A COMPARISON OF LANGUAGE SAMPLE ELICITATION METHODS FOR DUAL LANGUAGE LEARNERS

A Thesis
Submitted to
The Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
MASTER OF ARTS

---

by
Jacqueline Toscano
August 2017

Thesis Approvals:

Jamie Reilly, Advisory Chair, Communication Sciences and Disorders
Jodi Reich, Communication Sciences and Disorders
Felicidad García, Communication Sciences and Disorders

ABSTRACT

Language sample analysis has come to be considered the "gold standard" approach for cross-cultural language assessment. Speech-language pathologists assessing individuals of multicultural or multilinguistic backgrounds have been recommended to utilize this approach in these evaluations (e.g., Pearson, Jackson, & Wu, 2014; Heilmann & Westerveld, 2013). Language samples can be elicited with a variety of different tasks, and selection of a specific method by SLPs is often a major part of the assessment process. The present study aims to facilitate the selection of sample elicitation methods by identifying the method that elicits a maximal performance of language abilities and variation in children's oral language samples. Analyses were performed on Play, Tell, and Retell methods across 178 total samples and it was found that Retell elicited higher measures of syntactic complexity (i.e., TTR, SI, MLUw) than Play as well as a higher TTR (i.e., lexical diversity) and SI (i.e., clausal density) than Tell; however, no difference was found between Tell and Retell for MLUw (i.e., syntactic complexity/productivity), nor was there a difference found between Tell and Play for TTR. Additionally, it was found that the two narrative methods elicited higher DDM (i.e., frequency of dialectal variation) than the Play method. No significant difference was found between Tell and Retell for DDM. Implications for the continued use of language sample for assessment of speech and language are discussed.

ACKNOWLEDGMENTS

Firstly, I want to thank Dr. Jodi Reich and Professor Felicidad García for their endless support and encouragement, as well as the incredible amount of time and energy without which the completion of this thesis would not be possible. This thesis was also made possible by Dr. Carol Hammer whose longitudinal project not only sparked the selection of this topic of research but supplied the research with its sample. I also want to thank Dr. Reilly for his involvement and contribution, and especially for taking the time to work through the important decisions regarding design and statistical analysis. I want to thank Sonali Shah for her dedication and hard work during the data collection process, as well as Troy Vannucchi for his time and assistance. Finally, I would like to thank my family and friends for supporting me during this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAE……African-American English

CLD……Culturally and Linguistically Diverse

DLL……Dual Language Learner

DDM……...Dialect Density Measure

LSA…….Language Sample Analysis

MLU……...Mean Length of Utterance

MLUw……Mean Length of Utterance in Words

NDW……..Number of Different Words

NMAE……Non-Mainstream American English

SI……Subordination Index

SIE……..Spanish-Influenced English

SLP…….Speech-Language Pathologist

TTR…….Type-Token Ratio

CHAPTER 1
INTRODUCTION

Misdiagnosis of language impairment is more common in culturally and linguistically diverse (CLD) children than monolingual children of the same age due to the lack of unbiased standardized assessments, posing the need for alternative assessments (Bedore & Peña, 2008; Peña et al., 2006; Scheffner-Hammer et al., 2002). According to the United States Census Bureau (2000), there are 3,833,029 (18.6%) children between 5- and 9-years-old who speak a language other than English. Given that bi/multilingualism can be considered a component of the overall CLD label, under the assumption that speaking more than one language makes an individual linguistically diverse, the commonality of language impairment misdiagnosis for CLD children can be linked to these children who speak a language other than English. Furthermore, of the 18.6% of 5-to-9-year-old children mentioned above, 2,749,381 (71%) speak Spanish (U.S. Census Bureau, 2000). This high percentage of Spanish-English bilinguals poses a need for caution when considering diagnosis for this specific population.

Language sample analysis (LSA) has come to be known as the gold standard for speech-language pathologists (SLPs) in the assessment of CLD children (Pearson, Jackson, & Wu, 2014; Heilmann & Westerveld, 2013). LSA is a method of assessing language ability by eliciting spontaneous language in contexts that are more naturalistic than formal, standardized testing. It provides both clinicians and researchers with the ability to examine several features of an individual's expressive language in these contexts, as well as the relationships among these different features (Miller et al., 2006). In the category of "naturalistic contexts," there are several types of methods used to elicit

language samples. It has been shown that the observed linguistic skills and demands during LSA can vary based on the type of elicitation procedures implemented (Miller, Andriacchi, & Nockerts, 2016). While research indicates that LSA is most effective in revealing the authentic language abilities of bilingual children, there is no consensus on the best method of eliciting these language samples for analysis.

Researchers have compared different linguistic measures elicited via distinct tasks in hopes of allocating a "best method" (i.e., gold standard) of LSA. To bolster this research, one can refer to Wren (1985) who suggested three requirements when selecting a technique to elicit an efficient language sample. First, the procedure should elicit the type of language to be analyzed (e.g., syntax/complexity, semantics, linguistic productivity). Second, it must elicit a language sample that is representative of the child's typical language abilities. Finally, the procedure should provide some sort of comparability of the child's language abilities either in a criterion-referenced (e.g., comparing a child's performance at different time points) or normative manner if possible (i.e., comparing a child's performance to same-age peers; e.g., Mean Length of Utterance [MLU]; Brown, 1973). Since normative comparability is difficult to achieve with Dual Language Learners (DLLs) as there is a paucity of published cross-linguistic normative data, it is necessary that language samples be in the same context across assessment time points to facilitate comparable criterion-referenced results.

The current study focuses primarily on the Wren's (1985) first requirement for a LSA method to elicit a specific type of language (i.e., complex language characterized by specific microstructural measures), as well as it provides insight into the most efficient method for fulfilling this requirement. The language sample elicitation methods that were

analyzed and compared in this study are separated into three types: two methods of prompted fictional oral narratives (i.e., 'Story Retell' and 'Story Tell' conditions) and a pretend-play with an Elmo doll (i.e., 'Play' condition).

Assessment of Language Elicited with Narrative Tasks

Narrative samples are commonly used in language sample analysis (Arias & Friberg, 2016). Narrative sample elicitation is a useful and valid method to assess the expressive language abilities of CLD children. There are several types of narrative tasks including, but certainly not limited to, personal narratives (e.g., "Tell me about a time you won a competition"), script narratives (e.g., "Tell me all the things that happen when you go to the grocery story"; p. 17) and fictional narratives, such as story tell and story retell (Hughes et al., 1997). The current study will focus on story tell and retell.

Unlike standardized tests that tend not to be normed on this diverse population, narrative language assessments can be implemented in a wide range of naturalistic contexts across all languages and cultures, and have been shown to be educationally relevant (Lucero, 2015; Miller et al., 2006). The use of narratives has been further recommended because of its tendency to provide the individual being assessed with a meaningful context in which to apply cognitive and linguistic abilities when organizing his or her thoughts (Blank & Frank, 1971). The assessment of one's organization abilities is expected to include the structures that provide this organization (e.g., semantics, syntax, etc.). Within the category of narrative elicitation tasks, story retell and story tell are two methods that can be used to elicit these language samples for assessment of DLLs (Gutiérrez-Clellen, 1995; Muñoz et al., 2003).

In accordance with Wren's (1985) first requirement, when selecting a best method of elicitation for one's LSA, it is important to consider the linguistic measures to be addressed in the assessment. Typically, measures examined through LSA are divided into two categories: microstructure and macrostructure. Microstructure has been described as the "local" and "specific" measures of a narrative (p. 355; Hughes et al., 1997), such as MLU or Type-Token Ratio (TTR). Conversely, macrostructure would be the "global" measures that focus more on the organization of the narrative (p.355; Hughes et al., 1997), such as story structure or high point analysis (see Hughes et al., 1997 for details). For the purposes of the current study, the focus will be placed on microstructural rather than macrostructural measures. Especially with regards to narratives, microstructure has been reported to be variable depending on the type of narrative that was elicited (Allen, Kertoy, Petit, & Sherblom, 1994). For example, in a report by Mills, Watkins, and Washington (2013), school-aged, African-American English (AAE) speaking children (grades 2 – 5) demonstrated significantly higher lexical diversity (as measured by number of different words [NDW]) scores in their production of a personal narrative task compared to a fictional task. Additionally, expressive elaboration (a measure of narrative macrostructure) was seen to be greater in the fictional task condition (Mills et al., 2013).

Some research has shown evidence that story tells elicit more descriptive language for details, whereas story retells elicit longer and more reliably scored narratives overall (Merritt & Liles, 1989). In this particular study by Merritt and Liles (1989), it was suggested that having a clinician or examiner provide a model of the narrative (as seen in story retell conditions) is a possible benefit of the story retell task as it promotes scoring reliability. Ultimately though, they concluded that both story tell and

story retell are efficient methods of measuring narrative ability, regardless of their differences.

Other researchers, however, have specifically posited that the story retell condition provides an individual the opportunity to produce a variety of different structures and abilities, both semantically and syntactically as it is possible that the presentation of an adult model in the story retell condition can act as a linguistic scaffold as a child produces his or her narrative sample (Stalnaker & Creaghead, 1982). Comparatively, support for the story tell condition include that it might be more efficient for the analysis of a specific variable because it is a more "difficult" task (Liles, 1993; Merritt & Liles, 1987, 1989; Ripich & Griffith, 1988). This conclusion may be due to the consideration that in a story tell procedure, the child is not provided a model from which to structure his or her narrative response.

The distinction between methods of narrative elicitation is clinically relevant with regards to selecting valid and appropriate assessments for all children, especially CLDs, because, as evidenced above, specific measures of a child's expressive language performance may differ depending on which method is chosen. Therefore, additional research on the types of language elicited via particular methods can potentially aid the field in creating more effective and representative individualized speech-language assessments.

Assessment of Language Elicited Across Play and Narrative Tasks

Although narrative analysis has been closely studied, there are other valid methods of eliciting a representative language sample. In fact, much of the research in the

area of language sample analysis examines broader procedural comparisons such as fictional vs. personal narratives or structured vs. unstructured tasks. For example, one study investigating monolingual Afrikaans-speaking 5-year-old children revealed that narrative elicitation of a relatively structured story tell task elicits longer utterances than less structured tasks (i.e., what the authors titled "freeplay" and "conversation"), indicating story narrative procedures allow clinicians to observe this measure of "maximum behavior" (Southwood & Russell, 2004). In support of their claim that freeplay is an unstructured task, Southwood and Russell defined this condition as a session for which the participants were provided several varying toy items and a researcher to play parallel to each participant, providing minimal verbal prompts such as "What will happen next?" (p. 369). However, it is important to note here that, for the purposes of consistency with the current study, the variety of freeplay conditions referenced from the literature will hereafter be referred to as 'play' conditions (i.e., conditions in which the participants are presented with toy[s] and minimal prompts from researchers/adults).

Language sample elicitation through interview tasks has also been reported to elicit a greater amount of measurable language variables (Evans & Craig, 1992; e.g., MLU, TTR [a measure of lexical diversity]) than the less-structured play task in which participants selected a toy from a predetermined assortment and partook in child-directed parallel play with an adult. An earlier comparison can be linked to these results examining the linguistic differences between structured and unstructured elicitation tasks: Stalnaker and Creaghead (1982) revealed that a more structured story retell condition elicited a greater MLU score than the less structured play condition, though the play

condition elicited longer samples. Overall, these results support the potential that differences can exist in samples dependent on elicitation method, particularly between those with varying task "structure."

As evidenced in the reported literature, differences among elicitation methods can be analyzed in several different ways, especially since there are many different types of these methods and the kinds of support, stimuli, and scaffolding they can provide. In a similar manner to Wren (1985), the three conditions that were compared in the current study were Play, Story Retell, and Story Tell. Wren (1985) found that a story retell condition elicited more complex language as characterized by the "number of syntactic structures" (p. 92) produced, than play and story tell conditions. The latter conditions elicited the most typical (i.e., representative) language (Wren, 1985). The importance of relating Wren's work to the current study is not only the similarity in elicitation methods (e.g., her narratives also included visual stimuli, as did the current work), but also her statement regarding the importance of typical and "maximally complex" (p. 97) language in efficient LSA. In the current study, maximal performance was deemed the variable with which the elicitation methods would be ranked from most efficient to least.

Thus far, descriptions have been provided of the different language elicitation methods and major findings regarding the language complexity measures that can be predicted dependent on task. The information that follows is a description of a specific type of measure that can be implemented during LSA for populations similar to the participants included in the current study (i.e., CLD children).

Assessment of Dialectal Language across Tasks (The Dialect Density Measure)

The dialect density measure (DDM) is a measure that has been used in several studies of language to analyze dialectal variation in a language sample (e.g., Fabiano-Smith, Shuriff, Barlow, & Goldstein, 2014; Mills et al., 2013; Washington & Craig, 2002; Craig, Zhang, Hensel, & Quinn, 2009). DDM can be used to examine the quantity (i.e., frequency) and quality (i.e., type) of dialectal variations in language samples for the assessment of individuals who speak a Non-Mainstream American English (NMAE) dialect. It has been proposed that this measure can provide a wealth of quantitative information for assessment of an individual's rule system, as this can be especially important if that system is included under the label of NMAE for the purposes of distinguishing between difference and disorder (Oetting & McDonald, 2002). Furthermore, DDM is not only a measure of dialectal variation, but it has also been reported as correlated with language development, specifically narrative development (Mills et al., 2013; Ross, Oetting, & Stapleton, 2004; Gutiérrez-Clellen & Simon-Cereijido, 2009). For example, in one study, Ross, Oetting, and Stapleton (2004) found that individuals who presented with greater dialectal variations (as characterized by greater levels of African-American English utterances) also produced a better narrative structure than participants who did not present with as many African-American English (AAE) utterances. Another study reported that dialect variation can even be a helpful sign to predict early literacy abilities (Connor & Craig, 2006). Connor and Craig (2006) found that preschoolers with either high *or* low dialect density results demonstrated stronger fictional narratives predicting stronger early literacy abilities than the individuals who did not present with as strong levels (i.e., neither high nor low DDMs).

With the prevalence of misdiagnoses in the DLL population, a measure such as DDM can be very helpful in interpreting assessment performance for a more representative outcome. Gutiérrez-Clellen and Simon-Cereijido (2009) found that the quantity of dialectal differences among preschool and school-aged Spanish-English DLLs was significantly correlated with the number of misdiagnoses given to these NMAE speakers, meaning the greater the dialectal variation with which a child presented, the more likely he or she was to receive a misdiagnosis of language disorder (Gutiérrez-Clellen & Simon-Cereijido, 2009). This finding motivates the proposed study on the expectation that misdiagnoses can hopefully be reduced if it is determined which method of language sample elicitation provides a higher rating of DDM in order for this to be taken into account when determining the ultimate diagnosis (or lack thereof) of language impairment.

Given the significant DLL population and its implications in the speech-language assessment process (Bedore & Peña, 2008; U.S. Census Bureau, 2000), DDM analysis has been incorporated into comparisons of language sample elicitation procedures in past research. Some studies have found differences through these comparative analyses. For example, Fiestas and Peña's (2004) study analyzed cross-stimuli differences of story tell procedures to determine if the type of stimulus was a greater source of variability in language samples. This study of Spanish-English bilingual children (4;0 to 6;11) found more instances of Spanish-influenced English (SIE) utterances in a story tell task using a book as a visual aid compared to the same task using a single picture to prompt the narrative. Comparatively, Mills et al. (2013) did not find a significant difference of DDM between the two narrative conditions they compared. Despite these findings, it is

important to consider that DDMs have been shown to increase during the preschool years, suggesting that dialect density may have different results from different narrative elicitation tasks in the years prior to these later school years (Ross et al., 2004). The emerging literature's findings create the foundation and need for the current study to further compare language sample elicitation procedures using this dialectal variation measure.

Objectives

The purpose of this study is to supplement the research base in the area of assessment of the DLL/CLD population and address the need to streamline narrative language sample collections with quantifiable measures that may benefit all clinicians working with CLD groups. This was attempted by comparing language samples elicited by three procedures: two narrative elicitation procedures (Story Tell and Story Retell*)* and Pretend Play. All resulting language samples were compared with regards to a measure of dialectal variation in terms of quantity of SIE utterances (i.e., dialect density) and microstructural measures of narrative complexity (e.g., MLU in words [MLUw], TTR [Templin, 1957], subordination index). The elicitation procedure that resulted in a greater score of a particular microstructural measure than the others can be considered to indicate the comparative efficiency of this procedure in eliciting this specific measure. Results regarding the frequency of SIE utterances across Retell, Play and Tell procedures can provide information for speech-language pathologists to consider in their planning and analysis of language assessments for the DLL population by providing them with a sense of the level with which the child's language is influenced by the dialect spoken in

his or her community. As mentioned above, this influence can indicate prognosis for skills from narrative development to early literacy. Also, by taking dialectal variation into consideration for a comprehensive language assessment, an SLP would be providing a culturally sensitive, and likely more valid, judgment of a child.

Research Questions and Predictions

The first research question presented for the current study was the following: *Will one language sample method elicit more dialectal language features as evidenced by the presence of morphological and syntactic Spanish-influenced English utterances?* It may be expected that dialectal variation (i.e., DDMs) would be higher in the Play condition than the narrative conditions due to the predicted longer samples potentially providing the participants with more opportunities to produce SIEs (see Chapter One). However, it has been suggested that dialectal variation can be higher when elicited by more structured procedures than pretend-play (Washington et al., 1998). Across narrative contexts, it has been reported that dialect density does not differ (see Chapter One). Therefore, for the current study, it was predicted that DDMs would be higher in narrative conditions (i.e. more structured contexts) than the Play condition and there would not be a significant DDM difference across narrative contexts. It was expected that the results of this study would provide information that may assist the formation of individualized language sample assessments for DLL populations.

The second question was the following: *Will one language sample method elicit more complex microstructure as measured by greater TTR, SI, or MLUw?* Based on prior findings in the literature, the study's results were expected to reveal language sample

methods that elicit different quantities of a particular microstructure measure than each other, as well as eliciting comparative dialect densities across procedure. Regarding linguistic complexity, it was predicted that story retells will have greater scores of complex language than story tells as evidenced by higher results for MLUw, TTR, and SI, as the participants had a model from which to scaffold their narratives (see Chapter One). Since narrative methods have been shown to produce more complex productions than play methods, it was predicted that the results of the study would reveal a similar pattern (Wren, 1985).

CHAPTER 2

METHOD

In this study, the samples were analyzed for measures of narrative microstructure, including TTR (Templin, 1957), SI, and MLUw. Dialect density (DDM) was measured across three common language sample elicitation methods known as Story Tell, Story Retell, and Play to determine if one method elicits a more evident pattern of dialectal language as evidenced by the presence of morphological and syntactic Spanish-influenced English utterances.

Longitudinal Corpus

The language samples included in this study are a part of a longitudinal corpus of samples elicited from Puerto-Rican Spanish and English DLLs (Hammer, Lawrence, & Miccio, 2007, 2008; Hammer, Davison, Lawrence, & Miccio, 2009; Davison & Hammer, 2012; Cycyk, Bitetti, & Hammer, 2015)

Participants

The participants were recruited from Head Start programs in urban Central Pennsylvania. To be included, the participants had to qualify for the Head Start program for two years, have a mother who spoke Puerto-Rican Spanish dialect, be typically developing as concluded from a lack of parent and/or teacher concerns regarding their development as well as a score resembling the typical range on the Denver II

(Frankenburg et al., 1990), and pass a hearing screening (Hammer et al., 2007, 2008; Hammer et al., 2009).

Procedures

Language sample elicitation procedures for this study are explained below. All language samples were audio-recorded and transcribed into SALT Software (Miller & Iglesias, 2008) for further analysis.

Current Study

Participants

The samples used in the current study were selected from the original longitudinal corpus. The participants that provided these samples were in the spring term of their kindergarten year at the time of data collection. The spring time-point is the second of two elicitation time-points for that particular year. At this time point, all included DLLs attended kindergartens in which English was the language of instruction. The mean chronological age of these participants was 6.27 years (*SD* = .334). Of the individuals from which the samples were elicited, 40% were male and the remaining 60% were female. All individuals were typically-developing.

The rationale for selecting these samples that represented this kindergarten age range from the original longitudinal corpus is based upon several considerations in the literature. Due to the significant percentage of 5- to 9-year-old children speaking a language other than English in the United States (i.e., 18.6%; U.S. Census Bureau, 2000), studies regarding individualized assessment protocols to be used for this language group

are necessary. With specific regards to age range, although by age 5 children typically have a foundational understanding of their languages' basic syntactic rules, they are still developing their linguistic abilities with regards to complexity and application of these rules and structures they have acquired (Chomsky, 1969; Slobin, 1973); therefore, it is reported that children in the selected age range are experiencing a significant developmental period with regards to producing narratives (Pearson, 2002). Mills et al. (2013) informed that curricular education incorporates narration as a significant requirement. Assessment of oral language abilities provides insight into a student's eventual literacy abilities (Scarborough, 2001; Griffin, Burns, & Snow, 1998; Storch & Whitehurst, 2002); therefore, the use of narration as an assessment tool for this age group is an important indicator for future academic performance.

Language Exposure

Language exposure information was collected through parent questionnaires. A distinction consistent with Davison and Hammer (2012) was made between Spanish-speaking participants who were exposed to both English and Spanish at home since birth and those whose exposure to English began upon their entry into Head Start. For the 63 individuals included in this study, 98% (i.e., 62) of their caregivers provided information regarding age of language exposure. From these responses, it was concluded that 58% (i.e., 36) of the participants were reported as bilinguals exposed to English and Spanish since birth (i.e., 1 year of age or younger) and the remaining 42% (i.e., 26) of participants were reported as bilinguals who were first exposed to English at Head Start (i.e., either 3- or 4-years-old).

Only the English language samples were incorporated into data analysis as the purpose of the study was to analyze methods of assessment for Spanish-influenced English language. With regards to potential considerations when linking the results of this study to language samples in Spanish, it has been noted in the literature that English expressive language can predict DLL's readiness for reading in Spanish as well (Miller et al., 2006). Similarly, Fiestas and Peña (2004) found that linguistic complexity was equal across Spanish and English narrative samples, indicating that the results of this study may be mirrored in the Spanish samples. Therefore, though further studies are needed to generalize the results of this study to Spanish samples, the literature does present a connection between samples in both languages indicating the potential for this cross-linguistic generalization.

Inclusion Criteria

This study specifically used the English Story Tell, Story Retell, and Play language samples elicited during the spring term of the participants' kindergarten year. Criteria for inclusion in the current data set were available data, at least a Tell and a Retell sample for each participant (i.e., if the participant was not administered a Play sample, he/she was still included in data analysis; if a participant had a Play sample and either a Tell or Retell, he/she was excluded from data analysis), at least 50% of utterances in English (following research standards; e.g., Heilmann et al., 2008), and samples of at least 15 utterances in length. Once exclusion, data collection, and analysis were completed, the total number of language samples equaled 178 (i.e., 63 individual participants' samples). Of the 178 samples, there were 52 Play samples, 63 Retell

samples, and 63 Tell samples; therefore, the Retell and Tell were paired and 11 participants that had not received the Play condition at this kindergarten spring time-point were present in the collection sample. It is important to note that there was not a systematic reason for these 11 participants not being elicited a Play sample and that data analysis measures were taken in order to assure that the group that were administered all three conditions was not significantly different than the group of pairs that were only administered the two narrative conditions (see p. 21).

Description of the Three Elicitation Method Conditions

In the Story Retell condition, the examiner told the participant a scripted, adapted story created from Mercer Mayer's "frog story" wordless picture books called *Frog and His Friends Play Hide and Seek*. The participant was then shown the picture book and asked to retell the story to the examiner. The visual stimulus of the picture book pages was present throughout the Story Retell condition and the participant was prompted to flip through the pages as he or she retold the story. The Story Tell condition immediately followed the story retell. Each participant was shown a different wordless picture book, *Frog Goes to Dinner* (Mayer, 1974). In this condition, the participants were not provided with the initial examiner model; however, since this condition was immediately following the Story Retell condition, they did have previous exposure in completion of the task. The visual stimulus of the story pictures was present during this condition as well, as each participant was prompted to flip through the pages while telling the story in the same manner as in the Story Retell condition. In the Play condition, the participants were

presented with an Elmo doll and prompted to play with the doll. Minimal prompts were permitted during the Play condition.

Table 1: Description of the three conditions

	Pretend Play	Story Tell	Story Retell
Materials	Elmo doll	Wordless picture book: *Frog Goes to Dinner* (Mayer, 1974)	Wordless picture book: *Frog and His Friends Play Hide and Seek* (adapted materials from Mayer's *Frog* series)
Model	No model provided.	Prior modeling of unrelated story (*Frog and His Friends Play Hide and Seek*)	Prior, scripted modeling of same story
Prompts	Participant prompted to engage with Elmo doll during child-directed play.	Participant prompted to tell story while flipping through the book's pages.	Participant prompted to retell the story while flipping through the book's pages
Support	Real objects and verbal prompts.	Visual cues from book and minimal verbal prompts.	Visual cues from book and minimal verbal prompts.

Notes: Comparison of the three elicitation methods in their materials, presentation of models and prompts, as well as visual and/or verbal support provided throughout the task. "Materials" refers to the physical items presented to the child to complete each task. "Model" refers to the presence or absence of an oral presentation prior to the completion of each task. "Prompts" are the verbal assistance given by the examiner either as task instruction or as encouragement for the child to continue with the task. "Support" can be referred to as a blanket term for cues, prompts, and materials that facilitated the child in his or her completion of the task.

Description of the Data

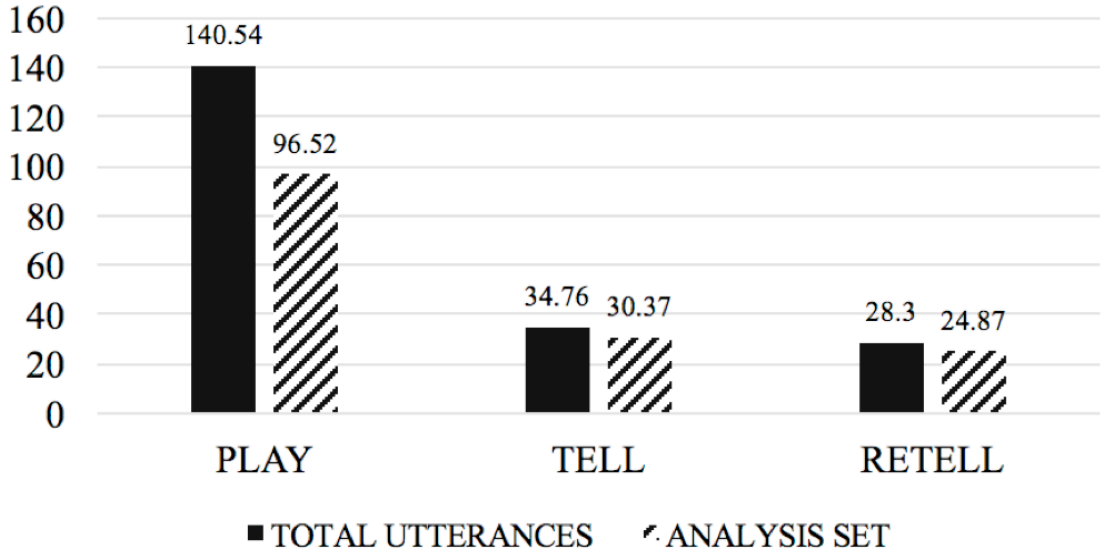
Sample Length

The samples were analyzed for length in total number of utterances and analysis set. The analysis set is a subset of the total number of utterances that includes only

complete, intelligible, and verbal utterances. The Play method elicited the longest samples at 140.54 total utterances (analysis set: 96.52). Tell samples were an average of 34.76 total utterances in length (analysis set: 30.37), and Retell samples were the shortest with an average of 28.30 total utterances (analysis set: 24.87). Although it is typically recommended to have 100 utterances when calculating MLU (Lahey, 1988; Retherford, 1993), the recommendation has been seen to vary in the literature, with some studies reporting 50 utterances as an acceptable length for MLU calculation (Miller & Chapman, 1981; Kemp & Klee, 1997).

Surveys of speech-language pathologists in the literature show reports of 25% (Hux et al., 1993) - 43% (Loeb et al., 2000) of SLPs using samples with fewer than 50 utterances. Kemp and Klee (1997) reported that only 15% of SLPs used the recommended 100-utterance sample when making these calculations. These reports support the notion that, although the average length of samples for the narrative elicitation conditions are shorter than the "standard" recommendation of 100 utterances, these results can still be generalized as it is likely that SLPs clinically use samples shorter than 100 utterances. It is important to note that small sample sizes have been reported to be pose an issue for test-retest reliability (Gavin & Giles, 1996); however, studies have found that shorter samples can provide results consistent with the actual language abilities of Spanish-English DLLs (Heilmann et al., 2008). Therefore, future research in this area should aim to elicit and analyze longer samples for this population for the purpose of greater generalization of results.

Figure 1: Length of samples across condition



Reliability

During preliminary data analysis, reliability was established between the researcher and a duplicate coder following a training on the SIE types and codes, as well as SALT Software's protocol for SI coding (Miller & Iglesias, 2008). Reliability was then maintained throughout actual data analysis for SI at 95.2% on 11% of the samples, and for DDM at 90.6% on 6% of the samples. Reliability ratings were calculated by dividing the number of agreements by the sum of agreements and disagreements (Oetting & McDonald, 2001; Craig et al., 2009).

Procedures for Language Sample Analyses

During the data collection for Hammer et al.'s original study, language samples were audio-recorded and transcribed into SALT Software (Miller & Iglesias, 2008). LSA for the current study included the coding and analysis of dialect density and three

different microstructure elements. The DDM was measured using a compilation of coded morphosyntactic SIE features that have been well-documented in the literature (Thompson, Craig, & Washington, 2004; Fiestas & Peña, 2004; Kayser, 1989; Wolfram, 1974; Politzer & Ramirez, 1973; see Appendix A for full list) and analyzed using both token- and, for post-hoc analysis, type-based methods (Oetting & McDonald, 2002; Washington & Craig, 2002). The token-based method is completed by dividing the sum of individual SIE tokens by the total number of utterances (Oetting & McDonald, 2002; Horton-Ikhard et al., 2005) and the type-based method takes into account the specific types of morphosyntactic SIEs based on a pre-determined compilation of well-documented SIE patterns (see Appendix A). It is important to note that "total number of utterances" in DDM analysis can be a different sum than the total number of utterances used in SALT's microstructural analysis since the DDM total number of utterances includes all utterances, whereas the SALT program automatically uses the analysis set that excludes incomplete, nonverbal, and unintelligible utterances when calculating microstructure measures, including MLUw and TTR (SALT Software). However, these differences were consistent across all participants, as well as they were accounted for during data analysis (see below).

Measures of child language through LSA can be performed at the macrostructural and microstructural level (Hughes et al., 1997). This study focused on the following variables. The microstructure variables analyzed across samples were TTR, SI, and MLUw. TTR is a measure of lexical diversity (i.e., number of different words divided by total number of words), SI is a measure of syntactic complexity via clausal density (i.e., number of clauses per C-unit; Mills et al., 2013), and MLUw has been reported as a

measure of grammatical productivity and sentence complexity that has been used for analysis across languages (Bedore et al., 2010; Parker & Brorson, 2005). For SI, the procedural instructions from SALT Software (Miller & Iglesias, 2008) was implemented to guide scoring. The analysis of microstructure variables is in accordance with SALT Software (Miller & Iglesias, 2008) formulas and the results for these microstructural measures have been generated through the Software.

Data Analysis

All analyses were conducted using Microsoft Excel for Mac (Microsoft, 2016). Descriptive statistics were calculated across elicitation methods (i.e., SI, TTR, MLUw, DDM). To determine the effects of elicitation procedure on each of the dependent variables, several paired samples *t*-tests were conducted with effect size estimates (i.e., Cohen's *d*).

Additional Analysis due to Uneven Pairs

It is important to note that in order to complete the *t*-tests and Cohen's *d* analyses, the conditions were paired. Since the criteria for inclusion in the current study was for each participant to have at least been administered a Tell and a Retell task, some of the participants (i.e., 11 of the total 63) were not administered the Play condition. Therefore, the Tell and Retell conditions were paired for the total 63 participants (i.e., 126 total narrative samples; 63 Retell and 63 Tell); whereas, the 11 participants who did not experience the Play condition led to exclusion of their Retell and Tell samples when completing data analyses for the condition pairs Play-Tell and Play-Retell across all

measures. This was done to ensure that each pair for the Play-Tell and Play-Retell analyses had a match (i.e., if the exclusion of those 11 Play samples was not done, the Tell and Retell samples for those participants would not be paired to Play samples). However, as a means of assuring that there are no confounding variables for the analyses following the exclusion of those 11 participants/samples, Table 4 in Chapter Three presents a side-by-side depiction of the *t*-test results for the Retell vs. Tell conditions, including the participants who completed all three tasks (i.e., Retell, Tell, and Play).

CHAPTER 3

RESULTS

Descriptive statistics are displayed in Table 2 and graphically represented in Figure 2. Table 3 lists the *p* values and Cohen's *d* effect sizes for each of the condition pairs.

Microstructure Variables

MLUw

MLUw was significantly lower for Pretend Play ($M = 4.52$, $SD = 0.86$) relative to Story Tell ($M = 6.82$, $SD = 1.30$), $t(51) = 13.11$, $p < .001$, $d = 2.1$ (large effect), and Story Retell ($M = 7.00$, $SD = 1.31$), $t(51) = 14.42$, $p < .001$, $d = 2.3$ (large effect). MLUw did not differ from Tell ($M = 6.82$, $SD = 1.30$) or Retell ($M = 7.00$, $SD = 1.31$), $t(62) = 1.31$, $p > .05$, $d = 0.1$ (negligible effect).

SI

SI was significantly different across elicitation methods with Retell ($M = 1.29$, $SD = 0.20$) eliciting a higher SI than Tell ($M = 1.22$, $SD = 0.17$), $t(62) = 3.51$, $p < .001$, $d = 0.4$ (small-medium effect) and Pretend Play ($M = 1.05$, $SD = 0.10$), $t(51) = 8.06$, $p < .001$, $d = 1.6$ (large effect). Tell elicited a higher SI ($M = 1.22$, $SD = 0.17$) than Play ($M = 1.05$, $SD = 0.10$), $t(51) = 6.88$, $p < .001$, $d = 1.3$ (large effect).

TTR

TTR was found to be significantly lower for the Play ($M = 0.35$, $SD = 0.07$) relative to the Retell condition ($M = 0.39$, $SD = 0.06$), $t(51) = 3.66$, $p < .001$, $d = 0.6$ (medium effect), but not the Tell condition ($M = 0.38$, $SD = 0.07$), $t(51) = 1.77$, $p > .05$, $d = 0.4$ (small-medium effect). Notably, the effect size for the Play-Tell pair was inconsistent with the *t*-test result. Possible explanations for this inconsistency will be considered in Chapter Four.

The two narrative contexts were significantly different from each other with Retell ($M = 0.39$, $SD = 0.06$) eliciting a greater TTR than Tell ($M = 0.38$, $SD = 0.07$), $t(62) = 2.28$, $p < .05$, $d = 0.2$ (small effect).

Dialect Density

DDM

Of the 178 samples analyzed, 173 (i.e., 97%) contained at least 1 feature of Spanish-influenced English listed on the compiled list of well-documented SIEs (see Appendix A). As analyzed by the token-based method (Oetting & McDonald, 2002), the DDM frequencies were significantly lower for Pretend Play ($M = 0.09$, $SD = 0.04$) as compared to Tell ($M = 0.17$, $SD = 0.12$), $t(51) = 4.51$, $p < .001$, $d = 1.0$ (large effect) and Retell ($M = 0.17$, $SD = 0.11$), $t(51) = 4.81$, $p < .001$, $d = 1.1$ (large effect). There was no significant difference found between the DDMs elicited by the Tell ($M = 0.17$, $SD = 0.12$) and Retell ($M = 0.17$, $SD = 0.11$), $t(62) = 0.03$, $p > .05$, $d = 0.0$ (no effect) conditions.

Table 2: Descriptive statistics for measures of language analyses

	M	*SD*	*Range*
Pretend Play			
DDM	0.09	0.04	0.20
SI	1.05	0.10	0.59
TTR	0.35	0.07	0.39
MLUw	4.52	0.86	4.70
Story Tell			
DDM	0.17	0.12	0.52
SI	1.22	0.17	0.95
TTR	0.38	0.07	0.40
MLUw	6.82	1.30	7.66
Story Retell			
DDM	0.17	0.11	0.52
SI	1.29	0.20	1.05
TTR	0.39	0.06	0.32
MLUw	7.00	1.31	5.77

Note. DDM = Dialect Density Measure, SI = Subordination Index, TTR = Type Token Ratio, MLUw = Mean Length of Utterance in Words

Figure 2: Comparison of measures of language analyses by elicitation method

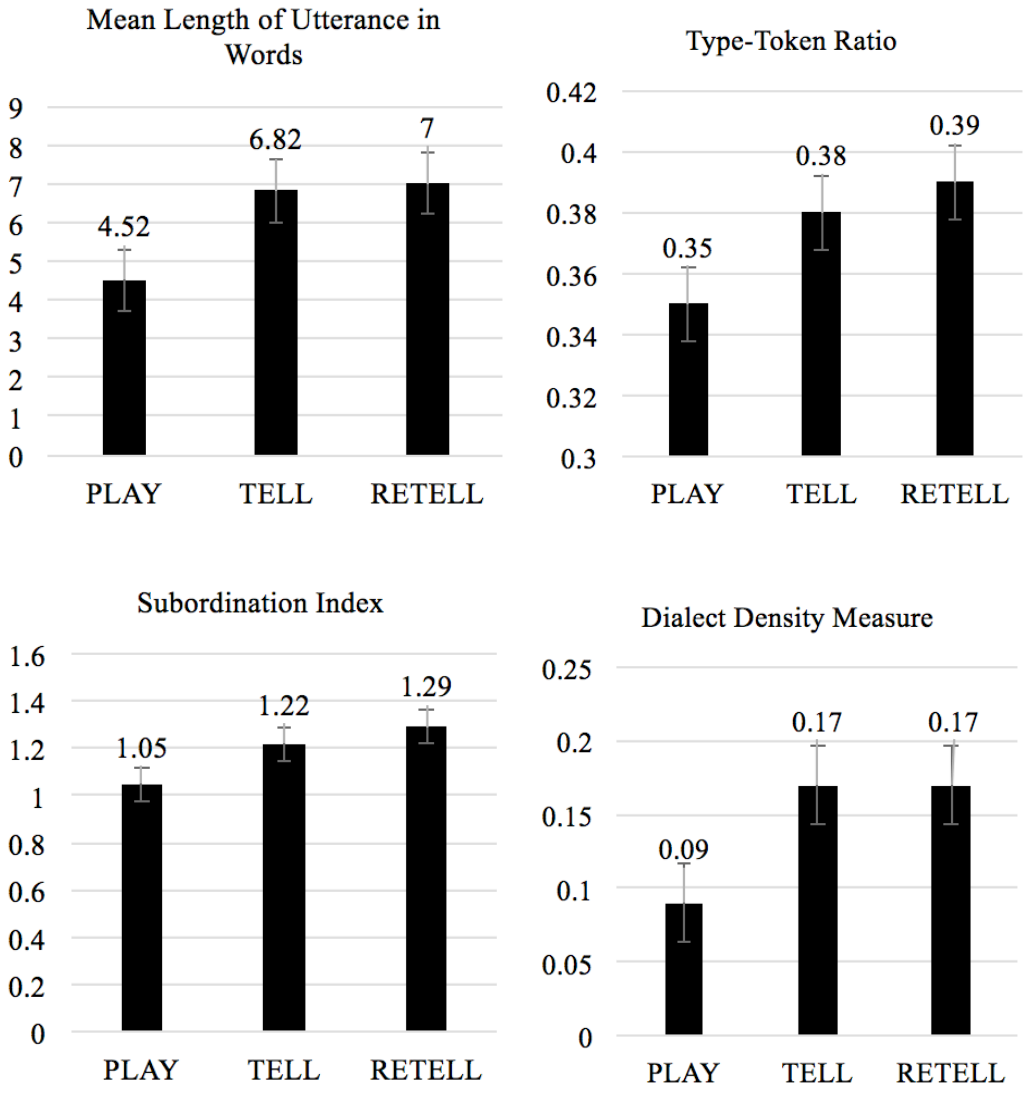


Table 3: Results of paired samples *t*-tests and Cohen's *d* effect size analysis

	p	Cohen's *d*
Story Tell – Story Retell DDM	0.9800	0.0
Story Tell – Story Retell SI	0.0008*	0.4
Story Tell – Story Retell TTR	0.0259**	0.2
Story Tell – Story Retell MLUw	0.1951	0.1

Play – Story Retell DDM	0.0000*	1.1
Play – Story Retell SI	0.0000*	1.6
Play – Story Retell TTR	0.0006*	0.6
Play – Story Retell MLUw	0.0000*	2.3
Play – Story Tell DDM	0.0000*	1.0
Play – Story Tell SI	0.0000*	1.3
Play – Story Tell TTR	0.0823	0.4
Play– Story Tell MLUw	0.0000*	2.1

Notes. *p<.001 **p<.05; Cohen's d results should be interpreted based on the following conventions: d = 0.2 is a small effect size; d = 0.5 is a medium effect size; d = 0.8 is a large effect size.

t-test Analysis for Uneven Pairs

See the Data Analysis section in Chapter Two for details regarding the purpose of these analyses. Table 4 shows the side-by-side results of the Retell vs. Tell comparisons in the following two groups: Group one includes those who were not administered the Play condition. Group two excludes those who were not administered the Play condition due to non-systematic reasons (e.g., timing considerations during the study). The analysis with the excluded participants shows functional equivalence between the group that was administered all three conditions and the group that was only administered the two narrative conditions, indicating that there should not be any concerns that the results from the current study reported above are not generalizable should the excluded participants have been administered the Play condition during the original study.

Table 4: *t*-test analysis for uneven pairs

	Group One	Group Two
MLUw	$p = 0.195$	$p = 0.096$
TTR	$p = 0.026$**	$p = 0.019$**
SI	$p = 0.001$*	$p = 0.001$**
DDM	$p = 0.980$	$p = 0.971$

Notes. *$p<.001$ **$p<.05$; Group One includes those who were not administered the Play condition and Group Two excludes them.

CHAPTER 4

DISCUSSION

The purpose of this study was to compare three microstructural measures of language complexity and productivity (i.e., MLUw, SI, TTR) and the frequency of dialectal variation across three elicitation methods (i.e., Play, Tell, Retell) on the population of kindergarten-age Spanish-English DLLs. This study was motivated by the desire to identify an elicitation method that results in a greater score of a particular microstructural measure than the other methods, as well as to identify a method that elicited the highest SIE frequency as a means of providing information for SLPs who work with this growing DLL population. This study was aimed toward SLPs in their assessment planning for DLLs as they select which method to use when analyzing the narrative language of their clients.

The first research question in this study addressed whether one language sample method would elicit a more evident pattern of dialectal language as evidenced by the frequency of morphological and syntactic SIEs. It was expected that DDMs would be higher in Play than Tell or Retell and that there would not be a significant difference between the two narrative contexts. Given the results of the current study, the hypothesis was supported: the two narrative methods elicited significantly higher DDMs than the Play method, and there was no significant difference in DDM found between the Tell and Retell conditions.

The second research question addressed whether one language sample method would elicit greater narrative microstructural results (i.e., TTR, SI, MLUw) as evidence of greater language complexity and productivity. Based on the literature, it was expected

that Retell would elicit greater TTR, SI, and MLUw than the Tell method. It was further hypothesized that both Tell and Retell would elicit more complex samples than Play as evidenced by greater results over all three microstructure variables. Following analysis, this prediction was partially supported. Retell was found to elicit some greater scores of productive and complex language than Tell as evidenced by higher scores for TTR and SI; however, MLUw was similar between the narrative contexts. Additionally, one of the narrative methods (i.e., Retell) elicited more complex language than Play when considering all three microstructure measures. However, Tell did not produce more complex language than Play with regard to the TTR measure of lexical diversity, though there was an small-medium effect size, indicating that sample length may have limited statistical power, inflating the likelihood of type II error. Therefore, effect size will be considered when discussing this measure. The results corresponding with each research question will also be further discussed below in the format of Tell vs. Retell, then Narrative vs. Play.

Comparison of Narrative Methods

Both narrative methods, Tell and Retell, appeared to present similar levels of utterance length and dialect density; however, Retell-elicited narrative samples elicited a higher level of syntactic complexity, as measured via SI, and lexical diversity, as measured via TTR, than the Tell narrative conditions. These findings are also in correspondence with the recent bilingual literature (Roch, Florit & Levorato, 2016). Roch et al. (2016) proposed that the greater narrative performance in their retell condition than the tell condition may have been evoked by a different level of cognitive difficulty

between the tasks, further indicating that the retell task may be cognitively less challenging than the tell task since the kindergarten-age children were asked to generate their own narratives from the visual support to produce the tell, whereas the retell task more likely followed a "reconstructive process" (p. 64). This is in accordance with Liles (1985, 1993) who reported that the tell task posed a challenge for the child, therefore lending to the child having more difficulty with producing complex and productive language, supporting the difference found between the two narrative conditions in eliciting diverse vocabulary (TTR) and syntactic complexity (SI). Merritt & Liles (1989) also posited that retell tasks may be found to be easier than the tell tasks without an initial model; however, the comparative efficiency between the two narrative tasks still remains in question (Lever & Sénécal, 2011).

Though it has been proposed that Retell involves an "easier" process of producing a narrative sample, it is also true that the presence of a narrative model allows for the participant to scaffold as a means of recreating the narrative he or she just heard (Liles, 1993; Stalnaker & Creaghead, 1982). While this may make the Retell condition less cognitively challenging, it also provides an opportunity for a more dynamic process of assessment. Dynamic assessment of DLLs has been supported in the literature as an effective method in the comprehensive evaluation process for this special population (Peña et al., 2006). It involves a level of learning that many assessment processes tend not to include. Dynamic assessment can be incorporated into the interpretation of the results of the current study with regards to the difference in TTR and SI across narrative conditions. In the Retell condition, the initial story model provided a scaffold for both lexical items and morpho-syntactic structures; therefore, the child had the opportunity to

"learn" these new items and structures and to implement them in the restructuring and retelling of his or her own story.

The lack of a significant difference between Tell and Retell in their elicitation of MLUw is supported by past claims that both of these narrative conditions are efficient in their role in comprehensive assessment (Merritt & Liles, 1989); however, the similarity in MLUw is interesting since the measurements for the other method of syntactic complexity (i.e., SI) revealed a significant difference between the narrative methods. In the interpretation of these MLUw results, it is important to reiterate that the Tell condition technically did have an indirect model for producing a narrative as this condition was presented immediately following the Retell condition; therefore, it is understandable that the results of language complexity and productivity were not more significantly different between the two narrative conditions. Even though the indirect model for the Tell condition was for an unrelated story, it may have primed the child to produce more complex language by reminding the child of the manner in which a narrative can be told. Additionally, the two wordless story books to which the children had access to in each condition (one per condition) did have a similar theme (i.e., frog story either by or based from Mayer, 1975). Therefore, in the case of the current study, the Tell task was not as "difficult" as it has been reported to be. Fergadiotis, Wright, and Green (2015) performed their conditions in a similar manner with the presence of an initial unrelated model that they titled a "warm-up task" (p. 842) to prepare the participants for each elicitation procedure.

As seen in the MLUw results, there was no significant difference revealed between Tell and Retell with regards to dialect density. These results are consistent with

the research (Mills et al., 2013). Though the narrative conditions implemented by Mills et al. (2013) were different than that between the Tell and Retell in the current study, it is still interesting that varying narrative conditions have been found to be similar in their elicitation of dialectally variant language. As in the current study, Mills et al. (2013) found this similarity in DDM while also finding significant differences between elicitation procedure in the narrative microstructural results. This and the results of the current study may indicate that DDM may not vary across narrative contexts, further posing the conclusion that, in analyzing narrative samples, clinicians should largely take into account that narrative contexts elicit an equally high level of dialect density and to let that influence an eventual diagnosis that is sensitive to the DLL individual.

Comparison of Narrative and Play Methods

In comparing the two narrative methods to the Play method, it was revealed that Play elicited significantly lower levels of MLUw, SI, TTR, and DDM than Retell and Tell (when including effect size analysis). The greater MLU and SI scores in the narrative contexts as compared to the Play contexts are supported by past findings (Stalnaker & Creaghead, 1982; Wren, 1985; Leadholm & Miller, 1992; MacLachlan & Chapman, 1988). Specifically, Wren (1985) reported that the retell condition elicited a greater number of syntactic structures than both the play and tell conditions, though she did also find that the less structured conditions (i.e., play and tell) elicited the language most correlated to the individuals' own abilities (i.e., most "representative"). The results are even further supported by a similar finding that the story tell task elicited longer utterances than a pretend play method, further indicating that the story narrative

procedures allow clinicians to observe the child's "maximum behavior" (Southwood & Russell, 2004).

The finding that the narrative methods (especially Retell) greatly elicited the most complex language as compared to the less structured Play condition is significant in that maximal performance can be a crucial component to consider in the process of a meaningful language sample analysis. For example, a clinician may want to discover the highest MLUw a child can produce. Given the results of this study, that clinician may not select Pretend Play as an option for that given scenario; however, it is important to still consider the representativeness of the sample in other assessment contexts. Similar conclusions can be drawn for the TTR scores across the three methods. Overall, the results indicate that narrative conditions with visual support and a direct or indirect model may elicit more maximal performance than less-structured play-based conditions with regards to measures of narrative microstructure.

It is possible that the less structured and less challenging nature of the Play condition may have provided an environment in which the participants were not obligated the produce their most complex language, as they were in the narrative contexts; therefore, the less complex language in the Play condition would present as a lack of structured opportunities to provide longer, more complex utterances with more diverse lexical items. Narrative samples may also be considered as presenting with denser (i.e., more syntactically and lexically complex) language as induced by the structure of the task itself; for example, when telling and/or retelling a story, description of the details and story elements (i.e., macrostructural elements) may require more detailed and comprehensive language. Therefore, it is possible that, in comparison with that of play-

based methods, the context of a narrative lends itself to longer utterances with more complex vocabulary and syntactic structures as a sort of "requirement" for completing the narrative; whereas, in the Play method, the participants did not receive a model nor the structured nature of a narrative context as an impetus to providing any particular level of language. Past literature in this area support these claims (Hadley, 1998; Evans & Craig, 1992). Hadley's (1998) review on language sample analysis for school-age children revealed that narrative samples tend to expose their maximal performance better than the alternatives, while Evans and Craig (1992) found a mirroring result in which the optimal linguistic complexity was found in the most structured condition (in the case of Evans & Craig, an interview format).

Regarding dialect density, both of the narrative conditions elicited a significantly higher DDM than the Play condition. The lesser frequency of dialectal features in the Play samples may be evidence of the less-structured nature of this condition. As has been reported in a book by Kohnert (2013), there is a point in dual language acquisition (at least for those experiencing L2 learning since birth) at which the vocabulary in each language is developed into two separate lexicons but the morphosyntax in both languages overlaps in a way that may present as a high frequency of dialectal variation in the L2 (i.e., English). In the present study, the narrative conditions may have elicited a higher frequency of dialectal features in English because of the linguistic demand on the participants to produce academic, narrative, and structured language in the form of a story Retell or Tell. The overlap may have exposed itself as the participants produced these more structured language samples as a result of the supposed higher frequency of

complex morphosyntactic structures, therefore providing the opportunity to produce these more frequent SIEs.

Clinical Implications

The current study has revealed ways in which its results can be utilized in the field of speech-language pathology. Firstly, with regards to the findings about DDM frequency (i.e., DDM was found to be more frequent in narrative tasks than in a Play task; no distinction for DDM frequency was found between Retell and Tell conditions), it is likely that implications can be made for clinical assessment and intervention with kindergarten-aged, Spanish-English DLLs. The finding that a higher frequency of dialectal features can present in narrative samples may provide SLPs with an awareness when assessing children of this population. Gutiérrez-Clellen and Simon-Cereijido (2009) found a significant correlation between dialectal variations and misdiagnoses for a CLD population. As mentioned in Chapter One, misdiagnoses are commonly given to individuals in diverse populations (Bedore & Peña, 2008; Peña et al., 2006; Scheffner-Hammer et al., 2002). It is possible that an awareness to the types of language sample procedures that elicit the most dialectal variance may facilitate clinicians to approach DLL language differently, perhaps by attributing certain SIE features as such rather than using these features as evidence for a potential misdiagnosis.

Another potential implication of the current research speaks to the microstructural language complexity and productivity findings. It may be concluded from the data that maximal performance can be best represented in narrative sampling tasks. More specifically, the findings suggest that maximal performance for discrete measures of

language may be best elicited by a Retell narrative, following a model and with visual support. It is important to note, however, that not all clinical assessment scenarios are in need of a measure for maximal performance; rather, a clinician might require a method that produces the most representative or natural sample. In those cases, it is possible that methods other than Retell or Tell (e.g., Play) are more suitable.

Limitations and Future Directions

Although the findings of the current study show differences between the conditions in their ability to elicit measure of linguistic complexity and dialectal variation, there were some factors that may have impacted the results. Sample length is included in these limiting factors. Due to the availability of data, samples were excluded if they had less than 15 utterances. It has been reported in the literature that a sample of 50 to 100 utterances in length is preferable; therefore, this must be considered when generalizing the results to longer samples. Further, the TTR measure, which is calculated as a ratio, is expected to have been impacted by sample length as revealed in its insignificantly difference between Play and Tell through *t*-test analysis, but small-medium effect size. Further research should attempt to include more controlled sample length.

Several potential future directions have resulted from the findings of the current study. As mentioned above, clinicians are not always searching for the assessment tool to elicit maximal performance; therefore, it would be interesting for future research to examine the differences of these three elicitation methods on the basis of other measures such as naturalness or representativeness. Another potential future direction would be to

increase the size of the participant sample to improve generalization to the population, as well as to control for length of language sample. In the current study, Play samples tended to have many more utterances than the Narrative samples; therefore, controlling sample length may have an effect on the results and permit insight into other distinctions, or lack thereof, between the elicitation methods. Additionally, the specific SIE types (rather than the frequency), less-documented SIE patterns (see Appendix B), appearance of other nonmainstream American English (NMAE) dialect patterns (e.g., African American English [AAE]) and grammatical errors were all recorded in this study for potential post-hoc analyses. It would be interesting for future research to explore effects of elicitation type on these variables, as well as to examine the types of SIEs more closely.

Future research may also want to examine the differences between elicitation methods for different language and dialectal populations. Much research in this area has been completed for the AAE population (e.g., Mills et al., 2013); however, more studies are merited, even involving dialects or variations of SIE other than the Puerto Rican-English portrayed in the current study's results. Similar to the suggestion to examine other languages, it may be worthwhile to take a closer look at the effect of language on the performance of Spanish-English DLLs with regards to DDM and microstructural complexity; for example, is there a difference in the dialect density for an individual of this population in samples produced in English as opposed to Spanish. Finally, it is recommended for future research that comparisons be made to monolingual English speakers and DLLs as a means of confirming results for the DLL population.

REFERENCES

Allen, M.S., Kertoy, M.K., Sherblom, J.C., & Pettit, J.M. (1994). Children's narrative productions: A comparison of personal event and fictional stories. *Applied Psycholinguistics, 15,* 149–176.

Arias, G. & Friberg, J. (2016). Bilingual language assessment: Contemporary versus recommended practice in American schools. *Language, Speech, and Hearing Services in Schools, 48*, 1-15.

Bedore, L.M., & Pena, E.D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1-29.

Bedore, L.M., Pena, E.D., Gillam, R.B., & Ho, T.H. (2010). Language sample measures and language ability in Spanish-English bilingual kindergarteners. *Journal of Communication Disorders*, *43*(6), 498-510.

Blank, M., & Frank, S. (1971). Story recall in kindergarten children: Effect of method of presentation on psycholinguistic performance. *Child Development, 42*, 229-312.

Brown, R. (1973). *A First Language: The Early Stages*. Cambridge: Harvard University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Connor, C.M., & Craig, H.K. (2006). African American preschoolers' language, emergent literacy skills, and use of African American English: A complex relation. *Journal of Speech, Language, and Hearing Research, 49,* 771-792.

Craig, H.K., Zhang, L., Hensel, S.L., & Quinn, E.J. (2009). African American English–speaking students: An examination of the relationship between dialect shifting and reading outcomes. *Journal of Speech, Language, and Hearing Research*, *52*(4), 839-855.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cycyk, L.M., Bitetti, D., & Hammer, C.S. (2015). Maternal depressive symptomatology, social support, and language development of bilingual preschoolers from low-income households. *American Journal of Speech Language Pathology*, 24(3), 411-425.

Davison, M.D. & Hammer, C.S. (2012). Development of 14 English grammatical morphemes in Spanish-English preschoolers. *Clinical Linguistics & Phonetics, 26(8)*, 728-742.

Evans, J.L. & Craig, H.K. (1992). Language sample collection and analysis interview compared to freeplay assessment contexts. *Journal of Speech, Language, and Hearing Research, 35(2)*, 343-353.

Fergadiotis, G., Harris Wright, H., & Green, S.B. (2015). Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research, 58,* 840-852.

Fiestas, C.E. & Peña, E.D. (2004). Narrative discourse in bilingual children language and task effects. *Language, Speech, and Hearing Services in Schools*, *35*(2), 155-168.

Frankenburg, WK.; Dodds, J.; Archer, P.; Shapiro, H.; Bresnick, B. DENVER II. Denver, CO: Denver Developmental Materials; 1992.

Griffin, P., Burns, M.S., & Snow, C.E. (Eds.). (1998). *Preventing reading difficulties in young children*. National Academies Press.

Guitart, J.M. (1997). Variability, multilectalism, and the organization of phonology in Caribbean Spanish dialects. F. Martinez-Gil & A. Morales-Front (Eds.), *Issues in the phonology and morphology of the major Iberian languages*, 515-536.

Gutiérrez-Clellen, V.F. (1995). Narrative development and disorders in Spanish-speaking children: Implications for the bilingual interventionist. *Bilingual Speech-Language Pathology: An Hispanic Focus*, 97-128.

Hadley, P.A. (1998). Language sampling protocols for eliciting text-level discourse. *Language, Speech, and Hearing Services in Schools, 29*, 132-147.

Hammer, C.S., Lawrence F.R., & Miccio, A.W. (2007). Bilingual children's language Abilities and early reading outcomes in Head Start and kindergarten. *Language Speech and Hearing Services, 38*(3), 237-248.

Hammer, C.S., Lawrence, F.R., & Miccio, A.W. (2008). Exposure to English before and after entry into Head Start1: Bilingual children's receptive language growth in Spanish and English. *International Journal of Bilingual Education and Bilingualism, 11*(1), 30-56.

Hammer, C.S., Davison, M.D., Lawrence, F.R., & Miccio, A.W. (2009). *Scientific Studies of Reading, 13*(2), 99-121.

Heilmann, J., Miller, J.F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K.D. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders*, *28*(2), 178-188.

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.

Heilmann, J.J., & Westerveld, M.F. (2013). Bilingual language sample analysis: Considerations and technological advances. *Journal of Clinical Practice in Speech-Language Pathology*, 87.

Horton-Ikhard, R., Weismer, S.E., & Edwards, C. (2005). Examining the use of standard language production measures in the language samples of African-American toddlers. *Journal of Multilingual Communication Disorders, 3(3),* 169-182.

Hughes, D., McGillivray, L. & Schmidek, M. (1997) *Guide to Narrative Language: Procedures for Assessment.* Pro Ed.

Kayser, H. (1989). Speech and language assessment of Spanish-English speaking children. *Language, Speech, and Hearing Services in Schools*, *20*(3), 226-244.

Kester, E.S., & Gorman, B.K. (1996). Spanish-influenced English: Typical semantic and syntactic patterns of the English language learner. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.516.3449&rep=rep1&type=pdf

Klee, T., Stokes, S.F., Wong, A.M.Y., Fletcher, P., & Gavin, W.J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1396-1410.

Kohnert, K. (2013). *Language Disorders in Bilingual Children and Adults, Second Edition*. San Diego, California: Plural Publishing, Inc.

Leadholm, B.J. & Miller, J.F. (1992). *Language Sample Analysis: The Wisconsin Guide.*

Lever, R. & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology, 108*, 1-24.

Liles, B.Z. (1985). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Language Research, 28,* 123-133.

Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A Critical Review of the Literature. *Journal of Speech, Language, and Hearing Research*, *36*(5), 868-882.

Lucero, A. (2015). Cross-linguistic lexical, grammatical, and discourse performance on oral narrative retells among young Spanish speakers. *Child Development, 86(5)*, 1419-1433.

MacLachlan, B., & Chapman, R. (1988). Communication breakdowns in normal and language-learning disabled children's conversation and narration. *Journal of Speech and Hearing Disorders, 53*, 2-7.

Mainess, K.J., Champion, T.B., & McCabe, A. (2002). Telling the unknown story complex and explicit narration by African American preadolescents: Preliminary examination of gender and socioeconomic issues. *Linguistics and Education, 13*, 151-173.

Mayer, M. (1974). *Frog goes to dinner*. New York: Dial.

Merritt, D.D., & Liles, B.Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. *Journal of Speech and Hearing Research, 30*, 539-552.

Merritt, D.D., & Liles, B.Z. (1989). Narrative analysis clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders*, *54*(3), 438-447.

Miller, J.F., Andriacchi, K., & Nockerts, A. (2016). Using language sample analysis to assess spoken language production in adolescents. *Language, Speech, and Hearing Services in Schools,* 1-14.

Miller, J.F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D.J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice*, *21*(1), 30-43.

Miller, J., & Iglesias, A. (2008). Systematic analysis of language transcripts (SALT), research version 2008 [Computer software]: SALT Software.

Mills, M.T., Watkins, R.V., & Washington, J.A. (2013). Structural and dialectal characteristics of the fictional and personal narratives of school-age African American children. *Language, speech, and hearing services in schools*, *44*(2), 211-223.

Muñoz, M.L., Gillam, R.B., Peña, E.D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language Speech Hearing Services in Schools*, *34(4),* 332-342.

Oetting, J.B., & McDonald, J.L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*, 207-223.

Oetting, J.B., & McDonald, J.L. (2002). Methods for characterizing participants' nonmainstream dialect use in child language research. *Journal of Speech, Language, and Hearing Research*, *45*(3), 505-518.

Parker, M.D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First language*, *25*(3), 365-376.

Pearson, B.Z. (2002). *Narrative competence among monolingual and bilingual school children in Miami.*

Pearson, B.Z., Jackson, J.E., & Wu, H. (2014). Seeking a valid gold standard for an innovative, dialect-neutral language test. *Journal of Speech, Language, and Hearing Research*, *57*(2), 495-508.

Peña, E.D., Gillam, R.B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research*, *49*(5), 1037-1057.

Politzer, R.L., & Ramirez, A.G. (1973). An error analysis of the spoken English of Mexican-American pupils in a bilingual school and a monolingual school. *Language Learning*, *23*(1), 39-61.

Ripich, D.N., & Griffith, P.L. (1988). Narrative abilities of children with learning disabilities and nondisabled children: Story structure, cohesion, and propositions. *Journal of Learning Disabilities, 21*, 165-173.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Applied Psycholinguistics, 37*, 49-67.

Ross, S.H., Oetting, J.B., & Stapleton, B. (2004). Preterite had plus V-ed: A developmental narrative structure of African American English. *American Speech, 79*, 167-193.

SALT Software, LLC. SALT transcription conventions.

Scarborough, H.S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. *Handbook of Early Literacy Research,* 97-110.

Scheffner-Hammer, C., Pennock-Roman, M., Rzasa, S., & Tomblin, J.B. (2002). An analysis of the Test of Language Development—Primary for item bias. *American Journal of Speech-Language Pathology, 11,* 274-284.

Simon-Cereijido, G., & Gutiérrez-Clellen, V.F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, *30*(02), 315-337.

Slobin, D.I. (1973). Cognitive prerequisites for the development of grammar. *Studies of Child Language Development*, *1*, 75-208.

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.

Stalnaker, L.D., & Creaghead, N.A. (1982). An examination of language samples obtained under three experimental conditions. *Language, Speech, and Hearing Services in Schools*, *13*(2), 121-128.

Storch, S.A., & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental psychology*, *38*(6), 934.

U.S. Census Bureau. (2000). *America speaks: A demographic profile of foreign-language speakers for the United States: 2000*.

Washington, J.A., & Craig, H.K. (2002). Morphosyntactic forms of African American English used by young children and their caregivers. *Applied Psycholinguistics*, *23*(02), 209-231.

Wolfram, W. (1974). Sociolinguistic Aspects of Assimilation: Puerto Rican English in New York City.

Wren, C.T. (1985). Collecting language samples from children with syntax problems. *Language, Speech, and Hearing Services in Schools*, *16*(2), 83-102.

APPENDIX A

COMPILATION OF SPANISH-INFLUENCED ENGLISH UTTERANCES USED IN DDM ANALYSIS

	Form	Example
POS	Non-obligatory possessive 's	Miguel ate Mario tacos.
BOD	Article used with body parts instead of possessive	I cut the finger.
PNM	Use of post noun modifier	This is the homework of my brother.
PLU	Non-obligatory plural –s	The girl are playing.
IPLU	Irregular plurals marked with –s	The childrens play all day.
REG	Non-obligatory regular past –ed	I talk to her yesterday.
IRR	Irregular past tense verbs marked with –ed Regularization of irregular past	He sanged a song. He singed a song.
COM	More frequent use of longer comparative form	He is more tall; He is most tall.
SNG	Nonobligatory regular third person singular present tense –s	She eat too much.
ART	Omitted articles; omitted articles with nationality, profession, etc.	I am going to store. Is American; Is teacher.
NO	Use of "no" before the verb to indicate negation	She no eat candy.
DON	No used for don't in negative imperatives	No throw stones.
INV	No use of noun-verb inversion for questions	María is going?
DO	Nonobligatory do- insertion in questions	You like ice cream?
HAV	Occasional use of have for copula	I have ten years; I have hunger.
SUB	Omitted subject pronouns (when subject has been identified in the previous sentence)	Father is happy. Bought a new car.
FUT	Use of go + to for the future tense; omission of "will" in future	I go to dance; The boy play.
SDIS	Subject dislocation, before or after the verb phrase	And the dog he left. He was screaming, the kid.
PREP	Switched prepositions 'in' 'onto' 'into' 'up to' 'at' and 'on'	The food is in the table.
OF	Possessive markers	The book of my sister.
ADJ	Adjective following noun; in agreement in number	The dress yellow; the dresses yellows.
ING	Use of simple present where English uses progressive	He sleep now.
DEF	Titles used with definite article	The Mr. Jones.
COM	Noun-compounding not used or order in error	The wife-house; the wife of the house

NEG	Double negative	He didn't do nothing.
SVD	Subject-verb disagreement	The apples was coming down.
WO	Word order	A leaf he got.
COP	Copula omission	He in the water.

Note. SIEs compiled from Fiestas & Peña (2004), Kayser (1989), Wolfram (1974), Politzer & Ramirez (1973), Kester & Gorman (1996)

APPENDIX B

COMPILATION OF LESS-DOCUMENTED SPANISH-INFLUENCED ENGLISH UTTERANCES

	FORM	EXAMPLE
FOR	Use of "for"	For I have got.
DID	Use of "did"	You could did it to yours? He did like ohmygod.
ODIS	Object dislocation	The man (obj.) he pinched him.

Note. SIEs compiled from Politzer & Ramirez (1973)