary is abstract (Beck et al., 2013), so this may suggest that emotion vocabulary is a prerequisite not only of emotional literacy but also successful academic literacy at school.

Early Language Environments and Emotion Vocabulary

A child's language environment predicts their vocabulary development (Snow, 2017). This input is largely shaped by parents and teachers. However, Ogren and Sandhofer (2021) report a lack of research on the emotion vocabulary children are exposed to, their frequency and diversity, which limits our understanding of how linguistic environments influence the development of emotion vocabulary and emotions more generally. Researchers have therefore begun to attend to this research gap.

For example, Streubel et al. (2020) examined the emotion vocabulary sizes and depth of knowledge of 86 children aged 4- to 9-year-old children and their similarity to adult vocabulary use. They also explored the contributions of emotion-specific and general vocabulary knowledge to children's emotion understanding. The researchers assessed emotion-specific vocabulary through a picture-naming task, in which children saw illustrated scenarios and were asked to name the emotions of characters (basic emotions such as *joy, anger, disgust, sadness, fear* or *surprise* and complex emotions derived from these, such as *pride, envy* etc.). Streubel et al. (2020) collected adult norms to measure the extent to which children's responses converged with adult usage, providing a proxy for children's

vocabulary "depth". General expressive vocabulary measures were also taken. The researchers found consistent, positive relationships between emotion understanding, general vocabulary size, emotion vocabulary size and depth. They report that older children were more similar to adult usage than younger children, suggesting a developmental path whereby children converge on the adult input in their language environment over time.

Grosse et al. (2021) hypothesized that some emotion words are acquired before others as function of frequency in children's language environments and "word specificity," which they define as consistency in use. Using the picture-naming elicitation task of Streubel et al. (2020), they had 4–11-year-old children (*N* = 123) describe the feelings of characters depicting the eight basic emotions proposed by Plutchik (1991) - *trust, anger, disgust, joy, sadness, anticipation, surprise* and *fear* - as well as 14 subordinate emotions (e.g., *pride* within *joy*). Through a comparison to 27 adults' speech samples extracted from the CHILDES data repository (transcripts containing adult-child interactions) (McWhinney, 2001), the researchers found a relationship between age of emotion-specific word understanding and the frequency of use in the corpus model of adult input. That is, higher frequency emotion-specific vocabulary in the input was more known by younger children, judged by their picture-naming performance. More specific and less frequent vocabulary was known better by older children. This, Grosse et al. (2021) concluded, reflected that emotion vocabulary development is influenced by the frequency and diversity of emotion vocabulary in a child's language environment.

Ogren and Sandhofer (2021) examined which emotion-specific words occurred in children's input (German), their frequencies, and developmental trends. They hypothesized that the diversity of emotion words and their frequencies would predict which words are learned and in which sequence. They examined approximately 2,000 transcripts from CHILDES containing 179 children aged 1;3 to 3;9. They found that the frequency and vocabulary types used by children and adults were correlated. For example, the 10 most frequent emotion words produced by children were (i.e., in frequency rank-order: *happy, love, scared, sad, angry, mad, afraid, hate, cranky, glad)*, almost identical to the child-directed speech input (i.e., *love, happy, sad, scared, angry, mad, afraid, glad surprised, cranky*) excepting slightly different rank-order and the additional emotion word *surprise*. Since picturebooks can alter the language input to children in terms of the vocabulary they are exposed to (Nation et al., 2022, Montag et al., 2018), it is worth extending such corpus-based research to characterize how a child's language environment might be (positively) altered in regard to emotion vocabulary through regular reading.

Baron-Cohen et al. (2010) developed a list of 336 emotion-specific words that children learn from ages 4 to 16. They selected words from a thesaurus that described emotional states, could complete test phrases such as *[subject] feels/looks/sounds [emotion word]*, or were epistemic words with an emotional dimension (e.g., "*doubting*"). Their final list included emotion-specific words such as *angry, shy, annoyed, cold, comfortable, determined, moody, playful, grumpy*, and *grouchy*. The researchers obtained responses from 377 children aged 4–16 (UK) on a survey asking for a response to each word: either "clearly understood," "not understood," or "possibly understood." Findings indicted an increase in age correlated with an increase in known emotion words. For example, 87 words were known by 50% or more of children aged 4–6;11, rising to 166 for children aged 9–11. Words like *sad* and *sorry* were the only ones known by 100% of the 4-6-year-olds, while 100% of 9–11 year-olds knew *sad, sorry, happy, love, hate, scared, nasty, friendly, frightened* and *excited*. While pedagogy is not the focus of Baron-Cohen et al. (2010), it has pedagogical implications further developed in the current paper and discussed the following section.

Vocabulary and Emotional Literacy Instruction in the Context of Picturebooks

When selecting instructional vocabulary, it is widely recommended that teachers (and parents involved in home literacy) choose "Tier 2" vocabulary (Beck et al., 2013). This is vocabulary that is age appropriate and common in children's print-language environments, such as picturebooks. For

any given age/grade level, instructional words should be neither generally unknown nor known by most children at that age/grade. Biemiller (2001) suggests to choose among "words being acquired" at particular age range, which he suggests are those typically known by about 25–75% of children. Baron-Cohen et al. (2010) similarly suggests that those known by more than 75% of children in their data fall into the category of "known by most." Thus, their study can be drawn from for age-appropriate "Tier 2" emotion vocabulary. Recomputing their results shows that for 4–6 year-olds (the age range most relevant for picturebook reading) there were 112 words known by 25–75% of children, thus a "words being acquired" pool of potential instructional vocabulary. Their data also show which words shift from generally unknown, i.e., fewer than 25% at ages 4–6, to being acquired at ages 9–11. There are 88 such words unknown by 4–6 year-olds but that move into the 25–75% known range for 9–11 year-olds. In the following study, therefore, models are developed to determine exposure children receive from picturebooks of the 112 emotion-specific words "being acquired" at ages 4–6, and the 88 emotion-specific words that shift to the "being acquired" stage at ages 9–11. This helps determine if regular reading facilitates their acquisition. Further, information about the distribution of these words in children's print input is developed into pedagogical wordlists.

Picturebooks have often been recommended for developing children's emotional literacy (Harper, 2016). Picturebooks can reflect and validate children's emotions, model dealing with emotions, show causes and effects, and potentially enhance sensitivity to other's emotions (Barton et al., 2019; Garces-Bacsal, 2022). Nikolajeva (2013) suggests that picturebooks offer heightened emotional experiences and facilitate rich dialogic episodes of joint attention during which parents/teachers mediate emotional experiences and language. Thus, picturebook reading presents an opportunity for emotion socialization (Eisenberg, 2020). Harper (2016) recommends selecting picturebooks with themes related to specific emotions to promote emotional literacy; yet, additionally, it is hypothesized in the current study that emotion vocabulary could be broadly available in picturebooks, beyond those specifically focused on emotion as a narrative theme. If so, this would provide some evidence for the general utility of picturebooks vocabulary and emotional literacy instruction.

Modelling the Impact of Picturebooks on Language Environments

In a recent review, Nation et al. (2022) called for increased investigation of the language within picturebooks to better understand how reading experience shapes language, literacy, and socio-emotional development. They note the limited research on the linguistic content of this print input. Arguably, a contributing factor to the lack of content analysis has been the lack of data sets to explore, a research gap recently addressed through the development of a large corpus of picturebooks by Green et al. (2024). Nation et al. (2022) noted a particular gap in studies on the affordances for emotional development.

The only study that has specifically examined emotion vocabulary in a corpus of children's picturebooks was Dawson et al. (2021). They examined 160 picturebooks (approximately 320,000 words) and compared the vocabulary to a reference corpus of child-directed speech (3.8 million words extracted from CHILDES). They computed the "*keywords*" of picturebooks, which they called "book words." The keyword procedure is a statistical process that computes which words occur more commonly in a target corpus (i.e., the picturebooks) than a large reference corpus (i.e., the proxy for child-directed speech) to a threshold of statistical significance. Dawson et al. (2021) computed keyness as the ratio of the normalized frequency in books to child-directed speech, adding a constant of 10 and generating a ratio score for each word. They ranked words by their keyness values and took the 500 most key as representing a pool of vocabulary most associated with books. Dawson et al. (2021) extracted for these keywords their emotional arousal and valence ratings (from Warriner et al., 2013). Arousal refers to the emotional intensity of a word (rated 1–9, least to more emotionally intense), and valence refers to how positive or negative the emotion is (rated 1–9 from negative to positive). When compared to the 500 words with least keyness (i.e., more associated with spoken language vocabulary), the researchers found significantly more emotional arousal in book keywords than speech. They

conclude that vocabulary in books conveys greater emotional intensity than typical words in child-directed speech.

Dawson et al. (2021) also developed a model to track lexical diversity exposure over time. They developed code that compared the number of word-types in random samples (100-word increments) from their picturebook corpus to the number of word types in a similar sized sample of words from their CDS proxy. They then modeled cumulative exposure to new types by a hypothetical child as a function of input volume. Their findings indicate that, regardless of sample size, picturebook input provides much greater exposure to different word types compared to child-directed speech. Green et al. (2024) demonstrated a similar trend in a much larger corpus of 2146 picturebooks. They modeled the additional lexical diversity contributed by 1 to 5 books per day over 1 year, with different daily input rates (approximately 8, 11, or 17 thousand words) (Hart & Risley, 2003). They found that lexical diversity gaps in the model, resulting from variation in the quantity of a child's input, could be narrowed or closed by reading 1–3 books per day, which equates to approximately 20–30 min of reading per day on average.

These modeling studies indicate potential benefits of regular reading to a child's language environment with regard to both quality (i.e., lexical diversity) and quantity of exposure. However, they did not specifically model the changes in the frequency and diversity of emotion vocabulary, which is therefore followed-up in the current study.

This Study

To address the limited research into emotion-specific and emotion-related vocabulary in children's picturebooks and to determine how regular reading might contribute to children's language environments, this study addresses the following:

(1) What emotion-specific and emotion-related words are present in picturebooks, what are their frequencies, emotional intensities, and which are "keywords" (i.e., associated with the print environment)?
(2) How does the quantity and diversity of emotion-specific and emotion-related words change with the addition of 1–3 picturebooks per day in models of children's daily language input?

The first research question is addressed through a comprehensive analysis of a large corpus of picturebooks, extracting statistical information for thousands of emotion-specific and emotion-related words. This information includes the frequency, intensity, and keyness (i.e., extent of association with print). The second research question is explored using a computational model that simulates daily input to a child and tracks changes in their language environment regarding emotion vocabulary when picturebooks are incorporated into child-directed speech. The model is designed for a hypothetical monolingual child exposed to 12,000 words per day from monolingual English-speaking adults. Furthermore, the diversity of emotion-specific and emotion-related vocabulary is analyzed in a dataset representing 1 year of child-directed speech, as well as in the same data supplemented by the additional input from 1, 2, or 3 picturebooks per day over the course of the year.

Methodology

Data

The picturebook corpus consisted of 2146 transcripts of presenters reading picturebooks aloud on social media. Its development is described in Green et al. (2024). The corpus was constructed by utilizing speech-to-text AI word recognition. As it was impractical to build a substantial dataset considering the resources required for sourcing, scanning, and applying optical character recognition

to a large number of picturebooks (which have minimal word count), previous studies have had to rely on smaller datasets consisting of 100–160 picturebooks (Dawson et al., 2021; Montag et al., 2018). In the current study, we have overcome this limitation and obtained approximately 1,120,778 words and 25,585 word-types from a proxy for 2146 different books. The vocabulary is available to other researchers in the Open Science Framework (https://tinyurl.com/4este73c). Examples of source titles include *How to Catch a Leprechaun; Pete the Cat, When Your Lion Needs a Bath; Lola Goes to School; Goldilocks and The Three Bears; The Rainbow Fish Book* etc. Considering social media is driven by popularity metrics, it is likely that the input model reflects popular picturebooks.

Transcripts were selected from YouTube channels dedicated to picturebook read-alouds, identified by search words "picture books, read aloud". Full details on the sources and all code is provided in the Open Science Framework here (https://tinyurl.com/yc8dnsfm). Data was cleaned by removing opening and closing statements by the online readers, isolating the book text in-between. It did not appear that readers deviated from verbatim readings in these online videos. Nonetheless, while noise in the data was minimized through cleaning, undoubtedly noise remains as in any webcrawled/AI-generated data. This is detailed in full in the limitations. For interpreting the following results, however, the limitations do not suggest any concern in capturing the signal (i.e., the real vocabulary of children's picturebooks), given potential noise (e.g., AI speech-to-text mistakes for some words, potentially words spoken by the readers not in the book). Evidence that this data represents a valid proxy for early print-language environments was reported in Green et al. (2024) by showing that the frequency of the vocabulary in these transcripts correlated with similar databases of children's picturebooks reporting word frequencies, but with improved predictions of children's age of acquisition norms.

To simulate a speech environment, as in previous studies, input was sampled from CHILDES. Since the emotion words identified by Baron-Cohen et al. (2010) were sourced from the UK, while the picturebook corpus was predominantly from American sources, CHILDES transcripts from both the United Kingdom (UK) and North American (NA) directories were pooled. All transcripts where the target child was aged 4–6;11 were extracted, and a random sample of 4.3. million words of child-directed speech taken and placed in a concatenated plain txt file. The corpus size was determined by aggregating 12,000 words per day over 365 days and ensuring it was four to five times larger than the book corpus, as recommended for keyword computation by Berber-Sardinha (2000).

Source Lists (Emotion-Specific and Emotion-Related Vocabulary)

Emotion-specific words were the 336-item wordlist of Baron-Cohen et al. (2010), chosen given its methodological soundness in defining and selecting emotion-specific words, as well as the availability of data regarding each item's age appropriateness for children around 4–6;11. Four search lists were developed. The first comprised the words exactly as published, without alterations (hereafter the "original" list). A second contained all words sharing the same root as the original list (hereafter "emotion word families"). Only using the original list might have underestimated exposure and learning affordances; for example, the original list contains "*dislike*" but not "*disliked*" or "*dislikes*," "*happy*" but not "*happiness*", etc. To generate root-related words, the original list was processed by the natural language processing tool *Lextutor* (*Familizer/Lemmatizer* function) (Cobb, & Laufer, 2021). This resulted in 2277 words sharing the same roots as the original list. A third list comprised the 112 words identified as being acquired at ages 4–6;11 according to the data from Baron-Cohen et al. (2010) (henceforth the "acquiring" subset), while a fourth contained the 88 words unknown to 4–6 year-olds (0–25% known), but which transitioned into the acquiring stage for 9–11-year-olds (25–75% known) (henceforth "next stage" words).

For emotion-related vocabulary, two sources from the field of sentiment and emotion mining were selected. One represented the valence and intensity of a large pool of words along a positive and negative scale. The other contained an extensive pool of English vocabulary categorized into eight basic emotions. Both were considered worth investigating, given a child's development involves both broad valence-related emotional understanding and progressively more differentiation among specific

Table 1. Source lists.

Baron-Cohen et al. (2010) Vocabulary	VADER Vocabulary
Emotion words (original): 336 words	Positive (High Intensity): 320 words
Emotion words (word families): 2277 words	Positive (Low Intensity): 2678 Words
Emotion words (acquiring): 112 words	Negative (Low Intensity): 3342 words
Emotion words (next stage): 88 words	Negative (High Intensity): 493 words
NRC-EIL Vocabulary	
Anger: 755 words	Sadness: 678 words
Anticipation: 444 words	Fear: 904 words
Joy: 653 words	Surprise: 305 words
Trust: 829 words	Disgust: 556

emotions as vocabulary grows. The Valence Aware Dictionary for Sentiment Reasoning/VADER List (Hutto & Gilbert, 2014) is one of the largest and widely used valance dictionaries available (S. A. Crossley et al., 2017). It contains approximately 7517 words (and emojis). Words are rated on a scale from 4 (extremely positive) to −4 (extremely negative). The scale represents intensity and valence, with words rounding to 0 neutral, with words above 1 having positive valence and below minus 1 negative valence. Intensity is indexed by a word's closeness to the ±4 upper limits of the scale. Ratings were crowdsourced on Amazon Mechanical Turk and represent the average of 10 raters who received training prior to rating and had been tested for college-level reading proficiency in the USA. In the current study, the VADER dictionary was additionally recoded into four categories: +4/+3 high Intensity, Positive Emotion Sentiment; +2/+1 Low Intensity, Positive Emotion Sentiment; −1/-2 Low Intensity, Negative Emotion Sentiment; −3/-4 high Intensity, Negative Emotion Sentiment. The rationale was that this organization would help educators with a simpler heuristic for selecting words. For example, the word *great* is in the high intensity positive sub list and *good* is low intensity positive, while *angry* is in the low intensity negative sublist and *furious* the high intensity negative list.

The NRC Emotion Intensity Lexicon (NRC-EIL) (Mohammad, 2017) contains approximately 10,000 words categorized into eight basic emotions (Plutchik, 1991): *trust, anger, disgust, joy, sadness, anticipation, surprise*, and *fear*. There has been debate over the categorization of emotions into eight basic types; however, no other resources representing alternative models fit for purpose in the current study goals, and the Plutchik (1991) categorizations have been used in recent related corpus work (Grosse et al., 2021). Ratings in NRC-EIL were obtained through crowdsourcing. Participants were given a pool of four words and asked to select which was most or least associated with the target emotion, for example, "*joy*", "*trust*", "*anger*" etc. (Mohammad, 2017). Results for each word were scaled 0 (least emotional intensity) to 1 (the most emotional intensity), with Mohammad (2017) suggesting that the pool of vocabulary rated .5 and above (the upper-half subset) are those words that best invoke the emotion. Therefore, in the following study the upper-half subset of each emotional category was used as search lists. All lists are summarized in Table 1.

Analysis

A simple model of a hypothetical child's daily speech environment was built. The number of words heard per day is variable, but research suggests approximately 12,000 words of child-directed speech is typical (Brushe et al., 2020; Gilkerson et al., 2017; Hart & Risley, 2003). This was therefore chosen as the model parameter for the input quantity. Code was designed to open a txt file which contained the child-directed speech corpus and randomly sample 12,000 words. The code was then fed the emotion vocabulary source lists as search words and computed how many occurred in the input sample.

The code next called a directory containing the picturebooks split into plain text files and randomly sampled one book and computed the emotion vocabulary that occurred in this book. The process was repeated for two and three books. The results were then added to the speech sample. This produced output for how many search words per day would be heard in the speech and how many were heard
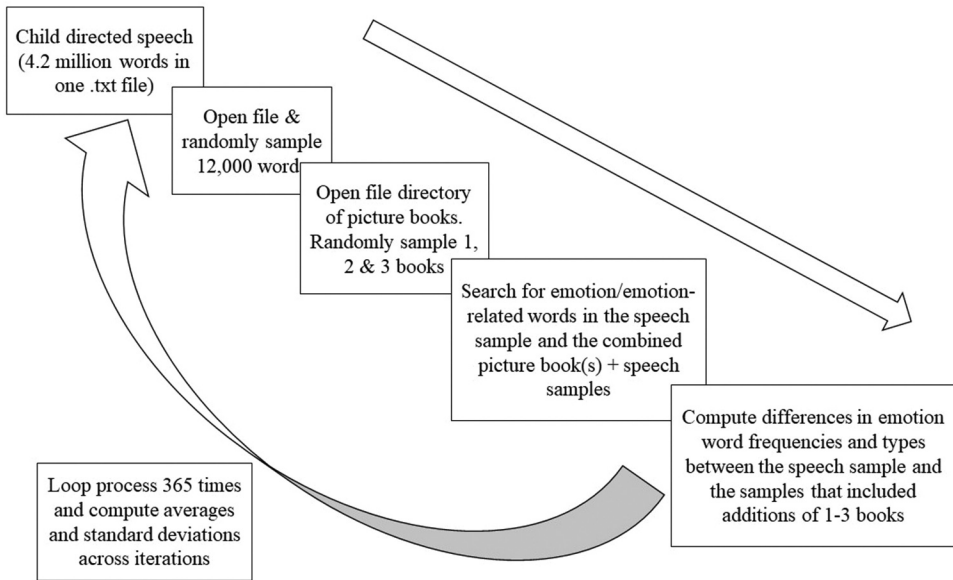
Figure 1. Processing pipeline for computing emotion vocabulary in speech and picturebooks.

once 1–3 books were added. Up to 3 books per day were modeled given recent survey research in the US that reported an average of 2–3 picturebooks per day for children who are read to on a regular basis (Patel et al., 2024). The code further computed how many unique emotion word types were added over and above the speech sample, i.e. that otherwise did not occur in the daily speech input. The process ran 365 times, and the average was computed to represent a "daily snapshot" of the approximate potential contribution of picturebooks to the daily oral language vocabulary environment of a child. All code is available in the Open Science Framework for checking, further analysis, and improvement here (https://tinyurl.com/4pvcb8b4). All other primary data, e.g., source lists and corpora, are also open access as cited above. A simplified visualization of the code's processing pipeline is provided in Figure 1.

To determine how many words from the source lists occurred in the 1.1-million-word picturebook corpus, the 4.3 million words of speech, and how the speech sample might be altered by regular reading, the search words and their frequencies were extracted from the picturebook and speech corpora. Following Montag et al. (2018), pools of book words were added to the speech sample (representing 365 days of input) by randomly adding 365, 730, 1095 books to create 3 more corpora, representing a year's speech to with either 1, 2, or 3 books' vocabulary added per day. The search words were recomputed in these corpora to evaluate changes to their vocabulary profile through the addition of books to the speech environment.

Following Dawson et al. (2021), the frequencies of all emotion-specific and emotion-related vocabulary in the corpus of picturebooks were compared to their frequencies in the speech sample to compute keywords. Words and their frequencies were imported into the UCREL log-likelihood (LL) calculator (https://ucrel.lancs.ac.uk/llwizard.html) (Rayson, 2008). The LL measure is a significance test, with values above 3.84 indicative of statistically significant differences at $p < .05$ between a target and larger reference corpus. The Log Ratio, an effect size (Hardie, 2014), was also computed. This measure is a binary log, easy to interpret: when words in different corpora are similar in relative frequency, the Log Ratio approximates 0; when a word is twice as common in books it approximates 1; four times as common it approximates 2, and so forth. The log ratio also indicates directionality, i.e., words can be significantly different between speech and print by virtue of being associated more with print or more with speech, and thus in the current study book keywords refer only to those words that had positive log ratios indicating the direction of print association.

Results

The following reports the averages of 365 simulations of daily input that included 12,000 words of child-directed speech and reading either 1–3 books per day. Table 2 begins with the emotion-specific words derived from Baron-Cohen et al. (2010).

Results in Table 2 indicate that both the quantity and diversity of emotion-specific vocabulary in a child's daily language environment is increased through picturebooks. The model suggests that a hypothetical child might hear approximately 87 of the Baron-Cohen's et al. (2010) emotion-specific words per day, which increases to 94 by adding 1 book, 101 with 2, and 108 with 3. Children hear many more words with the same root as these emotion words, as reflected in the word family results. In speech, 576 words (on average) with the same roots occurred, increasing to approximately 600, 628, and 657 with 1, 2, or 3 books, respectively. For emotion-specific words in the acquiring stage for 4–6 year-olds, exposure is raised through picturebook reading from hearing 43 in speech to 47, 50, and 53.

Beyond quantity, the model shows that books add several emotion-specific words that otherwise were not in the speech environment. For example, the vocabulary input of 1–3 books provided exposure to approximately 2 to 6.5 more emotion-specific word types from the original list that were otherwise not in children's daily input. For related words within a word family, this increase is more impressive, from 6 to 20 new emotion words that otherwise would not have occurred in the input without reading. For words in the acquiring stage at ages 4–6;11, approximately for every book an otherwise unheard emotion word enters their daily language environment. Small but potentially meaningful gains are also seen for more advanced next-stage words. Table 3 reports the results for the positive/negative emotion-related vocabulary (Hutto & Gilbert, 2014).

Table 3 indicates consistent increases in a child's exposure to vocabulary associated with positive and negative emotions. Adding a single book to the language environment adds approximately six more occurrences and 1–2 otherwise absent highly positive words. When the vocabulary of three books is added, the input contains approximately 16 more occurrences, 4–5 of which are unique. For lower intensity positive vocabulary, 1 book adds approximately 20 more occurrences, 5–6 of which are new, and this raises up to almost 66 more occurrences of lower intensity vocabulary, with 17 unique exposures. The language environment also changes in respect to negative words; for example, the

Table 2. Model of daily input for emotion-specific vocabulary.

Average per day over 365 days	Original	Word families	Acquiring	Next Stage
Speech frequency of emotion words	87.18 (9.26)	576.00 (23.64)	43.58 (6.41)	1.97 (1.50)
Frequency when 1 book added	94.41 (11.83)	600.48 (32.17)	47.24 (7.75)	2.43 (1.84)
Frequency when 2 books added	101.15 (13.53)	628.65 (38.61)	50.19 (7.92)	2.87 (2.05)
Frequency when 3 books added	108.46 (15.79)	657.35 (46.96)	53.47 (9.60)	3.23 (2.07)
Word-types added by 1 book	2.12 (2.56)	6.12 (6.32)	1.18 (1.43)	0.35 (0.76)
Word-types added by 2 books	4.29 (3.51)	13.94 (9.50)	2.30 (2.18)	0.66 (0.99)
Word-types added by 3 books	6.45 (4.00)	20.09 (10.38)	3.47 (2.72)	1.00 (1.19)

*Standard deviation in brackets.

Table 3. Model of daily input for positive/negative emotion-related vocabulary.

Average per day over 365 days	Positive (High Intensity)	Positive (Low Intensity)	Negative (Low Intensity)	Negative (High Intensity)
Speech frequency	41.94 (6.48)	669.62 (23.54)	228.69 (15.36)	17.59 (4.06)
Frequency when 1 book added	47.37 (8.58)	690.02 (27.45)	239.98 (19.01)	18.64 (4.62)
Frequency when 2 books added	51.63 (9.86)	714.54 (34.54)	248.53 (20.30)	19.73 (5.61)
Frequency when 3 books added	57.18 (11.59)	735.22 (41.23)	260.72 (25.14)	20.67 (5.75)
Word-types added by 1 book	1.54 (1.75)	5.50 (4.86)	4.43 (5.36)	0.49 (0.90)
Word-types added by 2 books	2.92 (2.61)	12.46 (8.39)	7.72 (6.08)	0.96 (1.48)
Word-types added by 3 books	4.39 (2.92)	17.58 (9.37)	11.72 (7.46)	1.37 (1.69)

*Standard deviation in brackets.

Table 4. Model of daily input for emotion-related vocabulary associated with eight emotions.

Average per day over 365 days	Anger	Trust	Surprise	Sadness
Speech frequency	37.32 (6.35)	115.57 (10.27)	17.72 (4.06)	35.48 (5.97)
Frequency when 1 book added	39.50 (7.31)	125.38 (14.77)	19.96 (6.24)	38.56 (7.06)
Frequency when 2 books added	42.19 (8.27)	133.59 (18.02)	22.06 (6.22)	41.74 (8.48)
Frequency when 3 books added	45.25 (10.37)	145.51 (21.87)	24.48 (7.08)	44.69 (9.46)
Word-types added by 1 book	0.97 (1.43)	2.73 (3.47)	1.13 (1.47)	1.42 (2.05)
Word-types added by 2 books	2.33 (2.48)	5.55 (4.74)	2.36 (2.24)	2.72 (2.81)
Word-types added by 3 books	3.56 (3.14)	8.66 (5.63)	3.61 (2.68)	4.11 (3.48)
	Joy	Fear	Disgust	Anticipation
Speech frequency	178.01 (12.29)	54.29 (7.29)	24.55 (4.55)	101.51 (9.51)
Frequency when 1 book added	194.84 (19.55)	58.58 (9.78)	26.18 (6.12)	109.05 (12.39)
Frequency when 2 books added	208.23 (25.34)	62.67 (11.54)	27.96 (7.73)	117.19 (13.72)
Frequency when 3 books added	220.90 (26.40)	67.14 (14.05)	28.95 (7.32)	124.16 (14.45)
Word-types added by 1 book	4.58 (4.36)	1.64 (2.40)	0.71 (1.32)	2.03 (2.41)
Word-types added by 2 books	8.13 (5.96)	3.32 (3.48)	1.36 (2.42)	4.17 (3.22)
Word-types added by 3 books	11.40 (6.28)	4.96 (4.40)	1.66 (1.66)	6.00 (3.58)

*Standard deviation in brackets.

input from 1 book increases low intensity negative words by approximately 11 occurrences, with 4–5 typically not present in the daily speech.

Table 4 reports the results for vocabulary related to the eight emotions of Anger, Trust, Surprise, Sadness, Joy, Fear, Disgust, Anticipation.

Table 4 shows that vocabulary eliciting the emotions *trust* and *joy* predominates compared to the other emotions, both in picturebooks and the speech environment generally. Vocabulary related to these emotions are also those that experience the most gains from adding picturebooks to the speech environment. For example, adding one book per day to the model raised the frequency of *joy* related words by almost 16 occurrences, 4–5 of which were typically not otherwise in the speech. Gains of almost 42 additional occurrences occurred from 3 books, with 11–12 of these being unique. While positive vocabulary predominates, simulated gains are evident in both quality and quantity across all the emotions, including negative ones. The general observation from Table 4 is that for *anger, sadness, fear*, gains typically range from about 2–3 more occurrences with each book added and 1–2 additional types over and above speech. The largest gains are for *fear*, probably reflecting the narratives of many picturebooks involving characters and contexts invoking the emotion (e.g., the big bad wolf), but likely also because children's literature often deals with overcoming fears.

Table 5 reports the total number of different emotion word-types on the source lists found in the picturebooks corpus, child-directed speech corpus, and the speech corpus with books added.

Table 5. Emotion vocabulary in the picturebook corpus and 1-year of child-directed speech (CDS).

	Books Types	CDS Types	CDS + 365_Books	CDS + 730_Books	CDS + 1095_Books
Emotion Words (Families)	1065	785	950	1017	1076
Emotion Words (Original)	269	227	260	270	276
Emotion Words (Acquiring)	105	95	102	104	105
Emotion Words (Next stage)	67	58	68	70	71
Anger	303	252	296	319	343
Anticipation	280	207	250	266	281
Disgust	222	183	210	223	235
Fear	406	327	370	402	434
Joy	413	307	381	404	416
Sadness	304	231	280	302	326
Surprise	173	123	152	163	175
Trust	467	325	419	448	476
Positive (High Intensity)	155	109	135	144	150
Positive (Low Intensity)	897	644	794	852	907
Negative (Low Intensity)	943	757	898	949	1014
Negative (High Intensity)	156	131	151	165	175

Table 5 indicates that more emotion vocabulary coverage occurs in the picturebook corpus than in the speech corpus, despite the latter being approximately four times larger. This confirms DDawson et al. (2021) proposal that the book language is a more concentrated language environment for emotion vocabulary than everyday child-directed speech. Since children receive much more speech than picturebook input, the columns showing the speech corpus with 1–3 books on average added per day over a year are informative about the added vocabulary beyond a baseline oral language environment. For example, adding only one book per day on average over a year adds 42 new words from the original Baron-Cohen et al. (2010) emotion-specific words and 280 with the same roots. These words did not occur otherwise in the large sample of 4.3 million words of child-directed speech, including words such as "*comforting*," "*watchful*," "*needy*," "*encouraging*," "*startled*," "*weary*," and "*overjoyed*".

For the eight emotions and negative/positive vocabulary, the addition of 1 book added approximately 61 additional emotion-related word types on average across each emotional category, 2 books added 86 more words, and 3 books added approximately 111 that otherwise would not have occurred in speech input, words such as "*wondrous*", "*villain*," and "*dismay*". Only a slight trend toward diminishing returns for up to three books is seen, that is, with every new book, a substantial number of new words are added into the speech environment. The trend, at least for up three books modeled here, might be summarized as the more that a child is read to, the more unique emotion-specific and emotion-related words they are exposed to that they would not otherwise hear on a daily basis. Note that with positive/negative vocabulary, results in Table 3 indicated much more frequent use of positive vocabulary than negative. However, Table 5 reveals that input contained a wider range of negative than positive words, with one exception (highly negative words in speech). Thus, these simulated children's language environments exhibit the Pollyanna effect, namely, a smaller set of positive words occurs with much greater frequency than a larger set of negative words used with much less frequency.

Moving focus to the words themselves, (1) below reports the most frequent and most key emotion-specific words (ranked by the log-ratio effect size and occurring more than 10 times in the corpus). Complete wordlists and statistical details are provided in the supplementary materials (https://tinyurl.com/4pvcb8b4)

1. The 10 most frequent and most key emotion-specific words in picturebooks

a. Emotion Words (Original):

Most Frequent: love, happy, sure, wanted, kind, cold, sad, sorry, warm, lost

Most Key: startled, distant, weary, terrified, grateful, caring, focused, courageous, puzzled, loyal

b. Emotion Words (Families):

Most Frequent: like, know, love, help, want, think, friends, thought, play, happy

Most Key: thankful, agreed, cheered, delight, friendship, listened, respect, excitement, cared, delighted

c. Emotion Words (Acquiring ages 4–6):

Most Frequent: sure, wanted, cold, warm, mean, close, brave, afraid, fine, excited

Most Key: thankful, cheered, delighted, caring, puzzled, joy, grateful, knowing, hurried, cheerful

d. Emotion Words (Next stage > ages 4–6):

Most Frequent: moved, curious, low, bold, settled, determined, respect, startled, thrilled, eager

Most Key: respect, startled, distant, weary, bold, focused, defeated, regret, foolish, begging

Results in (1) show that the most frequent emotion-specific words that children are exposed to in picturebooks include "*love*", "*help*", "*kind*", "*sad*," and "*sorry*", while the most key book words include "*weary*", "*courageous*," "*terrified*," and "*grateful*". The most frequent emotion words are therefore characterized by both positive and negative emotions and tend to be shorter and semantically broader when compared to keywords associated with book language. The latter are longer, more morphologically complex and semantically narrower. For the subset of emotion-specific words in the acquiring stage for 4–6;11-year-olds, those that children are most frequently exposed to include "*mean*," "*brave*," "*afraid*," and "*excited*," while those that stand out most from their speech input include *thankful, delighted, caring, puzzled*, and *cheerful*. Note that the keywords for the original list words overlap with the words that come to be known by 9–11-year-olds more widely, for example "*startled*", "*distant*",

"*courageous*" etc. What this suggests is that the emotion vocabulary associated with picturebooks tend to be slightly more advanced than the learner's current stage of development and their typical speech environment.

In the interest of space, the most frequent, key, and intense emotion-related vocabulary from the NRC-EIL and VADER sources are provided in Appendix A available in the OSF supplementary materials here: https://tinyurl.com/4pvcb8b4. This lists words that occurred more than 10 times per million in the corpus of picturebooks, a threshold common in pedagogical vocabulary wordlist analysis since, for example, a word that occurs only 1–2 times every million words of input is likely not worth teaching, regardless of keyness or emotional intensity (Hiebert, 2019). The vocabulary in Appendix A appears to be rather uncontroversially valuable learning targets for emotion-related vocabulary in the target domains. For example, consider sadness: the most frequent words include "*cried*", "*sad*", "*alone*", "*worry*", relatively short but important words related to the emotion, while those keywords associated with print include more complex expressions, such as "*dismay*", "*weary*", "*dreary*", "*defeated*". Those most intense in invoking sadness revolve around death, e.g., "*dying,*" "*grief;*" slavery, e.g., "*enslaved*", "*slavery*"; destruction, e.g., "*destroyed*" and extreme states of sadness, e.g., "*miserable*". For the emotion of trust, books most frequently present vocabulary, such as "*good,*" "*love,*" "*school,*" "*happy,*" "*true*", while keywords are multimorphemic abstract nouns *freedom, courage, diversity*, *gratitude, confidence*, and those most intense include "*honesty,*" "*faithful,*" "*wisdom,*" "*true*" and "*promise,*" and archetypes, i.e., "*hero*."

Discussion

Implications and Pedagogical Resources

Ogren and Sandhofer (2021) noted that fundamental to understanding the emotional development of children is understanding the emotion vocabulary input children receive. The results of this study show that picturebooks increase the frequency and diversity of emotion vocabulary in children's language environments. Such positive changes in the input would likely support emotional literacy development. Contemporary frameworks for understanding emotions such as Conceptual Act Theory (Barrett, 2017) propose that emotions are constructed as stored representations abstracted across multiple prior experiences. These experiences come to be labeled as children learn emotion vocabulary. Thus, picturebook reading could add critical value to children's emotion development, by enhancing their experiences with the language of emotions in diverse contexts, albeit fictional, and providing significantly more exposure to emotion vocabulary than they would otherwise receive. This could support children to build richer conceptual representations of emotions compared to children who are read to infrequently, or not at all.

This study has demonstrated in the modeling component that regular reading will enrich the emotion-specific and emotion-related vocabulary in a child's language environment. Beyond this important finding, the results of this study have additional direct pedagogical applications. The data underlying the results reported above have been organized into vocabulary resources in three Excel sheets for educators (and potentially parents given picturebook read-alouds that occur in home literacy environments). These resources, available here (https://tinyurl.com/4pvcb8b4), have been developed as easy-to-search wordlists. Figure 2 shows screenshots of the resources and their organization.

In Figure 2, column 1 of the resource contains the emotion-specific or emotion-related vocabulary in order of their frequency in picturebooks. Column 2 provides their frequency in the child-directed speech corpus (e.g., useful for seeing which words occur mostly, or only, in picturebooks). Columns 3 and 4 provide the log-likelihood and log ratio technical statistics used for identifying keywords. Those with greater log ratio values are more "key" (i.e., more book-associated vocabulary). The final column simplifies these keyness statistics for teachers, by simply labeling whether a word was a keyword of not. The emotion-related resources, illustrated in the bottom panel of Figure 2, also contain a column with

Baron-Cohen et al. (Emotion words)	Book_Freq	CDS_Freq	Log Likelihood	Log Ratio	Book Keyword
love	1756	1638	1593.493095	2.04019	Keyword
happy	1004	897	951.2565134	2.10242	Keyword
sure	728	2838	0.14807969	-0.02303	
wanted	707	1558	138.2725122	0.79992	Keyword
kind	536	3706	187.6783532	-0.84972	
cold	452	1087	64.94034908	0.67388	Keyword
sad	442	309	519.0905631	2.45628	Keyword
sorry	430	1122	43.46396267	0.55617	Keyword
warm	421	550	253.0577994	1.55422	Keyword
lost	313	689	61.42979113	0.80149	Keyword
mean	309	3803	542.0769327	-1.68162	
scared	292	805	22.17194057	0.47682	Keyword
close	279	848	10.93834632	0.33604	Keyword
safe	261	151	351.3103576	2.72934	Keyword
brave	255	40	588.239543	4.61226	Keyword

Sadness (Emotion)	Book_Freq	CDS_Freq	Log Likelihood	Log Ratio	Emotion-Intensity-Score	Book Keyword
cried	669	136	1440.7043	4.238236	0.719	Keyword
dark	463	434	418.05786	2.033153	0.545	Keyword
sad	442	309	519.09056	2.456276	0.844	Keyword
alone	338	372	255.21049	1.801557	0.6	Keyword
lost	313	689	61.429791	0.801495	0.625	Keyword
worry	263	265	219.8948	1.928907	0.641	Keyword
cry	202	292	103.71925	1.408223	0.75	Keyword
terrible	189	152	197.51974	2.254151	0.578	Keyword
sick	170	605	0.739169	0.108436	0.531	
worried	170	90	242.18867	2.857374	0.621	Keyword
mad	150	405	12.765088	0.506877	0.5	Keyword
hurt	134	1455	177.31523	–1.50088	0.688	
lose	108	216	28.06324	0.939836	0.509	Keyword
lonely	94	21	196.73185	4.102108	0.656	Keyword

Figure 2. Visualization of the pedagogical resources.

emotion intensity values. The resource developed from the Baron-Cohen et al. (2010) emotion-specific vocabulary has also been organized into the vocabulary that 4–6 year-olds are typically acquiring, and those typically being acquired by 9–11 year-olds. All word family variants are also provided. This organization should make it easier for pedagogical tasks, such as choosing Tier 2 words; for example, "*brave*", "*afraid*", "*excited,*" and "*proud*" are frequent picturebook words and in the acquiring phase for 4–6 year-olds so potentially good choices for instructional words. For more advanced learners, words like "*curious*", "*bold*", and "*determined*" are potential learning targets, among the more frequent picturebook words in the acquiring phase of children at ages 9–11.

Learning the words that are most frequent, most key, and most emotionally intense all have their value in contributing to a rich understanding of emotional states and supporting development. These are three different pieces of information relevant for teaching. Frequency information may be used as a guide to choose vocabulary that is going to be encountered commonly in print. This is the typical criterion for teachers in choosing "Tier 2" vocabulary (Beck et al., 2013). For example, a teacher may be planning a lesson around a book, such as managing negative emotions, and the resources allow for checking which vocabulary is related to this theme may be "Tier 2" instructional words because the resources provide general frequency expectations in this print environment. Beyond frequency, the intensity of the emotion may provide a reason for instruction. Emotional intensity can facilitate word acquisition (Dawson et al., 2021). Highly emotive words are typically not the most frequent but important for a child to add to their emotion vocabulary, for example, to understand the differences

between "*sad*" and "*miserable*", "*afraid*" and "*terrified,*" "*happy,*" and "*fabulous*" etc. Finally, the words that are key might be selected on the basis of importance to reading development, for example, in terms of understanding the emotional states of characters in texts, and developing vocabulary size more generally with regard to the emotion vocabulary associated with print.

The resources provided are a pool of words from which to choose, with information to support teachers' decision-making and the development of teaching materials. Teachers can relatively easily choose a picturebook to use for an emotional learning lesson objective, however they often need additional materials to accompany a book. With new AI tools, materials development using the emotion vocabulary resources can be achieved much more rapidly than in the past, reducing the burden on teachers. AI can be leveraged to provide more extensive reading input and increase the repetitions of target vocabulary. To demonstrate, we gave ChatGPT the following prompt: "Write a 150-word story for a five-year-old child with simple vocabulary. The theme is overcoming the emotion of fear. Include these target learning words: wonder, fear, thankful, calm, grateful, nervous, confused, disappointed. Repeat every target word at least twice. Make the story rhyme and set it in a theme park." The interaction with the AI produced the material in (2)

2. **Extensive reading input generated by AI**

Once upon a time, in a theme park so bright,
Lived a little girl named Lily, who was filled with fright.
She felt nervous and scared, of rides oh so tall,
But she wanted to conquer her fear, once and for all.

One sunny day, Lily went with her pals,
To the theme park, with its thrilling trails.
They rode the roller coaster, up and down,
Lily felt fear, she wore a frown.

But she closed her eyes and took a deep breath,
Wondering if she could overcome fear's depth.
As the ride went on, Lily opened her eyes,
Feeling calm and thankful, what a surprise!

She realized fear was just a feeling inside,
And she was grateful for the fun, that fear couldn't hide.
Lily was no longer confused or sad,
She conquered her fear and felt really glad.

From that day on, Lily rode every ride with glee,
Overcoming fear, she was finally free.
And she lived happily ever after, full of delight,
A brave little girl, shining so bright.

The input in (2) could serve as an additional read-aloud to a book with the same theme or setting. The material may not be high-quality children's literature; however, it has the advantage of being designable by the teacher according to the needs of their students. For example, teachers can specify emotion-specific or emotion-related words, their repetition rates, a genre, a setting of interest to their students, and so forth.

Limitations

There is a level of noise/error in any computational work. For example, the results clearly show the limitations of wordform matching, which is blind to semantic content. The word *bear* might represent friendly bears as well as scary bears in books (so not always *fear*), and words such as *warm* and *cold* are not only emotion words but also simply descriptions of temperature. Also, note that single word list-based analysis has noise (i.e., error in the signal) because it does not capture nuance, such as *good* being a positive word but not in *the porridge was not good*. Even without

computing negation, however, lexicon matching methods are robust to content analysis so likely the noise from this is acceptable (Catelli et al., 2022). Such issues, which are few but important, are easily spotted in the pedagogical resources and teachers can use their critical judgment. In terms of the modeling in the study, it means that there may be some inflation of emotion vocabulary in terms of quantity for lexical matches where the semantics does not match the intended emotion. The issue is likely minor and relatively trivial for the lexical diversity analyses since it is more likely that for most types, only the emotional version of words was captured, and only a minority has multiple meanings (i.e., polysemy tends to be associated with the high frequency words) (S. Crossley et al., 2010).

Another limitation is that the models are designed to represent the averages from a "daily snapshot." This means that while on average a certain number of new types occurred over and above the speech sample, it does not mean that over the year every day would introduce this number of types never heard before. It is also important that the data sources were predominantly US-based and thus there are potential biases in the social media transcripts limiting the generalization of these findings to different cultural contexts. With regard to pedagogical applications, the resources are not lists to be used "as is" nor without context. Despite their frequency or strong association with print, a teacher may feel some vocabulary is not appropriate for their cohort due to age, culture, religion, or proficiency. The resources simply offer more information to teachers about the emotion vocabulary that occurs in picturebooks, and this is valuable when selecting instructional targets.

Conclusion

This study investigated the emotion-specific and emotion-related vocabulary found within children's picturebooks, aiming to understand the influence of regular reading on the linguistic environment of young learners. Employing computational modeling and corpus analysis, it modeled the effects of daily input comprising both child-directed speech and reading of 1–3 books per day, with regard to children's emotion-specific and emotion-related vocabulary exposure. The findings highlight a notable augmentation in both the quantity and diversity of emotion vocabulary when picturebooks are integrated into a child's language environment. These books serve to introduce new emotion words absent or infrequent in child-directed speech, thereby enriching children's exposure with a broader spectrum of emotion expressions and experiences. Moreover, the study identifies the most frequent, book-associated, and emotionally intense words within picturebooks. The findings recommend regular reading of picturebooks, at home and school, to bolster children's emotional literacy development. Educators can leverage the emotion wordlists derived from this study for a range of pedagogical purposes, such as using AI to generate tailored materials that complement conventional resources, offering content aligned with specific learning objectives and emotional themes.

Disclosure Statement

Funding

ORCID

Clarence Green 🆔 http://orcid.org/0000-0003-3463-141X
He Sun 🆔 http://orcid.org/0000-0001-5729-8002

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available in the Open Science Framework (OSF) repository, https://tinyurl.com/4pvcb8b4.

Data Deposition

OSF persistent identifier for the data set: https://tinyurl.com/4pvcb8b4.

References

Baron-Cohen, S., Golan, O., Wheelwright, S., Granader, Y., & Hill, J. (2010). Emotion word comprehension from 4 to 16 years old: A developmental survey. *Frontiers in Evolutionary Neuroscience*, *2*, 109. https://doi.org/10.3389/fnevo.2010.00109

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Harcourt.

Barton, G., Baguley, M., Kerby, M., & MacDonald, A. (2019). Exploring how quality children's literature can enhance compassion and empathy in the classroom context. In G. Barton & S. Garvis (Eds.), *Compassion and empathy in educational contexts* (pp. 165–188). Palgrave. https://doi.org/10.1007/978-3-030-18925-9_9

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Berber-Sardinha, T. (2000, October). Comparing corpora with WordSmith tools: How large must the reference corpus be? In *Proceedings of the Workshop on Comparing Corpora* (pp. 7–13).

Biemiller, A. (2001). Teaching vocabulary. *American Educator*, *25*(1), 24–28.

Borghi, A. M. (2020). A future of words: Language and the challenge of abstract concepts. *Journal of Cognition*, *3*(1), 42–42. https://doi.org/10.5334/joc.134

Brushe, M. E., Lynch, J. W., Reilly, S., Melhuish, E., & Brinkman, S. A. (2020). How many words are Australian children hearing in the first year of life? *BMC Pediatrics*, *20*(1), 1–9. https://doi.org/10.1186/s12887-020-1946-0

Burchinal, M., Foster, T. J., Bezdek, K. G., Bratsch-Hines, M., Blair, C., Vernon-Feagans, L., & Family Life Project Investigators. (2020). School-entry skills predicting school-age academic and social–emotional trajectories. *Early Childhood Research Quarterly*, *51*, 67–80.

Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics*, *11*(3), 374. https://doi.org/10.3390/electronics11030374

Chan, M., Williams, A. I., Teng, Y. P. T., & Zhou, Q. (2023). Links between parent-child emotion talk and preschoolers' socioemotional behaviors in Chinese-heritage families. *Early Education & Development*, *34*(2), 366–386. https://doi.org/10.1080/10409289.2022.2048347

Cobb, T., & Laufer, B. (2021). The nuclear word family list: A list of the most frequent family members, including base and affixed words. *Language Learning*, *71*(3), 834–871.

Crawford, P. A., Roberts, S. K., & Lacina, J. (2024). Picturebooks and young children: Potential, power, and practices. *Early Childhood Education Journal*, *52*(7), 1–7. https://doi.org/10.1007/s10643-024-01701-0

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, *49*(3), 803–821. https://doi.org/10.3758/s13428-016-0743-z

Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, *60*(3), 573–605. https://doi.org/10.1111/j.1467-9922.2010.00568.x

Dawson, N., Hsiao, Y., Banerji, N., Tan, A. W. M., & Nation, K. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research*, *1*(1), 9–48. https://doi.org/10.34842/5we1-yk94

Eisenberg, N. (2020). Findings, issues, and new directions for research on emotion socialization. *Developmental Psychology*, *56*(3), 664. https://doi.org/10.1037/dev0000906

Garces-Bacsal, R. M. (2022). Diverse books for diverse children: Building an early childhood diverse booklist for social and emotional learning. *Journal of Early Childhood Literacy*, *22*(1), 66–95. https://doi.org/10.1177/1468798420901856

Gendron, M., & Barrett, L. F. (2018). Emotion perception as conceptual synchrony. *Emotion Review*, *10*(2), 101–110. https://doi.org/10.1177/1754073917705717

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Green, C., Keogh, K., Sun, H., & O'Brien, B. (2024). The Children's picture books lexicon (CPB-LEX): A large-scale lexical database from children's picture books. *Behavior Research Methods*, *56*(5), 4504–4521.

Grosse, G., & Streubel, B. (2024). Emotion-specific vocabulary and its relation to emotion understanding in children and adolescents. *Cognition & Emotion*, 1–10.

Grosse, G., Streubel, B., Gunzenhauser, C., & Saalbach, H. (2021). Let's talk about emotions: The development of children's emotion vocabulary from 4 to 11 years of age. *Affective Science*, *2*(2), 150–162. https://doi.org/10.1007/s42761-021-00040-2

Hardie, A. (2014). *Log ratio: An informal introduction*. http://cass.lancs. ac.uk/?p=1133

Harper, L. J. (2016). Using picture books to promote social-emotional literacy. *YC Young Children*, *71*(3), 80–86.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30-million-word gap by age 3. *American Educator*, *27*(1), 4–9.

Hiebert, E. H. (2019). *Teaching words and how they work: Small changes for big vocabulary results*. Teachers College Press.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web & Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology*, *6*, 121301. https://doi.org/10.3389/fpsyg.2015.00444

MacWhinney, B. (2001). From CHILDES to TalkBank. In B. MacWhinney, M. Almgren, A. Barreña, M. Ezeizaberrena, & I. Idiazabal (Eds.), *Research in child language acquisition* (pp. 17–34). Cascadilla Press.

Mohammad, S. M. (2017). Word affect intensities. *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, May 2018, Miyazaki, Japan.

Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early word learning environments. *Cognitive Science*, *42*(S2), 375–412. https://doi.org/10.1111/cogs.12592

Morris, C. A., Denham, S. A., Bassett, H. H., & Curby, T. W. (2013). Relations among teachers' emotion socialization beliefs and practices and preschoolers' emotional competence. *Early Education & Development*, *24*(7), 979–999. https://doi.org/10.1080/10409289.2013.825186

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

Ng, S. C., & Sun, H. (2022). Promoting social emotional learning through shared book reading in kindergarten classrooms. *Early Education & Development*, *33*(8), 1326–1346. https://doi.org/10.1080/10409289.2021.1974232

Nikolajeva, M. (2013). Picturebooks and emotional literacy. *The Reading Teacher*, *67*(4), 249–254. https://doi.org/10.1002/trtr.1229

Ogren, M., & Sandhofer, C. M. (2021). Emotion words in early childhood: A language transcript analysis. *Cognitive Development*, *60*, 101122. https://doi.org/10.1016/j.cogdev.2021.101122

Patel, T. N., Marasli, Z. B., Choi, A., & Montag, J. L. (2024). An online survey of picture book reading practices with children between the ages of 0 and 30 months. *Language Learning and Development*, 1–27. https://doi.org/10.1080/15475441.2024.2372245

Plutchik, R. (1991). *The emotions*. University Press of America.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519–549. https://doi.org/10.1075/ijcl.13.4.06ray

Sabater, L., Ponari, M., Haro, J., Fernández-Folgueiras, U., Moreno, E. M., Pozo, M. A., Ferré, P., & Hinojosa, J. A. (2023). The acquisition of emotion-laden words from childhood to adolescence. *Current Psychology*, *42*(33), 29280–29290. https://doi.org/10.1007/s12144-022-03989-w

Satpute, A. B., & Lindquist, K. A. (2021). At the neural intersection between language and emotion. *Affective Science*, *2*(2), 207–220. https://doi.org/10.1007/s42761-021-00032-2

Sharp, P. (2014). *Nurturing emotional literacy: A practical for teachers, parents and those in the caring professions*. Routledge.

Snow, C. E. (2017). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 179–193). https://doi.org/10.1111/b.9780631203124.1996.00007.x

Streubel, B., Gunzenhauser, C., Grosse, G., & Saalbach, H. (2020). Emotion-specific vocabulary and its contribution to emotion understanding in 4-to 9-year-old children. *Journal of Experimental Child Psychology*, *193*, 104790. https://doi.org/10.1016/j.jecp.2019.104790

Sturrock, A., & Freed, J. (2023). Preliminary data on the development of emotion vocabulary in typically developing children (5–13 years) using an experimental psycholinguistic measure. *Frontiers in Psychology*, *13*, 982676.

Sun, H. (2019). Home environment, bilingual Preschooler's receptive mother tongue language outcomes, and social-emotional and behavioral skills: One stone for two birds?. *Frontier in Psychology*, *10*, 1640. https://doi.org/10.3389/fpsyg.2019.01640

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Weimer, A. A., Sallquist, J., & Bolnick, R. R. (2012). Young children's emotion comprehension and theory of mind understanding. *Early Education & Development*, *23*(3), 280–301. https://doi.org/10.1080/10409289.2010.517694

Widen, S. C., & Nelson, N. L. (2022). Differentiation and language acquisition in children's understanding of emotion. In D. Dukes, E. Walle, & A. Samson (Eds.), *The Oxford handbook of emotional development* (pp. 174–185). OUP. https://doi.org/10.1093/oxfordhb/9780198855903.013.23