

Do children with language impairment tell depressing and anxious stories? An emotion-based narrative language sample analysis of over 1000 children

Scott R. Schroeder^{a,*} , Max R. Freeman^b, Glennys A. Cuatlacatl^b, Journee Cunningham^b,
Natalie M. Jordan^b , Alexandra K. Occhiuto^b

^a Department of Speech-Language-Hearing Sciences, Hofstra University, USA

^b Department of Communication Sciences and Disorders, St. John's University, USA

ARTICLE INFO

Keywords:

Language impairment
Children
Depression
Anxiety
Vocabulary

ABSTRACT

Children with language impairment (LI) experience increased mental health challenges, which may be reflected in the words they use to tell stories. We examined whether children with LI produced more emotionally negative narratives than typically developing (TD) children, and whether lower lexical diversity (a proxy for vocabulary size) was associated with more negativity. Language samples were analyzed from four corpora, totaling 1197 children, ages 3–15. In the two larger corpora, there was evidence for increased negativity in LI children compared to TD children. However, this increased negativity in LI children was not observed across all measures of emotional valence in the two larger corpora, nor was it observed in any measure in the two smaller corpora (and was in the opposite direction on one measure in one of the smaller corpora). There were many associations within both the LI and TD groups between lower lexical diversity and more negativity, indicating that children with smaller vocabularies produced narratives with elevated negativity, but these associations were generally small and sometimes absent or in the reverse direction. In sum, the negative emotions that are known to be experienced by children with weaker language abilities may manifest in their stories, but the results were inconclusive, and more research is needed. Implications for speech-language pathology and clinical psychology are discussed.

In recent years, an increasing number of children have experienced mental health challenges, imposing significant burdens on the affected children, their families, and society (Ghandour et al., 2019). Children with a communication disorder, including language impairment (LI), are especially prone to comorbid mental health challenges (Bernard et al., 2022; McAllister et al., 2023). The current study aims to further elucidate the concerning relation between childhood LI and the negative emotions that accompany mental health challenges. Specifically, we conduct language sample analyses across four corpora to examine the emotion words that children with LI and typically developing (TD) children use to tell stories. We hypothesize that, if children with LI experience increased negative emotions, leading in many cases to clinical-level anxiety and depression, then such children may narrate events in a more emotionally negative manner than TD children. We also examine, in both children with LI and TD children, whether lower vocabulary skills (as measured by lexical diversity) are associated with more emotionally negative narratives.

1. Elevated negative affect in children with language impairment

There is now considerable evidence associating LI in adolescence and childhood to mental health disorders, such as anxiety and depression (Botting et al., 2016; Forrest et al., 2021; van den Bedem et al., 2018, but see e.g., Zupan et al., 2022). For adolescents with a history of Specific Language Impairment (SLI), 39 % reported depressive symptoms that reached a clinical level compared to 14 % without such a history, and 12 % with a history of SLI compared to 2 % without reached the clinical threshold for anxiety (Conti-Ramsden & Botting, 2008). The increased prevalence of anxiety and depression among *adolescents* with a history of LI also applies to *children* with LI. For anxiety, 81 % of parents to 6- to 12-year-old children with Developmental Language Disorder (DLD) indicated that their child experienced anxiety, compared to 58 % of parents to TD children (Burnley et al., 2023). The parents of children with DLD also rated their children significantly higher on both

* Correspondence to: Dept. of Speech-Language-Hearing Sciences, 1000 Hempstead Turnpike, Hempstead, NY 11549, USA.

E-mail address: scott.r.schroeder@hofstra.edu (S.R. Schroeder).

<https://doi.org/10.1016/j.laheal.2026.100079>

Received 22 June 2025; Received in revised form 6 January 2026; Accepted 10 January 2026

Available online 20 January 2026

2949-9038/© 2026 The Author(s). Published by Elsevier B.V. on behalf of Shandong University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Generalized Anxiety Disorder symptoms and Social Anxiety Disorder symptoms. For depression, 8- to 16-year-olds with DLD reported higher levels of depression on the Child Depression Inventory than those without DLD (van den Bedem et al., 2018). In addition to internalizing disorders, such as anxiety and depression, there is also evidence that children with LI have increased externalizing behaviors, such as aggression or hyperactivity (Beitchman et al., 1996; Carson et al., 1998).

Parents' descriptions of their children with LI provide qualitative evidence for mental health challenges. For example, one parent stated about her son: "The fact that things go over his head, that is causing a certain level of uncertainty, because he can't grasp it. And then of course it is only logical to slip into anxiety" (Burnley et al., 2023). Adults with LI likewise describe the psychological difficulties of having a LI. For example, an adult with LI stated: "Stress, worry, frustration, and daily anxiety trying to; reading information correctly or not, understand what people are saying to you, remembering what people have said, express your thoughts and ideas across and for the other person to understand what you are saying – it is a tiring process and I don't think people actually understand how much effort we have to go through on daily basis" (Wilmot et al., 2024).

While both quantitative and qualitative data indicate a strong association between LI and psychological health challenges, the cause of the association is unknown. However, relevant models have been proposed, such as the Social Adaptation Model and the Social Deviance Model. According to the Social Adaptation Model (Redmond & Rice, 1998), children with LI do not inherently experience socioemotional difficulties, however, such difficulties emerge as a result of adapting to having a communication impairment in social contexts. According to the Social Deviance Model (Redmond & Rice, 1998), children with LI do inherently experience socioemotional difficulties, which co-occur with but are not the result of impaired communication. Regardless of the model, children with LI are likely to have co-occurring socioemotional difficulties, which increase the risk of anxiety and depression.

1.1. Negative emotional valence is detected in language sample analyses

Word choice reflects the negative emotional valence in individuals with anxiety and depression, as well as the decreases in negative emotional valence (i.e., more positive emotional valence) that occur through interventions such as psychotherapy (Bernard et al., 2016; Jarrold et al., 2011; Rook et al., 2022; Rude et al., 2004; Sonnenschein et al., 2018; Tølbøll, 2019). Specifically, increased use of negative emotion words (e.g., *sad* and *angry*) reflects poorer psychological health status, such as anxiety and depression in individuals without LI (Meyerhoff et al., 2023; Rook et al., 2022; Sonnenschein et al., 2018; Tølbøll, 2019). As evidence, in a meta-analysis that examined oral and written language samples produced by depressed participants (diagnosed through clinical assessments, such as the Beck Depression Inventory and the Hamilton Depression Rating Scale) as well as control participants without depression, the depressed participants produced significantly more negative emotion words than control participants (Tølbøll, 2019). Language sample analysis of emotionally negative words also demonstrated promise in classifying individuals with and without Generalized Anxiety Disorder, with satisfactory diagnostic accuracy (Rook et al., 2022).

Moreover, in psychotherapy for anxiety and depression, certain word choices that reflect psychological distress (e.g., a lack of linguistic distancing from negative stimuli) change over the course of therapy, and changes in these word choices are associated with reduced symptoms (Nook et al., 2022). Moreover, when reappraising a negative event, which is a crucial therapeutic strategy, the type of language used during reframing is associated with emotional regulation success (Nook et al., 2020).

This association between word choice and negative emotion also extends to individuals without clinical anxiety or depression. For example, in a recent study (Knuppenburg & Fredericks, 2021),

researchers first induced either a negative mood state or a positive mood state in college students. The negative mood state was elicited by having the participants imagine (and then write a narrative about) "a time when [they] felt mood states of hostility, aversiveness, anger, contempt, guilt, fear, and nervousness," whereas the positive mood state was elicited by having participants imagine (and then write about) "a time when [they] felt active, alert, enthusiastic, happy, and engaged in pleasurable activities." When researchers analyzed the subsequently produced narratives, the narratives generated after the negative mood state contained more words categorized as negatively valenced than the narratives generated after the positive mood state. In sum, previous research examining the narratives of individuals with anxiety and depression, individuals in psychotherapy, and individuals without a clinical diagnosis but in a temporarily negative emotional state suggests that language sample analyses provide pertinent information about a speaker's psychological state. Furthermore, language sample analyses can be a powerful tool to detect symptoms of anxiety and depression.

1.2. Current study

The main question in the current study (*Research Question 1*) is whether children with LI produce words in narratives that are more negative in emotional valence, as compared to TD children. When narrating a story – for example, re-telling the story of a boy whose pet frog goes missing, and the adventurous search for the frog that ensues – children with LI may produce narratives that are more negatively valenced, as compared to TD children, given that language impairment is associated with poorer mental health. To illustrate, in the frog story, a child with LI may be more likely to discuss the negative situation of the pet frog leaving, whereas a TD child may talk more about the positive situation of finding the frog in the end. If both the LI and TD child emphasize the positive situation of finding the frog, the child with LI may nevertheless use a less positive (i.e., more negative) word, such as *good* instead of *great*.

In addition to our primary measure, emotional valence, we also compare LI and TD children's narrative language samples on their degree of emotional intensity, and on their expression of the specific negative emotions of anger, anxiety, and sadness. For emotional intensity, we examine whether children with LI produce words that are lower or higher in emotional intensity than TD children (*Research Question 2*). Emotional intensity, also known as emotional arousal, is typically measured alongside valence, as they are the two main components of emotion. We include intensity in the current investigation given its relevance to mental health; in certain mental health conditions, emotional intensity is heightened (Nalipay & Mordeno, 2016), and in others, it is reduced (Kerns et al., 2008). Thus, children with LI may produce more or less intense narratives relative to TD children.

For specific negative emotions, we assess whether children with LI, as compared to TD children, express more anger, anxiety, and sadness in their narratives (*Research Question 3*). The specific emotions of anger, anxiety, and sadness are measured because they provide a more fine-grained emotional assessment than general negative valence, and all three of these emotions are linked to poor mental health (Lee et al., 2008). Children with LI may thus produce angrier, more anxious, and sadder narratives than TD children.

Lastly, we examine, in both children with LI and TD children, whether a child's lexical diversity (as a proxy of vocabulary size) is related to their narrative's emotional valence (*Research Question 4*). There is previous evidence that children with smaller vocabularies experience more emotional difficulties (Westrupp et al., 2020). Therefore, children with LI (and TD children) who have a reduced lexical diversity score may produce more emotionally negative narratives.

To summarize and specify our research questions and hypotheses:

- a) *Research Question 1 (RQ1)*: Do children with LI produce narratives that are more negative in emotional valence than those produced by

Table 1
Participants and Narrative Tasks in Corpora.

Corpus	N	Age	Narrative Task
Gillam (Gillam & Pearson, 2004)	LI = 171 (100 M, 71 F) TD = 497 (245 M, 252 F)	5.0–11.9	Retell & Formulation
ENNI (Schneider et al., 2006)	LI = 77 (47 M, 30 F) TD = 300 (150 M, 150 F)	4.0–9.9	Formulation
EisenbergGuo (Eisenberg & Guo, 2013)	LI = 17 (10 M, 7 F) TD = 17 (10 M, 7 F)	3.0–3.9	Formulation
Conti-Ramsden (Wetherell et al., 2007)	LI = 19 (14 M, 5 F) TD = 99 (38 M, 61 F)	13.1–15.9	Retell

TD children? We measure the average emotional valence of words in a narrative using both the Evaluative Lexicon and the Affective Norms for English Words, each of which provides valence scores for many English words based on Likert-scale ratings from large samples of speakers. *Hypothesis 1:* We hypothesize that children with LI, compared to TD children, will produce narratives with a more negative valence.

- b) *Research Question 2 (RQ2):* Do children with LI produce narratives that are more or less emotionally intense than those produced by TD children? We quantify a narrative's average emotional intensity again using both the Evaluative Lexicon and the Affective Norms for English Words, which provide intensity scores for many words derived from Likert-scale ratings. *Hypothesis 2:* We hypothesize that children with LI will produce narratives with more or less (but not the same) emotional intensity compared to TD children.
- c) *Research Question 3 (RQ3):* Do children with LI produce narratives that are emotionally more angry, anxious, and sad than those produced by TD children? The degrees of anger, anxiety, and sadness in a narrative are assessed using Likert-scale ratings derived from artificial intelligence (ChatGPT-4). *Hypothesis 3:* We hypothesize that children with LI, relative to TD children, will tell angrier, more anxious, and sadder narratives.
- d) *Research Question 4 (RQ4):* In both children with LI and TD children, does a child's lexical diversity relate to the average emotional valence of their narrative? We measure a child's lexical diversity by computing the Moving-Average Type-Token Ratio of the narrative, which is then correlated with the emotional valence scores derived from the Likert-scale ratings (as described above). *Hypothesis 4:* We hypothesize that, in both children with LI and TD children, lower lexical diversity will correlate with more negatively valenced narratives.

2. Methods

2.1. Transparency and openness

The data analyses and hypotheses were preregistered; the preregistration can be found on the study's Open Science Framework website: https://osf.io/mjchn/?view_only=29d1b673ed0844faae50690fc9deb4f2. Deviations from the preregistration are indicated below by labeling certain analyses as unpreregistered. All data used in the data analyses are available on the study's Open Science Framework website. All language samples included in this study are available through the CHILDES TalkBank website: <https://childes.talkbank.org>.

2.2. Participants

There were 1197 children included in the analyses. The children's language samples were obtained from four corpora within the CHILDES Clinical English TalkBanks (MacWhinney & Snow, 1990): Gillam Corpus (Gillam & Pearson, 2004), ENNI Corpus (Schneider et al., 2006), EisenbergGuo Corpus (Eisenberg & Guo, 2013), and Conti-Ramsden Corpus (Wetherell et al., 2007). These four corpora were selected because they contained data from children with language impairment

(LI) and age-matched typically developing (TD) control children, all of whom performed the same narrative elicitation tasks within each corpus. The full datasets of the four corpora were included in the analyses. No participants were removed from the full dataset, with the sole exception of one subsample analysis that was conducted to equate the LI and TD groups in sample size and gender. Participants in these four corpora are described below in Table 1 and in prose in the subsequent paragraphs.

The Gillam Corpus consisted of 171 children with LI (100 males, 71 females) and 497 TD children (245 males, 252 females). The children with LI had a mean age of 8.0 years old (range=5.2–11.5 years), and the TD children had a mean age of 8.2 years old (range=5.0–11.9 years). The LI and TD groups did not differ in age, $t(666) = 0.91$, $p = .36$. However, the groups did differ in gender, $\chi^2(1, N = 668) = 4.30$, $p = .04$, with a higher proportion of males in the LI group. The children in this corpus were identified as LI or TD through the norming of the Test of Narrative Language (Gillam & Pearson, 2004).

The ENNI Corpus contained 77 children with LI (47 males, 30 females) and 300 TD children (150 males, 150 females). The children with LI were on average 7.1 years old (range=4.0–9.9 years), and the TD children were on average 7.1 years old (range=4.2–9.8 years). The two groups did not differ in age, $t(375) = 0.06$, $p = .95$, but there was a marginally significant difference in gender, $\chi^2(1, N = 377) = 2.99$, $p = .084$, reflecting a higher proportion of males in the LI group. The children with LI were diagnosed with SLI based on a score of lower than 85 on the Clinical Evaluation of Language Fundamentals-3 (CELF-3) (for children older than 6 years) or CELF-Preschool (CELF-P) (for children younger than 6 years) and based on SLPs having no concern about the child's cognitive abilities. Eighty-eight of the TD children completed the full CELF-3 or CELF-P and on average scored above 100; the rest of the TD children completed two subtests from the CELF-3 (Concepts & Directions and Recalling Sentences) or CELF-P (Linguistic Concepts and Recalling Sentences in Context).

The EisenbergGuo Corpus included 17 children with LI (10 males, 7 females) and 17 TD children (10 males, 7 females). The children with LI had a mean age of 3.4 years (range=3.1–3.8 years), and the TD children had a mean age of 3.5 years (range=3.0–3.9 years). The LI and TD groups did not differ in age, $t(32) = 0.32$, $p = .75$, or in gender, $\chi^2(1, N = 34) = 0.00$, $p = 1.0$. Among the 17 children with LI, 11 were previously diagnosed with SLI at age 2, though at the time of testing, many were scoring in the typical range on the Structured Photographic Expressive Language Test-P2 (SPELT-P2). The other 6 children with LI were diagnosed with SLI at the time of testing, scored more than $-.86$ standard deviations below the mean on the SPELT-P2, and were rated below the typical range on at least 4 items in a 14-item parent rating scale of children's language development. The 17 TD children scored better than $-.86$ standard deviations below the mean on the SPELT-P2 and were rated within the typical range on at least 12 of the 14 items on the parent rating scale.

The Conti-Ramsden Corpus, specifically Corpus 4, consisted of 19 children with LI (14 males, 5 females) and 99 TD children (38 males, 61 females). The children with LI had a mean age of 14.3 years (range=13.3–15.3 years), and the TD children had a mean age of 14.5 years (range=13.1–15.9 years). There was not a statistically significant

difference between groups in age (though note that the p -value was approaching marginal significance), $t(116) = 1.60, p = .11$. However, there was a statistically significant difference in gender ratio, $X^2(1, N = 118) = 8.06, p = .005$, due to the LI group having a larger proportion of males. All children with LI had a history of SLI, defined as achieving at one point in time a non-verbal IQ score of 85 or above and scoring at least 1 standard deviation below the mean on at least 1 standardized language assessment, such as the CELF. However, at the time of testing, some children were no longer in the typical range for non-verbal IQ, and some were now in the typical range for language.

3. Materials and procedure

3.1. Narrative tasks in language sample corpora

Participants across all four corpora completed narrative retell and/or formulation tasks. In the narrative retell tasks, a participant related a story after it was initially presented to the participant by another individual (e.g., *Frog, Where Are You?* story). In the narrative formulation tasks, a participant generated a story based on pictures that were shown to them.

In the *Gillam Corpus*, children performed three narrative tasks. One task was a narrative retell task (the McDonald's Restaurant story). In this story, children and their mother go to McDonald's and place an order, but then realize that their mother forgot her purse at home. Another task was a narrative formulation task that was elicited from a picture sequence and that followed an adult model (the Late for School story). In this story, a boy wakes up, eats breakfast, breaks a shoelace, misses his bus, and therefore arrives late to school. The last task was a narrative formulation task from an individual picture also following an adult model (the Aliens story). In this story, children discover an alien spaceship. All three tasks were compiled into one language sample per participant.

In the *ENNI Corpus*, children formulated two sets of narratives based on sequences of pictures. All narratives included two animals (such as a giraffe and an elephant) as main characters. Before the task, children first completed a training narrative. After the training, each child was presented with a picture-based narrative one page at a time. Then, again the child was presented with the picture-based narrative one page at a time, but this time the child was asked to narrate. The examiner did not tell the story during the first page-turning, so it was not a story retell task.

In the *EisenbergGuo Corpus*, children formulated a narrative for each of 15 individually presented pictures. These pictures included, for example, a beach scene with two boys fighting over a sand bucket and a woman running at them, and a kitchen scene in which a father is making toast, a mother is pouring milk, and two girls are nearby. For each picture, four elicitation prompts were provided and then further prompting occurred if the child did not produce an on-topic response. The first prompt was: "What is happening in the picture?" The second prompt was: "What else is happening in the picture?" For the third prompt, the examiner said, "Now I'll start the story and you finish it" (such as "The boy is trying to get the cookies and then..."). The fourth prompt was: "Tell me one more thing about the story."

In the *Conti-Ramsden Corpus*, children completed a narrative retell task. Specifically, children listened to and then re-told the *Frog, Where Are You?* story by Mercer Mayer. In this 24-page wordless picture book, a boy and his dog search for the boy's pet frog that escaped. The escapades include the dog falling out of a window, a gopher biting the boy on the nose, the dog barking at bees and the beehive falling down, an owl knocking down the boy, the boy climbing onto a deer's antlers, and the boy and the dog falling off the edge of a cliff.

3.1.1. Emotion measures

Valence of Emotions. A language sample's emotional valence was assessed through both the Evaluative Lexicon of the Lexical Suite

(Rocklage et al., 2018) and the Affective Norms for English Words from the Sentiment Analysis and Cognition Engine (Crossley et al., 2017). The Evaluative Lexicon identifies and assigns a valence score to adjectives in a language sample. Each of the 1541 adjectives in the database has a valence score that was derived from several native English-speaking raters who rated each word on a 0-very negative to 9-very positive scale. The average valence score of all identified adjectives in the sample is the valence score that was used in the current study's analyses. For example, if a child's sample included only two adjectives, *terrible* (valence=0.83) and *amazing* (valence=7.97), their valence score would be 4.4 (0.83 + 7.97 / 2). The Affective Norms for English Words identifies, in addition to adjectives, nouns and verbs in a sample. The 1034 words in the Affective Norms database have a valence score that was derived from ratings from several American university students; these judges rated words on a scale from 1-very unpleasant depicted by a frowning, unhappy figure to 9-very pleasant depicted by a smiling, happy figure. The average valence of all the identified words was the valence score used in the analyses in the Results section below. For instance, if a child's sample included the words *terrible* (valence=1.93), *cake* (valence=7.26), and *hurting* (valence=1.90), the valence score would be 3.70 (1.93 + 7.26 + 1.90 / 3). Fig. 1 presents a portion of a language sample from a child with LI and indicates words in the sample that have a valence in the Affective Norms database.

Intensity of Emotions. Emotional intensity was also assessed through the Evaluative Lexicon and the Affective Norms. Intensity was measured in a similar, scaled manner as valence (see above). The Evaluative Lexicon used a 0-not at all emotional to 9-very emotional scale to rate adjectives. The Affective Norms used a 1-calm to 9-excited scale to rate nouns, verbs, and adjectives. A language sample that contains only two words recognized by these dictionaries, namely *upset* (Affective Norms intensity=5.86; Evaluative Lexicon intensity=7.06) and *angry* (Affective Norms intensity=7.17; Evaluative Lexicon intensity=7.72), would have an intensity score of 6.52 (5.86 + 7.17 / 2) for the Affective Norms and 7.39 (7.06 + 7.72 / 2) for the Evaluative Lexicon.

Specific Emotions (Anger, Anxiety, and Sadness). We also measured the anger, anxiety, and sadness expressed in the language samples. The degree of these specific emotions was determined through ChatGPT-4 (OpenAI, 2023). ChatGPT has been shown to accurately assess emotions in a text by asking it simple questions without any pre-teaching,







and by the time he got to his school the teacher was very mad	  
angry he he was late	
and he was in big trouble	
so the teacher had to call his mom	
and the mom was very upset because of him being late	

Fig. 1. Valence Words in Narrative Language Sample from a Child with Language Impairment. Note. This figure depicts part of a language sample from a child with LI in the Gillam Corpus. In this part of the language sample, the child is telling the Late for School story. All the bolded words (in red and blue) are words in the Affective Norms database. The words are marked in red with a frown underneath if their valence is 3 or less on the 1-very unpleasant to 9-very pleasant scale. The words are marked in blue with a neutral face below if the valence is higher than 3.

such as “is this text negative?” (Rathje et al., 2024; see also Fatouros et al., 2023; Fu et al., 2024; Hur et al., 2024; Maharjan et al., 2025; Park & Kim, 2024; Yang et al., 2025 for additional evidence that ChatGPT can accurately identify emotional valence). For each language sample, the following three questions were posed to GPT, with the sample itself entered below each question: (1) “How angry is the person below on a scale of 1–7, with 1 being the least angry and 7 being the most angry?”, (2) “How anxious is the person below on a scale of 1–7, with 1 being the least anxious and 7 being the most anxious?”, and (3) “How sad is the person below on a scale of 1–7, with 1 being the least sad and 7 being the most sad?”. ChatGPT inferred the relative degree of the emotion in question (namely, anger, anxiety, or sadness) from linguistic cues (such as lexical and pragmatic patterns), and then mapped that inferred emotionality onto the 1–7 ordinal scales, relative to the scale anchors (1 =least, 7 =most).

3.1.2. Lexical diversity measure

The Moving-Average Type-Token Ratio (MATTR) was used to index lexical diversity (Covington & McFall, 2010). The MATTR calculates the Type-Token Ratio (i.e., the number of unique words divided by the total number of words) in a window that moves through the entire sample. The current MATTR used a window size of 20 words, so a Type-Token Ratio was computed for words 1 through 20, then 2 through 21, then 3 through 22, until the sample ended. Then the mean of all Type-Token Ratios was computed. The Moving-Average Type-Token Ratio is superior to the regular Type-Token Ratio and its variants, such as log TTR, because it is unaffected by varying language sample lengths (Covington & McFall, 2010). As would be expected, in all four corpora, the LI group had a smaller lexical diversity score than the TD group, all $p < .001$.

4. Results

The results are organized according to the four research questions (RQs) stated at the end of the Introduction. The headings below correspond to each RQ and whether the analysis was preregistered. Note that, as indicated in the preregistration, it was decided a priori that each corpus would be analyzed separately, rather than in an omnibus analysis, since the corpora differ from each other in several significant ways, including the types of narratives produced (i.e., retell versus formulation), the specific stories that were told, the number of participants, the ages of the children, and the criteria used to classify the children as LI or TD. In the regression analyses below, both gender and age are included in the models to control for any differences between language groups (i.e., LI vs TD) in gender and/or age, and to identify interactions among language group, gender, and age.

RQ1. Do children with LI produce more emotionally negative stories than TD children?

4.1. Comparing LI & TD in emotional valence: preregistered

First, we examined whether LI and TD children differed in the emotional valence of their narratives. In the *Gillam Corpus*, the largest corpus of the study, children with LI ($N = 171$) ages 5–12 produced narrative samples that were more emotionally negative than those produced by TD children ($N = 497$). Specifically, children with LI demonstrated a significantly lower valence (i.e., more negativity and less positivity) score than TD children on both the Evaluative Lexicon metric, $t(555) = 3.40$, $p < .001$, Cohen’s $d = .34$, and the Affective Norms metric, $t(666) = 2.52$, $p = .01$, Cohen’s $d = .22$. (Note that the degrees of freedom in the above analyses are smaller for the Evaluative Lexicon than for the Affective Norms because the Evaluative Lexicon analyzes only adjectives, and not all children produced relevant adjectives.)

While increased negativity was observed in children with LI in the *Gillam Corpus*, it was not evident in the other three, smaller corpora

(ENNI, EisenbergGuo, and Conti-Ramsden). In the *ENNI Corpus*, children with LI ($N = 77$) and TD children ($N = 300$) did not differ in valence on either the Affective Norms or the Evaluative Lexicon, both $t < 1.5$ and $p > .1$. Similarly, in the *EisenbergGuo Corpus*, children with LI ($N = 17$) and TD children ($N = 17$) did not differ in valence on either the Evaluative Lexicon or the Affective Norms, both $t < 1.5$ and $p > .1$. However, in the *Conti-Ramsden Corpus*, surprisingly, children with LI ($N = 19$) demonstrated significantly higher valence scores than TD children ($N = 99$) on the Affective Norms, $t(116) = 2.10$, $p = .04$, Cohen’s $d = .53$, but there was no difference on the Evaluative Lexicon metric, $t(91) = .80$, $p = .43$, Cohen’s $d = .22$. The full set of analyses for emotional valence are provided in Appendix A below.

Because the *ENNI Corpus* in particular *both* failed to reveal a significant difference between groups *and* had very unequal sample sizes between groups (which can meaningfully reduce statistical power and mask a true effect), we decided to conduct additional analyses with matched sample sizes. Specifically, in these post-hoc, unregistered analyses of the ENNI Corpus dataset, we randomly sampled from the larger group ($N = 300$) to attain a group size that is identical to the smaller group’s size ($N = 77$) – by doing so, we were also able to fully match the groups in gender (47 males and 30 females in both groups). Then, we compared the groups on the Affective Norms and the Evaluative Lexicon. In this matched subsample, the LI group had increased negativity compared to the TD group on the Affective Norms, $t(152) = 2.79$, $p = .006$, Cohen’s $d = .45$, but this group difference did not reach significance on the Evaluative Lexicon, $t(150) = 1.52$, $p = .13$, Cohen’s $d = .25$.

4.2. Comparing LI & TD in emotional valence with age & gender in model: unregistered

To account for effects of age and gender in the above analyses, linear regressions were then conducted, with language group (LI vs. TD), gender (M vs. F), and age (continuous and centered) as the predictor variables, and with valence from the Evaluative Lexicon and the Affective Norms as the dependent variables.

In the *Gillam Corpus*, the language group effect, whereby children with LI produced more negative stories than TD children, remained significant for the Evaluative Lexicon, $\beta = .15$, $p = .01$, but fell short of significance for the Affective Norms, $\beta = .09$, $p = .13$.^{1,2} Fig. 2 below displays valence scores from the Evaluative Lexicon broken down by language group as well as age and gender.

In the *ENNI Corpus* and *EisenbergGuo Corpus*, there were no significant effects, as was observed above without gender and age in the model (though note that the ENNI corpus did produce a significant difference in the subset analysis above that matched groups on sample size and gender). In the *Conti-Ramsden Corpus*, with age in the model (gender was not in the model because gender data were not available), the unexpected higher positive valence for children with LI on the Affective Norms metric remained significant, $\beta = .19$, $p = .045$, with no other results reaching significance.

RQ2. Do children with LI and TD children differ in the emotional intensity of their stories?

¹ In the *Gillam* models, there was also an age by gender interaction for the Evaluative Lexicon score, which was marginally significant, $\beta = .28$, $p = .07$. Importantly, there was no 3-way language group by age by gender interaction.

² Even though the language group differences in the Evaluative Lexicon score remained significant when gender was entered into the model, to fully rule out the possibility that the gender distribution difference between language groups was a meaningful confounding variable, we conducted two additional *t*-tests – in one, LI males were compared to TD males, and in the other, LI females were compared to TD females. In both analyses, LI children produced significantly increased negativity in their narratives relative to their TD counterparts ($p = .017$ in the males-only analysis and $p = .016$ in the females-only analysis).

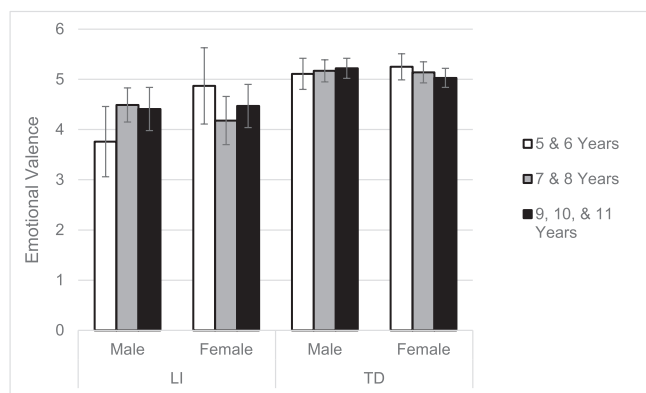


Fig. 2. Emotional Valence by Language Group, Gender, and Age. Note. This figure displays emotional valence scores from the Evaluative Lexicon in the Gillam corpus categorized by language group (LI vs. TD), gender (Male vs. Female), and age (5- and 6-year-olds vs. 7- and 8-year-olds vs. 9- and 10-year-olds). A lower score corresponds to a more negative narrative. Error bars indicate the standard error of the mean.

4.2.1. Comparing LI & TD in emotional intensity: preregistered

Next, we analyzed emotional intensity, which produced mixed results. In the Gillam Corpus, children with LI produced *less* emotionally intense narratives than TD children on the Affective Norms metric, $t(666) = 3.43$, $p < .001$, Cohen's $d = .30$, but there was no significant difference on the Evaluative Lexicon metric, $t(555) = 0.26$, $p = .80$, Cohen's $d = .03$. However, in the Conti-Ramsden Corpus, children with LI produced *more* emotionally intense narratives than TD children on the Affective Norms, $t(116) = 2.20$, $p = .03$, Cohen's $d = .55$, but there was no significant difference on the Evaluative Lexicon metric, $t(91) = 0.04$, $p = .97$, Cohen's $d = .01$.

In the ENNI Corpus and EisenbergGuo Corpus, there were no significant differences between children with LI and TD children on the Affective Norms or Evaluative Lexicon, all four t s < 1.5 and p s > 0.1 . The entire set of analyses for emotional intensity can be seen in Appendix B.

4.2.2. Comparing LI & TD in emotional intensity with age & gender in model: unregistered

To control for gender and age in the emotional intensity analyses, linear regressions were then conducted, with language group, gender and age as predictor variables and intensity from the Evaluative Lexicon and the Affective Norms as the dependent variables.

The *decreased* emotional intensity seen in the children with LI on the Affective Norms metric in the Gillam Corpus remained significant, $\beta = .14$, $p = .02$, as did the *increased* emotional intensity seen in the children with LI in the Conti-Ramsden Corpus, $\beta = .25$, $p = .01$. No other significant results related to language group in either of these two corpora or in ENNI Corpus and EisenbergGuo Corpus.³

RQ3. Do children with LI produce sadder, angrier, and more anxious stories than TD children?

4.2.3. Comparing LI & TD in sadness, anger, & anxiety: preregistered

Regarding the specific emotions of sadness, anger, and anxiety, in the Gillam Corpus, children with LI, relative to TD children, produced language samples that, according to ChatGPT, were significantly sadder, $t(666) = 2.73$, $p < .01$, Cohen's $d = .24$, and angrier, $t(666) = 4.71$, $p < .001$, Cohen's $d = .42$. Moreover, the anxiety metric approached

³ Two other, marginally significant results emerged. In the ENNI Corpus, gender had a marginal effect on the Affective Norms, $\beta = .24$, $p = .08$, and in the EisenbergGuo Corpus, there was a marginal age by gender interaction on the Evaluative Lexicon, $\beta = .73$, $p = .097$.

marginal significance, $t(666) = 1.62$, $p = .11$, Cohen's $d = .14$.

However, in the three other corpora (i.e., the ENNI Corpus, Eisenberg Corpus, and Conti-Ramsden Corpus) the sadness, anger, and anxiety measures did not significantly differ between children with LI and TD children, all t s < 1.75 and p s > 0.1 . All analyses for sadness, anger, and anxiety can be found in Appendix C.

4.2.4. Comparing LI & TD in sadness, anger, & anxiety with age & gender in model: unregistered

With age and gender included in linear regression analyses of sadness, anger, and anxiety, in the Gillam Corpus, children with LI's increased sadness, $\beta = .13$, $p = .02$, and anger, $\beta = .20$, $p < .001$, remained significant, but the anxiety result was no longer approaching marginal significance, $\beta = .04$, $p = .44$.

In the Gillam Corpus, there was also a language group by age interaction for sadness, $\beta = .35$, $p = .01$; a marginal language group by age by gender interaction, $\beta = .24$, $p = .051$, for anger; and a marginal language group by age by gender interaction, $\beta = .25$, $p = .054$, for anxiety.⁴ These language group interactions appeared to be driven largely by especially high sadness, anger, and anxiety scores in LI males in the younger years (5–6 years old) and middle years (7–8 years old).

In the ENNI Corpus and EisenbergGuo Corpus, there were no significant results related to language group.⁵

In the Conti-Ramsden Corpus, there was a language group by age interaction, $\beta = .51$, $p = .0496$, for sadness; a marginal effect of language group, $\beta = .16$, $p = .07$, and a language group by age interaction, $\beta = .80$, $p < .01$, for anger; and a language group by age interaction, $\beta = .80$, $p < .01$, for anxiety, $\beta = .55$, $p = .04$. The language group interactions for sadness, anger, and anxiety seemed to generally reflect the TD children having higher sadness, anger, and anxiety scores at ages 13 and 14 with the pattern reversing at age 15, at which point children with LI had higher sadness, anger, and anxiety scores.

RQ4. Is Lexical Diversity Associated with Emotionally Negative Stories?

4.2.5. Correlating lexical diversity and emotional valence within LI & TD groups: unregistered

In the final set of analyses, we examined whether our primary measure (i.e., valence) was associated with lexical diversity scores, as indexed by the Moving-Average Type-Token Ratio. Correlations were computed between lexical diversity and valence from both the Evaluative Lexicon and the Affective Norms separately for the LI and TD groups and separately for the four corpora. Three correlations reached statistical significance. All were positive, indicating that higher lexical diversity was associated with positive emotion, or, in other words, lower lexical diversity was associated with negative emotion. In the Gillam Corpus, in the TD group, there was a significant positive correlation of $r(495) = .10$, $p = .03$ with the Affective Norms measure. In the ENNI Corpus, there was a significant positive correlation in both the TD group, $r(295) = .16$, $p < .01$, and the LI group, $r(73) = .29$, $p = .01$, with the Evaluative Lexicon measure. None of the other correlations reached significance.

While not reaching conventional statistical significance ($p > .05$), we observed some positive correlations with at least small effect sizes (Pearson's r s $\geq .1$), suggesting at least a minor association between lower lexical diversity and more negative emotion. Pearson's $r \geq .1$ was chosen because it is typically considered the lower limit for a small correlation. These correlations were as follows. In the TD group, there was a

⁴ In the Gillam Corpus, there was also an age effect, $\beta = .41$, $p < .01$, for anger; an age effect, $\beta = .28$, $p = .04$, and an age by gender interaction, $\beta = .27$, $p = .04$, for anxiety.

⁵ In the Eisenberg Corpus, there was a marginal effect of age, $\beta = .64$, $p = .09$, for anxiety.

Table 2
Positive Correlations between Lower Lexical Diversity and Lower Emotional Valence.

Corpus	LI	TD
Gillam		☑
ENNI	☑	☑
EisenbergGuo	☑	☑
Conti-Ramsden	☑	

Note. This table indicates, through a check mark, that there was a positive correlation between lexical diversity and emotional valence (either from the Evaluative Lexicon or the Affective Norms) that was either Pearson's $r \geq .1$ or statistically significant ($p < .05$). That is, there was evidence for an association between reduced lexical diversity and more emotional negativity in 7 out of the 8 groups (i.e., 4 corpora, each with LI and TD children).

correlation of $r = .39$ in the EisenbergGuo Corpus with the Evaluative Lexicon measure. In the LI group, there was a correlation of $r = .15$ in the EisenbergGuo Corpus with the Affective Norms measure, and a correlation of $r = .23$ in the Conti-Ramsden Corpus with the Evaluative Lexicon measure. [Table 2](#) below indicates, for each corpus and for both LI and TD groups, whether there was evidence for an association between lower lexical diversity and more negative emotion, defined as a positive correlation between lexical diversity and emotional valence (either the Evaluative Lexicon or the Affective Norms) that was either Pearson's $r \geq .1$ or statistically significant ($p < .05$).

It should also be noted that, though not statistically significant, there were also two negative correlations of $r \geq .1$; in the TD group, there was a $r = -.14$ in the EisenbergGuo Corpus with the Affective Norms measure, and in the LI group, there was a $r = -.24$ in the EisenbergGuo Corpus with the Evaluative Lexicon measure. The full set of correlations is provided in Appendix D.

Discussion

The current investigation examined the use of emotion words in narratives produced by children with LI and TD children. The first and primary research question (RQ1) was whether children with LI narrated stories in a more emotionally negative manner than TD children. In the largest corpus (Gillam), containing more than half of the children in the current study (668 out of the 1197 children), children with LI narrated events more negatively than TD children according to the Evaluative Lexicon, but not the Affective Norms when gender and age were considered. In the second largest corpus (ENNI), there was also evidence that children with LI narrated events more emotionally negatively than TD children, but only according to the Affective Norms and only in a subset analysis that equated the groups in sample size and gender. In the two smaller corpora, increased negativity in children with LI was not observed and surprisingly this effect was reversed in one (Conti-Ramsden) of these two corpora (though only on the Affective Norms measure). The second research question (RQ2) examined whether children with LI differ from TD in the emotional intensity of their narratives. Results were mixed for emotional intensity, with higher intensity, lower intensity, and similar intensity seen in the LI group compared to the TD group, dependent on the corpora. The third research question (RQ3) asked whether children with LI produced sadder, angrier, and more anxious narratives. The children with LI demonstrated higher scores for these specific negative emotions in the largest corpus, but language status interacted with age and gender, and results were absent or mixed in the other corpora. The fourth and final question (RQ4) was whether a child's lexical diversity (as a proxy for vocabulary ability) was associated with negatively valenced narratives. There was some evidence of associations within both the LI group and the TD group between lexical diversity and emotional negativity, with less lexical diversity correlating

with more emotional negativity. These correlations, however, were mostly small and sometimes non-existent or in the opposite direction.

Why in some of our analyses did we find that children with LI produced narrative samples that were more negative? An explanation that aligns with previous research is that children with LI are more likely to experience clinical or sub-clinical levels of anxiety or depression (e.g., [Burnley et al., 2023](#); [Conti-Ramsden & Botting, 2008](#); [van den Bedem et al., 2018](#)). If these children are experiencing a disproportionately high degree of negative emotions in their daily life, then they may perceive events less favorably and use negative words more frequently, such as *angry*, *trouble*, and *hate*. There are, however, other possible accounts for the negative narratives. For example, a language testing context is likely to be stressful for children with LI, resulting in a feeling of acute unhappiness in this situation. Thus, narratives for such children may not be negative when they are in a different context, such as telling a story to friends at school. Relatedly, regardless of context, producing multiple sentences to form a coherent and cohesive story could be frustrating for children, especially those with LI, and this frustration may lower a child's mood. Therefore, children with LI's stories may be negative in all narrative contexts, but they may not have low mood in non-narrative contexts. These speculations require follow-up studies.

Why in some of our other analyses did children with LI *not* produce overly negative language samples? The two corpora in which there was no evidence for increased negativity in LI children (namely, the EisenbergGuo and Conti-Ramsden corpora) may not have provided sufficient data. The EisenbergGuo and Conti-Ramsden corpora contained fewer participants (17 and 19 children with LI, respectively), more than 8 times less than the largest corpus. Moreover, in these corpora, some children in the LI group no longer had an LI diagnosis at the time of testing, and thus classification into groups was not entirely accurate. In addition to small samples and misclassification, there were several other factors that could have accounted for the lack of an LI negativity effect in the EisenbergGuo and Conti-Ramsden corpora, including the types of narratives produced, the ages of the participants, and the testing context.

Interestingly, there was also some evidence of both increased and decreased emotional intensity in children with LI. While future work is required to replicate and better understand these findings, both intense anger and flattened affect can be symptoms of mental health issues and may be reflected in children with LI's language samples. Indeed, both fury and anhedonia can be part of major depressive disorder ([Busch, 2009](#); [Treadway & Zald, 2011](#)). The decreased emotional intensity observed in the largest corpus is more likely to be replicable than the increased emotional intensity demonstrated in the much smaller corpus, given that small sample sizes can produce unreliable results. Future work is needed to assess reliability.

The relation between language abilities and emotional negativity in the current study was observed in some children with LI, *but also in some*

TD children, as lower lexical diversity in both the TD and LI groups was associated, in certain correlational analyses, with more emotional negativity in narratives. This association suggests that children with lower vocabularies and by extension lower language abilities – even if such children are typically developing and do not have LI – may experience more emotional negativity, and that such negativity may be reflected in their narration of events. This finding is consistent with previous research showing that children with smaller vocabularies demonstrate increased emotional symptoms, hyperactivity/inattention, and conduct issues (Westrupp et al., 2020). However, this association was not observed in several of the analyses, which may partly reflect the imprecision of measuring vocabulary size and emotional negativity from a short language sample.

While the current study is the first to compare children with LI to TD children in the emotional valence of narratives, there are relevant prior studies, and findings from the current study are both consistent and contrastive with findings from these prior studies. Specifically, our findings of more emotional negativity in LI children are consistent with two bodies of research. They are consistent with research showing that children with LI experience more anxiety and depression, as both anxiety and depression are highly correlated with increased negativity (see Botting et al., 2016; Forrest et al., 2021; van den Bedem et al., 2018). Moreover, our findings of increased negativity in LI can also be interpreted as broadly consistent with prior studies showing that negatively valenced words in language samples accurately reflect the negative emotions associated with anxiety and depression (see Bernard et al., 2016; Jarrold et al., 2011; Rook et al., 2022; Rude et al., 2004; Sonnenschein et al., 2018; Tølbøll, 2019). However, we failed to find increase negativity in LI children in some of our analyses, and these particular null findings might be interpreted to contradict either or both of these previous bodies of research – namely, that LI children have increased anxiety and depression and/or that language samples accurately reflect anxiety and depression.

4.3. Limitations, future directions, and implications

The current study has multiple limitations that should be addressed through future research. One critical limitation is that measures of mental health were not included, as we were restricted to secondary data analysis of existing datasets. In future work, researchers should determine whether mental health assessments, such as the Behavior Assessment System for Children, correlate with the negativity expressed in children with LI's narratives. Another limitation is that the emotional language measures were crude. The primary measures covered the very broad concept of emotional negativity, though sadness, anger, and anxiety were specifically assessed. Follow-up studies can determine which kinds of negative emotions – jealousy, loneliness, shame, contempt, helplessness, etc. – are more frequent in children with LI's narratives. Moreover, negative emotion was measured in the current study by adjective, noun, and verb word choice. However, negative emotion, as in the case of depression, can also be expressed in less variability in pitch, a slower rate of speech, and more use of first-person singular pronouns (Koops et al., 2023). Furthermore, while we focused on word choice in the current study, some important characteristics of words, such as word frequency, were not considered but could provide valuable information. For example, the effect of emotional intensity on word recognition has been found to depend on how rare the word is (Kuperman et al., 2014), and individuals with and without LI likely differ in their use of rare words. Thus, in future research, additional speech and language variables, such as prosody and word frequency, should be considered. Another significant limitation involves the unequal sample sizes between groups in most of the corpora, with considerably more children in the TD groups than the LI groups. While an important limitation, the statistical tests used in the current study are generally robust to unequal sample sizes, especially when the sample sizes are large. Moreover, for one of the corpora (which had very

unbalanced groups), we conducted a post-hoc analysis on a small subset of the participants in order to match the groups in sample size. Nevertheless, large groups that are similar in size are ideal, and though it is often difficult to find a large enough clinical population to match the typically developing group, future work should aim to test large and more balanced groups.

There are several implications of the current work for both clinical psychology and speech-language pathology. For the field of clinical psychology, an emotion-based narrative language sample analysis can potentially be used in assessment protocols for identifying and potentially diagnosing anxiety and depression in children with LI, as well as for measuring symptomatic improvement related to anxiety and depression. Self-report of mental health symptoms may not always be valid in children (Perrin & Last, 1992), even more so in children with LI. Thus narrative language sample analyses could help determine whether a child has symptoms of anxiety or depression. For speech-language pathology, therapists who are conducting language sample analysis to diagnose LI or to track progress in therapy could refer children with negatively valenced narratives to a mental health counselor for further evaluation. Speech-language pathologists themselves can obtain training on counseling services, such as cognitive behavioral therapy (e.g., Menzies et al., 2009). Thus, speech-language pathologists' intervention approaches could aim to simultaneously enhance both communicative skills and emotional health in children with LI. Indeed, narrative-based therapies have been found to not only enhance language abilities (Hessling & Schuele, 2020; Petersen et al., 2010), but also to enhance emotional intelligence (Beaudoin et al., 2016; Koivula et al., 2020). For instance, in a recent study (Tarshis et al., 2024), when bilingual children and adolescents diagnosed with Developmental Language Disorder participated in a 10-session oral narrative intervention, their language skills (e.g., story coherence) improved, as did certain skills related to emotional intelligence (e.g., social competence).

5. Conclusion

By conducting emotion-based language sample analyses of children with LI and TD, we aimed to gain a better understanding of how child language ability relates to negative emotions. Ultimately, elucidating the relation between language and emotion in children will allow speech-language pathologists to be better equipped to identify emotional challenges earlier on in treatment with children, and to provide appropriate referrals to and more effectively collaborate with mental health counselors to maximize treatment outcomes. Emotions that are expressed through language and identified via language sample analysis can provide valuable insights into communication and mental health disorders.

CRedit authorship contribution statement

Scott R. Schroeder: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Max R. Freeman:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Glennys A. Cuatlaacatl:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Journeyne Cunningham:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Natalie M. Jordan:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Alexandra K. Occhiuto:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Scott R. Schroeder reports a relationship with Hofstra University that includes: employment. Max R. Freeman reports a relationship with St

John's University that includes: employment. If there are other authors, they declare that they have no known competing financial interests or

personal relationships that could have appeared to influence the work reported in this paper

Appendix A

All Valence Results

Comparison	Corpus	Measure	t	df	p	Cohen's d
LI vs. TD	Gillam	Evaluative Lexicon	3.40	555	< .001	.34
LI vs. TD	ENNI	Evaluative Lexicon	1.33	370	.18	.17
LI vs. TD	EisenbergGuo	Evaluative Lexicon	1.34	30	.19	.47
LI vs. TD	Conti-Ramsden	Evaluative Lexicon	0.80	91	.43	.22
LI vs. TD	Gillam	Affective Norms	2.52	666	.01	.22
LI vs. TD	ENNI	Affective Norms	1.58	373	.12	.20
LI vs. TD	EisenbergGuo	Affective Norms	0.39	32	.70	.13
LI vs. TD	Conti-Ramsden	Affective Norms	2.10	116	.04	.53

Appendix B

All Intensity Results

Comparison	Corpus	Measure	t	df	p	Cohen's d
LI vs. TD	Gillam	Evaluative Lexicon	0.26	555	.80	.03
LI vs. TD	ENNI	Evaluative Lexicon	1.28	370	.20	.17
LI vs. TD	EisenbergGuo	Evaluative Lexicon	.11	30	.91	.04
LI vs. TD	Conti-Ramsden	Evaluative Lexicon	0.04	91	.97	.01
LI vs. TD	Gillam	Affective Norms	3.43	555	< .001	.30
LI vs. TD	ENNI	Affective Norms	0.56	373	.58	.07
LI vs. TD	EisenbergGuo	Affective Norms	0.02	32	.98	.01
LI vs. TD	Conti-Ramsden	Affective Norms	2.20	116	.03	.55

Appendix C

All Sadness, Anger, and Anxiety Results

Comparison	Corpus	Measure	t	df	p	Cohen's d
LI vs. TD	Gillam	Sadness	2.73	666	.006	.24
LI vs. TD	ENNI	Sadness	0.24	375	.81	.03
LI vs. TD	EisenbergGuo	Sadness	0.14	32	.89	.05
LI vs. TD	Conti-Ramsden	Sadness	0.37	116	.71	.09
LI vs. TD	Gillam	Anger	4.71	666	< .001	.42
LI vs. TD	ENNI	Anger	1.53	375	.13	.20
LI vs. TD	EisenbergGuo	Anger	0.00	32	1.00	.00
LI vs. TD	Conti-Ramsden	Anger	1.38	116	.17	.35
LI vs. TD	Gillam	Anxiety	1.62	666	.11	.14
LI vs. TD	ENNI	Anxiety	0.20	374	.84	.03
LI vs. TD	EisenbergGuo	Anxiety	0.59	32	.56	.20
LI vs. TD	Conti-Ramsden	Anxiety	0.46	116	.64	.12

Appendix D

Correlations between Lexical Diversity and Emotional Valence

Corpus	LI		TD	
	Evaluative Lexicon	Affective Norms	Evaluative Lexicon	Affective Norms
Gillam	.05	.09	.05	.10*
ENNI	.29*	-.04	.16*	.03
EisenbergGuo	-.24*	.15	.39	-.14*
Conti-Ramsden	.23	.03	-.06	.05

Note. The bold font indicates a positive correlation that is $r \geq .1$. Asterisks indicate statistical significance ($p < .05$).

Data availability

All of the data used in the data analyses are available on the study's Open Science Framework website: https://osf.io/mjchn/?view_only=29d1b673ed0844faae50690fc9deb4f2

References

- Beaudoin, M. N., Moersch, M., & Evare, B. S. (2016). The effectiveness of narrative therapy with children's social and emotional skill development: an empirical study of 813 problem-solving stories. *Journal of Systemic Therapies*, 35(3), 42–59.
- Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., Inglis, A., & Lencee, W. (1996). Long-term consistency in speech/language profiles: II. Behavioral, emotional, and social outcomes. *Journal of the American Academy of Child Adolescent Psychiatry*, 35(6), 815–825.
- Bernard, J. D., Baddeley, J. L., Rodriguez, B. F., & Burke, P. A. (2016). Depression, language, and affect: An examination of the influence of baseline depression and affect induction on language. *Journal of Language and Social Psychology*, 35(3), 317–326.
- Bernard, R., Hofslundengen, H., & Frazier Norbury, C. (2022). Anxiety and depression symptoms in children and adolescents who stutter: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, 65(2), 624–644.
- Botting, N., Toseeb, U., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2016). Depression and anxiety change from adolescence to adulthood in individuals with and without language impairment. *PLoS One*, 11(7), Article e0156678.
- Burnley, A., St. Clair, M., Bedford, R., Wren, Y., & Dack, C. (2023). Understanding the prevalence and manifestation of anxiety and other socio-emotional and behavioural difficulties in children with Developmental Language Disorder. *Journal of Neurodevelopmental Disorders*, 15(1), 17.
- Busch, F. N. (2009). Anger and depression. *Advances in Psychiatric Treatment*, 15(4), 271–278.
- Carson, D. K., Klee, T., Perry, C. K., Muskina, G., & Donaghy, T. (1998). Comparisons of children with delayed and normal language at 24 months of age on measures of behavioural difficulties, social and cognitive development. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health*, 19(1), 59–75.
- Conti-Ramsden, G., & Botting, N. (2008). Emotional health in adolescents with and without a history of specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, 49(5), 516–525.
- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type – token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100.
- 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, 803–821.
- Eisenberg, S. L., & Guo, L. Y. (2013). Differentiating children with and without language impairment based on grammaticality. *Language, Speech, and Hearing Services in Schools*, 44(1), 20–31.
- Fatouros, G., Soldatos, J., Kouroumalis, K., Makridis, G., & Kyriazis, D. (2023). Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, 14, Article 100508.
- Forrest, C. L., Gibson, J. L., & St Clair, M. C. (2021). Social functioning as a mediator between developmental language disorder (DLD) and emotional problems in adolescents. *International Journal of Environmental Research and Public Health*, 18(3), 1221.
- Fu, Z., Hsu, Y. C., Chan, C. S., Lau, C. M., Liu, J., & Yip, P. S. F. (2024). Efficacy of ChatGPT in Cantonese sentiment analysis: comparative study. *Journal of Medical Internet Research*, 26, Article e51069.
- Ghandour, R. M., Sherman, L. J., Vladutiu, C. J., Ali, M. M., Lynch, S. E., Bitsko, R. H., & Blumberg, S. J. (2019). Prevalence and treatment of depression, anxiety, and conduct problems in US children. *The Journal of Pediatrics*, 206, 256–267.
- Gillam, R.B., & Pearson, N.A. (2004). *TNL: Test of narrative language*. Austin, TX: Pro-ed.
- Hessling, A., & Schuele, C. M. (2020). Individualized narrative intervention for school-age children with specific language impairment. In *Language, Speech, and Hearing Services in Schools*, 51 pp. 687–705.
- Hur, J. K., Heffner, J., Feng, G. W., Joormann, J., & Rutledge, R. B. (2024). Language sentiment predicts changes in depressive symptoms. In *Proceedings of the National Academy of Sciences*, 121.
- Jarrold, W., Javitz, H. S., Krasnow, R., Peintner, B., Yeh, E., Swan, G. E., & Mehl, M. (2011). Depression and self-focused language in structured interviews with older men. *Psychological Reports*, 109(2), 686–700.
- Kerns, J. G., Docherty, A. R., & Martin, E. A. (2008). Social and physical anhedonia and valence and arousal aspects of emotional experience. *Journal of Abnormal Psychology*, 117(4), 735.
- Knuppenburg, R. M., & Fredericks, C. M. (2021). Linguistic affect: Positive and negative emotion words are contagious, predict likability, and moderate positive and negative affect. *Inquiries Journal*, 13(03).
- Koivula, M., Turja, L., & Laakso, M. L. (2020). Using the storytelling method to hear children's perspectives and promote their social-emotional competence. *Journal of Early Intervention*, 42(2), 163–181.
- Koops, S., Brederoo, S. G., de Boer, J. N., Nadema, F. G., Voppel, A. E., & Sommer, I. E. (2023). Speech as a biomarker for depression. *CNS Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS Neurological Disorders)*, 22(2), 152–160.
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, 143(3), 1065–1081.
- Lee, B. T., Seok, J. H., Lee, B. C., Cho, S. W., Yoon, B. J., Lee, K. U., & Ham, B. J. (2008). Neural correlates of affective processing in response to sad and angry facial stimuli in patients with major depressive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 32(3), 778–785.
- MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language*, 17(2), 457–472.
- Maharjan, J., Jin, R., Zhu, J., & Kenne, D. (2025). Psychometric evaluation of large language model embeddings for personality trait prediction. *Journal of Medical Internet Research*, 27, Article e75347.
- McAllister, J., Skinner, J., Hayhow, R., Heron, J., & Wren, Y. (2023). The association between atypical speech development and adolescent self-harm. *Journal of Speech, Language, and Hearing Research*, 66(5), 1600–1617.
- Menzies, R. G., Onslow, M., Packman, A., & O'Brian, S. (2009). Cognitive behavior therapy for adults who stutter: A tutorial for speech-language pathologists. *Journal of Fluency Disorders*, 34(3), 187–200.
- Meyerhoff, J., Liu, T., Stamatis, C. A., Liu, T., Wang, H., Meng, Y., & Mohr, D. C. (2023). Analyzing text message linguistic features: Do people with depression communicate differently with their close and non-close contacts? *Behaviour Research and Therapy*, 166, Article 104342.
- Nalipay, M. J. N., & Mordeno, I. G. (2016). Emotional intensity of trauma memory as moderator of the relationship between posttraumatic cognitions and PTSD symptoms. *Journal of Pacific Rim Psychology*, 10, Article e7.
- Nook, E. C., Hull, T. D., Nock, M. K., & Somerville, L. H. (2022). Linguistic measures of psychological distance track symptom levels and treatment outcomes in a large set of psychotherapy transcripts. In *Proceedings of the National Academy of Sciences*, 119.
- Nook, E. C., Vidal Bustamante, C. M., Cho, H. Y., & Somerville, L. H. (2020). Use of linguistic distancing and cognitive reappraisal strategies during emotion regulation in children, adolescents, and young adults. *Emotion*, 20(4), 525.
- OpenAI. (2023). *ChatGPT (Mar 14 version) [Large language model]*. (<https://chat.openai.com/chat>).
- Park, C., & Kim, J. (2024). Exploring affective representations in emotional narratives: An exploratory study comparing ChatGPT and human responses. *Cyberpsychology, Behavior, and Social Networking*, 27(10), 736–741.
- Perrin, S., & Last, C. G. (1992). Do childhood anxiety measures measure anxiety? *Journal of Abnormal Child Psychology*, 20, 567–578.
- Rathje, S., Mirea, D. M., Sucholutsky, I., Marjeh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. In *Proceedings of the National Academy of Sciences*, 121.
- Redmond, S. M., & Rice, M. L. (1998). The socioemotional behaviors of children with SLI: Social adaptation or social deviance? *Journal of Speech, Language, and Hearing Research*, 41(3), 688–700.
- Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior Research Methods*, 50, 1327–1344.
- Rook, L., Mazza, M. C., Lefter, I., & Brazier, F. (2022). Toward linguistic recognition of generalized anxiety disorder. *Frontiers in Digital Health*, 4, Article 779039.
- Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition Emotion*, 18(8), 1121–1133.
- Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton narrative norms instrument. *Journal of Speech Language Pathology and Audiology*, 30(4), 224.
- Sonnenschein, A. R., Hofmann, S. G., Ziegelmayer, T., & Lutz, W. (2018). Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. *Cognitive Behaviour Therapy*, 47(4), 315–327.
- Tarshis, M. A., McGrath, M. S. S., LSW, K., Barresi, M. A., Battino, M. S., Henderson MSED, M. S., ... Bonuck PhD, K. (2024). Impact of a Narrative Language Intervention on Language, Behavior, and Self-Concept among Bilingual Children with Developmental Language Disorder. *Developmental Disabilities Network Journal*, 4(2), 10.
- Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neuroscience Biobehavioral Reviews*, 35(3), 537–555.
- Tølbøll, K. B. (2019). Linguistic features in depression: a meta-analysis. *Journal of Language Works-Sprogvidenskabeligt Studentertidsskrift*, 4(2), 39–59.
- van den Bedem, N. P., Dockrell, J. E., van Alphen, P. M., de Rooij, M., Samson, A. C., Harjunen, E. L., & Rieffe, C. (2018). Depressive symptoms and emotion regulation strategies in children with and without developmental language disorder: a longitudinal study. *International Journal of Language Communication Disorders*, 53(6), 1110–1123.
- Westrupp, E. M., Reilly, S., McKean, C., Law, J., Mensah, F., & Nicholson, J. M. (2020). Vocabulary development and trajectories of behavioral and emotional difficulties via academic ability and peer problems. *Child Development*, 91(2), e365–e382.
- Wetherell, D., Botting, N., & Conti-Ramsden, G. (2007). Narrative in adolescent specific language impairment (SLI): A comparison with peers across two different narrative genres. *International Journal of Language Communication Disorders*, 42(5), 583–605.
- Wilmot, A., Boyes, M., Sievers, R., Leitao, S., & Norbury, C. (2024). Impact of developmental language disorders on mental health and well-being across the lifespan: Lived experience perspectives.
- Yang, K., Deng, R., Wei, Y., & Wang, S. (2025). The power of ChatGPT in processing text: Evidence from analysis and prediction in the exchange rate markets. *Financial Innovation*, 11(1), 118.
- Zupan, B., Hutchings, S. M., Everitt, L. E., & Gupta, C. (2022). Language disorder and internalizing mental health problems in youth offenders: A systematic review. *International Journal of Language Communication Disorders*, 57(6), 1207–1228.