5-6-2021

# Language sampling with older school-aged children: Examination of analyses and sampling contexts

Michelle Lenhart
*James Madison University*

LANGUAGE SAMPLING WITH OLDER SCHOOL-AGED CHILDREN:

EXAMINATION OF ANALYSES AND SAMPLING CONTEXTS

Michelle H. Lenhart

A dissertation submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

Department of Communication Sciences and Disorders

May 2021

---

FACULTY COMMITTEE:

Committee Chair: Geralyn R. Timler

Committee Members/ Readers:

Stacey L. Pavelko

Carol C. Dudding

Dannette A. Bronaugh

## Acknowledgments

It is with profound gratitude (and a little disbelief) that I write this acknowledgements section because reaching this point in my education has taken a collective of many individuals. First, I would like to thank the members of my dissertation committee whose guidance and support not only facilitated but enabled the completion of this project. I sincerely thank Dr. Geralyn Timler for her patience, encouragement, and feedback as chair of this committee. Dr. Stacey Pavelko is thanked for her guidance throughout the project, but especially in the early phases of formulating these questions and making community partnerships. I am grateful to Dr. Carol Dudding for her support and friendship for eight years prior to and during this program. Additionally, my thanks go out to Dr. Dannette Allen Bronaugh for her support and particularly for her sharp eye for detail which helped this document reach its current state.

Although this committee of individuals provided the brunt of the academic and research support for this project, the entire James Madison University Communication Sciences and Disorders community, including faculty, staff, and students played a role in this accomplishment. This included via collegiality, instilling critical thinking in a Master's student, access to community contacts, or being an ear off which to bounce ideas. Special thanks go out to my fellow doctoral students, especially Danielle, Julian, Lakshmi, Danika, Shiree, Linda, Lindsay, and Daniel. Additional thanks go out to the students at JMU who worked to process the data, including Hannah Schroyer, Mary Farinholt, Emma Sprance, Katherine Krese, and Samantha Slater.

I express my thanks to the many school systems, organizations, and families who allowed me into their facilities, lives, or homes for the purposes of collecting this data. Particular organizations to thank include Culpeper County Public Schools, Epiphany Catholic School, Ottobine Elementary School, and the Boys and Girls Club of Staunton and Waynesboro, among others. Without these organizations and the cooperation of some kind children, this research would not be possible. Additional heartfelt thanks go out to the educators in my life who helped stoke my love of learning, curiosity for how things work, and encouraged me to continue to grow, learn, and discover.

The doctoral education process was not without its twists and turns, including but not limited to the birth of a child, a global pandemic, and accepting a new position during my schooling. This new position at Radford University blessed me with another tribe of colleagues and students to thank. To the faculty and staff in the Communication Sciences and Disorders department at Radford University, you have my heartfelt thanks for the support before, during, and after the dissertation defense process. The Radford family is not limited to the individuals on payroll, but the undergraduate and graduate students offered kind words or actions of encouragement that will not be forgotten. Particular gratitude goes to Stephanie Wood and Molly Maurin who worked thoroughly on preparing the samples for data analysis.

Finally, but perhaps most importantly, I would like to thank the people that I call my family. Thank you to my parents, Mike and Theresa Harris, for instilling values such as perseverance and encouraging personal growth and development, without which this journey would not have been pursued. I offer gratitude to the other parent figures in my life, including my in-laws, grandparents, aunts, and uncles, with special mention to Gayle

Harris. Thank you to my support system, which includes my brother, Timmy Harris, dear friends, especially Calyn Witz and Dr. Nathaniel Mason among many others. I struggle with the words to adequately thank the final two individuals, my husband, Zachary Lenhart, and the light of our lives, Paul Lenhart. Thank you for being you, and making sacrifices so that I can be a better version of myself. You mean the world to me.

## Table of Contents

## List of Tables

## List of Figures

**Abstract**

Purpose: Conversation, narrative, and expository language sampling contexts are recommended for school-aged children (Pezold et al., 2020), and multiple ways to analyze these samples have been promoted in the clinical literature. This dissertation addressed two gaps in the literature related to analyses and sampling contexts. The purpose of study one was to examine differences in two commonly-used language sample analysis methods, Sampling Utterances Grammatical Analysis Revised (SUGAR) and Systematic Analysis of Language Transcripts (SALT). The purpose of study two was to examine the presence of age-related changes in conversation, narrative, and expository contexts for older school-aged children.

Method: Conversational, narrative, and expository language samples were elicited from 85 typically developing children ages 8-11. In study one, the conversation language samples were transcribed and analyzed using both the SUGAR and SALT conventions for all children ages 8-10 (n=68). Four paired language performance metrics were calculated including (a) Total Number of Words (TNW) and Number of Total Words (NTW), (b) Mean Length of Utterance SUGAR ($MLU_S$) and Mean Length of Utterance in Morphemes (MLUM), (c) Words Per Sentence (WPS) and Mean Length of Utterance in Words (MLUW), and (d) Clauses Per Sentence (CPS) and Subordination Index (SI). Z-scores were calculated by comparing the computed values to their respective normative databases. Differences in z-scores were examined for each paired metric. Next, to determine if SUGAR and SALT conventions would classify children differently based on language status, these z-score values were compared to thresholds suggestive of typical development versus suggestive of language impairment.

In study two, the conversation, narrative, and expository language samples were transcribed and analyzed for mean length of utterance in words (MLUW) and subordination index (SI), a measure of complex language, using SALT conventions.

Results: Results for study one revealed statistically significant differences in mean z-score values for all paired language sample analysis metrics. The SUGAR z-scores were significantly lower for the TNW/NTW, $MLU_S$/MLUM, and WPS/MLUM comparisons, but SUGAR z-scores were significantly higher for the CPS/SI comparison. In addition, while no children fell below the thresholds suggestive of language impairment using SUGAR, four children fell below the threshold using the SALT conventions. Study two revealed no age-related changes in MLUW or SI. Sampling contexts did yield significant differences. Narrative and expository language samples elicited longer utterances and more complex language than conversation samples.

Conclusions: Comparisons between different normative databases for language sample analyses may yield differing clinical conclusions. Although age-related changes were not detected in any context, the results indicate that narrative and expository contexts may be more appropriate language sampling contexts to examine complex sentence use in school-aged children ages 8-11 years.

# Introduction

Approximately 7% of school-aged children (Tomblin et al., 1997) have a language impairment (LI), usually characterized by delayed morphosyntactic development including production of shorter sentences, more grammatical errors, and reduced production of complex sentences than same-age typically developing peers. An important tool for the identification of LI in school-aged students is the collection and analysis of language samples. Language sample analysis (LSA) has been found to be diagnostically accurate for the identification of language impairment in school-aged children and adolescents (Heilmann et al., 2010a; Pavelko & Owens, 2019). Furthermore, LSA offers a functional and more authentic view into a child's speech and language skills and minimizes cultural/linguistic bias (Paul et al., 2018; American Speech-Language-Hearing Association, 2016). Despite the benefits of LSA, practicing SLPs report using this clinical tool infrequently (Pavelko et al., 2016). One key barrier to LSA use is the existence of multiple methods for performing the analyses within the literature. This variety of analysis methods leads to ambiguity about which method is the most appropriate to use, which may cause SLPs to use unstandardized methods or cite lack of training as reasons to avoid language sampling (Pavelko et al., 2016). An additional problem with current LSA use is that practicing SLPs report primarily using conversational contexts (Pavelko et al., 2016), despite the scholarly literature showing that sampling contexts such as narration and exposition have been found to be more sensitive to developmental changes, particularly in adolescents (Nippold, et al., 2008).

Less is known about whether these developmental changes can also be observed in school-aged children between the ages of 8 and 11. Further examination of analysis methods and the effects of sampling contexts on the identification of language impairment in school-aged students is needed to inform assessment guidelines. The two studies described here focus on these gaps in the clinical literature. In study one, two commonly-used LSA methods will be compared. In study two, age-related changes across multiple language sampling contexts will be examined. In the sections that follow, relevant literature for each study is reviewed followed by the research questions and hypotheses addressed within each study.

**Study One: Analysis Methods**

One reason for the stated lack of clinician expertise (Pavelko et al., 2016) with language sampling could be due to the variety of language sample analysis procedures. For example, one common metric of LSA is the calculation of the mean length of utterance (MLU). MLU is most commonly defined as the average number of morphemes within an utterance. A related measure is the mean number of words per sentence (WPS) or per utterance (MLUW) (Pavelko & Owens, 2017; Miller et al., 2019). Multiple methods are now available to calculate MLU and a variety of other LSA metrics (Brown, 1973; Casby, 2011; Miller et al., 2019; Pavelko & Owens, 2017; Retherford, 2007). The use of computer-based transcription and analysis technology can not only increase the diversity of analyses available to aid in clinical decision-making (Miller & Iglesias, 2019), but may also improve the accuracy of the analyses (Long, 2001). Two commonly-used computerized LSA methods, the Sampling Utterances Grammatical Analysis Revised (SUGAR; Pavelko & Owens, 2017) and the Systematic Analysis of Language Transcripts (SALT; Miller et al., 2019) have both been shown to accurately identify children with a language impairment (Heilmann et al., 2010a; Pavelko & Owens, 2019), but the conventions for utterance segmentation and calculation of metrics for analysis differ. SUGAR and SALT methods are summarized below.

SUGAR is a newly developed computerized language transcription method. This method describes a protocol for eliciting a robust conversational speech sample from children ages three through 10;11 (Pavelko & Owens, 2017; Owens & Pavelko, 2020). SUGAR was developed in response to many of the barriers to use of LSA in clinical practice (Pavelko et al., 2016). Specifically, the sampling protocol is designed to be easy to elicit, quick to transcribe, and it does not require any special equipment or software. Although SUGAR is new to the language sample analysis literature, it has been shown to be diagnostically accurate (Pavelko & Owens, 2019), appropriate for a variety of ages (Pavelko & Owens, 2017; Owens & Pavelko, 2020), and to support the development of treatment goals (Owens et al., 2018).

SALT (Miller & Iglesias, 2019) is a computerized LSA software available for purchase. The SALT website and SALT clinician guides (Miller et al., 2019) offer elicitation protocols, transcription conventions, and a variety of analyses for use in LSA. After sample transcription and coding, SALT can automatically tally the codes of interest. The SALT databases provide norms for a plethora of language sampling contexts and diverse ages. Databases are available for play, conversation, narrative story generation and retell, expository, and persuasive samples elicited from monolingual English-speaking children. Narrative databases are also available for bilingual Spanish/English speakers as well as monolingual Spanish speakers. The databases include samples for children as young as 2;8 to young adults aged 18;9 (SALT Software, 2020). SALT has been used extensively in language development and disorders research and has been found to accurately identify language impairment and has served as a progress monitoring tool (Heilmann et al., 2010a; Miller et al., 2013).

LSA offers the opportunity to examine and quantify many aspects of syntax and semantics. Namely, both SALT and SUGAR provide metrics to quantify the total number of words (TNW for SUGAR; NTW for SALT) within an entire sample. At the sentence level, SUGAR provides mean words per sentence (WPS in SUGAR) to measure the average sentence length in words. In SALT, similarly to WPS, the mean length of utterance in words (MLUW) quantifies the mean number of words within a communication unit. Drilling down further from the number of words to the number of morphemes, both SUGAR and SALT provide normative values for the mean length of an utterance in morphemes ($MLU_S$ and MLUM, respectively). When examining syntactic development, LSA methods quantify the use of increasingly advanced syntax by using measures of syntactic complexity. When considering syntactic complexity, SUGAR uses clauses per sentence (CPS; Pavelko & Owens, 2017), and SALT offers a similar metric, the Subordination Index (SI; Miller & Iglesias, 2019). These metrics (CPS and SI) quantify the clausal density of a given unit, or the number of clauses within a given unit (e.g. a sentence or C-unit).

Although the names of these metrics may be similar, the values are not identical because SUGAR and SALT differ in both rules for morpheme counting and utterance segmentation (Miller et al., 2019; Pavelko & Owens, 2017). Major differences between the SUGAR and SALT methods are described in Table 1. Lack of clarity regarding the appropriate methods of segmentation and analysis to use may lead to clinicians abstaining from language sample analysis or using self-designed protocols that lack empirical testing (Pavelko et al., 2016). These differences in utterance segmentation and morpheme counting are likely to affect the raw totals in terms of the mean length of utterance and number of utterances from a given language sample.

In a recent tutorial on language sampling, Pezold and colleagues (2020) transcribed and analyzed the same two adult-child play language samples from preschoolers using both SUGAR and SALT conventions. The authors then calculated the mean length of utterance ($MLU_S$ and MLUM respectively) in standard deviation (SD) units from the corresponding database mean. For example, the same sample using SUGAR was 2.41 units below the mean for $MLU_{S,}$ whereas using SALT, it was 1.93 SD below the mean for MLUM. With nearly a half standard deviation of differences between the two metrics when examining the Z-score, this work is illustrative of the potential differences between the two analysis conventions. However, examination of a single metric with two preschool children is insufficient to inform clinical practice, so a larger sample is needed. These children were also preschoolers, so no information is available about the differences in school-aged children. Notably, the sampling context was play-based, which is not the recommended language sampling context for SUGAR (Pavelko & Owens, 2017). The norms for SUGAR were developed from adult-child conversation samples without the use of toys. This context difference could be a reason for standard score differences between the SALT normative database and the SUGAR normative sample because play-based samples typically elicit less complex language in conversational samples (Southwood & Russell, 2004). In order to effectively use language sample analysis to inform a diagnostic decision, it is imperative that clinicians are using identical sampling methods and analysis procedures as the database to which they are comparing.

***Research Questions Study One:***

It remains unclear whether use of SUGAR or SALT transcription conventions and analysis rules yield differences in the clinical conclusions that stem from each analysis. As such, this study addressed the following research questions about conversation samples collected from a sample of typically developing 8;0 to 10;11-year olds:

1) Does the choice of LSA method affect z-score values (i.e. standard deviation differences from the mean) for measures of language productivity (TNW/NTW, WPS/MLUW, and $MLU_{S/}$MLUM)?
2) Does the choice of LSA method affect z-score values (i.e. standard deviation differences from the mean) for a measure of language complexity (CPS/SI)?
3) Are there differences in classification (suggestive of typical development versus suggestive of language impairment) for metrics of language productivity and complexity using the SUGAR and SALT methods of language sample analysis?

It is hypothesized that there will be no significant differences between SUGAR and SALT z-scores for each of these metrics. It is hypothesized that, because the sample consists solely of typically developing children, there will be no classification differences using the two analysis methods.

**Study Two: Sampling Contexts**

The best language sampling contexts to use with children change with age (Pezold et al., 2020). This is because one of the purposes of language sampling is to elicit the most complex language possible for children's developmental age while remaining developmentally appropriate. There are five commonly used sampling contexts: play, conversation, narration, exposition, and persuasion. Developmental information on the use of syntax across language contexts for young children and for adolescents is available in the clinical literature but is limited for older school-aged children between the ages of 8 and 11. The recommended language sampling contexts for elementary aged students (as seen in Figure 1) are conversation, narration, and exposition (Pezold et al., 2020). Relevant literature for each of these three sampling contexts is reviewed.

### *Conversation Samples*

Conversation, or a dialogue between individuals in which each contributes by making statements and responding to questions, is appropriate for preschool and elementary aged children. Conversation is the primary context used by practicing SLPs, as reported in a survey by Pavelko and colleagues (2016). Use of only conversational contexts is not recommended for older children. Nippold and colleagues (2008, 2014) found that conversational contexts may not be adequately sensitive to detect changes in complexity and productivity of children by the time they reach adolescence. In a cross-sectional study, Nippold and colleagues (2005) found some age-related changes in productivity (total T-units) and complexity (as measured by mean length of T-unit) in a conversational context between the age groups of 8- and 44-year old typically developing individuals, but the growth patterns were not linear and the data is presented within wide age ranges, such as 8 and 11, making it difficult to extrapolate how and if productivity and complexity performance changes within this age range. Pavelko and Owens (2017) found that using a robust conversational sampling context in which the use of process questions and commenting are maximized and yes/no questions are minimized yielded age-related changes in clauses per sentence and a measure of mean length of utterance ($MLU_S$) up through age 7;11 (Pavelko & Owens, 2017). Furthermore, measures of mean length of utterance SUGAR ($MLU_S$), total number of words, and words per sentence yielded significant age-related changes up through age 10;11 (Owens & Pavelko, 2020). A study examining both older students with and without language impairment found that differences in mean length of T-unit and nominal, relative, and adverbial clause use were no longer present in conversational speech in adolescents ages 12 through 15 (Nippold et al., 2008). Although by adolescence it is clear that conversation is no longer recommended, *when* the age-related differences in conversation diminish is not clearly established.

### *Narrative Samples*

Narrative, or storytelling, is a recommended language sampling context for preschoolers all the way through adolescence (Pezold et al., 2020). For adolescents, narrative contexts have been found to elicit more complex language, such as higher clausal density, than conversation alone (Nippold et al., 2014, 2015). In a seminal language sampling study, Scott and Windsor (2000) found differences between students with language learning disabilities, defined in the study as students receiving special education services for reading and spoken language difficulties, and typically developing peers in nine to twelve-year old children using a narrative context, comprised of viewing and then retelling the contents of a video depicting a story. They also gave this task to younger children who were language-aged matched with the individuals with language learning disabilities. Students with language learning disabilities produced fewer communication units and had fewer words per unit on average than age-matched children, and students with language learning disabilities produced more grammatical errors than both age- and language-matched children. Use of video, while offering standardization of sample collection, provides limited application for clinical practice. Firstly, the videos are not commercially available. Secondly, each video is approximately 15 minutes in duration and thus collection and analysis of these samples may be too time-consuming for standard clinical practice. Because lack of time is commonly cited as a reason that LSA is not used (Pavelko et al., 2016), the time needed to replicate this narrative task reduces the clinical utility of this research. Additionally, Scott and Windsor's groups of younger typically developing children (7;9-10;3) and older typically developing children (9;10- 12;11) consisted of overlapping age ranges, making it difficult to use the data to make age-related clinical comparisons for children ages 9 and 10. More recent research has documented the presence of age-related changes for individuals ages four to eighteen in a narrative generation task using a wordless picture book, but the data is not available in a method allowing for age-matched comparisons, as the authors did not report mean performance for each age (Channell et al., 2018).

Narrative retell language sampling has also been examined with short texts in adolescents, ages 12;10-14;11. Namely, Nippold and colleagues (2014, 2015, 2017) completed multiple narrative language sampling studies using retellings of Aesop's fables. Narrative language samples consisting of these stories elicited more syntactically complex language, as measured by clausal density and mean length of communication unit (MLCU), than conversational samples. Furthermore, in a follow-up, many of the participants were able to answer critical thinking questions as it relates to these fables.

These protocols have only been examined with adolescents, despite the reading level for several of the fables being appropriate for school-aged children.

### *Expository Samples*

Exposition is used to impart information on a communication partner and is referred to as the "language of the curriculum," making it especially important for school-aged children to be able to understand and produce expository discourse (Ward-Lonergan, 2010). Again, for adolescents, expository contexts have been found to elicit more complex language, as measured by mean length of T-unit (MLTU) and clausal density, than conversation alone (Nippold et al., 2008). Scott and Windsor (2000) also found differences between students with language learning disabilities and typically developing peers in nine to twelve-year old children when completing an expository language sample task in response to a short video describing a desert, as described above.

A commonly used expository context within the limited literature is the favorite game or sport (FGS) task. This is a procedural expository task in which participants are asked to describe the rules and strategies of a favorite game or sport. Westerveld and Moran (2011) compared the performance of 6- and 7-year olds to 11-year olds. Both groups were able to complete this task, and age-related changes were found in total T-units, but not the length of T-unit (MLTU). Nippold and colleagues (2005) used an expository task to examine language productivity and complexity across a range of childhood and adulthood ages but did not report age-related changes by years as the research question focused on differences between children and adults across a wide range. In a later study using the FGS task, Nippold and colleagues (2008) found that mean length of T-unit values differed between adolescents with typical language development and those with language disorders.

There are multiple subtypes of expository discourse. The six most common subtypes are procedural, descriptive, enumerative, cause/effect, compare/contrast, and problem/solution (Lundine & McCauley, 2016). There is evidence that the type of expository discourse can affect language performance measures (Lundine et al., 2018). Lundine and colleagues investigated adolescents' performance in summarizing cause-and-effect and compare-and-contrast expository samples and found differences in performance across these differing macrostructures. For example, overall summary quality, MLU, and subordination index (SI) were higher in the cause-and-effect context than in the compare-and-contrast context. While the favorite game or sport task is a procedural exposition context, research is limited regarding changes in development for these other expository contexts via language sampling.

Nippold and colleagues (2007) examined the use of a problem-solution expository context known as the peer conflict resolution task. Here, participants are given two social scenarios in which a problem occurs. They are asked to retell the scenarios and answer problem-solving questions regarding how to handle the situations. This task was found not only to elicit more complex language from the adolescents than conversation (Nippold et al., 2007), but performance was also found to differ between adolescents with typical development and those with language disorders (Nippold et al., 2009). Use of peer conflict tasks in older school-aged children has not yet been examined.

Language sampling is considered best practice to assess expository skills of children (Nippold, 2014), but more evidence-based protocols and analysis techniques need to be developed in order to guide clinicians through this process (Lundine & McCauley, 2016). While conversational, narrative, and expository sample collection is recommended for the assessment of school-aged students, it is unlikely that school-based SLPs have the time to collect and analyze three language samples for every student. More research is needed to determine which of the three sampling contexts is most likely to reveal developmental changes across the older school-aged years. To date, no single study has examined age-related changes for all three contexts within this age group. This study will be the first within-group examination of changes in language complexity across three sampling contexts in school-aged students. School-aged children are often asked to perform complex language tasks in their schoolwork (Scott & Balthazar, 2010), but the presence of age-related changes in these sampling contexts has not been clearly indicated. Furthermore, expanding the normative data available for narrative and expository contexts may increase the clinical utility of the previously established sampling contexts in the literature by establishing what levels of complexity are typical at different ages.

***Research Questions Study Two:***

The gaps in the clinical literature regarding the relationship between sampling context (i.e. conversation, narration, and exposition) and developmental change in school-aged children ages 8 to 11, as measured by utterance length and complex syntax use, motivated the research questions for this study.

1) Does utterance length, as measured by mean length of utterance in words (MLUW), vary by language sampling context? If yes, which language sampling context elicits the largest MLUW within each age group?

2) Does complex syntax production, as measured by the Subordination Index (SI) vary by language sampling context? If yes, which language sampling context elicits the largest SI within each age group?

It is hypothesized that there will be an interaction between age and sampling context on mean utterance length (MLUW) and complex syntax production (SI). Namely, it is hypothesized that as age increases, these measures will increase in expository and narrative contexts, whereas little or no increase will be observed in these metrics in the conversation context, especially among the older children,

3) For children ages 8 to 11, can age-related changes in complex syntax use, as measured by the subordination index, be detected in:

a) a robust conversational context?

b) a narrative retell context?

c) a problem-and-solution expository context?

4) For children ages 8 to 11, can age-related changes in utterance length, as measured by mean length of utterance in words, be detected in:

a) a robust conversational context?

b) a narrative retell context?

c) a problem-and-solution expository context?

It is hypothesized that there will be increases in sentence length and complex syntax use with age for the narrative and exposition contexts, but the conversation context may not be sensitive to age-related changes, especially amongst the oldest age groups.

# Methods

## Participant Recruitment

The recruitment procedures and methods for this study were approved by the Institutional Review Board of the first author's university. A convenience sampling method was used to recruit children through schools, childcare centers, and word of mouth. Recruitment flyers stated that the research team was seeking 8-12-year-old children without known developmental or cognitive disabilities, but students with diagnosed learning disabilities and/or language disorders were eligible for a second portion of the study that is not addressed in this manuscript. Recruited participants without known disabilities included 106 children between the ages of 8 and 12 years of age. Due to limited recruitment of 12-year old children, these participants were excluded from the study. Parent report was used to determine whether participants had a history of special education or related services, including speech-language pathology, occupational therapy, or physical therapy services. Six parents reported that English was not the child's first language. These participants were excluded from the sample. Participants were required to pass a hearing screening and score within normal limits on two language measures (see details below).

## Study One Participants

Study one, which focused on SUGAR and SALT analyses of conversation samples, included 68 children between the ages of 8;0- 10;11. Note that 11-year-olds were excluded from this study because the upper range of SUGAR normative information is 10;11 (Owens & Pavelko, 2020). The sample consisted of 30 males and 38 females. Seventy-six percent (*n* = 52) were white, 4% (*n* = 3) Asian, 10% (*n* = 7) African American, and 5% (*n* = 4) were multiracial. Six percent (*n* = 4) of the participants were of Hispanic or Latinx ethnicity. Ethnicity information was not reported for 1% (*n* = 1) of the participants. Information regarding maternal (or primary caregiver's) education was obtained as a proxy for socioeconomic status. The educational history of these families consisted of the following, 13% (*n* = 9) high school diploma or equivalent, 25% (*n* = 17) some college, 26% (*n* = 18) Bachelor's degree, 33% (*n* = 23) Graduate degree. Information on caregiver education was not reported for 1% (*n* = 1) of the caregivers. Although all participants were native speakers of English, 6% (*n* = 4) also spoke another language. Demographic information for study one participants is displayed in the 8;0 through 10;11 columns in Table 2.

**Study Two Participants**

Study two focused on examination of age-related changes across conversational, narrative, and expository samples. Participants included 85 children between the ages of 8;0- 11;11. Twenty 8;0-8;11 year olds, twenty-six 9;0-9;11 year olds, twenty-two 10;0-10;11 year olds, and seventeen children ages 11;0-11;11 were included. The sample consisted of 37 males and 48 females. Although all participants were native speakers of English, 5% (*n* = 4) of the sample also spoke another language.

Within this sample, 76% (*n* = 65) were white, 3% (*n* = 3) Asian, 11% (*n* = 9) African American, and 7% (*n* = 6) were biracial or multiracial. Five percent (*n* = 4) of the participants were of Hispanic or Latinx ethnicity. Ethnicity information was not reported for 1% (*n* = 1) of the participants. Information regarding maternal (or primary caregiver's) education was obtained as a proxy for socioeconomic status. The educational history of these families consisted of the following, 18% (*n* = 15) high school diploma or equivalent, 26% (*n* = 22) some college, 23% (*n* = 20) Bachelor's degree, and 32% (*n* = 27) Graduate degree. Information on caregiver education was not reported for 1% (*n* = 1) of the caregivers. Participant demographic information can be found in Table 2.

## Procedures and Materials (Studies 1 and 2)

### *Test Sessions*

Participants completed one 40- to 60-minute testing session. The testing session occurred in a quiet, private room in the school, childcare facility, university lab, or personal residence. Each participant completed a hearing screening first. After the completion of the hearing screening, remaining data collection activities were counterbalanced. Half of the participants completed language screening measures first and half of the participants completed the language sampling tasks first so as to avoid order effects. All language screening and sampling tasks were audio-recorded using a digital voice recorder. The first author, a licensed and certified speech-language pathologist, conducted all test sessions.

### *Screening Measures*

1) A pure-tone hearing screening at 25 dB HL at 500, 1000, 2000, and 4000 Hz. Passing performance was required for inclusion in this study. One participant did not pass the hearing screening and was subsequently excluded from both studies.

2) The Structured Photographic Expressive Language Test- Third Edition (SPELT-3; Dawson et al., 2003), a measure of syntax and morphology, was administered to all participants. For individuals 9;11 and younger, passing performance was defined as a

standard score performance of 95 or higher, as this cut score demonstrated adequate sensitivity (90%) and specificity (100%) in previous empirical research (Perona et al., 2005). For individuals 10;0 and older, a raw score of 45 or above (equivalent to a standard score of 97 for 9;11 year olds) was required for inclusion. Six participants did not achieve a passing score on this measure and were subsequently excluded from both studies.

3) The Social Communication (SC) subtest of the Test of Integrated Literacy and Language Skills (TILLS; Nelson et al., 2016) was administered to all participants. This subtest was chosen because there was a statistically significant difference between typically developing children and children with language and literacy disorders on this subtest. Furthermore, it examined different areas of language (vocabulary and pragmatics) than the SPELT-3, in order to get a broad view of language functioning. Passing performance as indicated by a standard score of 7 or higher was required. Six participants did not achieve a passing score on this measure and were subsequently excluded. Four of these participants were excluded from both studies; two were excluded only from study two.

An additional two participants failed both language screening measures and were excluded from both studies.

### *Language Sampling Tasks*

All three language sampling tasks described below were elicited during the testing session, but study one focused only on the analysis of the conversational language samples to examine SUGAR and SALT outcomes. Study two focused only on all three language sampling contexts using analyses obtained from SALT.

1) Completion of a 10-minute conversational language sample. Samples were collected using the SUGAR language protocol. The conversational language sample followed the elicitation techniques described in Pavelko and Owens (2017). For example, care was taken to facilitate an interactive conversation by matching the length of the child's turn, matching the child's interest, and using anticipatory body language and commenting to encourage further elaboration from the child. Additionally, use of process questions and narrative elicitations ("Tell me about a time when…") were frequent, whereas use of questions that can be answered with a single word were limited as much as possible.

2) Completion of two fable narrative retells based on Nippold and colleagues (2017). This study replicated the methods of using a fable retell from Nippold et al., (2014) while using the fables from a later study (Nippold et al., 2017). The fable retell activity was based on two of Aesop's Fables, *The Oak and the Reed* and *The Fox and the Crow.* These stories were selected out of the fables used in the literature because they had the easiest reading levels, according to the Flesch-Kincaid readability statistic (Flesch, 1948) as provided by Microsoft Word (2020). The reading grade level for *The Oak and the Reed* was 3.4, and it was 4.3 for *The Fox and the Crow.* First, the examiner placed a copy of the fable and an image of the characters on an 8" by 11" piece of paper in front of the child and read the story aloud while the child followed along. Following the first reading, the child was asked if he or she wanted to hear the story a second time. The child was then instructed to tell the story back to the examiner as best he or she could with the image remaining and the text covered up. Notably, this activity differed from the fable retellings described in Nippold et al.'s (2017) study because it consisted solely of the

fable retell and did not include the critical thinking questions that accompanied those retellings.

3) Completion of an expository language sample in which the participants were asked to resolve peer conflicts as described in Nippold et al. (2007). The expository sample consisted of the peer conflict resolution (PCR) task. The participant listened as the examiner read two scenarios in which a conflict arises between two individuals. They are entitled *The Science Fair* and *The Fast-Food Restaurant.* Following the first reading, the child was asked if he or she wanted to hear the story a second time. After the presentation of each scenario, the participant was asked to retell the scenario and answer a series of questions posed by the examiner about the problem as how to go about solving it.

4) A brief interview in which the participant was asked to compare and relate to the narrative and expository sampling contexts. Specifically, the child was asked which context was easier and why, which story or scenario was their favorite, and to relate each story to an event in his or her own life. Analysis of these data is not included in this manuscript.

## Language Sample Transcription and Analysis

A trained research assistant transcribed all language samples in Microsoft Word on a personal computer. The research assistants included five undergraduate students, three graduate students, and one licensed speech-language pathologist. Only child utterances were transcribed. The investigator then listened to the transcript, verified the accuracy of the transcription, and made any corrections.

### *Interrater Reliability*

To examine transcription and utterance segmentation reliability, a research assistant and the investigator independently listened to, transcribed, and segmented 10% of the conversation samples ($n$ = 7) using the respective conventions. Word-by-word agreement exceeded 95% and segmentation agreement exceeded 90% in the examined SUGAR samples. Word-by-word agreement and segmentation agreement exceeded 95% in the examined SALT samples.

## Study 1

First, each sample was transcribed and analyzed according to SUGAR conventions. Only the first 50 utterances of the language sample were used for sample analysis, as consistent with SUGAR conventions (Pavelko & Owens, 2017). Next, the SUGAR sample was adapted to fit SALT conventions, and the first 50 C-units were used for analysis. Note that in most cases, fewer utterances were needed to meet the 50 C-unit criteria because of slight differences in utterance segmentation between SUGAR and SALT. SUGAR and SALT conventions, including utterance segmentation, are described below.

### *SUGAR Conventions*

Utterances were deemed terminated if pausing or ceasing to speak occurred for two seconds or more. Additionally, intonation contours suggesting that the utterance was completed were also used to determine utterance boundaries (Pavelko & Owens, 2017). Sentences that were interrupted by the examiner or abandoned by the child also marked the end of an utterance, as defined in Casby (2011). If an utterance contained more than three unintelligible words, it was excluded. Unintelligible words within utterances were designated by the symbol “XX” and remained in the sample if there were two or fewer unintelligible words within an utterance.

Utterances that met the operational definition described above, may have included multiple dependent and independent clauses. A notable exception to this rule is that no more than two independent clauses could be joined by “and.” If an utterance contained three or more clauses separated by “and”, it was separated into multiple utterances with the omission of “and” at the beginning of the next utterance. Following transcription, each sample was analyzed for the four SUGAR metrics: total number of words (TNW), mean length of utterance SUGAR ($MLU_S$), words per sentence (WPS), and clauses per sentence (CPS).

Following SUGAR conventions, the total number of words was calculated by counting the number of words produced by the child within the sample. Mean length of utterance SUGAR was calculated by dividing the total number of morphemes by the number of utterances (i.e., 50). $MLU_S$ counts select derivational and all inflectional morphemes as separate morphemes. Table 3 provides a list of these morphemes. Additionally, contractions were counted as two morphemes as were “wanna”, “hafta”, and “gotta”. The word “gonna” was counted as three morphemes.

After calculation of TNW and $MLU_S$, the sentences in the 50-utterance sample were identified. A sentence is operationally defined as an utterance containing a subject and a verb that can stand alone. Additionally, responses to examiner questions that include at least the verb in the response with ellipsis of the subject were counted as sentences. For example, if the examiner asked, “What are you doing?”, and the child answered, “playing basketball,” this utterance counted as a sentence using SUGAR conventions. Sentences consisting of verb phrases where the subject is omitted because of an imperative construction were also counted. Utterances that are not sentences were deleted for the remaining analyses. Words per sentence (WPS) was calculated by dividing the total number of remaining words by the number of sentences. Next, clauses per sentence (CPS) was calculated by dividing the number of clauses by the number of sentences. Clauses must contain a subject and a verb, but they may include dependent clauses that cannot stand alone. A sample of each calculation can be found in Table 4.

*SALT Conventions*

The fifty-utterance sample described above was then edited to align with SALT conventions and entered into the SALT software program (Miller & Iglesias, 2019). Instead of utterance-level segmentation as described above, the sample was segmented into Communication Units (C-Units). C-Units must contain a subject or a verb, unless they are in response to an examiner question, in which case elliptical responses are included. For example, when asked “How do you think they will both feel?” and the participant answers “happy”, this is retained even though it does not contain a subject and a verb within the utterance. In SUGAR, elliptical responses are included in the transcript, but only elliptical responses containing a verb phrase are included when examining sentences (e.g. when calculating WPS). Whereas SUGAR allows for utterances containing up to two unintelligible words to be included in the sample, SALT conventions require that utterances containing one or more unintelligible words are excluded.

Grammatical morphemes were segmented according to the SALT conventions of using a slash followed by the grammatical morpheme. Although SALT includes all eight of the inflectional morphemes in English in its MLUM calculation, derivational morphemes are not separated in SALT conventions, so that words such as "really" and "faster" are counted as one morpheme. The subordination index (SI) is created by coding the number of clauses within each communication unit. After following these coding conventions, SI, MLUM, NTW (Number of Total Words), and MLUW (mean length of C-unit) were calculated by the software program. A sample of each calculation can be found in Table 5.

*Z-score calculations*

Each SUGAR and SALT metric was compared to its respective age-matched normative sample. For each participant on each metric, z-scores (TNW/NTW, $MLU_S$/MLUM, WPS/MLUW, CPS/SI), were calculated as the deviation from the mean (i.e., in standard deviation units) of the SUGAR and SALT normative samples. SUGAR provides two age ranges within this age group (7;0-8;11 and 9;0-10;11), so SUGAR values were derived by comparing to the respective age group. SALT z-scores were calculated by comparing to samples that were age-matched within 12-months of the participant's age and equated by length to the 50-C-unit samples. Note that the number of samples against which SALT values were derived varied by age, ranging from as few as 28 samples for older children (e.g. children 10;8 and above) to 170 samples for younger children (e.g. eight-year-olds).

**Study 2**

For study two, only SALT conventions were used in order to be consistent with previous studies of narrative and expository sampling analyses and because currently, the SUGAR method has been examined with conversational samples only. For the conversational sample, only the first 50 Communication Units (C-Units) were transcribed. Communication units consist of a main clause and any subordinate clauses that are attached to it. C-Units must contain a subject or a verb unless they are in response to an examiner question, in which case elliptical responses are included. Utterances containing any unintelligible words were excluded. Consistent with previous research, all utterances were transcribed for the narrative and expository contexts.

The investigator entered all transcripts into the SALT software program (Miller & Iglesias, 2019). Mean length of utterance in words (MLUW) was derived from the total number of words divided by the total number of C-units within each sampling context. C-units were coded for the subordination index (SI). Consistent with the subordination index conventions, only finite clauses were counted. For example, the C-unit "I decided to go to the store" consists of only 1 finite clause ("I decided…") and one non-finite clause (infinitive clause "to go"). Using the subordination index (Miller et al., 2019), this C-unit has a subordination index of one. Subordination Index was calculated by the SALT software program. This represents the number of clauses per communication unit.

## Power Analysis

### *Study 1*

An a priori power analysis was conducted using G*Power to determine if the sample size is adequate for the planned statistical analyses (Faul et al., 2007). A power analysis for a two-tailed paired samples t-tests given an effect size of $d = .4$, an alpha level of .05, and power of .8 was conducted. The recommended total sample size is 52. Given the current sample of 68, power should be adequate to detect an effect should there be one.

***Study 2***

An a priori power analysis was conducted using G*Power to determine if the sample size of the data is adequate (Faul et al., 2007). A power analysis was run using a mixed ANOVA approach with two groups, e.g. younger school-aged children and older school-aged children, (as in Owens & Pavelko, 2020) and three measurements (for the three sampling contexts). Given a medium effect size ($d$ =.29, as calculated from Owens & Pavelko, 2020), an alpha level of .05, and power of .8, and a .5 correlation between measurements, the recommended total sample size is 66. Although a different statistical analysis approach was ultimately used for the analysis, the alpha level for the new analysis was also set at .05. As such, the recruited sample of 85 participants is likely adequate for detecting differences if differences exist.

## Results

### Study 1

Each SUGAR and SALT metric was compared to its respective age-matched normative sample. For each participant on each metric, z-scores (TNW/NTW, $MLU_S$/MLUM, WPS/MLUW, CPS/SI), were calculated as the deviation from the mean (i.e., in standard deviation units) of the SUGAR and SALT normative samples. Next, mean z-scores were calculated. Means and standard deviations for each of these z-score values can be found in Table 6.

Differences for SUGAR and SALT mean z-scores were examined using a paired-samples t-test. For example, the mean SUGAR z-score for TNW and the mean SALT z-score for

NTW were compared. There was a significant difference between z-score values for TNW and NTW, $t(67) = -13.52$, $p < .001$, $d = 1.24$, 95% CI [-1.19, -.88]. The z-score for SUGAR TNW values was significantly lower than for SALT's NTW.

For mean length of utterance values ($MLU_S$ and MLUM), there was a significant difference between mean z-score values, $t(67) = -14.33$, $p < .001$, $d = 1.20$, 95% CI [-1.16, -.88]. $MLU_S$ was significantly shorter than MLUM. When examining WPS and MLUW, again, the WPS z-score value was significantly lower than that for WPS, $t(67) = -13.90$, $p < .001$, $d = 1.40$, 95% CI [-1.31, -.98]. Finally, when comparing z-score values for CPS and SI, CPS was significantly higher than SI values, $t(67) = 6.30$, $p <. 001$, $d = .076$, 95% CI [.500882, .965353].

The third research question examined classification differences (suggestive of typical development or language impairment) based on the SUGAR and SALT metrics for each participant. The SUGAR research team found acceptable sensitivity and specificity for diagnostic accuracy using a set of two cut scores (i.e., the combination of -1 SD for $MLU_S$ and -1.25 SD for CPS) for samples from children ages 3-7;11 with and without language impairment (Pavelko & Owens, 2019; Plante & Vance, 1994). The SALT team has used discriminant function analysis with a series of LSA metrics to classify children based on typical versus atypical language status, but cut scores associated with these diagnostic accuracy values have not been reported (Heilmann et al., 2010a). The SALT software automatically flags values more than 1 standard deviation above or below the mean. As such, for the purpose of this study, any SALT z-score value more than 1 standard deviation below the mean was used to denote "suggestive of language impairment". None of the 68 participants were classified as suggestive of language impairment using the recommended SUGAR diagnostic accuracy cut score values of – 1 SD for $MLU_S$ and -1.25 SD for CPS. In contrast, results from the SALT analyses revealed that four participants had subordination index (SI) z-scores below 1 SD, potentially suggestive of language impairment.

**Study 2**

A mixed model approach was adopted for this analysis for three reasons. First, in mixed model approaches, observations are clustered within an individual participant and are therefore not independent. Secondly, sphericity was violated in this dataset. Sphericity is a required assumption for a mixed ANOVA, but it is not required for mixed model approaches. Third, a mixed model approach was adopted because it has been found to reduce Type I error rates and account for participant bias (Judd et al., 2012). Thus, the planned mixed ANOVA approach was abandoned for this analysis. A mixed model approach permits the researcher to account for the non-independence by treating each subject's mean value as a random effect. The experimental variables of interest (e.g. context, age) are treated as fixed effects.

Models were constructed for each of the dependent variables (i.e., SALT MLUW and SI metrics). In each of these models, the participant ID was treated as a random factor, and a random intercept was created to account for the non-independence within each individual's values. Each of the fixed factors (context, age, and the interaction of context and age) were added into subsequent models to see if these variables controlled for significantly more variance above and beyond the participant values. The optimal or best-fit model contained only the random effect for participant and the fixed effect for context. In the models below, age was treated as a categorical variable. Note that two different ages groupings were examined. The first grouping examined age by individual years (e.g. 8;0-8;11, 9;0-9;11, 10;0-10;11, and 11;0-11;11). Given the visual trends in the data and the extant literature documenting the slowing of language development within this age band, (Pavelko & Owens, 2020; Rice et al., 2010), the models were re-run a second time with the age variable collapsed into two categories (8;0-9;11) and (10;0-11;11) instead of four categories.

The full model for examining MLUW contained a random intercept to account for participant differences, fixed effects of context, age, and the interaction between context and age. Age was reported as a categorical variable with each 12-month age band (e.g. 8;0-8;11) as a category. In this model, main effects were found for context, $F(2,162) = 81.507$, $p <. 001$. Age $F(3,81) = 1.226$, $p = .306$, and the interaction term, $F(6,162) = .412$, $p = .870$ were not significant. The interaction term was therefore removed from the model.

In the model containing only the random intercept and context and age as fixed effects, a significant main effect was found for context, $F(2,168) = 84.329$, $p = <.001$, but no main effect was found for age, $F(3,81) = 1.226$, $p = .301$. Pairwise comparisons using this model indicated that there were no significant differences in MLUW between the narrative and expository contexts, $t(168) = -.636$, $p = .525$), but there were differences in MLUW between conversation and both narrative $t(168) = -11.58$, $p < .001$ and expository contexts, $t(168) = -10.94$, $p < .001$,. The conversation context elicited significantly shorter utterances than both the narrative and expository contexts.

In the best fit model, only the random effect for participant and fixed effect for context were included. A significant main effect was found for context $F(2,168) =$

84.329, $p$ = <.001. Pairwise comparisons, using a Bonferroni correction to adjust for multiple comparisons, revealed significant differences in MLUW between conversation and both narrative $t(168) = -11.57$, $p<.001$ and expository contexts $t(168) = -10.94$, $p$ <. 001, but no differences between the narrative and expository contexts, $t(168) = .64$, $p$ = .999. Descriptive information for MLUW by context and age can be found in Table 7. Mean values for MLUW by context and age can be found in Figure 2.

When examining subordination index, the same approach was used. The full model for examining SI contained a random intercept to account for participant differences, fixed effects of context, age, and the interaction between context and age. Age was reported as a categorical variable with each 12-month age band (e.g. 8;0-8;11) as a category. In this model, no main effects were found for context, $F(2,166) = .490$, $p$ = .614, age $F(1,83) = 1.34$, $p = .246$, or the interaction term, $F(2,166) = 1.63$, $p = .199$. When the interaction term was removed from the model, a significant main effect was found for context, $F(2,168) = 64.527$, $p$ = <.001, but no main effect was found for age, $F(1,83) = 1.364$, $p = .246$. Pairwise comparisons, using a Bonferroni correction, found that there were no significant differences between the narrative and expository contexts, $t(168) = 1.17$, $p = .739$, but that conversation elicited less complex syntax than both narrative $t(168) = -10.47$, $p < .001$ and expository $t(168) = -9.30$, $p < .001$ contexts.

In the model for examining SI, where only the random effect for participant and fixed effect for context were included, a significant main effect was found for context, $F(2,168) = 64.527$, $p < .001$. Pairwise comparisons, using a Bonferroni correction, indicated that there were no significant differences between the narrative and expository context, $t(168) = 1.17$, $p = .739$, but that conversation elicited significantly less complex language than the both narrative, $t(168) = -10.48$, $p < .001$) and expository contexts, $t(168) = -9.30$, $p < .001$. Table 8 and Figure 3 display the mean values for SI by age and context.

As mentioned above, due to visual trends in the data, the models were rerun with the age variable collapsed into two categories (8;0-9;11 and 10;0-11;11). For MLUW, the interaction was not significant in the full model, $F(2, 166) = .410$, $p = .664$, and was therefore removed from the model. The model containing the fixed effects for context and age as well as a random effect for participant resulted in a significant main effect for context, $F(2,166) = 84.048$, $p < .001$ and approached significance for age, $F(1,83) = 3.546$, $p = .063$. The younger age group produced shorter utterances than the older group, although as noted, this difference only approached significance. As above, conversation elicited shorter utterances than both the narrative and expository contexts. Figure 4 displays the mean values for MLUW by context using two age bands.

When examining SI, the interaction was not significant, $F(2, 166) = 1.054$, $p = .351$ in the full model and was therefore removed. The model containing the fixed effects for context and age as well as a random effect for participant resulted in a significant main effect for context, $F(2,168) = 64.527$, $p < .001$ and approached significance for age, $F(1,83) = 3.527$, $p = .064$. The younger age group produced less complex utterances than the older group, but this difference only approached significance. As above, conversation elicited less complex utterances than both the narrative and expository contexts. Figure 5 displays the mean values for SI by context using two age bands.

# Discussion

## Study 1

In Study 1, conversation language samples were collected from 68 participants ages 8;0-10;11 and were analyzed using the SUGAR and SALT methods. The results revealed significant differences in mean z-scores between all four paired SUGAR and SALT metrics. Z-scores for the SUGAR metrics (TNW, $MLU_S$, WPS) were significantly lower than the three of the four corresponding paired SALT metrics (NTW, MLUM, MLCU). Z-scores for the SUGAR CPS metric were significantly higher than the corresponding paired SALT SI metric.

Because significant differences were noted, potential classification differences were examined. None of the participants' metrics met the SUGAR recommended cut scores indicative of language impairment (i.e., a combination of z-scores at or below -1.25 SD for CPS and -1 SD for $MLU_S$; Pavelko & Owens, 2019). As noted above, The SALT research team has examined diagnostic accuracy (Heilmann et al., 2010a), but there are not yet established cut scores suggestive of language impairment. Given that the SALT software flags values at ±1 SD of the mean, a score of -1 SD on any of the four metrics was used to classify a participant's performance as suggestive of impairment (Miller & Iglesias, 2019). Four participants had subordination index values suggestive of language impairment. These results suggest that although SUGAR and SALT methods provide similar metrics, the procedures used to transcribe the samples and then to calculate each metric yield significantly different values. In addition, and perhaps, most importantly, differences in identification of impairment may occur when using these two analysis methods and recommended cut-scores.

This the first study to examine the same language sample using both the SALT and SUGAR methods of analysis in a group of school-aged children. The results of the current study have important clinical implications. Although both SUGAR and SALT sampling and analysis methods are recommended within the clinical literature (e.g., Pezold et al., 2020), it is imperative that clinicians are familiar with their chosen method of analysis and follow it with fidelity. Crucially, following the transcription and metric analysis procedures for one method and then using normative data developed for the other method could lead to differences in clinical decision-making. For example, SUGAR and SALT calculate mean length of utterance values using different sets of morphemes, so comparing values gathered using SALT conventions to SUGAR norms (or vice versa) could change the interpretation of the child's performance. To illustrate, MLU values are higher in SUGAR than SALT because SUGAR counts more inflectional and derivational morphemes than SALT does. If the value using SUGAR morphemes were compared to the SALT normative information, a child's performance would be inflated because of the additional morphemes that are included. On the other hand, if the SALT value was compared to the SUGAR normative information, it would deflate the participant's values.

For illustrative purposes, consider the results from a nine-and-a-half-year-old participant who was one of the lowest-performing children in the study. The SUGAR analysis of his conversation sample revealed performance more than two standard deviations below the mean on measures of TNW, $MLU_S$, and WPS and nearly one standard deviation below the mean for CPS. This student did not meet the SUGAR diagnostic accuracy criteria for language impairment because his CPS was not below

1.25 *SD* below the mean (Pavelko and Owens, 2019). The participant's SALT analysis revealed performance within one standard deviation of the mean on all measures except for the subordination index (SI), on which he was over 2 standard deviations below the mean for his age. Notably, this participant did not produce a single complex sentence within the conversation sample. Given the criteria established for this study, this participant's performance could be suggestive of impairment using the SALT method of analysis.

This case example highlights several points of interest. First, diagnostic accuracy cut scores are a powerful clinical tool that can aid in the decision-making process when assessing a child with a suspected language impairment, but even good diagnostic accuracy values do not indicate that every child who has a disorder will be identified by that assessment tool. For example, the SUGAR team (Pavelko & Owens, 2019) report sensitivity values of 97.22% and specificity values of 82.96% using the cut scores described above. Note that values above 90% are considered as excellent, and values above 80% are considered fair (Plante & Vance, 1994). These values are clinically useful, especially for a freely-available criterion-referenced instrument, but this still means that 2.5% of children with language impairment were not identified and approximately 17% of children who do not have language impairment would be flagged by this instrument. While language sample analysis is a valuable clinical tool, this highlights that it should not be the only clinical tool used to identify impairment in a child. When examining the case above, for instance, although the child did not meet the SUGAR diagnostic accuracy criteria, further assessment would be recommended due to his proximity to the cut-offs and low scores across all metrics. Comprehensive language assessment and use of clinical judgment are key components of evidence-based assessment practices.

This study is not without limitations. Namely, only typically developing school-aged children were recruited for this study. Given that children with typical language skills are expected to perform within the average range, classification differences (i.e., typical versus impaired) were not expected. The results support further study of potential classification differences between these two methods. Recruitment of a clinical population, such as school-aged children with developmental language disorder, could highlight key differences in both z-score values and classification using SUGAR and SALT methods.

An additional weakness is that the number of samples against which these z-scores were calculated varied widely by age, especially when using SALT. Although the SALT conversation database contains samples from children ages 2;9-13;3, there are no fourth- or sixth-graders within the sample, which means there were limited samples to which to compare the older students when calculating z-scores. As a result, z-scores calculated for younger children in this study were compared against a sufficient number of samples (e.g. 150 samples) in the SALT database, whereas for the older children, values were derived by comparing against relatively few samples (e.g. 28 samples). A larger sample of age-matched peers, especially for the older children in the sample, are needed.

The z-score values and subsequent comparisons were calculated using a 50-utterance sample, as this is the only length for which normative information is available in this age group using the SUGAR method. The SALT conversation database contains longer conversation samples (e.g. approximately 100 utterances), but the database enables calculation of z-scores based on sample length. Although reliability between shorter and longer samples appears to be fairly consistent across differing sample lengths (Heilmann et al., 2010b; Pavelko et al., 2020), further study is needed to examine potential classification difference (i.e., typical or suggestive of language impairment) in samples of different lengths. A future direction may be to examine differences in z-scores when the sample length corresponds with the length used within the normative database for each analysis method.

More work is needed to determine diagnostically accurate cut scores for conversation samples from this age group for both SUGAR and SALT. Although z-score values were based on age-matched samples for SUGAR, the cut scores associated with diagnostic accuracy for SUGAR were developed using samples from younger children (ages 3;0-7;11). SALT researchers (Heilmann et al., 2010a) have examined diagnostic accuracy within this age range (ages 6;0-9;11 and ages 10;0-13;6), but cut scores for these metrics were not reported for clinical use. More research is needed to provide clinicians with specific cut scores for school-aged children that are suggestive of language impairment for both SUGAR and SALT.

The implications of this study are limited to SUGAR and SALT analyses, two of the most popular methods of language sampling analyses. These are not, however, the only computer-assisted methods of completing a language sample analysis. One other common, and freely available language sample analysis tool is Computerized Language Analysis (CLAN) (MacWhinney, 2000). CLAN is a computerized tool that works with the Child Language Data Exchange System (CHILDES) to allow clinicians to analyze a language sample compared to this publicly-funded corpora of data. Recent research has demonstrated that CLAN could identify individuals with language impairment from a group of African American English speakers (Overton et al., 2021). Given that the SUGAR and SALT normative databases consist predominantly of white children, future research investigating all three of these language sample analysis methods, especially in the face of an increasingly pluralistic society, is warranted.

**Study 2**

In study 2, conversation, narrative, and expository language samples were collected from 85 participants ages 8;0-11;11. These samples were transcribed and analyzed for differences in utterance length (MLUW) and syntactic complexity (SI) using the SALT conventions. Age-related changes in measures of complex syntax (SI) and utterance length (MLUW) were not found in any context, nor was there an interaction between age and context on utterance length or syntactic complexity. Importantly, context did show significant differences. Conversation elicited significantly shorter and less complex utterances than both narrative and expository contexts across all age groups. There were no differences in utterance length or syntactic complexity when comparing narrative and expository contexts.

This is the first study to simultaneously examine conversation, narrative, and expository language sampling contexts within this age group of children. The lack of age-related changes in utterance length and syntactic complexity was an unanticipated result. While it has been found that language development slows during this period (Rice et al., 2010), some studies had documented age-related changes in language performance measures within a similar age range (Owens & Pavelko, 2020, Scott & Windsor, 2000).

The results indicate that the wide range of individual variability in language performance measures may obscure any quantitative growth in these measures that occur within this age range for these tasks.

This study demonstrated the feasibility of using specific language sampling tasks with school-age children that were previously only used for older children and adolescents. For example, fable retells have been utilized in previous research as a narrative elicitation method with adolescents (Nippold et al., 2014, 2017); feasibility with older school-aged children had not yet been examined. Although the language elicited in this study was less complex than in Nippold and colleagues' previous work with older participants (2014, 2017), most participants retold the fables completely, indicating that this elicitation task is feasible for this age group. The elicitation of these two fables lasted approximately 6-8 minutes, making it a fairly time-efficient elicitation method.

The expository task examined in this study, the peer conflict resolution task, has been utilized in previous research with older children. Namely, this task has been used with older children, adolescents, and adults (Nippold et al., 2007), but it had never been used with children younger than 11 years old. Similarly to the fable retell tasks, the expository task elicited comparable utterance length and syntactic complexity values to both the previous research (Nippold et al., 2007). This elicitation method was also relatively brief, requiring approximately 6 to 8 minutes for administration and sample collection. These findings suggest that this task is a time-efficient method for eliciting an expository language sample from children within the 8- to 11-year old age range.

Conversation, while the most common language sampling procedure used by school-based SLPs (Pavelko et al., 2016), elicits less syntactically complex language than narrative and expository language sampling contexts (Nippold et al., 2008; Nippold et al., 2014). For example, analysis of conversation samples reveals no significant differences in language complexity measures between adolescents with and without language impairment (Nippold et al., 2008). The findings in this study echo much of that research as it relates to older school-aged children. Namely, the conversational context elicited the shortest and least syntactically complex communication units in older school-aged children.

These findings underscore the importance of utilizing narrative and/or expository contexts, rather than conversational contexts, for analysis of syntax in older school-aged children. Both narrative and expository tasks are common within the educational curriculum in this age group, making both authentic assessment tasks (Scott & Balthazar, 2010; Ward-Lonergan, 2010). Narrative and expository samples did not differ in utterance length or syntactic complexity, but elicited lengthier and more complex utterances than conversation. Alternatively, conversation samples, while eliciting shorter and less complex utterances, may still have clinical utility in this age range, if the primary concern is to document pragmatic skills rather than syntactic skills.

This study is not without limitations. Namely, only typically developing school-aged children were recruited for this study. Although it has been established that conversation may not be adequately sensitive to pick up differences between adolescents with and without language impairment, this question could not be investigated due to the nature of the participant sample. Because narrative language is culturally mediated (Goldstein, 2000 as cited in Paul et al., 2018), and SLPs practice within an increasingly pluralistic society, investigating these same questions with a more diverse group of children would be beneficial. Due to the nature of using a convenience sample for recruitment, the variability of socioeconomic status (SES) within this sample may also be a limitation. In young children, there are links between language performance measures and socioeconomic status (Hart & Risley, 1995, Rowe et al., 2012) In this sample, many of the younger children came from families with higher socioeconomic status than the families of the older children. Little is known about the long-term impacts of low SES in school-aged children. One study was found that looked specifically at SES in school-aged children, but the sample consisted of English Learners (ELs). Alt and colleagues (2016) found that SES did significantly correlate with language performance measures in 907 ELs in kindergarten and second grade. Although there is limited research about the impact of SES on utterance length and syntax within monolingual speakers of this age, future research in which this variable is controlled may limit its potential influences.

The results of this study indicate that within this age group, narrative and expository tasks elicit equivalent language in terms of utterance length and syntactic complexity. In the future, investigating these contexts within a sample containing both typically developing children and children with developmental language disorder may provide insight into whether one of these contexts is more effective at differentiating between the two groups. Although utterance length and syntactic complexity did not show age-related changes, another avenue of research may be the presence or absence of age-related changes in other language performance measures found to increase in adolescence, such as lexical diversity and increased use of metacognitive verbs.

Previous research (Nippold et al., 2014, 2017) has investigated whether or not adolescents could relate to the morals of fables and relate them to their life circumstances. Another future direction may include investigating the relationship between familiarity with a task's content and language production variables, such as language productivity and complexity. Nippold (2009) found that there were no differences in total T-units, mean length of T-unit (MLTU), or subordinate clause use when comparing the language of experienced versus novice chess players, but this was within the same tasks. Investigating whether these language performance measures relate to familiarity across language sampling contexts (e.g. fables narrative tasks versus peer conflict resolution expository task) may provide insight into the role of background knowledge in language performance within this age group.

Finally, the curricular expectations within this age group shift to include an increased focus on written language (Common Core State Standards, 2021), but normative information for written language sampling narrative and expository tasks are scarce. Further exploration may include the viability of these tasks as both oral and written language sampling tasks, for the purposes of building a normative database for practitioners to use in their evaluations.

## Conclusion

Language sampling is commonly recommended as an authentic and less-biased assessment tool when evaluating children's language (Costanza-Smith, 2010; Gutiérrez-Clellen et al., 2000, Nippold 2014). When conducting language sampling, both the method of language sample analysis and the context of the elicited language sample are important considerations. This study investigated two common methods of language sample analysis as well as the recommended language sampling contexts for school-aged children.  The presence of significant differences in both language group classification and differing values when comparing two LSA methods to respective normative data underscore the importance of conducting language sample analysis consistent with the conventions to which the results are being compared. Electing to arbitrarily select pieces and parts of differing methods may result in differing clinical interpretations.

Results indicated that there were no significant age-related language growth changes in the conversation, expository, and narrative contexts in the school-aged years. However, both narrative and expository contexts elicited longer and more complex utterances than the conversation context, which school-based clinicians most commonly use for language sample analysis (Pavelko et al., 2016). Based on the results of this study, clinicians should opt to use either narrative or expository contexts when performing language sample analysis with school-aged children. Language sample analysis should continue to be used within this age band because even though language performance growth was not evident within this age band, these contexts elicit language that more closely aligns with the academic expectations required to be successful learners and communicators, making them authentic evaluation tasks. One limitation for analysis of narrative and expository samples is the lack of a large normative database. To support increased use of these contexts, clinical researchers need to establish normative guidelines so that clinicians are more likely to collect narrative and expository samples from school-aged children.

## Tables

Table 1

*Major differences between Systematic Analysis of Language Transcripts (SALT) and Sampling Utterances Grammatical Analysis Revisited (SUGAR) transcription conventions*

| **Feature** | **SALT** | **SUGAR** |
|---|---|---|
| Utterance Segmentation | C-Unit (Communication unit segmentation; independent clauses are separated)<br>Example: I went to the store<br>And I bought some bananas. (2 C-units) | Segmentation by utterance (may contain 2 independent clauses joined by "and")<br>Example: I went to the store and I bought some bananas (1 utterance) |
| Morpheme Count | Grammatical morphology only<br>Example: She run/3s faster now. | Grammatical and select derivational and inflectional morphology<br>Example: She run s fast er now. |
| Mazes | Transcribed and placed in parentheses<br>Example: (I really um) I really like that. | Omitted<br>Example: I really like that. |
| Unintelligible Words | Utterances containing unintelligible words are not included in comparisons to normative data | Utterances containing up to 2 unintelligible words remain |

Table 2

*Participant demographic information, disaggregated by age category.*

| Demographic variable | 8;0-8;11 | 9;0-9;11 | 10;0-10;11 | 11;0-11;11 |
|---|---|---|---|---|
| **Race** | | | | |
| Asian | 1 (5%) | 1 (4%) | 1 (5%) | 0 (0%) |
| Black/African American | 1 (5%) | 3 (11%) | 3 (14%) | 2 (12%) |
| Multiracial | 1 (5%) | 1 (4%) | 2 (9%) | 2 (12%) |
| White | 17 (85%) | 20 (77%) | 15 (68%) | 13 (76%) |
| Missing/Not reported | 0 (0%) | 1 (4%) | 1 (5%) | 0 (0%) |
| **Ethnicity** | | | | |
| Hispanic/Latinx | 0 (0%) | 4 (15%) | 0 (0%) | 0 (0%) |
| Missing/Not reported | 0 (0%) | 1 (4%) | 0 (0%) | 0 (0%) |
| **Maternal/Primary Caregiver's Education level** | | | | |
| Missing/Not reported | 1 (5%) | 0 (0%) | 0 (0%) | 0 (0%) |
| High school diploma /equivalent | 2 (10%) | 4 (15%) | 3 (14%) | 6 (35%) |
| Some college | 2 (10%) | 8 (31%) | 7 (32%) | 5 (29%) |
| Bachelor's degree | 5 (25%) | 8 (31%) | 5 (23%) | 2 (12%) |
| Graduate degree | 10 (50%) | 6 (23%) | 7 (32%) | 4 (24%) |

*Note:* Study 1 participants include 8;0-8;11, 9;0-9;11, and 10;0-10;11 columns. Study 2 participants include all participants listed.

Table 3

*Bound morphemes included in SUGAR LSA* **(**Table from Pavelko & Owens, 2017, with permission)

| Morpheme | Example |
|---|---|
| *–ing* | |
| adjective | smiling girl |
| gerund | I love hiking |
| progressive verb | He is jumping |
| *–s* | |
| plural | cats |
| possessive | Mommy's shirt |
| third-person singular | He walks |
| *–ed* | |
| adjective | Is this powdered sugar |
| regular past tense | He scared me |
| dis– | dislike |
| *–er* | |
| comparative | bigger<br>faster<br>older |
| *–est* | |
| superlative | biggest<br>easiest |
| *–ful* | thoughtful<br>spoonful |
| *–ish* | foolish<br>bluish |
| *–ly* | really<br>mostly<br>usually<br>accidentally |
| *–ment* | entertainment |
| re– | redo<br>refill |
| –sion | discussion<br>mission |
| –tion | constellation<br>invitation<br>audition |
| un– | unhappy<br>unlock |
| –y | |
| adjective | bumpy<br>bouncy |

*Note.* The table presented here appears courtesy of the authors. Copyright © 2017 The authors.

Table 4

*Sample SUGAR transcription conventions metric calculations*

| Metric | Example | Calculation |
|---|---|---|
| TNW | I snuck up on my parents and then I just sat down and started crying<br>It was funny<br>And they were like "How did you get out of your crib?"<br>Stuff like that<br>It was really funny | Number of words<br><br>TNW: 37 |
| $MLU_S$ | I snuck up on my parent s and then I just sat down and start ed cry ing<br>It was funn y<br>And they were like "How did you get out of your crib?"<br>Stuff like that<br>It was real ly funn y | Number of morphemes/number of utterances<br><br>$MLU_S$: 43/5 = 8.6 |
| WPS | I snuck up on my parents and then I just sat down and started crying<br>It was funny<br>And they were like "How did you get out of your crib?"<br>It was really funny | Number of words/number of sentences<br><br>WPS: 34/4 = 8.5 |
| CPS | I snuck up on my parents<br>and then I just sat down and started crying<br>It was funny<br>And they were like<br>"How did you get out of your crib?"<br>It was really funny | Number of clauses/number of sentences<br><br>CPS: 6/4 = 1.5 |

Table 5

*Sample SALT transcription conventions metric calculations*

| Metric | Example | Calculation |
|---|---|---|
| NTW | I snuck up on my parents.<br>and then I just sat down and started crying.<br>It was funny.<br>And they were like “How did you get out of your crib”?<br>It was really funny. | Number of words<br><br>TNW: 34 |
| MLUM | I snuck up on my parent/s.<br>and then I just sat down and start/ed cry/ing.<br>It was funny.<br>And they were like “How did you get out of your crib”?<br>It was really funny. | Number of morphemes/number of utterances<br><br>MLUM: 37/5 = 7.4 |
| MLUW | I snuck up on my parents.<br>and then I just sat down and started crying.<br>It was funny.<br>And they were like “How did you get out of your crib”?<br>It was really funny. | Number of words/number of C-units<br><br>MLCU: 34/5 = 6.8 |
| SI | I snuck up on my parents [SI-1].<br>and then I just sat down and started crying [SI-1].<br>It was funny [SI-1].<br>And they were like “How did you get out of your crib” [SI-2]?<br>It was really funny [SI-1]. | Number of clauses/number of C-units<br><br>SI: 6/5 = 1.2 |

Table 6

*Descriptive Statistics for Language Sampling Measures (n = 68)*

| Measure | | Raw Score *M* | Raw Score *SD* | Z-score *M* | Z-score *SD* |
|---|---|---|---|---|---|
| Pair 1 | TNW | 416.71 | 55.63 | .13 | .86 |
| | NTW | 377.49 | 125.62 | 1.16 | .80 |
| Pair 2 | $MLU_S$ | 9.58 | 1.25 | .20 | .85 |
| | MLUM | 8.44 | 1.28 | 1.22 | .84 |
| Pair 3 | WPS | 9.43 | 1.05 | -.01 | .80 |
| | MLCU | 7.55 | 1.15 | 1.14 | .84 |
| Pair 4 | CPS | 1.54 | .13 | 1.19 | .89 |
| | SI | 1.28 | .13 | .46 | 1.04 |

Table 7

*Mean Length of Utterance in Words (MLUW) by Age and Context*

| Age (years) | Context | *n* | *M* | *SD* |
|---|---|---|---|---|
| 8;0-8;11 | CON | 20 | 7.42 | 1.22 |
| | EXPO | 20 | 9.29 | 1.67 |
| | NAR | 20 | 9.13 | 1.12 |
| 9;0-9;11 | CON | 26 | 7.46 | 1.01 |
| | EXPO | 26 | 9.17 | 1.16 |
| | NAR | 26 | 9.59 | 2.16 |
| 10;0-10;11 | CON | 22 | 7.78 | 1.26 |
| | EXPO | 22 | 9.82 | 1.32 |
| | NAR | 22 | 9.84 | 1.38 |
| 11;0-11;11 | CON | 17 | 7.60 | 1.45 |
| | EXPO | 17 | 9.77 | 1.53 |
| | NAR | 17 | 9.86 | 1.39 |
| Total | CON | 85 | 7.56 | 1.21 |
| | EXPO | 85 | 9.49 | 1.41 |
| | NAR | 85 | 9.60 | 1.61 |

*Note:* CON = conversational language samples, EXPO = peer conflict resolution task, and NAR= fable retellings.

Table 8

*Subordination Index (SI) by Age and Context*

| Age (years) | Context | n | *M* | *SD* |
|---|---|---|---|---|
| 8;0-8;11 | CON | 20 | 1.27 | .12 |
| | EXPO | 20 | 1.49 | .18 |
| | NAR | 20 | 1.45 | .15 |
| 9;0-9;11 | CON | 26 | 1.28 | .14 |
| | EXPO | 26 | 1.45 | .17 |
| | NAR | 26 | 1.51 | .26 |
| 10;0-10;11 | CON | 22 | 1.30 | .12 |
| | EXPO | 22 | 1.52 | .18 |
| | NAR | 22 | 1.57 | .17 |
| 11;0-11;11 | CON | 17 | 1.27 | .17 |
| | EXPO | 17 | 1.53 | .21 |
| | NAR | 17 | 1.55 | .19 |
| Total | CON | 85 | 1.28 | .13 |
| | EXPO | 85 | 1.49 | .18 |
| | NAR | 85 | 1.52 | .20 |

*Note:* CON = conversational language samples, EXPO = peer conflict resolution task, and NAR= fable retellings.

## Figures

Figure 1.

*Recommended language sampling contexts by age* (figure from Pezold et al., 2020, used with permission)

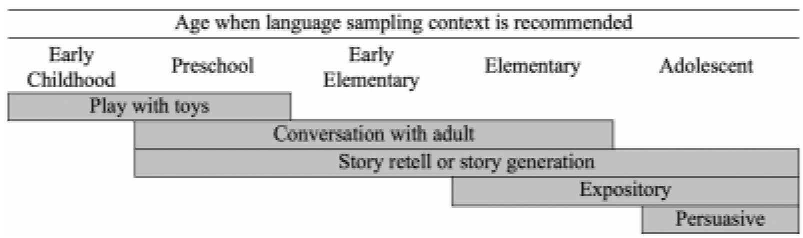
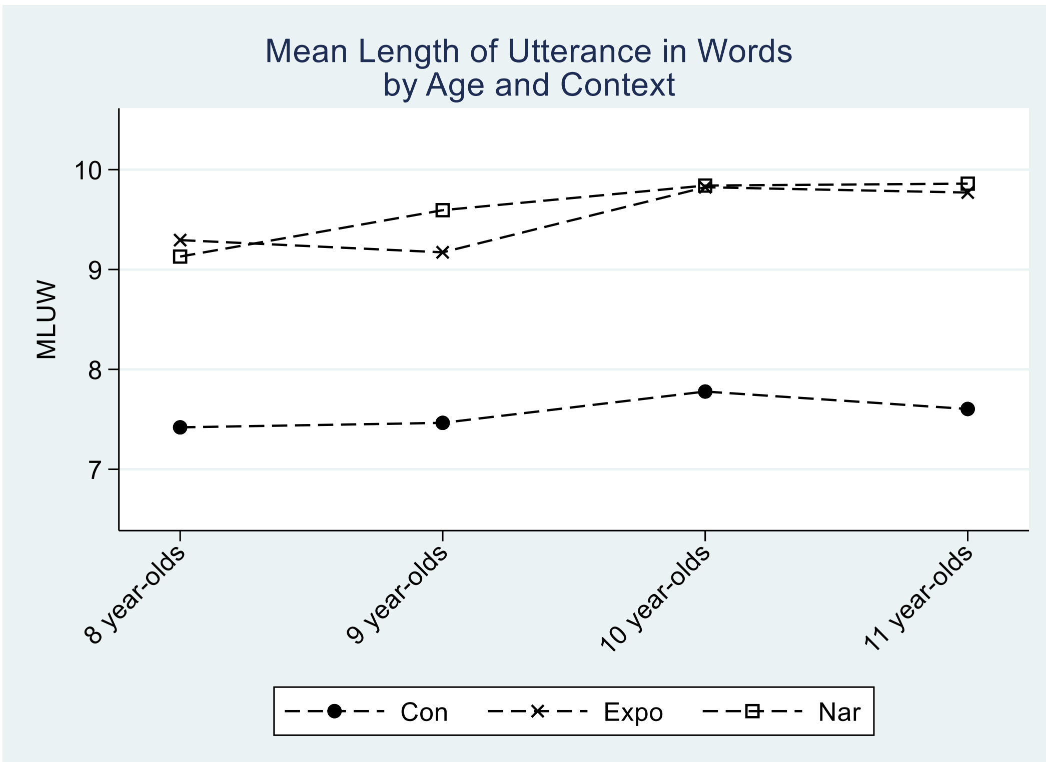


Figure 2.

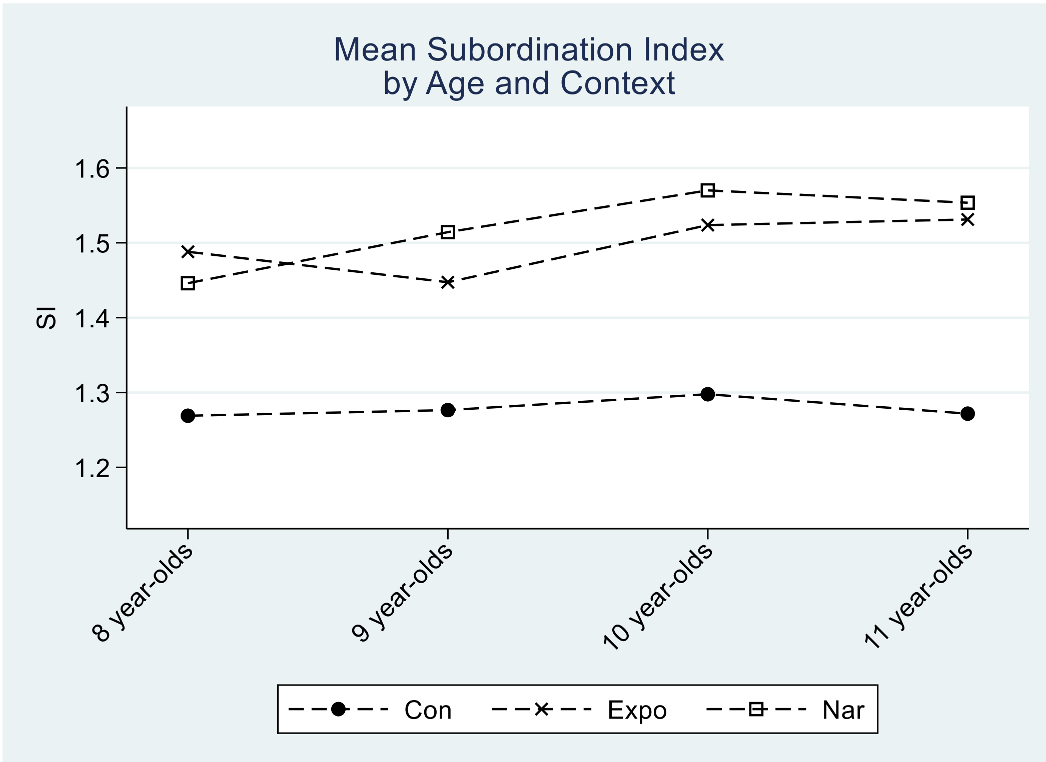
*Mean length of utterance in words (MLUW) by age and context using single-year age bands*




*Note:* CON = conversational language samples, EXPO = peer conflict resolution task, and NAR= fable retellings.

Figure 3.

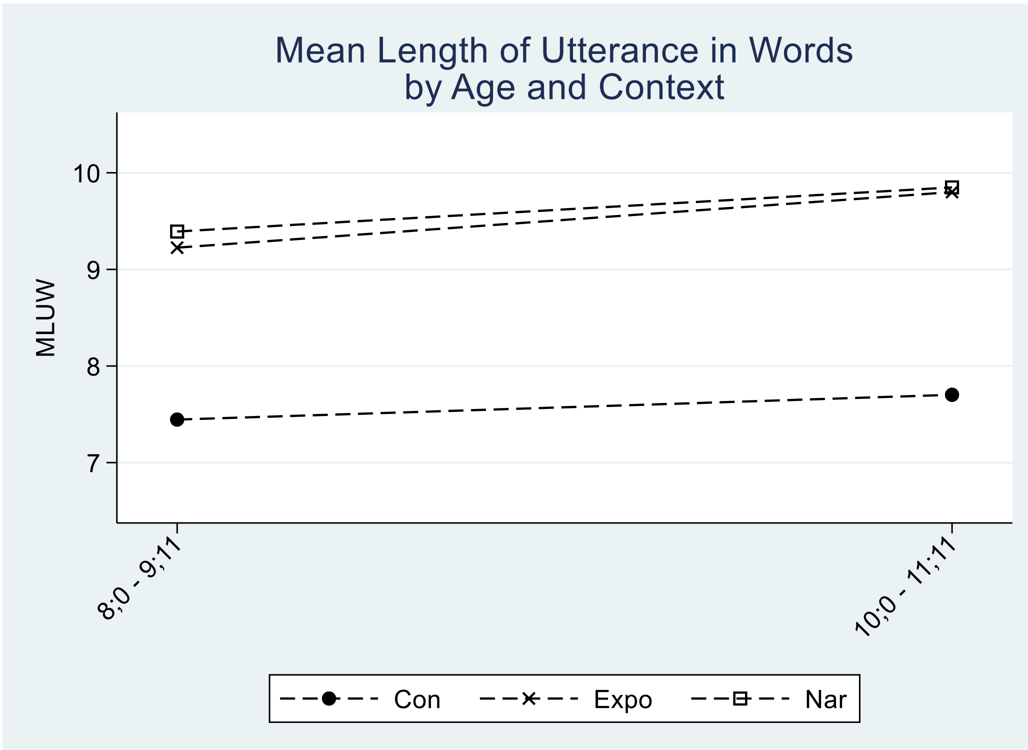
*Mean subordination index (SI) values by age and context using single-year age bands*




*Note:* CON = conversational language samples, EXPO = peer conflict resolution task, and NAR= fable retellings.

Figure 4.

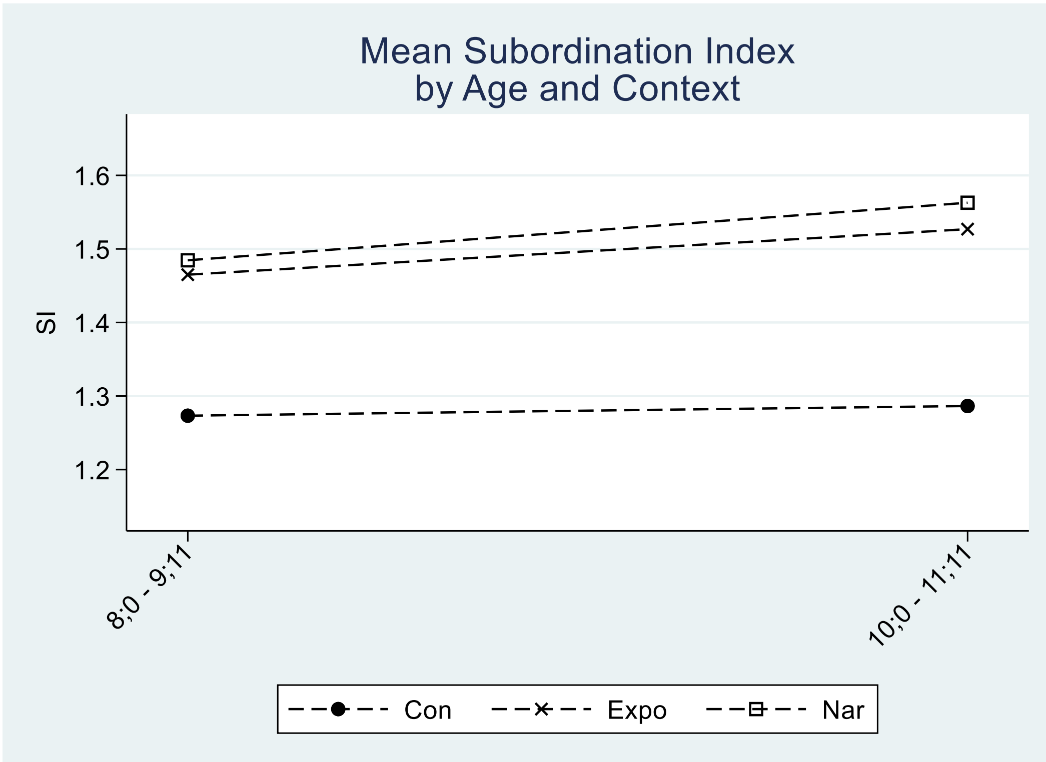
*Mean length of utterance in words (MLUW) by age and context using two-year age bands*




*Note:* CON = conversational language samples, EXPO = peer conflict resolution task, and NAR= fable retellings.

Figure 5.

*Mean subordination index (SI) values by age and context using two-year age bands*




*Note:* CON = conversational language samples, EXPO = peer conflict resolution task, and NAR= fable retellings.

**References**

Alt, M., Arizmendi, G. D., & DiLallo, J. N. (2016). The role of socioeconomic status in the narrative story retells of school-aged English language learners. *Language, Speech, and Hearing Services in Schools*, *47*(4), 313-323. https://doi.org/10.1044/2016_LSHSS-15-0036

American Speech-Language-Hearing Association. (2016). *Scope of practice in speech-language pathology* [Scope of Practice]. Available from www.asha.org/policy/.

Brown, R. (1973). A first language: The early stages. Cambridge, MA: Harvard University Press.

Casby, M. W. (2011). An examination of the relationship of sample size and mean length of utterance of children with developmental language impairment. *Child Language Teaching and Therapy*, 27, 286–293. https://doi.org/10.1177/0265659010394387

Channell, M. M., Loveall, S. J., Conners, F. A., Harvey, D. J., & Abbeduto, L. (2018). Narrative language sampling in typical development: Implications for clinical trials. *American Journal of Speech-Language Pathology*, *27*(1), 123-135. https://doi.org/10.1044/2017_AJSLP-17-0046

Costanza-Smith, A. (2010). The clinical utility of language samples. *Perspectives on Language Learning and Education*, *17*(1), 9-15. https://doi.org/10.1044/lle17.1.9

Common Core State Standards Initiative (2021). *English language arts standards.* Retrieved from: http://www.corestandards.org/ELA-Literacy/

Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *Structured Photographic Expressive Language Test: Third Edition.* DeKalb, IL: Janelle Publications.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233. https://doi.org/10.1037/h0057532

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools*, *31*(1), 88-98. https://doi.org/10.1044/0161-1461.3101.88

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools,* 41(1), 84–95. https://doi.org/10.1044/0161-1461 (2009/08-0075)

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the transcript matter?. *Language, Speech, and Hearing Services in Schools, 41*(4), 393-404. https://doi.org/10.1044/0161-1461(2009/09-0023)

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54-69. https://doi.org/10.1037/a0028347

Long, S. H. (2001). About time: A comparison of computerized and manual procedures for grammatical and phonological analysis. *Clinical Linguistics & Phonetics*, *15*(5), 399-426. https://doi.org/10.1080/02699200010027778

Lundine, J. P., & McCauley, R. J. (2016). A tutorial on expository discourse: Structure, development, and disorders in children and adolescents. *American Journal of Speech-Language Pathology*, *25*(3), 306-320. https://doi.org/10.1044/2016_AJSLP-14-0130

Lundine, J. P., Harnish, S. M., McCauley, R. J., Blackett, D. S., Zezinka, A., Chen, W., & Fox, R. A. (2018). Adolescent summaries of narrative and expository discourse: Differences and predictors. *Language, Speech, and Hearing Services in Schools*, *49*(3), 551-568. https://doi.org/10.1044/2018_LSHSS-17-0105

Miller, J., Andriacchi, K., & Nockerts, A. (2019). *Assessing language production using SALT software: A clinician's guide to language sample analysis (3rd edition)*. Madison, WI: SALT Software.

Miller, J., Andriacchi, K., & Nockerts, A. (2013). SLPs turn to language sample analysis software to document progress on the Common Core State Standards. *Advance for Speech-Language Pathologists & Audiologists.* Online publication, July 1, 2013.

Miller, J., & Iglesias, A. (2019). *Systematic analysis of language transcripts* (SALT), Research Version 20 [Computer Software]. Middleton, WI: SALT Software LLC.

Nelson, N. W., Plante, E., Helm-Estabrooks, N., & Hotz, G. (2016). *Test of integrated language and literacy skills (TILLS).* Baltimore, MD: Brookes.

Nippold, M. A. (2014). *Language sampling with adolescents: Implications for intervention* (2nd ed.). San Diego, CA: Plural.

Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C., & MacKinnon, M. (2014). Conversational and narrative speaking in adolescents: Examining the use of complex syntax. *Journal of Speech, Language, and Hearing Research, 57*(3), 876-886. https://doi.org/10.1044/1092-4388(2013/13-0097)

Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C., & MacKinnon, M. (2015). Critical thinking about fables: Examining language production and comprehension in adolescents. *Journal of Speech, Language, and Hearing Research*, *58*(2), 325-335. https://doi.org/10.1044/2015_JSLHR-L-14-0129

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse. *Journal of Speech, Language, and Hearing Research*, 48(5), 1048-1064. https://doi.org/10.1044/1092-4388(2005/073)

Nippold, M. A., Mansfield, T. C., & Billow, J. L. (2007). Peer conflict explanations in children, adolescents, and adults: Examining the development of complex syntax. *American Journal of Speech-Language Pathology*, *16*(2), 179-188. https://doi.org/10.1044/1058-0360(2007/022)

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. *American Journal of Speech-Language Pathology*, *17*(5), 356-366. https://doi.org/10.1044/1058-0360(2008/07-0049)

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic development in adolescents with a history of language impairments: A follow-up investigation. *American Journal of Speech-Language Pathology*, *18*(3). 241-251. https://doi.org/10.1044/1058-0360(2008/08-0022)

Nippold, M. A., Vigeland, L. M., Frantz-Kaspar, M. W., & Ward-Lonergan, J. M. (2017). Language sampling with adolescents: Building a normative database with fables. *American Journal of Speech-Language Pathology*, *26*(3), 908-920. https://doi.org/10.1044/2017_AJSLP-16-0181

Owens, R. E. & Pavelko, S. L. (2020). Sampling utterances and grammatical analysis revisited (SUGAR): Quantitative values for language sample analysis measures in 7- to 11-year-old children. *Language, Speech, and Hearing Services in Schools, 51*(3), 734-744. https://doi.org/10.1044/2020_LSHSS-19-00027

Owens, R. E., Pavelko, S. L., & Bambinelli, D. (2018). Moving beyond mean length of utterance: Analyzing language samples to identify intervention targets. *Perspectives of the ASHA Special Interest Groups*, *3*(1), 5-22.

Overton, C., Baron, T., Pearson, B. Z., & Ratner, N. B. (2021). Using free computer-assisted language sample analysis to evaluate and set treatment goals for children who speak African American English. *Language, Speech, and Hearing Services in Schools*, *52*(1), 31-50. https://doi.org/10.1044/2020_LSHSS-19-00107

Paul, R., & Norbury, C. F., & Gosse, C. (2018). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (5th ed.). St. Louis, MO: Elsevier.

Pavelko, S. L., Price, L. R., & Owens Jr, R. E. (2020). Revisiting Reliability: Using Sampling Utterances and Grammatical Analysis Revised (SUGAR) to Compare 25-and 50-Utterance Language Samples. *Language, Speech, and hearing services in schools*, *51*(3), 778-794. https://doi.org/10.1044/2020_LSHSS-19-00026

Pavelko, S. L., & Owens Jr, R. E. (2019). Diagnostic accuracy of the sampling utterances and grammatical analysis revised (SUGAR) measures for identifying children with language impairment. *Language, Speech, and Hearing Services in Schools*, *50*(2), 211-223. https://doi.org/10.1044/2018_LSHSS-18-0050

Pavelko, S. L., & Owens Jr, R. E. (2017). Sampling utterances and grammatical analysis revised (SUGAR): New normative values for language sample analysis measures. *Language, Speech, and Hearing Services in Schools*, *48*(3), 197-215. https://doi.org/10.1044/2017_LSHSS-17-0022

Pavelko, S. L., Owens Jr, R. E., Ireland, M., & Hahs-Vaughn, D. L. (2016). Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, Speech, and Hearing Services in Schools*, *47*(3), 246-258. https://doi.org/10.1044/2016_LSHSS-15-0044

Perona, K., Plante, E., & Vance, R. (2005). Diagnostic accuracy of the structured photographic expressive language test. *Language, Speech, and Hearing Services in Schools, 36*(2), 103-115. https://doi.org/10.1044/0161-1461(2005/010)

Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language, Speech, and Hearing Services in Schools, 51*(1), 103-114. https://doi.org/10.1044/2019_LSHSS-18-0148

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*(1), 15-24. https://doi.org/10.1044/0161-1461.2501.15

Retherford, K. (2007). *Guide to analysis of language transcripts* (3rd ed.). Austin, TX: PRO-ED.

Rice, M.L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53*(2), 333-349. https://doi.org/10.1044/1092-4388(2009/08-0183)

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, *83*(2), 508–525. https://doi.org/10.1111/j.1467-8624.2011.01710.x

SALT Software (2020). *SALT software: Computerized language sample analysis* [Product Brochure]. SALT Software, LLC. https://www.saltsoftware.com/media/wysiwyg/ordering/SALT_20_Brochure.pdf

Scott, C. M., & Balthazar, C. H. (2010). The grammar of information: Challenges for older students with language impairments. *Topics in Language Disorders*, *30*(4), 288. https://doi.org/ 10.1097/TLD.0b013e3181f90878

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, *43*(2), 324-339. https://doi.org/10.1044/jslhr.4302.324

Southwood, F., & Russell, A. F. (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research*, *47*(2), 366-376. https://doi.org/10.1044/1092-4388(2004/030)

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245-1260. https://doi.org/10.1044/jslhr.4006.1245

Ward-Lonergan, J. M. (2010). Expository discourse intervention: Helping school-age children and adolescents with language disorders master the language of the curriculum. In M. A. Nippold & C. M. Scott (Eds.), *New directions in communication disorders research: Integrative approaches. Expository discourse in children, adolescents, and adults: Development and disorders* (p. 241–273). Psychology Press.

Westerveld, M. F., & Moran, C. A. (2011). Expository language skills of young school-age children. *Language, Speech, and Hearing Services in Schools*, *42*(2), 182-193. https://doi.org/10.1044/0161-1461(2010/10-0044)