Examining School-Age Children's Use of Rare Words in Expository Discourse: Applications of Computerized Analysis

Serena Jae Pedroza

University of Texas at El Paso

Follow this and additional works at: https://scholarworks.utep.edu/open_etd

Part of the Other Languages, Societies, and Cultures Commons, Speech and Hearing Science Commons, and the Speech Pathology and Audiology Commons

Recommended Citation
Pedroza, Serena Jae, "Examining School-Age Children's Use of Rare Words in Expository Discourse: Applications of Computerized Analysis" (2024). Open Access Theses & Dissertations. 4132. https://scholarworks.utep.edu/open_etd/4132

This is brought to you for free and open access by ScholarWorks@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.
EXAMINING SCHOOL-AGE CHILDREN’S USE OF RARE WORDS IN EXPOSITORY DISCOURSE: APPLICATIONS OF COMPUTERIZED ANALYSIS

SERENA JAE PEDROZA
Master’s Program in Speech-Language Pathology

APPROVED:

________________________________________
Stacy Wagovich, Ph.D., Chair

________________________________________
Vanessa Mueller, Ph.D.

________________________________________
Iva Ivanova, Ph.D.

________________________________________
Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School
Dedication

For Javy, Melissa, Kayla, Deeandra, and Tiger Pedroza – all I am is because of you.
EXAMINING SCHOOL-AGE CHILDREN’S USE OF RARE WORDS IN EXPOSITORY DISCOURSE: APPLICATIONS OF COMPUTERIZED ANALYSIS

by

SERENA JAE PEDROZA, B.S.

THESIS

Presented to the Faculty of the Graduate School of The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Speech, Language and Hearing Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

May 2024
Acknowledgments

I would like to express sincere gratitude and appreciation for Dr. Stacy Wagovich who has been an exceptional mentor. Thank you for believing in me and trusting me throughout this entire process; you have taught me more about research and leadership than I ever could have anticipated. Dr. Wagovich has a work ethic unlike any other person I’ve ever met, and I am confident that she will continue to inspire me throughout my career. I would also like to thank my brilliant committee members, Dr. Iva Ivanova and Dr. Vanessa Mueller for their constructive feedback and support. Additionally, this project would not have been possible without the hard work of Ms. Anna Guilkey who so graciously allowed me to use data she collected for her own master’s thesis.

There are not enough words to properly express how grateful I am for the love and support of my family. Mom and Dad, thank you for always making me feel as though I could accomplish anything and instilling the values of humility and hard work in us. Deeandra and Kayla, thank you for setting the bar so high. Paxson, thank you for making every day brighter!

Lastly, thank you to the most beautiful, loving, and supportive friends Stacy Alcantar, Kristen Apodaca, and Ariana Valdez. Thank you for always being there to listen to me complain, help me decompress, and share a laugh.
Abstract

In the school-age years, children use varied discourse styles to convey information. In addition to conversational and narrative discourse, they learn to use expository discourse to explain complex topics. Previous studies have found that compared to conversation, expository discourse samples are generally more syntactically complex (Lundine et al., 2018; Nippold et al., 2005). In addition, the vocabulary used in expository discourse may be more advanced than that in other discourse types. One way to explore vocabulary use in discourse is to measure the rarity of words within samples of discourse. The purpose of this study was to compare typically developing school-age children’s use of rare words between expository generation, expository retell, and conversational samples and examine if their rare word use was correlated with their scores on standardized language assessments. Language samples were collected from a convenience sample of twenty-six school-age children who were considered typically developing. Nippold’s favorite game or sport (FGS) task was used to elicit expository generation samples (Nippold et al., 2005). For expository retell, the children were shown a scientific video and asked to explain its content. The Wordlist for Expressive Rare Vocabulary Evaluation (WERVE; Mahurin-Smith et al., 2015) and a second rare word list generated from the word frequency dataset SUBTLEX (Brysbaert & New, 2009), a movie subtitle corpus, were used to measure lexical rarity. Results showed that across measures, expository retell samples had a significantly greater number of rare words compared to generation and conversational samples. There was no significant correlation between rare word use and language assessment scores. Overall, the findings indicated that children used more advanced vocabulary when retelling complex content from a video, compared to sharing information about a familiar activity or participating in conversation.

Keywords: expository, discourse, rarity, school-age
Table of Contents

Dedication ................................................................................................................. ii
Acknowledgments ...................................................................................................... iv
Abstract .................................................................................................................... v
Table of Contents ....................................................................................................... vi
List of Tables ............................................................................................................. viii
List of Figures ........................................................................................................... ix
Chapter 1: Introduction ............................................................................................. 1
    Vocabulary Development in the school-age years ................................................. 1
    Expository Discourse ............................................................................................ 2
    Transcription and Segmentation of Language Sample Data ............................... 5
    Lexical Measures of Expository Discourse .......................................................... 6
        Lexical Diversity .............................................................................................. 6
        Lexical Rarity ................................................................................................ 6
        Approaches to Measurement of Lexical Rarity .............................................. 7
    Computer-Assisted Measures of Rare Vocabulary ............................................. 9
    Purpose of the Study and Research Questions .................................................. 13
Chapter 2: Methods .................................................................................................. 15
    Participants ........................................................................................................... 15
    Experimental Stimuli .......................................................................................... 17
        Expository Generation ................................................................................... 17
        Expository Retell ............................................................................................ 18
        Conversation .................................................................................................. 19
    Procedures .......................................................................................................... 19
    Data Preparation ................................................................................................ 20
    Low Frequency Vocabulary Assessment .......................................................... 21
    Data-Checking and Reliability ............................................................................ 23
    Statistical Analysis .............................................................................................. 23
<table>
<thead>
<tr>
<th>Chapter 3: Results</th>
<th>..........................................................</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Statistics</td>
<td>.........................................................................................</td>
<td>24</td>
</tr>
<tr>
<td>Lexical Rarity Using WERVE</td>
<td>..................................................................................</td>
<td>25</td>
</tr>
<tr>
<td>Lexical Rarity Using SUBTLEX</td>
<td>.................................................................</td>
<td>26</td>
</tr>
<tr>
<td>Relationship Between Lexical Rarity Scores and Language Test Scores</td>
<td>................................</td>
<td>27</td>
</tr>
<tr>
<td>Chapter 4: Discussion</td>
<td>..................................................................................</td>
<td>28</td>
</tr>
<tr>
<td>Summary of Findings</td>
<td>..................................................................................</td>
<td>28</td>
</tr>
<tr>
<td>Lexical Rarity</td>
<td>..................................................................................</td>
<td>29</td>
</tr>
<tr>
<td>Relations Between Language Test Scores and Lexical Rarity</td>
<td>................................</td>
<td>33</td>
</tr>
<tr>
<td>Limitations</td>
<td>..................................................................................</td>
<td>34</td>
</tr>
<tr>
<td>Conclusions and Future Directions</td>
<td>...........................................................................</td>
<td>36</td>
</tr>
<tr>
<td>References</td>
<td>..................................................................................</td>
<td>39</td>
</tr>
<tr>
<td>Appendix A</td>
<td>..................................................................................</td>
<td>45</td>
</tr>
<tr>
<td>Vita</td>
<td>47</td>
<td></td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1: Standardized Test Scores ................................................................. 16
Table 3.1: Descriptive Statistics ........................................................................ 25
Table 3.2: Correlation Between the Proportion of Rare Words and Language Assessment Scores ................................................................. 27
Table 4.1: Comparison Between WERVE and SUBTLEX .................................. 32
List of Figures

Figure 3.1: Median Proportion of Rare Words Across Sample Types WERVE ....................... 26
Figure 3.2: Median Proportion of Rare Words Across Sample Types SUBTLEX .................... 27
Chapter 1: Introduction

Vocabulary Development in the School-Age Years

Children display immense growth in their language skills during their school-age years. It has been estimated that by the age of ten, children know as many as 40,000 words (Anglin, Miller, & Wakefield, 1993). As children enter their school-age years, there is an apparent shift in the way that they use and learn language. This can be attributed to a variety of factors such as more complex word exposure in school, and an increase in communication partners. For instance, school-age children are expected to use academic vocabulary to convey knowledge in the classroom and when talking with others about learned content. Not only do children learn vocabulary from their instructors, but they also experience incidental word learning through reading (e.g., Hill, Wagovich, & Manfra, 2017; Konopak et al., 1987; Nagy, Anderson, & Herman, 1987; Wagovich, Hill, & Petroski, 2015; Wagovich & Newhoff, 2004). While literacy requires sufficient lexical knowledge, as a child develops their reading skills, independent reading itself can enhance word learning (Nippold, 1998). In addition, through exposure to words in oral communication with peers, instructors, and family members, children are exposed to vocabulary that is more abstract and less common during the middle childhood years (Reggin, Muraki, & Pexman, 2021). The ability to use and understand abstract language presents children with the opportunity to discuss more complex topics.

For school-age children, models of complex language use come from parents and teachers in a variety of settings and using different types of discourse, including expository discourse. The quality of the expository interactions a child encounters (e.g., a parent explaining a complex task of interest to the child) can influence the child’s language learning and, in particular, the child’s lexical knowledge (Nippold, 2010). Complex explanations often provide introductions to rare words that then become a part of the child’s vocabulary. Therefore, thoroughly measuring a child’s vocabulary skills requires not just receptive and expressive tests of vocabulary knowledge, but examination of vocabulary use in discourse. In particular, as
pertains to this study, a child’s use of rare words in a complex discourse genre, such as expository discourse, may provide a sensitive measure of a child’s lexical knowledge, indicating their ability to use advanced vocabulary to discuss complex topics.

**Expository Discourse**

In the school-age years, children are expected to use a range of discourse styles to convey information appropriately at school and across settings. Conversational and narrative skills continue to develop in their sophistication over childhood. In addition, during this period, expository discourse skills develop considerably, as children use language to share information about complex topics (Nippold, 2010). This can be seen, for example, in a student explaining how the discovery of antibiotics impacted American society or the functions of various organelles. Conversely, conversational discourse is a less formal genre, used in social settings in interactions with others (Nippold, Hesketh, Duthie, & Mansfield, 2005). For school-age children, conversation can occur in various situations, such as sharing opinions about preferred movies or sports. Of course, discourse types can overlap in everyday communication. For example, a conversation may involve conveying a story using narrative conventions, or it may involve explaining a procedure using expository elements. Nevertheless, it is useful to consider the demands and language outcomes of different discourse types separately.

While it is understood that both conversational and expository genres require cognitive and linguistic resources to organize and communicate ideas, expository discourse may be more taxing on the language system, perhaps owing to the need to convey content by using greater linguistic complexity (Nippold, 2010). Lundine and colleagues (2018) note that compared to narrative samples, expository production requires “higher-level thinking and planning skills” (p. 553) making it more cognitively demanding than other language sample types. Similarly, Nippold and colleagues (2005), compared expository discourse to conversational discourse, finding that expository samples were more syntactically complex than conversational samples.
across typically developing individuals aged 7-49. The authors concluded that “complex thought drives the use of complex language” (Nippold et al., 2005, p. 1058). For clinicians and researchers, this is a key consideration: Conversational language samples likely require speakers to use less sophisticated language than expository samples. Therefore, if a language sample is being collected, an expository sample is likely to provide a more accurate representation of the upper limits of a child’s language skills.

In studies of expository discourse production, there are several types of expository tasks that have been administered. Two common ones are expository generation and expository retell. The first, generation, requires the participant to produce a sample on a provided topic, without any model (Nippold & Scott, 2010). A common task of this type is the Favorite Game or Sport (FGS) task. Introduced by Nippold and colleagues (2005), the FGS task requires participants to describe and explain an activity they enjoy. An advantage of this approach is that it is likely highly motivating to discuss a favorite activity about which one is extremely knowledgeable. Nippold and Scott noted that increased interest and motivation can improve the quality of a verbal or written product. Expository generation is also likely more ecologically valid, especially when the topic is based on the child’s interests; this may lead to a sample with richer language compared to expository retell samples. Furthermore, the individualization of generation tasks increases the meaning of the activity for the child, who may be inherently interested in the selected topic. On the other hand, a limitation of expository generation methods like FGS is that a child’s sample could range in its conceptual complexity (e.g., it could be a simple game/sport or a complex one), making it difficult to compare expository ability among the participants completing the task (Heilmann & Malone, 2014).

The second task, expository retell/summarization, involves giving the participant an expository model and asking them to either retell or summarize what was presented. Lundine and colleagues (2018) note that expository retell requires a different level of cognitive demand. Specifically, retell tasks require speakers to continuously monitor their understanding of the material being retold and rapidly resolve inconsistencies in their own understanding (Nippold,
An advantage of retell tasks is that the content of the samples is standardized across participants, as they are all presented with an equally complex discourse model and are asked to provide a retelling. By providing the same prompt to all participants, retell tasks are perhaps more similar to what children are expected to do in school than generation tasks, in summarizing learned content. A possible drawback of this approach, however, is that children may be memorizing and repeating information from the retell source, which does not allow researchers to accurately determine if they truly understand the words they are using. In addition, if a child does not have a good understanding of the content, they may not produce a sample that is truly reflective of their expository abilities.

Acknowledging these surface-level strengths and limitations between the two elicitation types, research has been conducted to quantify the differences in performance using different expository sampling methods. To compare expository generation and expository retell, Westerveld and Moran (2013) examined the syntactic complexity, verbal fluency (maze behaviors), and verbal productivity across sample types among typically developing school-age children and adolescents. They found that the expository generation task resulted in longer samples with a higher number of different words (NDW); however, retell samples contained longer utterances and greater clausal density. That is, the expository retell tasks produced shorter but more syntactically complex samples than expository generation. Expanding on previous findings, Guilkey and Wagovich (2022) sought to determine which sampling method would produce the most robust, syntactically complex, and lexically diverse sample when children were provided with some scaffolding to support the production of expository discourse. With the hypothesis that retelling new, unfamiliar content may require more complex thought than generating an expository sample based on familiar content, the researchers compared expository generation and retell samples. Their findings showed that expository generation samples were longer and more substantial in content; however, expository retell samples contained greater lexical diversity. No difference was observed in syntactic complexity between sample types. These findings were partially consistent with those of Westerveld and Moran (2013) in that both
studies revealed that expository generation resulted in more language output than retell, and that expository retell resulted in greater complexity, either lexically (Guilkey & Wagovich) or syntactically (Westerveld & Moran).

TRANSCRIPTION AND SEGMENTATION OF LANGUAGE SAMPLE DATA

Due to the technical nature of language sample analysis, samples are transcribed online using software programs such as Computerized Language Analysis (CLAN) (MacWhinney, 2000) and Systematic Analysis of Language Transcripts (SALT) (Miller & Iglesias, 2015). These systems allow researchers and clinicians to derive quantitative data through specific programs that perform computations such as mean length of utterance (MLU). Language sample analysis systems require samples to be manually transcribed, and the specific coding conventions differ for each system. In the case of SALT, users must segment the sample into utterances and identify individual morphemes. Similarly, when using CLAN, the user is responsible for segmenting the sample into utterances. However, the program performs any morphological parsing and coding of parts of speech. As relates to this project, CLAN also enables the user to evaluate lexical features of the samples, including specialized programs, as described below. Research questions and clinician/researcher preferences dictate which system is most appropriate.

In both CLAN and SALT, for language samples of school-age children, utterances are segmented using either t-units (generally for monologues or written text) or c-units (for conversational language). A t-unit includes an independent clause and its dependent clauses and/or nonclausal structures that are embedded in it (Hunt, 1966; Nippold et al., 2005). A c-unit is defined similarly but can include utterances without an independent clause, such as an utterance that is a response to a question (Loban, 1976; Nippold et al., 2005). The length (number of words) of each t-unit or c-unit is calculated by the language sample analysis system, which then computes the sample’s mean length of t-units or c-units. This measure can serve as a
rough indicator of a sample’s syntactic complexity, much like calculating MLU in morphemes does for children in the preschool years.

**LEXICAL MEASURES OF EXPOSITORY DISCOURSE**

**Lexical Diversity**

A variety of measures can be used to examine a spontaneous language sample’s lexical diversity. Researchers have analyzed spontaneous language through tools such as number of different words (NDW; Watkins et al., 1995), type-token ratio (Templin, 1957), and Moving Average Type-Token Ratio (MATTR; Covington & McFall, 2010). Each of these estimates the lexical diversity of discourse. However, if a child is using more rare, sophisticated vocabulary within their language sample, this information may be lost in a simple analysis of lexical diversity, requiring a more nuanced analysis. For example, if a child uses five rare vocabulary words four times each in providing a complex explanation as part of an expository sample, doing so does not necessarily lead to a greater lexical diversity value for the sample. Rare word use is not well captured through lexical diversity measurement. Instead, it would be captured by an analysis of the number or proportion of rare words contained in the sample.

**Lexical Rarity**

Lexical rarity measures have been developed to analyze the relative rarity of words in a given language sample (e.g., Beck, McKeown, & Kucan, 2013; Mahurin-Smith, DeThorne, & Petrill, 2015). During the school-age years, the use of more sophisticated language is expected of children in the classroom and among other mature language users, despite the fact that low frequency words are harder to retrieve, more decontextualized and abstract, and conceptually more complex (Mahurin-Smith et al., 2015). Studying lexical rarity and the development of lexically rare vocabulary is critical, in that rare vocabulary exposure has been linked to academic success and increased performance on language assessments. Notably, children with language
disorders will likely show fewer lexically rare words in their language samples. Treatment goals that focus on vocabulary enrichment can address this issue. However, as a first step to understanding the use of rare words by school-age children, it is necessary to explore the issue through analysis of language samples of children with typical language, to guide us in our further exploration with children with language weaknesses. Moreover, given the literature that suggests that expository discourse, in particular, tends to lead to more complex language (e.g., Nippold et al., 2005), the exploration of lexical rarity in expository discourse compared to conversation may provide important insight into the contexts in which children produce lexically rare words in the school-age years.

**Approaches to Measurement of Lexical Rarity**

Various approaches have been used to characterize rare vocabulary. Beck, McKeown, and Kucan (2013) placed words in categories based on a multidimensional system that considers a word's usefulness, frequency of use, and age of acquisition. They developed a three-tier model that has served as a tool for researchers and instructors to identify the types of words children are using. Tier 1 includes basic vocabulary words that are familiar and connected to prior knowledge. These would be words like *dog*, *pink*, and *apple*. Tier 2 words are high-frequency words in the language of adult speakers. These words should be included in classroom instruction because they contribute to verbal communication across multiple domains. Some Tier 2 words are *fortunate*, *require*, and *perform*. Tier 3 words are low-frequency words that are usually domain-specific and may require teaching in some instances. For example, *photosynthesis*, *beaker*, and *electrolyte* are all Tier 3 words. Rare words in this system come from Tiers 2 and 3. Thus, the tiered system by Beck and colleagues is one way to explore lexical rarity in children’s discourse.

The Academic Word List (AWL; Coxhead, 2000) is another tool that defines rare vocabulary based on words’ frequency of use in academics. The list was derived by examining
written academic texts. Words occurring frequently in academic texts but that were not one of the 2,000 most frequently used English words were included on the AWL. The list consists of 570 words divided into ten groups, with Group 1 containing the more frequent words and Group 10 containing the rarest words.

Another approach to exploring lexical rarity is to examine the age at which a word is typically learned. The Dale-Chall list (Chall & Dale, 1955) is comprised of 3,000 words understood by fourth grade students. If at least 80% of fourth graders indicated that they knew a word, it was added to the list. Researchers have used this list to measure rarity by classifying a word as rare if it is not on that list (Rowe, 2012). However, this list is mainly used for determining the difficulty of reading materials, not for analyzing lexical rarity of spoken language samples.

These were excellent steps toward studying rarity, but they are limited in that they are time-consuming, requiring substantial by-hand analysis, and some lists may be antiquated or not tailored to school-age children today. For example, the Dale-Chall list was created in 1995 and includes words such as “postage” and “savage” that are not typically found in the lexicon of present-day school-age children. This list is also based on the readability of words and not children’s ability to use the words in spoken language. Furthermore, the AWL (Coxhead, 2000) was created using samples from academic texts written by adults. And, the tiered approach of Beck et al. (2013), while still widely in use today, does not provide a concrete definition of rarity. This approach provides useful information about a child’s vocabulary, but it does not specifically address lexical rarity as a construct.

Therefore, while the previous methods discussed above are reasonable ways to measure lexical skills, computer-mediated analysis that uses large word frequency datasets has enabled the development of more robust tools of lexical rarity, as discussed in the section that follows. Since lexical rarity is a relatively underreported measure of children's language sample performance, it is imperative that researchers first establish the expected use of rare words for typically developing children. For example, we might ask, do most typically developing school-
age children use a range of rare vocabulary when they are describing a favorite activity? Does their use of rare vocabulary depend upon the formality of the discourse - expository versus conversational? Addressing these types of questions about lexical rarity among typically developing children will lead to studies that can inform research on children with language impairments and comparisons to the typically developing population.

**COMPUTER-ASSISTED MEASURES OF RARE VOCABULARY**

The most recent tool for analyzing lexical rarity is a proposed analysis from Mahurin-Smith and colleagues (2015) called the Wordlist for Expressive Rare Vocabulary Evaluation (WERVE). These authors identified the need for a new, updated measurement system for lexical rarity, highlighting several reasons that innovations in lexical rarity measurement are needed. First, they note that subtle changes in children’s language have been observed since the formulation of older lists, making them less suitable for analysis of children’s language today. Second, some previous measures were formulated using spoken and written language samples of adults. Words that are rare for a child are not necessarily rare for an adult, so measures that take into account children’s vocabulary over their development are needed. Finally, analysis of rarity using the previous methods (e.g., looking up each spoken word in a word frequency book) would be extremely time-consuming, requiring substantial by-hand coding. This would be impractical both in conducting research and in applying lexical rarity measurement to clinical work.

WERVE was developed from data collected for the Western Reserve Reading Project (WRRP), a study by Petrill and colleagues (2006) of environmental and genetic influences on children's mathematics and reading skills. These researchers studied 438 pairs of same-sex twins longitudinally, beginning when they were in kindergarten. The WERVE samples were selected from this larger dataset, from the second, third, and fifth annual home visits which were around
two hours each. During the second home visit, the mean age of participants was 7.19 years. During this visit, a conversational language sample and reading, math, and cognitive assessments were conducted for the WRRP study. The third home visit was completed about a year later and only included measures of language and reading skills, including a conversational language sample. In the fifth home visit, the researchers conducted multiple subtests from the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Wiig, Semel, & Secord, 2003) for their 2015 study, including Recalling Sentences, Understanding Paragraphs, Word Classes (Receptive), and Word Classes (Expressive). All conversational language samples were collected during a play activity using modeling clay. Examiners who interacted with the children during their conversational language samples had completed training on Leadholm and Miller’s (1992) language sampling guidelines. To elicit quality language samples, examiners were directed to avoid questions that only warrant a one-word answer. Examiners were also instructed to avoid asking many questions to children who were quieter during the interactions, allowing silence in their interactions with the children. To transcribe, trained assistants at Pennsylvania State University and the University of Illinois used the Systematic Analysis of Language Transcripts software (SALT, Version 8.0; Miller & Iglesias, 2015). Utterances were divided into Communication Units (C-units), as defined by Nippold’s (1998) guidelines (Loban, 1976). The Computerized Language Analysis program (CLAN) was later used to produce an alphabetical list of 1.2 million words from 1,437 language transcripts from twins in the WRRP study.

Using these data, Mahurin-Smith and colleagues (2015) then performed a frequency analysis in which any word that was used 15 times or fewer across the entire corpus was selected for their Low-Frequency Vocabulary (LFV) list. Thus, results for individual children varied, in that some children used no LFV words, and some used many. Prior to identifying rare words in
the individual children's samples, the list was edited to remove kinship terms like nana and opa, numbers, proper nouns, sound effects, and typographic errors. In addition, variations of morphemes were collapsed for the analysis (e.g., “borrowing” and “borrowed” were condensed to “forms of borrow” and considered multiple instances of the same root word). After this step, it was evident that some words on the list, despite only being used 15 times or less, were early developing words that shouldn’t be considered rare in the general population of children. Therefore, the MacArthur-Bates Communicative Development Index (Fenson et al., 2007), a list of words said by at least 15% of 2.5-year-olds in a normative sample, was used to identify words that should be removed from the LFV list. That cross-check led to the elimination of 61 words; the remaining 2079 low-frequency words comprised the final WERVE list.

Few studies have been conducted using WERVE as yet. To our knowledge, there are three published studies of rare vocabulary that used the WERVE list, two of which were authored by Mahurin-Smith (Mahurin-Smith et al., 2021; Mills et al., 2017), and the other of which was authored by a different research team (Kelley & Poholik, 2023). Mahurin-Smith and colleagues (2021) used WERVE to examine rare vocabulary use in narratives from school-age children from low-income communities. They found that rare vocabulary significantly correlated with language sample measures like the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4; Dunn & Dunn, 2007) and NDW, providing evidence of WERVE’s concurrent validity with other measures of vocabulary. A second study focused on the application of WERVE as a type of predictive assessment. Mills and colleagues (2017) used WERVE to determine if rare vocabulary use was a predictor of giftedness as indicated by children’s placement in a gifted classroom. Personal and fictional narrative samples were collected, and findings showed that rare word density, a measure that controls for sample length, was greater in personal narratives than
in fictional narratives. Additionally, rare word use, as defined by WERVE, was a predictor of
giftedness for personal narratives as gifted children produced more rare vocabulary than their
peers. However, in fictional narratives, there was not a difference in rare word use across
participants possibly due to the nature of the fictional narrative sampling procedures (i.e. use of a
standard wordless picture book for all participants). These findings are relevant in that they
provide evidence that personal narrative tasks may be preferable to fictional narratives for rare
word analysis of children’s language. Understanding the connection between rarity and
giftedness also has the power to influence the assessment of students for gifted programs in that
rare word density can be a “dialect-neutral” way of assessing children’s language content (Mills
et al., 2017, p. 521). Specifically, these authors observed that there was no correlation between
rare word density and dialect variation. Together, these two studies, both focusing on narratives,
provide a good base, enabling the use of this measure to evaluate lexical rarity in expository
discourse samples. The third study, by Kelley and Poholik (2023), focused on preschool
children, exploring whether the use of WERVE in play-based language samples correlated with
other vocabulary scores. Indeed, WERVE positively related to standardized vocabulary
measures, as well as lexical diversity measures from the same samples. These findings provide
evidence of the concurrent validity of WERVE as a measure of lexical skill for children in the
preschool years.

Brysbaert and New (2009) have argued that studies that involve word frequency data
have suffered from small corpuses of words from which to derive frequencies. This criticism
applies to lexical rarity studies, as well. The authors suggest that, at a minimum, a word
frequency corpus should consist of 16 million word tokens. They developed a corpus,
SUBTLEX, that far exceeded this cutoff, with 51 million tokens derived from movie and
television subtitles. To assemble the corpus, they downloaded subtitle transcripts of U.S movies and television shows from www.opensubtitles.org. After eliminating transcripts with greater than 2.5% typing errors, they had a total of 8,388 files. SUBTLEX is free to download as an Excel spreadsheet and includes various measures that allow the list to serve multiple functions in research. Measures that may be relevant to rarity studies include FREQcount, the total number of times a word is used; CD count, the number of films in which a word is used; and SUBTLwf, the word frequency per million words. While the authors developed this dataset with the intention that it could be used as a word frequency measure in research on cognitive processing and memory, it can also be used to generate wordlists based on their frequency for measuring rarity. By using movie and television subtitles, this list more accurately represents language used in typical conversation than other rare word measures like the Dale-Chall list that was derived from academic texts.

Studying rarity can help us gain a fuller understanding of children’s lexical skills – including the leading edge of vocabulary they can use appropriately in context. Lexical rarity provides different information from lexical diversity because the focus is not on all words in a sample but only those words that are relatively rarer. From a practical perspective, lexical rarity research may influence vocabulary teaching strategies, encouraging emphasis of instruction in the use of rare words, both in the classroom and across contexts, ultimately making children better communicators.

**PURPOSE OF THE STUDY AND RESEARCH QUESTIONS**

The purpose of this study was to compare typically developing school-age children’s use of rare words between two types of expository discourse samples: expository generation and expository retell. In addition, we compared how rare word use in expository samples differs from
rare word use in conversational samples. We chose to use the SUBTLEX corpus (Brysbaert & New, 2009) in addition to the WERVE corpus (Mahurin-Smith et al., 2015), analyzing rarity across the two measures.

A second purpose was to explore the extent to which standardized vocabulary and overall language test scores corresponded to children’s lexical rarity across the sample types, as a relation might be expected based on the prior literature.

The research questions are as follows:
1. Does rare word use, as measured by WERVE, differ across types of language samples, specifically expository generation, expository retell, and conversation, in typically developing school-age children?
2. Does rare word use, as measured by the SUBTLEX word list, differ across types of language samples, specifically expository generation, expository retell, and conversation, in typically developing school-age children?
3. Across measures, is there a relation between children’s rare word use and their standard scores on vocabulary and overall language tests?

We hypothesized that expository retell samples would contain proportionally more rare words than expository generation or conversation samples, because the expository retell sample involved learning new content, taking notes on content, and then conveying it aloud. In contrast, the other two samples involved discussing topics with which participants were likely very familiar. We also hypothesized that the use of rare vocabulary, across samples, would be correlated with children’s expressive and receptive vocabulary scores.
Chapter 2: Methods

Participants

The data analyzed in this study were collected by Guilkey and Wagovich (2022) for a study on expository discourse in school-age children. All procedures were reviewed and approved by the University of Missouri Health Sciences Institutional Review Board and later by the University of Texas at El Paso’s Institutional Review Board (Approval #1910696). A convenience sample of 26 children (8 females and 18 males), ages 9;1 (years; months) to 12;10, was collected. The mean age of participants was 10;7 with a standard deviation of 12 months. This age range was selected because expository discourse is a key skill in the later school-age years, predicting academic success. To recruit participants, researchers contacted the parents of children who had qualified for previous studies and had expressed interest in being included in future research projects. The children of these parents had previously been part of a study involving word learning, not associated with this study but conducted in the same laboratory. Children were deemed eligible to participate in the study if they scored within normal limits (with a standard score higher than 82, corresponding to a cutoff of 1.25 SD below the test mean) on standardized assessments of vocabulary and overall language testing. Eligibility was also reliant on the child passing a hearing screening and a cognitive screening. Standardized tests and screenings are described in the Procedures section. In addition, participants were required to have no history of a language disorder or attentional deficits, be typically developing overall. English was their first or only language learned and in all cases was their dominant language. Means and standard deviations for participants’ testing are provided in Table 1.
Table 2.1: Standardized Test Scores

<table>
<thead>
<tr>
<th>Language Assessment</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>CELF-5 (CLS)</td>
<td>110.46</td>
<td>9.37</td>
<td>85</td>
<td>125</td>
</tr>
<tr>
<td>CELF-5 (RLI)</td>
<td>108.65</td>
<td>7.82</td>
<td>95</td>
<td>125</td>
</tr>
<tr>
<td>CELF-5 (ELI)</td>
<td>107.8</td>
<td>11.9</td>
<td>85</td>
<td>133</td>
</tr>
<tr>
<td>PPVT-4</td>
<td>115.96</td>
<td>12.7</td>
<td>91</td>
<td>138</td>
</tr>
<tr>
<td>EVT-2</td>
<td>111.92</td>
<td>9.35</td>
<td>89</td>
<td>124</td>
</tr>
<tr>
<td>TONI-4</td>
<td>104.3</td>
<td>7.71</td>
<td>91</td>
<td>125</td>
</tr>
</tbody>
</table>

*Note. CELF-5 = Clinical Evaluation of Language Fundamentals-Fifth Edition (Wiig, Semel, & Secord, 2013); CLS = Core Language Score; RLI = Receptive Language Index; ELI = Expressive Language Index; PPVT-4 = Peabody Picture Vocabulary Test-Fourth Edition (Dunn & Dunn, 2007); Expressive Vocabulary Test-Second Edition (EVT-2; Williams, 2007); TONI-4 = Test of Nonverbal Intelligence-Fourth Edition (Brown, Sherbenou, & Johnsen, 2010);*

The sample size of the study was selected based on a power analysis, applying the means and standard deviations from Heilmann and Malone’s (2014) study. Results indicated that a sample size of 23 would yield 80% power to detect a clinically meaningful (medium to large) effect at p < 0.05.

Socioeconomic status was estimated using the Hollingshead Four Factor Index of Social Status (Hollingshead, 1975). This measure considers the education and occupations of each parent in the household. Of the participants in the study, 73% came from upper class families, 15% from upper middle-class families, 4% from middle-class families, and 8% from lower middle-class families. Based on parent report, 84.6% of participants identified as Caucasian, 3.85% as Asian, 3.85% as Hispanic, and 7.7% as more than one race.

A primary researcher informed both children and parents of the study's purpose and the activities involved in participation. A parent signed the informed consent document, and the child signed an assent form, agreeing to participate in the study.
Experimental Stimuli

Expository Generation

Nippold's Favorite Game or Sport Task was used to elicit expository generation samples. The examiner had a script to present instructions and prompts to the child. Participants were asked to describe their favorite game or sport. For a detailed description of the task and its instructions, see Nippold et al. (2005). A scaffolded note-taking sheet created by Heilmann and Malone (2014) was given to participants to plan their explanation. This sheet contained three columns. The first listed each component of a strong expository sample (i.e., object [of the game/sport], preparations, start, course of play, rules, scoring, duration, and strategies). The second column provided a description of what is covered by each component. The third column was a notes section for the child to make notes to assist them during their expository sample. Heilmann and Malone considered these integral components of a quality expository language sample.

A script from Heilmann and Malone was used to instruct children on how to complete the Favorite Game or Sport Task, including how to use the notetaking sheet. The script started by instructing the participants to imagine that the researcher is unfamiliar with their chosen game/sport and would like to learn more about it. Then, they were told the notetaking sheet's purpose was to help them organize their thoughts and include all necessary details. Children were asked to make their sample as detailed as possible and at least five minutes long. The time limit was not strictly enforced as it was meant to encourage substantial, detailed samples. The script then encouraged the participants to ask any remaining questions about the procedures. There was no set time limit for note-taking/planning. However, if a notes section was left blank, the researcher could prompt with the phrase, “Please do some planning for [topic choice].” The script concluded by asking the child to begin their explanation using their notes. When they were done speaking, the researcher asked, “Is there anything else you can tell me?”
Expository Retell

Expository retell samples were collected using similar sampling methods. Rather than choosing a topic, students were shown a 3-minute 43-second scientific video titled, “Hoover Dam and Hydroelectric Power,” and were asked to provide a retelling based on what they saw. A script and note-taking sheet specific to the selected materials were created by Guilkey and Wagovich (2022) using Heilman and Malone’s procedures as a model. Specifically, the script followed the same previously described format, but Favorite Game or Sport references were replaced with references to the video. Before playing the video, participants were informed that it would be played one time and paused in two places to give them the opportunity to take notes. Participants were encouraged to take notes and only write key phrases as a reminder of what they wanted to say. The note-taking sheet included relevant headings and guiding questions such as “purpose of large dams” with questions like, “What is hydroelectric power?” During retelling, the children could use their notes so that their samples reflected their expository skill rather than merely their ability to remember the newly presented content. The video selected was age-appropriate, containing curriculum-specific (Tier 3) vocabulary. In theory, exposure to a complex topic as the subject of a retelling should lead the child to produce more complex language than a simpler topic might (e.g., see Nippold & Scott, 2010, for a discussion). Moreover, as relates to this project, a more complex topic may facilitate rarer, more diverse vocabulary, as well. After collecting the expository retell sample, video content comprehension was assessed via a 20-question multiple choice test in which children were not permitted to use their notes. Questions were based on facts explicitly stated in the video. For example, one of the questions was as follows:

Which river in the Pacific Northwest also has a large dam (the Grand Coulee Dam) on it?

a. The White River
b. The San Juan River
c. The Green River

d. The Columbia River

**Conversation**

To elicit conversational samples, the examiner sat with the child and initiated conversation about engaging topics from a list created by the researchers. Topics included plans for the holidays, school-related interests and activities, etc. In general, samples were approximately 5 minutes in length.

**Procedures**

Each participant spent approximately 110 minutes (about 2 hours) in the study across one to two sessions. Sessions were conducted in the Language and Fluency Laboratory at the University of Missouri. The room used was quiet and attempts were made to minimize distractions. Aside from the participant, two researchers (a primary researcher and lab assistant) were present for each session. The language sampling and testing procedures were video recorded for future analysis. For language samples, recordings were used for transcription and language analysis. Recordings of testing procedures were collected in case a particular response needed to be reviewed for scoring purposes. To be considered eligible for the study, a binaural pure tone hearing screening at 1000, 2000, and 4000 Hz at 20 dB HL was administered. In addition, participants underwent formal language assessment and cognitive screening. Specifically, the *Clinical Evaluation of Language Fundamentals-Fifth Edition* (CELF-5; Wiig, Semel, & Secord, 2013), a norm-referenced assessment that provides an overall estimate of a child’s language skills, was administered. Receptive and expressive vocabulary skills were estimated using the *Peabody Picture Vocabulary Test-Fourth Edition* (PPVT-4; Dunn & Dunn, 2007) and the *Expressive Vocabulary Test-Second Edition* (EVT-2; Williams, 2007). The cognitive screening performed was the *Test of Nonverbal Intelligence-Fourth Edition* (TONI-4;
Brown, Sherbenou, & Johnsen, 2010), which is a language-free test that provides age-based standard scores. Although assessments were administered to each participant in a fixed order, the three experimental tasks were counterbalanced across participants. The order of the measures (tests, screenings, and experimental tasks) was organized by alternating tests/screenings with experimental tasks.

**Data Preparation**

The expository samples were transcribed and coded by the original researchers (Guilkey & Wagovich, 2022), using the CHILDES language analysis system (MacWhinney, 2000). CHILDES stands for Child Language Data Exchange System. CHILDES requires that samples be transcribed according to CHAT (“Codes for the Human Analysis of Transcripts”) format, so that language can be analyzed by the suite of programs referred to as CLAN (“Computerized Language Analysis”).

For the present study, conversational samples were transcribed and coded, and all three sets of samples (expository generation, expository retell, and conversation) were checked to ensure that they were coded appropriately for the proposed lexical analyses. In particular, I checked word spellings by running the FREQ program in CLAN, which provided an alphabetized list of each different word in a sample, so that misspellings and transcription errors could be identified efficiently and corrected. Samples were also checked to ensure that mazes (including repetitions, revisions, and fillers) were excluded. For example, if a child said, “You need to, you need to shuffle,” the underlined portion was removed from analysis, so that only the intended message, without repetition, was considered in the analyses.

The utterances within each sample were segmented into C-units (Loban, 1976). A C-unit is defined as either (a) an independent clause and its dependent clauses and/or nonclausal structures that are embedded in it, or (b) an incomplete clause that is a response to a question. This segmentation method is standard for analysis of school-age children’s spontaneous
language (Nippold et al., 2010). C-unit analysis enables measurement of mean length of c-unit, as an estimate of syntactic complexity.

**Low Frequency Vocabulary Assessment**

To address Research Question 1, I analyzed the expository retell, expository generation, and conversational samples for children’s use of low frequency vocabulary, by using WERVE and the CLAN software program. The WERVE rare word list contains 2079 rare words provided by Mahurin-Smith and colleagues (2015). CLAN analyzed each transcript against the rare word list and generated a list of rare words from each child’s sample. The analysis is performed using the FREQ command. The specific command used to enable CLAN to identify the rare words within each sample is as follows (referred to here as FREQrare):

```
freq +t*CHI +s@WERVE.cut *.cex > lowfreq.cut
```

This command instructs the program to take Mahurin-Smith’s list of rare words, included in the file WERVE.cut, and tabulate which of the words appears in the children’s samples, providing rare word types and the number of times each word appears (tokens). This analysis provided information about the number of Mahurin-Smith and colleagues’ rare words the children in our study produced. Once the list of words was generated for each sample, each item was reviewed by hand by reviewing the context in which the word had occurred. A rare word was considered semantically appropriate if it was used in a way that aligned with its definition. For example, the word “tortilla” spoken to mean the name of a family pet, would not be considered rare because of the semantic context. If the word was not rare given its context, it was not included in the total number of rare words the child produced in the sample.

To address Research Question 2, a similar approach was used. However, as a first step, I generated a rare word list from the SUBTLEX corpus including all words that had a word
frequency index (SUBTLWF) of 1.00 or less, meaning that it was used less than or equal to once per million tokens. This resulted in a rare word list of 43,004 words. Examples of SUBTLEX words that met this criterion are provided in Appendix A. Following creation of the rare word list, the same procedure as above was performed.

1. The same command as above was run, but this time, using the rare word list generated from the SUBTLEX corpus:

   freq +t*CHI +s@Subtlex.txt > lowfreq.cut *.cha

2. All words identified as “rare” using the SUBTLEX list were examined in context to be certain that the word was functioning as a rare word and was semantically appropriate. Context analysis was the same as that of the WERVE procedure outlined above. For the SUBTLEX list, base words were examined for word variations with inflectional morphemes to ensure that they met the cutoff of once per million units. Any word that contained a derivational morpheme was considered without removing the affix. This is because derivational morphemes typically change the part of speech of a word. For example, if a child said the word “cleverly”, the whole word would be considered in determining its rarity, and it would be considered rare, with a SUBTLWF value of 0.88. In contrast, if a child said “castles”, CLAN would identify the inflected form as rare (SUBTLWF = 0.88), but it would be removed manually from the list of the child’s identified rare words, because the root form has a rarity value outside of the cutoff of 1.0 (SUBTLWF = 21.55).

3. Finally, rare words from SUBTLEX were tabulated for each sample type per participant for statistical analysis.
To reduce the impact of sample length, the proportions of different rare words produced were calculated by dividing the number of rare word types in each sample by the total number of word tokens in the sample. We focused on rare word types rather than rare word tokens, because types may be more sensitive in detecting advanced vocabulary use. Mahurin-Smith and colleagues (2015) also limited analyses to rare word types.

**Data-Checking and Reliability**

Transcriptions were checked by a second individual trained in CHAT/CLAN transcription. Any disagreements identified were resolved by consensus and the appropriate corrections were made to the identified rare words in each sample.

To determine agreement on rare words pre-identified by CLAN that were subsequently selected for inclusion or exclusion based on one of the reasons identified (see section above), a second research assistant reviewed approximately 23% of the participants’ samples (18 samples, 6 of each type) and rare word WERVE and SUBTLEX data (i.e., CLAN output and the context surrounding each rare word used). The research assistant was asked to indicate which words should be excluded based on the criteria (e.g., proper nouns, idiosyncratic forms, kinship terms, typographical errors, etc.) versus which words should be retained. Overall agreement for the inclusion/exclusion of words was 95.9% across raters.

**Statistical Analysis**

Because of violations to the assumption of normality (substantial positive skew), nonparametric statistics were used to address each of the research questions. Specifically, to address Research Questions 1 and 2, which explored lexical rarity scores across the three sample types, Friedman tests were performed, with p-values reflecting Bonferroni corrections for multiple comparisons. The third research question was addressed using Spearman correlations to examine the relation between rare word use and language scores.
Chapter 3: Results

Analysis of rarity in expository generation, expository retell, and conversational samples was performed for all 26 participants. In this chapter, descriptive statistics for all variables are introduced. This is followed by comparisons of each of the two lexical rarity analyses across language sample types, using inferential statistics. Lastly, correlations between children’s overall lexical rarity and their standard scores on language tests are presented.

Descriptive Statistics

Table 3.1 displays the descriptive statistics of the presented variables. Some key observations were made from these data. First, the median total number of rare words used across sample types (Total SUB and Total WERVE) was greater for the WERVE analysis than the SUBTLEX analysis. Additionally, across measurement tools, the median proportion of rare words used was significantly greater in expository retell samples compared to the other sample types. As noted in Chapter 1, sometimes in conversation, a speaker may incorporate elements of expository discourse, when they offer an explanation, for example. Therefore, I reviewed all samples and coded utterances containing expository elements in them. Findings were that only 7.7% of all utterances in the conversational samples contained instances of exposition.
Table 3.1: Descriptive Statistics

<table>
<thead>
<tr>
<th>WERVE Variables</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>4.23</td>
<td>3.59</td>
<td>3</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Retell</td>
<td>4.23</td>
<td>2.69</td>
<td>4</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Conversation</td>
<td>2.77</td>
<td>1.9</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Generation Prop.</td>
<td>0.009</td>
<td>0.007</td>
<td>0.009</td>
<td>0</td>
<td>0.027</td>
</tr>
<tr>
<td>Retell Prop.</td>
<td>0.023</td>
<td>0.008</td>
<td>0.022</td>
<td>0.005</td>
<td>0.037</td>
</tr>
<tr>
<td>Conversation Prop.</td>
<td>0.009</td>
<td>0.006</td>
<td>0.009</td>
<td>0</td>
<td>0.009</td>
</tr>
<tr>
<td>Total WERVE</td>
<td>11.23</td>
<td>5.7</td>
<td>11</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>WERVE Prop.</td>
<td>0.012</td>
<td>0.004</td>
<td>0.012</td>
<td>0.005</td>
<td>0.021</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUBTLEX Variables</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>1.88</td>
<td>3.09</td>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Retell</td>
<td>1.85</td>
<td>1.62</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Conversation</td>
<td>1.12</td>
<td>1.82</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Generation Prop.</td>
<td>0.004</td>
<td>0.004</td>
<td>0.003</td>
<td>0</td>
<td>0.012</td>
</tr>
<tr>
<td>Retell Prop.</td>
<td>0.009</td>
<td>0.006</td>
<td>0.009</td>
<td>0</td>
<td>0.019</td>
</tr>
<tr>
<td>Conversation Prop.</td>
<td>0.003</td>
<td>0.004</td>
<td>0.003</td>
<td>0</td>
<td>0.015</td>
</tr>
<tr>
<td>Total SUBTLEX</td>
<td>4.85</td>
<td>5.65</td>
<td>3</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>SUBTLEX Prop.</td>
<td>0.005</td>
<td>0.003</td>
<td>0.004</td>
<td>0.001</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Note. Generation = rare words in expository generation task; Retell = rare words in expository retell task; Conversation = rare words in conversation task; Generation/Retell/Conversation Prop. = proportion of total words that were rare in each sample type; Total WERVE/SUBTLEX = total rare words across sample types; WERVE/SUBTLEX Prop. = proportion of total words that were identified as rare for each rarity analysis.

Lexical Rarity Using WERVE

The overall Friedman test comparing rare word use as analyzed using the WERVE procedure was significant, $\chi^2(2) = 18.69, p < .001$. Pairwise comparisons with a Bonferroni correction for multiple comparisons revealed statistically significant differences in rare words produced between the retell (Mdn = .0225) and the generation (Mdn = .0092) conditions, $p = .001$, and between the retell and conversation (Mdn = .0099) conditions, $p = .001$. There was no significant difference between the generation and conversation conditions. Figure 3.1 displays the differences in the medians among the three sample types.
Lexical Rarity Using SUBTLEX

Similarly, the overall Friedman test comparing rare word use as analyzed using the SUBTLEX procedure was significant, $\chi^2(2) = 16.98$, $p < .001$. Pairwise comparisons with a Bonferroni correction for multiple comparisons revealed statistically significant differences in rare words produced between the retell (Mdn = .0087) and the generation (Mdn = .0030) conditions, $p = .017$, and between the retell and conversation (Mdn = .0026) conditions, $p < .001$. There was no significant difference between the generation and conversation conditions.
Relationship Between Lexical Rarity Scores and Language Test Scores

To estimate the relations between rare word use and vocabulary and overall language test scores, Spearman correlations were performed. The data met the assumptions for the use of the Spearman (continuous data and paired observations). For this analysis, the total number of rare words across the three sample types divided by the total number of words across samples served as the dependent variable, with standard scores for the three language tests as independent variables. No significant relations were observed. The correlations are provided in Table 3.2.

Table 3.2: Correlation Between the Proportion of Rare Words and Language Assessment Scores

<table>
<thead>
<tr>
<th>Language Measures</th>
<th>WERVE $r_s$</th>
<th>SUBTLEX $r_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CELF-5 CLS</td>
<td>-0.217</td>
<td>0.076</td>
</tr>
<tr>
<td>PPVT-4</td>
<td>-0.196</td>
<td>0.104</td>
</tr>
<tr>
<td>EVT-2</td>
<td>0.165</td>
<td>0.102</td>
</tr>
</tbody>
</table>
Chapter 4: Discussion

Summary of Findings

Language sampling is a critical tool in language analysis, as it provides clinicians and researchers with a holistic view of a child’s language skills (Klatte, Van Heugten, Zwitserlood, & Gerrits, 2022) However, it is important to consider the fact that not all language sampling methods provide children with ample opportunity to demonstrate the upper limits of their vocabulary. For example, a child might use less advanced language during casual conversation than during an explanation of a specific topic (Nippold et al., 2005). One way to adjust sampling methods to promote the use of more advanced vocabulary is through use of expository discourse tasks as part of a language assessment. Expository discourse samples can provide different information about an individual’s language skills by encouraging the discussion of more complex topics than typically arise in conversation (e.g., Lundine et al., 2018; Nippold, 2010). Furthermore, expository discourse is an essential skill, especially during the school-age years when children are expected to use their oral and written language to demonstrate an understanding of learned material (Westerveld & Moran, 2011).

When analyzing language samples, lexical rarity is a measure that can be used to understand the upper limit of a child’s vocabulary use. Rarity can be measured through preidentified rare word lists, by identifying words in the samples that appear on the rare word lists. It can also be measured using the subtraction method with a common word list, identifying the words in the children’s samples that do not appear in the list. Both of these methods are feasible using freely available computerized language sample tools such as CLAN (MacWhinney, 2000). For this study, I used the first method to identify rare words in the language samples of school-age children. I used two preexisting rare word/word frequency lists.
The first, WERVE, was a rare word list developed by Mahurin-Smith and colleagues (2015). As described in Chapter 1, these authors derived their rare word list from a corpus containing 1.2 million words. They selected as “rare” those words that occurred 15 times or fewer. For the second lexical rarity analysis, Brysbaert and New’s (2009) SUBTLEX word frequency corpus was used. This corpus differed from WERVE in that it contained over 51 million tokens. In order to compile the list of rare words for the present study, I selected a cutoff of words that were used once per million units or less as “rare”. This cutoff resulted in a list of 43,004 different words. The decision to use the once-per-million cutoff, while arbitrary, resulted in a list of words that seemed appropriately rare for school-age children. Appendix A provides examples of words that met this cutoff and were therefore considered “rare”.

When comparing expository retell, expository generation, and conversational language samples, expository retell samples contained a higher proportion of rare words, in using both the WERVE analysis and the SUBTLEX analysis. In addition, there were no significant correlations between children’s rare word use and their scores on standardized language measures. Each of these results is discussed in greater detail below.

**Lexical Rarity**

Research questions 1 and 2 sought to examine the differences in lexical rarity across language sample types using different measures. When measuring with both WERVE and SUBTLEX, expository retell samples had a significantly higher proportion of rare words used than conversation and generation samples. During expository retell, the children were asked to watch a video about hydroelectric power and then provide an explanation of the content. Since the content of the video was more advanced, we hypothesized that children would use more rare
words in these samples. Our hypothesis was supported by the findings; however, it should be noted that other retell tasks that involve less conceptually complex topics may produce samples with fewer rare words.

In addition, since the children were encouraged to take notes while watching the video and to use those notes in their retelling, they may have written down how the novel words were used in the video and then repeated the words – or even the phrases in which they were embedded – in their expository retellings. This could mean that while the children used the words in the correct context, they did not truly understand the full meanings of the words. If this occurred often, it could impact the children’s lexical rarity scores and provide an inaccurate representation of their actual rare word use. For this study, we did not compare the context of each rare word used to the words and their context within the video. However, because the children were limited in their ability to take notes in full sentences (i.e., they were only given two opportunities to take notes with the video paused, and rewinding of the video was not permitted), it seems unlikely that their notes included robust context of rare vocabulary from the video. Therefore, it seems reasonable to assume that the novel words that were encountered in the video and used properly were in fact understood by the participants. Although retell tasks should be used to obtain expository samples, as they may enable children to display the upper limits of their vocabulary, it is important for researchers/clinicians to be cognizant of the fact that the stimulus content used can impact the types of words children produce in their samples.

These findings align to some extent with those of Mills and colleagues (2017), who found that rare word use when measured using WERVE differs across language sample types, specifically, fictional and personal narratives. In particular, they observed greater rare word
density in children’s personal narratives than in fictional narratives. Their study was of narrative discourse rather than expository discourse, and therefore, the studies are not directly comparable. However, it is worth noting that the rarity findings by Mills et al. favored discourse containing personal information over non-personal (i.e., fictional) – whereas the present findings favored non-personal (video retell) over personal (favorite game/sport and conversation). This difference in findings across studies may be due to differences in discourse types (narratives vs. exposition) or task demands. Specifically, Mills and colleagues used the “Frog, Where Are You?” (Mayer, 1969) picture book to elicit fictional narratives. Each child was asked to tell a story based on the pictures in the book. For the personal narrative task, participants were instructed to teach the examiner a card game, and then during that card game, the examiner presented a story prompt to elicit the narrative. The fictional narrative task had simple pictures, and the participants were potentially limited in producing their narratives by the pictures and the simplicity of the storyline. In contrast, in the personal narrative task, participants had the freedom to expand on a simple prompt. This likely provided a greater opportunity for advanced vocabulary use in the personal narratives. In the present study, the stimuli for the non-personal discourse was a scientific video, providing more opportunities for children to show the upper limits of their vocabulary compared to the Mills et al. fictional narrative task.

The rarity findings using WERVE vs. SUBTLEX were similar, in contrast to the hypothesis. Our hypothesis was that SUBTLEX would identify a greater proportion of rare words because, among other factors, it was a much larger, more robust list of rare words. Table 4.1 provides a descriptive comparison of the two rarity measures, which may be used to understand the strengths and limitations of each approach. Perhaps the biggest weakness of SUBTLEX is that the dataset is not based on child language but on movie and television
subtitles; therefore, the frequencies used to derive lexical rarity may not be appropriate. On the other hand, the SUBTLEX data set is considerably larger, increasing the likelihood that rare words the children used would be captured in this dataset.

To identify rare words using the SUBTLEX rare words list, we established a criterion of once per million units as “rare”. This cutoff was arbitrary, however, and therefore, may not have been the most appropriate considering the corpus was not generated from samples of children’s language. Had a less strict criterion been selected, there would have been a higher number of rare words identified by the SUBTLEX list. For example, if the cutoff was words used five times per million units some words on the list would be artistic, overwhelming, and evaluation. In addition, if the cutoff had been words that were used ten times per million units, words like invasion, primary, determine, and instrument would have been on the list. Future research using the SUBTLEX database may evaluate the use of different cutoffs, depending on the age of the children in the study.

Table 4.1: Comparison Between WERVE and SUBTLEX

<table>
<thead>
<tr>
<th>WERVE</th>
<th>SUBTLEX</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td><strong>Cons</strong></td>
</tr>
<tr>
<td>Created with the purpose of studying rarity.</td>
<td>The children were much younger than our study participants.</td>
</tr>
<tr>
<td>Has been studied previously.</td>
<td>Created using conversational samples.</td>
</tr>
<tr>
<td>Only contains about 1.2 million words.</td>
<td>Contains word frequency calculations.</td>
</tr>
</tbody>
</table>
Relations Between Language Test Scores and Lexical Rarity

In their 2021 study, Mahurin-Smith and colleagues used the WERVE list to examine school-age children’s rare vocabulary use in narratives. They found that rare vocabulary use was significantly correlated with PPVT-4 scores. This led us to hypothesize that rarity would be significantly and positively correlated with language assessment performance of our participants, as well. Conversely, in the present study, rarity was not significantly correlated to performance on any of the standardized language tests: the CELF-5 Core Language Score, the PPVT-4, or the EVT-2. One reason our findings may have differed relates to the sample demographics across studies. Mahurin-Smith et al. studied children from low-income communities exclusively. In contrast, the sample for the present study contained only two of 26 participants who, based on the Hollingshead (1975) Four Factor Index of Social Status, were classified as “lower-class”. Subjectively, most of the children in this study were recruited from the university community and did not necessarily represent the larger population of children within their community. The children tended to perform in the high-average to above-average range, as can be seen in Table 2.1. Note that the sample means for all tests exceeded the tests’ standard score mean of 100. Therefore, the difference in correlational results between studies may have been due to the participants recruited for each study.

Similar to Mahurin-Smith et al. (2021), Kelley and Poholik (2023), found that in conversational play samples of preschool children, lexical rarity as measured by WERVE positively related to standardized vocabulary measures. Their findings possibly differed because the children in their study were younger than those used to generate the WERVE list, while those in this study were older. Kelley and Poholik also used similar sampling methods to Mahurin-Smith and colleagues (2015). Therefore, it is likely that WERVE served as a more accurate
representation of rarity for children in that study. An additional factor to consider is that Kelley and Poholik calculated rare word density somewhat differently – as the total number of rare words divided by the total number of utterances. In contrast, we used the total number of rare words divided by the total number of words in the sample. These calculation differences may account for the differences in results.

Another possible explanation of the differing findings across studies is that the Core Language Score of the CELF-5 is intended to be a composite score of overall receptive and expressive language, and this score may not reflect the underlying skills needed to use rare words in context. If our analyses had included specific subtest standard scores, rather than just a composite score, results may have differed. For example, the Formulated Sentences subtest on the CELF-5 evaluates the student’s ability to formulate complete grammatically correct spoken sentences with varying complexity (Wiig et. al., 2013). Children who used more rare words in exposition and conversation may have had higher scores on this subtest since it focuses on their use of semantics and other language factors.

Limitations

The data for this project were collected by Guilkey and Wagovich (2022) for a thesis project at the University of Missouri. Convenience sampling was used; therefore, most participants were white, middle-class children. The sample of 26 children was relatively small but deemed sufficient to detect large, clinically meaningful effects. However, together, these factors might impact the generalizability of these findings.

In addition, there is no prior research using SUBTLEX to measure lexical rarity. As a result, the cutoff criterion was selected based on a perceptual judgment that, for children in the school-age years, words that occurred once per million could reasonably be considered rare. Of
course, a limitation of all lexical rarity analyses that are based on word frequency data is the inherent assumption that, because a word occurs less frequently in the external dataset, it is necessarily rare in the children’s vocabulary. There can be other reasons that a word does or does not appear in an external dataset besides the word being truly “rare” within the language. Nonetheless, the larger the word frequency dataset, the less likely this limitation would be to impact the analysis of rarity. SUBTLEX is a large dataset containing a highly diverse set of subtitles across movies and television and, as such, it deserves additional study as a tool for both word frequency analysis and lexical rarity analysis.

Although the present study captured lexical rarity using the proportion of different rare words a child used, another way to measure rarity is the total proportion of rare words used (including rare words that occurred multiple times in the sample). It may be valuable in future studies to analyze the proportion of total rare words, as well, as Kelley and Poholik (2023) did, because this could provide information about how comfortable children are using rare vocabulary in discourse (i.e., repetitions of the same rare word may point to greater familiarity or comfort with the word).

Although we use proportional data in the present study, another approach is to use the raw number of rare words included in the samples (rather than proportions), while controlling statistically for the length of the samples. Using this approach may have addressed the issue of non-normal distributions for many of the variables in the dataset.

Lastly, exposition is simply providing an explanation; therefore, some of our conversational samples did include instances of exposition. However, these moments of exposition only made up a small percentage of the samples. In a post-hoc analysis, approximately 7.7% of utterances overall in the conversational samples were deemed expository-
like. Moreover, while conversation involves dialogue between partners, expository discourse is (at least in our tasks) produced as monologue. Therefore, we consider conversation a separate type of discourse for the purposes of this study.

Conclusions and Future Directions

This study sought to compare school-age children’s rare word use in two types of expository discourse (retell and generation) and conversation, to provide insight into lexical rarity across sampling methods. Rarity was measured using WERVE, a rare word list created by Mahurin-Smith and colleagues, and a novel word list developed using Brysbaert and New’s (2009) SUBTLEX corpus of words collected from subtitles of movies and television. The results showed that across rarity measurement tools, the proportion of lexical rarity was greatest in expository retell samples. Contrary to previous findings, the results also indicated that there was no significant correlation between the children’s rare word use and their standardized language and vocabulary assessment scores. This was possibly because rarity is a specific analysis of one facet of language (and one facet of vocabulary), while the standardized assessments, particularly the CELF-5, measured overall language skills. Our findings may suggest that children’s language abilities overall do not necessarily predict their rare vocabulary knowledge and use.

In future studies and clinical practice, WERVE can be used as a tool for measuring rarity; however, it would be best applied to conversational samples of children of a similar age to those who contributed to the data on which WERVE is based. The WERVE list was generated from conversational samples of children 6 to 8. Other sampling methods like expository retell and generation may yield rare words that should be considered rare but are not on the WERVE list because of differences across contexts. For example, children in our study frequently used the
word “hydroelectric” in their expository retell samples as it related to the stimulus materials. While this is an advanced vocabulary word, it was not identified as rare by WERVE.

The SUBTLEX list is an excellent, robust resource that can be adapted as a rarity measure across a range of developmental levels. Because of the size of the data set, different rarity criteria can be set as appropriate for the age of the children. If enlisted for rarity analysis in the future, users should be mindful in selecting a cutoff for their analysis that would be most appropriate for the age of the participants. It should be noted that, because the SUBTLEX corpus is so large, it is not as vulnerable to the context-specific issues noted above. For example, “hydroelectric” appears in the corpus and was identified as a rare word, with a SUBTLWF value of 0.10.

Future research should also continue to examine the relation between rarity and language knowledge and skills. In addition, examination of lexical rarity may augment vocabulary test data, providing a more detailed picture of the child’s overall vocabulary skills. For example, a child’s vocabulary testing may demonstrate average range performance, but rarity analysis may suggest that the child is using some words that are well above developmental expectations. This information would lead to a more detailed understanding of the child’s vocabulary capabilities in discourse.

Overall, lexical rarity measures have the potential to provide a better understanding of an individual’s language skills. Similarly, expository discourse samples provide individuals the opportunity to convey complex content, using sophisticated vocabulary in their explanations. Therefore, best practices in assessing language should not just be limited to standardized measures of language but should include language sampling of children’s production of complex
content, while analyzing lexical skills in discourse, including the use of lexically rare forms. This study adds to the growing literature that highlights the importance of measuring rare word use in children in the school-age years.
References


https://doi.org/10.2307/1166112


https://doi.org/10.3758/BRM.41.4.977


https://doi.org/10.2307/3587951


Hollingshead, A.B. (1975). Four factor index of social status. Unpublished manuscript. Yale University, New Haven, CT.


https://doi.org/10.1044/2016_AJSLP-15-0157


https://doi.org/10.2307/1162893


https://doi.org/10.1044/1092-4388(2005/073)


https://doi.org/10.4324/9780203848821


https://doi.org/10.1177/00222194060390010501

https://doi.org/10.3389/fpsyg.2021.686478

https://doi.org/https://doi.org/10.1111/j.1467-8624.2012.01805.x


http://dx.doi.org.utep.idm.oclc.org/10.1044/2014_AJSLP-14-0046

https://doi-org.utep.idm.oclc.org/10.1044/1058-0360(2004/032)


Appendix A

Examples of SUBTLEX Words that Met Rarity Criteria (≤1 per million units)
<table>
<thead>
<tr>
<th>Word</th>
<th>Word</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>abduction</td>
<td>leisurely</td>
<td>slider</td>
</tr>
<tr>
<td>advantageous</td>
<td>neuron</td>
<td>surpass</td>
</tr>
<tr>
<td>comforter</td>
<td>pacifier</td>
<td>tactile</td>
</tr>
<tr>
<td>favoritism</td>
<td>pillowcase</td>
<td>tsunami</td>
</tr>
<tr>
<td>holistic</td>
<td>pterodactyl</td>
<td>unbeatable</td>
</tr>
<tr>
<td>incline</td>
<td>radiologist</td>
<td>validate</td>
</tr>
<tr>
<td>knelt</td>
<td>smoothie</td>
<td>whimsical</td>
</tr>
</tbody>
</table>
Vita

Serena Jae Davis graduated from Texas Woman’s University in 2018 with a Bachelor of Science degree in Communication Sciences and Disorders with a Biology minor. During her undergraduate experience, Serena was a member of the Honors Scholar Program for which she completed a capstone research project entitled “Occupational Noise-Induced Hearing Loss: Insights and Modifications for Current Hearing Conservation Programs”. She then went on to pursue her master’s degree in Speech-Language Pathology at the University of Texas at El Paso in August of 2022. While at UTEP, Serena worked closely with her advisor, Dr. Stacy Wagovich, to develop and execute her thesis project on school-age children’s rare word use in expository discourse. She also supervised members of the Language and Fluency Laboratory to conduct reliability checks for her thesis. Serena presented her thesis findings as a primary presenter at the Texas Speech-Language-Hearing Association’s (TSHA) annual conference in February 2024. Serena is expected to graduate with her Master of Speech-Language Pathology degree in May 2024.

Contact Information: serena.jae16@gmail.com