WORKING MEMORY AND SYNTACTIC PROCESSING IN BILINGUAL AND MONOLIGUAL CHILDREN

by

Carla I. Orellana

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:

Ronald Gillam, Ph.D. Major Professor	Sandra Gillam, Ph.D. Committee Member
Sonia Manuel-Dupont, Ph.D. Committee Member	Kathleen A. J. Mohr, Ed.D. Committee Member
Sarah Schwartz, Ph.D. Committee Member	Richard S. Inouye, Ph.D. Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2019

ABSTRACT

Working Memory and Syntactic Processing in Bilingual and Monolingual Children

by

Carla I. Orellana, Doctor of Philosophy

Utah State University, 2019

Major Professor: Dr. Ronald Gillam
Department: Communicative Disorders and Deaf Education

The purpose of this study was to examine the relationship between complex auditory working memory, syntactic knowledge, and canonical and noncanonical sentence comprehension in bilingual and monolingual children using both offline (behavioral) and online (eye-tracking) measures. There were 19 children in the monolingual group and 19 children in the bilingual group with an average age of 11 years. The children listened to four different sentence types while looking at a screen with three images representing the three nouns in the sentence. The children were instructed to select the agent of the sentence. Their eye movements were recorded as they completed this task. The four sentence types were: subject verb object (SVO), subject relative (SR), passive (PAS), and object relative (OR). Both groups of children had better sentence comprehension accuracy of SVO and SR sentences than PAS and OR sentences. Children with higher working memory tended to obtain better scores than children with lower working memory. This effect was strongest in the PAS and OR sentences.

Additionally, for PAS and OR sentences, bilingual children with similar levels of working memory as the monolingual children obtained lower scores of sentence comprehension. For both groups, children with higher working memory were slower to respond than children with lower working memory, especially when they chose incorrectly. Bilingual children tended to select the agent more quickly than monolinguals. Children with high working memory focused on the agent less than children with low working memory. Bilingual children had mixed results relating to their focus of attention.

(116 pages)

PUBLIC ABSTRACT

Working Memory and Syntactic Processing in Bilingual and Monolingual Children

Carla I. Orellana

The purpose of this study was to examine the relationship between complex auditory working memory, syntactic knowledge, and complex sentence comprehension in bilingual and monolingual children using both offline (behavioral) and online (eye-tracking) measures. There were 19 children in the monolingual group and 19 children in the bilingual group with an average age of 11 years. The children listened to sentences, while looking at a screen with three images of the three nouns in the sentence. They were instructed to select the doer of the action (agent). Their eye movements were recorded as they completed this task. The four sentence types were: subject verb object (SVO), subject relative (SR), passive (PAS), and object relative (OR). Both groups of children had better sentence comprehension accuracy of SVO and SR sentences than PAS and OR sentences. Children with higher working memory tended to obtain better scores than children with lower working memory. This effect was strongest in the PAS and OR sentences. Additionally, for PAS and OR sentences, bilingual children with similar levels of working memory as the monolingual children obtained lower scores of sentence comprehension. Children with high working memory were slower to respond. Bilingual children selected the answers more quickly than the monolingual children. Children with high working memory focused on the agent less than children with low working memory. Bilingual children had mixed results relating to their focus of attention.

ACKNOWLEDGMENTS

I would like to thank Dr. Ron Gillam for including me on the team of investigators who collected data on language and literacy in monolingual and bilingual children, funded by an internal SPARC grant from Utah State University. I would especially like to thank my committee members, Drs. Sandi Gillam, Sonia Manuel Dupont, Kit Mohr, and Sarah Schwartz, for their support and assistance throughout the entire process.

I give special thanks to my family, friends, and colleagues for their encouragement, moral support, and patience as I worked my way from the initial proposal writing to this final document. I could not have done it without all of you.

Carla I. Orellana

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Difficulty with sentence comprehension is one of the hallmark deficits of children with developmental language disorders (DLD) (Adams, 1990; Bishop, Bright, James, Bishop, & Van Der Lely, 2000; Norbury, Bishop, & Briscoe, 2002; van der Lely, 1996). Children with DLD tend to understand simple active sentences but have difficulty with complex sentences (Bishop et al. 2000; Montgomery & Evans, 2009). Bilingual children with DLD would be expected to have language difficulties in both languages as opposed to just one language (Kohnert, 2010). Diagnostic tools of DLD in Spanish-English bilinguals have been designed around specific skills, such as narrative production, morphosyntactic productions, and vocabulary and word-learning (Dollaghan & Horner, 2011) that have been found to be informative as clinical markers. However, there is a need for more research relating to the sentence comprehension of bilinguals with DLD (Gutiérrez-Clellen, Restrepo, & Simón-Cereijido, 2006). To inform clinical decisions about assessing and treating sentence comprehension difficulties in children with DLD, we first need to understand sentence comprehension processes in monolingual and bilingual children who are developing typically. During auditory sentence comprehension, a listener must derive meaning from a fleeting auditory signal. The listener creates a mental model of the sentence by recognizing the words in the sentence and assigning meaning to the syntactic and semantic relationships of these words. These relationships can be determined by constraints that are defined by the grammar of the language, which is accessed from long-term memory (LTM). Morphosyntactic and contextual information available in an auditory signal also provides clues about the relevant semantic and syntactic relationships. However, once an initial meaning representation of a sentence is generated, memory of the specific sounds and words in the sentence begins to decay and is eventually lost (Sachs, 1967) unless the information is held in an active state and/or updated by succeeding comprehension processes.

Sentence Comprehension

One model of sentence comprehension, known as *chunk-and-pass* processing is proposed to explain how the language system deals with what Christiansen and Chater, (2016a) and Christiansen and Chater (2016b) call the *Now-or-Never* bottleneck. This bottleneck occurs because listeners must make meaning from a deluge of incoming information very quickly due to the fleeting nature of memory and speed of oral communication. When listening to a sentence, a person may encode the auditory signal into phonemes. As an example, the sentence, "*The elephant stepped on the vehicle*," contains about 23 phonemes. Once the auditory input has passed, it can no longer be recovered. In order to process these 23 phonemes, the brain engages in *chunk-and-pass* processing. Phonemes from the speech signal are recoded into chunks of a more meaningful abstract level. These chunks could be syllables, morphemes, or word concepts. Words are compressed or further chunked into phrases. Any information that is not recoded will be forgotten. This process continues and can be taken up as far as discourse-level abstractions. An interesting component of *chunk-and-pass* modeling is that once items are chunked and passed to a higher level of abstraction, the chunk is at minimal risk of interference from subsequent items at the lower level. That is, if phonemes are chunked into words, incoming phonemes are unlikely to interfere with the word. Similarly, when words are chunked into phrases and sentences, that information, which is activated in LTM, is minimally susceptible to interference from additional, incoming information. That makes it possible for listeners to retain information from previously heard words as more words come in, increasing the amount of information that can be comprehended.

An additional component of *chunk-and-pass* processing is *anticipation,* in which the brain uses prior knowledge to recode information more quickly by anticipating or predicting future input (Christiansen & Chater; 2016a, 2016b). Having more knowledge and familiarity with specific verbs would result in greater probability of accurately predicting the subsequent noun, allowing for more efficient processing. Chunking also occurs incrementally and information is only processed in parallel to the extent that conflicts in encoding are resolved. That is, information cannot be chunked into higher levels of abstraction unless encoding conflicts are resolved. Christiansen and Chater propose that in typical language use there are sufficient clues in the environment to resolve such conflicts.

This model has important implications for bilingual children who may have semantic and syntactic knowledge in their first language (L1) that may not be readily available in their second language (L2). When listening to sentences in L2, these children may not automatically chunk vocabulary and sentence structures in their second language. Instead, they may allocate information-processing skills to translating the meaning into L1. Furthermore, for a child with limited syntactic knowledge in L2, lexical and syntactic anticipation is less likely to occur. Therefore, the chunking process will not be as efficient or quick, costing processing time. These extra cognitive processes could make incoming information vulnerable to interference from succeeding phonemes and words. If listeners are busy trying to figure out the meaning of incoming words rather than automatically chunking word meanings into phrases and sentences, that extra mental processing could increase the interference effect of L1. Interference could interrupt the ability to store information in LTM. We would expect bilingual children with lower working memory capacity to be more susceptible to this interference because they would not be able to hold as much information in a state of activation, resulting in decreased sentence comprehension. As a result, we would expect L1 sentence comprehension to be better than L2 sentence comprehension for bilingual children with lower WM capacity.

Another model of sentence comprehension, the *good-enough model* of sentence processing (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007; Karimi & Ferreira, 2016) suggests that the early representations created while interpreting sentences are typically shallow and incomplete. The listener's linguistic representations are likely to be underspecified and "good-enough" for the moment. These representations only become more specific or elaborated as a function of additional input. An example of *good-enough* processing is the *Moses illusion*, in which readers fail to notice the inconsistency when asked, "How many animals of each kind did Moses take on the Ark?" (Erickson & Mattson, 1981). A response of "two," indicates that the reader did not notice the substitution of Moses for Noah in the question. Because of the many shared semantic features between the characters, Moses is shallowly processed and makes a good-enough representation for Noah. The *chunk-and-pass* processing model has some basic similarities to the *good-enough* processing model. However, an important difference is that *chunk-and-pass* also emphasizes the need to get it right the first time because errors due to underspecification will increase the processing demand and time, putting the information at risk to interference from subsequent incoming auditory input.

These models of sentence comprehension were developed to explain comprehension processes in adults. Montgomery, Evans, Fargo, Schwartz, and Gillam (2018) and Gillam, Montgomery, Evans, and Gillam (2019) created a model of sentence comprehension for monolingual children. Montgomery and colleagues (2018) administered a variety of cognitive and linguistic tasks to 234 children between the ages 7-11 (117 children with DLD and 117 with typical language development). They then used confirmatory factor analysis to select the smallest number of latent variables (groups of measurements representing a construct) that represented the cognitive processes that were critical for comprehending canonical and noncanonical sentences. The four constructs that represented independent variance were: 1) fluid reasoning, 2) controlled attention, 3) complex working memory, and 4) language knowledge in long-term memory. Montgomery et al. then used structural equation modeling (SEM) to evaluate the nature of the relationships between sentence comprehension and the four cognitive processes of interest. The resulting model, referred to as the GEM model (Gillam-Evans-Montgomery model), makes specific predictions about the nature and extent of the structural relationships between cognitive processing and linguistic knowledge and their effects on sentence comprehension. Specifically, the GEM model proposes that working memory plays an important role in mediating the relationship between fluid reasoning and language knowledge for sentence comprehension. These authors believe that working memory may be more important for sentence comprehension in children than in adults because children are in the process of learning complex syntax. Because the organization of syntactic information in LTM is less well established in children, they may need to rely on working memory to support sentence comprehension to a greater extent than adults. I will elaborate on this notion later in this chapter.

Working Memory

Working memory is the retention of a small amount of information in a readily accessible form that facilitates planning, comprehension, reasoning, and problem solving (Cowan, 2014). According to Cowan, working memory is comprised of two critical components: the focus of attention and activated long-term memory. The focus of attention has a limited capacity of three to five meaningful items in adults (Cowan, 2001). Items held in the focus of attention are resistant to interference or forgetting. As incoming information is processed and integrated, items are chunked in long-term memory, allowing for additional information to be held in the focus of attention.

For example, nine items (dog, cat, bird, horse, pig, goat, car, bus, bike) can be chunked into three groups, (dog, cat, bird), (horse, pig, cow), and (car, bus, bike), that are easily held in the focus of attention. Unlike the focus of attention, activated long-term memory is not capacity limited. Instead, it is time limited (Cowan, 1999). Information that has already been processed, but is no longer in the focus of attention, remains in activated long-term memory for a longer period of time. Even though much more information can be held in this long-term activated memory, it is less prone to decay and/or interference, especially when it has been organized in a meaningful way. Once this information has been sufficiently integrated with prior knowledge, it can be offloaded into long-term memory, where it is less prone to interference.

Cowan’s model of working memory relates to both the *chunk-and-pass* and the *good-enough* models of sentence comprehension. Both Cowan’s model of working memory and *chunk-and-pass* have a similar concept of clustering smaller units of information into chunks that resist interference and enhance of recall. The focus of attention in working memory most closely relates to the encoding portion of *chunk-and-pass* because they both are time limited. In Cowan’s model, chunked information goes into activated long-term memory where it is at minimal risk of interference and decay, which is consistent with the *chunk-and-pass*’s proposal that chunked items are at less risk of interference. The *chunk-and-pass* model specifies that chunks at different levels of abstraction should not interfere with each other.

It may be helpful to think of chunks as meaningful abstract concepts. For example, once phonemes are encoded or chunked into words, the chunk becomes the abstract concept of that word. Similarly, the words can be chunked into a sentence of some abstract concept. One may recall the example sentence from earlier as, “*The elephant squashed the car*.” Note that the concept of the sentence was retained, but that the smaller units of information (the individual words) decayed and were not retained. When the specific words are part of the focus of attention, as proposed by Cowan, they are susceptible to decay. But once their higher meaning had been extracted and integrated with prior knowledge, as proposed in the *chunk-and-pass* model, they are less susceptible to interference and can be available for activation. Additionally, we can see how the final concept recalled is “*good-enough*” to represent the original meaning.

Both Cowan’s model of working memory and *chunk-and-pass* allow for decay. I envision these two models overlapping with encoding of chunks occurring in the focus of attention and chunks being held in activated long-term memory until they are needed again for encoding and subsequent chunking, such that items and chunks are moving in and out of the focus of attention. While linguistic information is held in activated long-term memory, incoming speech signals can be encoded in the focus of attention into a chunk and then this chunk is offloaded into activated long-term memory.

Because bilinguals may have less language experience in their L2, they may not have the prior knowledge about syntactic structures necessary to *anticipate* the information to follow, thus they would not get the facilitation effects for more efficient encoding as a person with more language experience. As a result of inefficient chunking, they may have to hold more smaller units of meaning. A bilingual with smaller working memory capacity, would be taxed and perhaps unable to synthesize the correct final concept. Thus, both syntactic knowledge and working memory play a role in comprehension.

In the next section, I discuss studies that explore the relationship between syntactic knowledge and sentence comprehension. I also discuss how working memory relates to sentence processing in monolinguals. Finally, I discuss how bilinguals and their diverse language experiences relate to sentence comprehension.

Working Memory, Syntactic Knowledge, and Sentence Comprehension

In the preceding section, I made the case that WM and LTM play important roles in sentence comprehension. Recall that once information is integrated into LTM, it is less susceptible to interference or decay, resulting in greater retention of information. We know that semantic and syntactic knowledge are often well-established in LTM. The following section explores the ways in which semantic and syntactic knowledge may play critical organizing roles in sentence comprehension.

The relative importance of WM, syntactic knowledge, and semantic knowledge for sentence comprehension have been studied by comparing participants' ability to identify the agent and/or patient of either canonical or noncanonical utterances (syntactic knowledge) that are either plausible or implausible (semantic knowledge). The comparison of active and passive forms is important because WM plays a different role in comprehension when the canonical order of English (the first noun as the agent) is maintained compared to when it is reversed, as in noncanonical passive sentences, in which the first noun is the patient. To comprehend a passive sentence (e.g., *The cheese was eaten by the mouse*), participants must hold the first noun in an active state in WM until it is clear that it is the patient of the action (the second noun). Plausibility is important because it can facilitate sentence comprehension through the facilitation effect of *anticipation*. In the previous example, existing semantic knowledge about cheese and mice would allow a person to predict *mouse* as the upcoming word for quicker encoding and comprehension. An implausible sentence (e.g., *The cheese was eaten by the chair*), would not be consistent with existing semantic knowledge. Therefore, comprehension of such a sentence would rely on syntactic knowledge and any faciliatory anticipation effect would depend on whether there was sufficient linguistic experience and syntactic knowledge of that structure.

To determine whether listeners would maintain misinterpretations of sentences heard, Christianson, Hollingworth, Halliwell, and Ferreira (2001) presented monolingual adults with temporarily ambiguous sentences (presented visually) that would initially elicit an incorrect interpretation, followed by a correct interpretation after reanalysis. An example of such a sentence is: *While Anna dressed the baby spit up on the bed*. A misinterpretation would interpret *baby* as the object of the verb *dressed*, whereas upon completing reading of the sentence, the *baby* must be interpreted as the subject of the sentence for correct reanalysis. If the misinterpretation persists despite correct reanalysis, then participants should answer yes to both of the following questions: *Did Anna dress the deer?* and *Did the baby spit up on the bed?* Consistent with the tenets of the *good-enough* model, participants answered yes to both questions for 57.3% of the ambiguous sentences compared to 11.5% of unambiguous sentences, indicating that misinterpretations did persist even after correct reanalysis. These results may represent a priming effect from the questions asked after reading the sentences. That is, because the questions were a forced-choice paradigm, it is possible that participants accepted misinterpretations only once they were forced to reevaluate the sentence in the manner suggested by the question. It is also possible that participants did not have a final interpretation until the question was asked, priming them with the concept supplied in the question.

Patson, Darowski, Moon, and Ferreira (2009) conducted a follow-up study with a similar design. However, instead of asking the yes/no questions, they asked participants to paraphrase the sentence. In their paraphrases, the participants tended to include two possible interpretations. For example, the sentence, *While Anna dressed the baby spit up on the bed*. may have been paraphrased as, *Anna dressed the baby and it spit up on the bed*. The results were similar to Christianson et al. (2001), in that participants persisted with the misinterpretation (Anna dressed the baby rather than herself) despite also arriving at a correct interpretation that the baby spit up on the bed. The authors argue that this supports the *good-enough* model of sentence processing. I believe that participants who maintained both representations were able to encode two chunks or meanings but did not engage in the next step of resolving the encoding conflict. This could have occurred for two reasons: participants did not activate the long-term knowledge necessary to notice that there was a conflict or participants were unable to resolve the conflict. Some participants were able to fully reanalyze the sentence and arrived at the correct interpretation. Because this sentence structure is syntactically not plausible, working memory limitations may have contributed to the high proportion of incomplete conflict resolution.

Ferreira (2003) examined adults' ability to interpret canonical (active) and noncanonical (passive) sentences that contained either plausible agent/patient relationships or implausible agent/patient relationships. Participants were instructed to identify the agent and patient of sentences that were heard. There were plausible reversible sentences (e.g., *the dog bit the man*) in which the first noun (the agent) and the second noun (the patient) were animate nouns that could play either role in the sentence, plausible reversible sentences (e.g., *the man bit the dog*) in which the agent was unlikely to have done the action to the patient, plausible non-reversible sentences (e.g., *the mouse ate the cheese*) in which only the agent was an animate noun, implausible non-reversible sentences (e.g., *the cheese ate the mouse*) in which the agent was an inanimate noun, and symmetrical sentences (e.g., *the woman visited the man, the man visited the woman*) in which exact agents and patients are reversed. All the sentences were presented in active (SVO) and passive forms. The participants in this study performed more poorly on passives than active sentences and even more poorly on implausible passive sentences than plausible passive sentences. During misinterpretations, world knowledge and thus semantic relationships were used to determine meaning rather than syntactic knowledge. Thus, WM, knowledge of syntax in LTM (especially knowledge of word order), and knowledge of plausible subject-verb relationships all play a role in sentence comprehension.

Traxler (2007) extended the study of syntactic and semantic knowledge to relative clauses and plausibility by having 96 native adult speakers of English read three types of sentences while recording their eye movements. Inanimate objects were used to control for plausibility. In the first two types of sentences, the relative clause attachment was either to the first noun, "*The writer of the letter/ that had/ blonde hair/ arrived this/ morning*." or the second noun, "*The letter of the writer/ that had/ blonde hair/ arrived this/ morning.*" The third sentence type (e.g., *The sister of the writer/ that had/ blonde hair/ arrived this/ morning.)* was completely ambiguous and the relative clause could be attached to either noun 1 or 2. Traxler found that participants had more difficulty (as indicated by longer reading times) with the unambiguous sentences than the ambiguous sentence types, although there was no difference between the two unambiguous sentence types. Additionally, participants' working memory capacity did not moderate online processing performance (all ts < 1.35, all ps $< .18$). However, working memory appeared to affect noun attachment preference, such that increases in working memory increased preference for noun 1 attachment rather than noun 2 attachment in the unambiguous sentences. This suggests that individuals with higher working memory had more resources to maintain the more distant Noun 1 active in memory, whereas individuals with less working memory were limited in this resource, therefore preferring the more local Noun 2. Traxler also suggests the possibility that participants with higher working memory also had more knowledge from reading experience resulting in the expectation of Noun 1 attachment, which resulted in more difficulty and longer reading times of the Noun 2 attachments. Another possibility is participants were required to hold and compare two interpretations in mind in order to resolve the ambiguity. Again, for individuals with lower working memory, this could be more taxing resulting in underspecification of sentence meanings. The role of LTM here is speculative because it was not measured in this study.

The studies I have discussed thus far concerned sentence comprehension in adults. However, children’s comprehension of sentences varies by age and type of sentence (Montgomery, Evans, Gillam, Sergeev, & Finney, 2016; Montgomery, Gillam, Evans, & Sergeev, 2017). Montgomery et al. (2016) examined typically developing monolingual children’s ability to comprehend different types of aurally presented sentences. The purpose of the study was to evaluate word-order sensitivities in children of varying ages. Specifically, they examined the children's understanding of the agent-patient relationship in canonical and noncanonical sentences using semantically implausible sentences (e.g., *The chair that the bread had splashed under the square was new*). In order to isolate the children's use of syntactic knowledge for interpretation, the researchers used inanimate objects, which removed semantic cues and thus probability cues. They asked participants to identify the agent in canonical (SVO and object relative) and noncanonical (passive and subject relative) sentences. They found that older children (mean age of 10;8) outperformed younger children (mean age of 8;1) on all sentence types presented. Both groups of children performed better on the canonical sentence types than on the noncanonical sentence types. However, unlike the older group, the younger children performed more poorly on noncanonical sentences with the object relative clause than on noncanonical sentences that did not contain a relative clause.

Montgomery et al., (2018) expanded on this study by assessing the structural relationships between several constructs and sentence comprehension. Using the task from Montgomery et al. (2016), they measured the sentence comprehension of 117 typically developing monolingual children (mean age of 9.5). Additionally, they grouped pairs of correlated measures to represent these various constructs. These latent variables, created to minimize the measurement error of each construct, included: fluid reasoning, controlled attention, phonological short-term memory, processing speed, complex working memory, and language knowledge in long-term memory. They utilized confirmatory factor analysis to determine the minimal set of variables that best represented the data on children's sentence comprehension. The resultant four latent variables were fluid reasoning, controlled attention, complex working memory, and language knowledge in long-term memory. Subsequently, they utilized structural equation modeling to assess the direct and indirect relationships of those constructs. They found that working memory mediated the effects of fluid reasoning and language knowledge in long-term memory on sentence comprehension, but not controlled attention. The findings indicate that working memory functioned as the underlying mechanism through which fluid reasoning and language knowledge in long-term memory indirectly facilitated the comprehension of the canonical and noncanonical sentences.

Together, these studies exemplify the interplay between long-term memory knowledge of syntax (word order) and plausibility. Participants tend to use and rely on semantic knowledge and the plausibility or semantic-syntactic relationships in sentences to facilitate comprehension of both canonical and noncanonical sentences. Both children and adults utilize long-term memory knowledge and working memory work together to comprehend sentences. Next, we will see how bilinguals, who tend to have varied language experiences use their knowledge of two languages to comprehend sentences.

Bilingual Sentence Comprehension

Syntactic knowledge, especially knowledge of grammatical constraints within a language, plays a critical role in all models of sentence comprehension. This may be especially true for bilinguals, particularly in cases in which the grammar of L1 and L2 do not correspond closely. In English, an example of a grammatical constraint is that a sentence must contain a subject (e.g., *I* in *I kicked the ball*). Thus, a sentence such as, "*Kicked the ball."* is ungrammatical. Spanish, a pro-drop language, does not have this constraint, making the subject optional. Thus, "*Pateé la pelota*." is grammatically correct in Spanish because the verb *pateé* contains information or cues about the subject, which allows the listener to determine who kicked the ball.

The English example sentence above also exemplifies the canonical word order of English, subject-verb-object (SVO). This type of word order lends itself well to the idea that sentences are comprehended in serial order. Even when adding a relative clause to the subject (e.g., *The boy, who was wearing a red shirt, kicked the ball.*), the sentence maintains its canonical word order and the first noun would be correctly identified as the agent of the sentence. However, noncanonical sentences in English (such as passives and object relatives) cannot be interpreted in serial order for correct interpretation of the sentence. For example, interpreting the first noun as the agent of the passive sentence, "*The girl was seen by the boy."* or the object-relative sentence, "*The girl that the boy saw was happy."* would lead to an incorrect interpretation that it was the girl that did the seeing. Additionally, Spanish verbs carry additional information about the subject, which allows for greater word order flexibility. Spanish word orders include SVO, VOS, OSV, SOV, OVS, and VSO (Lahousse & Lamiroy, 2012).

There have been a number of studies regarding the manner in which syntactic knowledge in one language affects a second language. Much of what is known about bilingual sentence processing is based on studies of adults who acquired a second language either after puberty or during adulthood. In such cases, it is clear that bilinguals experience either linguistic interference or linguistic transfer from one language when performing specific language tasks in the other language, depending on the type of task performed as well as other factors relating to their bilingualism (e.g., language proficiency, dominance, age of acquisition).

Some of these studies examined more closely the role that semantic and syntactic relationships play in sentence comprehension. Controlling for cues (i.e., noun verb agreement, animacy, and word order), Hernandez, Bates, and Avila (1994) found that English monolinguals demonstrated faster sentence comprehension when sentences followed an SVO pattern (followed by noun verb agreement and animacy), whereas Spanish monolinguals were faster for sentences with noun verb agreement (followed by animacy and word order). The Spanish-English bilinguals, while similar to both monolingual groups, showed more sensitivity to word order than the monolingual Spanish group, indicating influence of the second language (English) on first language (Spanish) processing. In contrast, Kilborn (1989) found that the participants' first language influenced processing of their second language. It is noteworthy to mention that the bilingual participants in Killborn's (1989) study were more dominant in their first language, whereas the Hernández et al. (1994) participants were more dominant in their second language, exemplifying the complexity of factors contributing to bilingual sentence processing.

Morett and Macwhinney (2013) explored the issues of syntactic knowledge and language dominance by having native English speakers with Spanish as a second language complete sentence interpretation tasks with varying levels of cues (i.e., common to both languages, English-specific, and Spanish-specific). Less advanced learners of Spanish relied less heavily on animacy than the more advanced learners. Both groups were approaching native-like interpretations, but results of latencies to selection showed increased time for less advanced learners, indicating more processing time for cues available in both languages. There was some evidence of transfer from second language to first language, though not as strong as in Hernández et al. (1994).

Other studies (Dussias, 2003; Dussias & Sagarra, 2007; Fernández, 2003) have explored daily exposure to a language as it relates to sentence processing in bilinguals, again with similar findings of cross-linguistic transference as the dominance studies (i.e., more exposure to L2 relates to greater L2-like syntactic parsing). Specifically, Dussias and Sagarra (2007) recruited native Spanish speakers with extensive English exposure, native Spanish speakers with limited English exposure, and functionally monolingual

Spanish speakers. They presented participants with temporarily ambiguous sentences containing a relative clause, which attached either to the first or second noun phrase, as in the following examples:

> Noun phrase 1 (NP1) attachment: *El policía arrestó a la hermana del criado que estaba enferma desde hacía tiempo.* [The police arrested the sister of the (male) servant who had been ill (fem) for a while.]

> Noun phrase 2 (NP2) attachment: *El policía arrestó al hermano de la niñera que estaba enferma desde hacía tiempo*. [The police arrested the brother of the (female) babysitter who had been ill (fem) for a while.]

Eye measurements were recorded for each sentence read and the fixation time was extracted for the critical juncture. The critical juncture was the adjective in the relative clause because it contained the gender cue necessary to disambiguate the sentence. Monolingual speakers of Spanish had slower reading times for NP2 attachment than NP1 attachment. Bilinguals with limited exposure to English also had slower reading times for NP2 than NP1 attachment, whereas those with extensive English exposure had faster reading times for NP2 than NP1 attachment. These results are similar to those of Fernández (2003), who found that Spanish speakers prefer NP1 attachment with longer relative clauses.

One study that examined syntactic processing in both adults and children was conducted by Jasinska and Petitto (2013). These authors administered a syntactic judgment task and measured neural processes using functional near infrared spectroscopy (fNIRS), a neuroimaging technology that indirectly assesses neural activity by measuring changes in oxygen levels in the blood vessels of the brain. Jasinska and Petitto recruited both children (ages 7-10) and adults with the purpose of determining: 1) if there were differences in the neural activation patterns in the developing monolingual and bilingual brain during language processing tasks; 2) if there were similar or different patterns of activation between early-exposed bilingual learners and later-exposed bilingual learners; and 3) whether bilingualism is mostly a language-specific activity or cognitive-general activity. They studied three groups of children: monolingual, early-exposed bilinguals (from birth), and later-exposed bilinguals (ages 4-6); and two groups of adults: monolingual and early-exposed bilingual adults. Their bilinguals all spoke English and one another language (e.g., Russian, Spanish, Arabic, Cantonese). Jasinska and Petitto administered a grammatical judgment task with four types of relative clause sentences: OS plausible (e.g., *The light-house guided the sailor that piloted the boat*), OS implausible (e.g., **The sailor guided the light-house that piloted the boat*), SO plausible (e.g., *The sailor that the light-house guided piloted the boat*) and SO implausible (e.g., **The light-house that the sailor guided piloted the boat*). Syntactic processing was measured in the native language (English) only; the second language was not assessed.

Behaviorally, reaction times and accuracy effects were evident between age groups and between the two sentence types, OS and SO. The adults were faster than the children and all participants were faster on OS vs SO. There were no significant differences in response time or accuracy between monolingual and bilingual groups or between later-exposed and early-exposed bilinguals. However, greater neural activation was seen in later-exposed bilinguals compared to early-exposed bilinguals and monolinguals. Neural activation was greater for SO vs OS for both adults and children, with no difference between bilinguals or monolinguals. Though no difference in activation was present between monolinguals or bilinguals, there was an interaction of sentence type and age of acquisition. Later-exposed bilinguals showed greater activation in the dorsolateral prefrontal cortex (DLPFC), left inferior frontal gyrus (LIFG), and supratemporal gyrus (STG). There was also a main effect of language group on whole-brain activation, such that bilinguals activated more than monolinguals, and later-exposed bilinguals had greater changes in hemoglobin concentration than early-exposed bilinguals. Additionally, children showed greater activation in the medial temporal gyrus (MTG) vs the LIFG compared to the adults, indicating that syntactic processing is continuing to mature in the children. Also notable, was the finding that despite no significant differences in the accuracy of comprehension, neural activity was significantly different between the different groups, indicating that there were processing differences even with similar outcomes. This study is unique in addressing various ages of acquisition and for including children, although there was no later-exposed bilingual adult group to compare.

Gutiérrez-Clellen, Calderón, and Weismer (2004) investigated the verbal working memory ability in bilingual children with varying levels of proficiency. Their goal was to determine whether language experience affected performance on a working memory task and whether there were cross-linguistic effects. They recruited 44 bilingual children (average age of 8 years) with typical development and divided them into three groups: children proficient in both English and Spanish (n = 22), children with limited English proficiency (n = 11), and children with limited Spanish proficiency (n = 11). These children all completed a listening span task, administered in English and Spanish, known as the *Competing Language Processing Task* (CLPT; Gaulin & Campbell, 1994). This measure, which was adapted from the reading span task by Daneman and Carpenter (1980), required children to listen to groups of very simple sentences, make judgments about the truthfulness of each sentence immediately after it was presented (the comprehension portion of the task), and then recall the last word in each sentence after a group of sentences had been administered. Seven items (4 in English, 3 in Spanish) were removed from the analysis because the majority of children consistently missed those items. The children also completed the Dual Processing Comprehension Task (DPCT), in which the children reenacted sentences heard simultaneously in each ear. They found no significant group differences between the children proficient in English and those proficient in both Spanish and English on the English DPCT and on the English CLPT in either the recall or comprehension portion of the task. Similarly, there were no significant group differences between the children proficient in Spanish and those proficient in both languages on their performance on the Spanish DPCT and Spanish CLPT in either the recall or comprehension portions of the task. Note that the children who had limited proficiency were not compared to either of the two language-proficient groups on either version of the tasks. Thus, we do not know how limited proficiency impacted performance on either of the tasks. Finally, within the group of children proficient in both languages, there was no significant difference between performance on the Spanish and English versions of the tasks. Language experience did not appear to influence performance on this working memory task. The Spanish CLPT was moderately correlated with the English CLPT (r = .44, p =.03) and the Spanish DCPT was also moderately correlated with the English DPCT (r = .48, p = .02). Though the language of the tasks were correlated with each other, the patterns were different across languages. The Spanish DPCT was highly correlated with the Spanish CLPT ($r = -.70$, $p < .0001$). However, the English DPCT was not correlated with the English CLPT ($r = -.31$, $p > .05$). The authors concluded that their results did not support that bilinguals have enhanced, reduced, or increased control of processing and that the lack of correlation between the two English tasks suggests that performance on measures of verbal working memory is not independent of language skill.

There have been very few studies of sentence processing in children, and none found that focused specifically on bilingual Spanish-English children. As noted above, (Montgomery, et al. (2018) and Gillam et al. (2019) found that the effects of fluid reasoning, controlled attention, and language knowledge in long-term memory on sentence comprehension were mediated by working memory. These outcomes indicate that, for monolingual children, working memory likely functioned as the underlying mechanism through which fluid reasoning, controlled attention, and language knowledge in long-term memory indirectly affected the comprehension of the sentences.

However, the GEM model, as written, entails at least three limitations. One limitation is that it used a global measure of language knowledge in long-term memory rather than a specific measure of syntactic knowledge. Language knowledge was represented in the model as a latent variable comprised of the comprehension and production portions of the Test of Narrative Language (Gillam & Pearson, 2004). During the comprehension portion of this assessment, children answered explicit and implicit comprehension questions after listening to three different narrative scripts. For the production portion of this task, children produced three narratives: a narrative retelling of a single scene, a narrative produced from a sequence of pictures, and a fictional narrative produced from a single scene. This assessment can be considered a global measure of language knowledge because it involves comprehension and production of morphosyntactic, semantic, and discourse elements of language. However, the syntactic elements of the assessment cannot be parsed out to determine the child's knowledge of syntax specifically.

A second limitation of the GEM model is that it has been applied only to monolingual children. Bilingual children were not included in the development of the GEM model and thus its applicability to bilingual children is unknown. With 22.5 % of children between the ages of five estimated to speak a language other than English in the United States (U.S. Census Bureau, 2017), it is important to make efforts to understand sentence processing in this population as well.

The third limitation of the GEM model is that it was based only on behavioral data. To truly make predictions about the underlying processes of sentence comprehension, a combination of online processing measures and offline measures (e.g., behavioral measures) is needed. Offline measures of comprehension, including responses to a question or forced-choice selections, occur well after a sentence has been processed. The importance of offline measures for research is clear, as most measures of language are offline measures of processing. However, online measures have the potential to provide additional information about sentence processing, as such data are collected throughout the duration of a sentence and up to the point of the offline observation.

The use of eye tracking is one method of collecting online data during the course of sentence processing. Recording the eye's gaze during stimulus presentation provides instantaneous and continuous reflection of processing demands and attention allocation. Eye tracking has been used to measure cognitive load during processing (Qian, Garnsey, & Christianson, 2018; Schluroff, 1982) and attentional allocation (Cooper, 1974). Using both offline and online measures opens the possibility of exploring differences in processing even when outcomes are similar.

Our understanding of bilingual and monolingual children's sentence comprehension is absent a model that describes the structural relationship between syntactic knowledge, working memory, and canonical and noncanonical sentence comprehension using both online and offline processing measures. A better understanding of the nature of the relationship between syntactic knowledge, complex working memory, and complex sentence comprehension in monolingual and bilingual children could provide preliminary information about the extent to which the GEM model holds for the relationship between these three constructs and the extent to which it applies to both monolingual and bilingual children.

The purpose of this study was to examine the relationship between complex auditory working memory, syntactic knowledge, and canonical and noncanonical sentence comprehension in bilingual and monolingual children using both offline (behavioral) and online (eye-tracking) measures.

Research Questions

1. What is the nature of the relationship between complex auditory working memory and canonical and noncanonical sentence processing of English sentences in monolingual and bilingual children?

a. To what extent does complex working memory account for response accuracy (selecting the correct agent) and response-time measures of canonical and noncanonical comprehension of English sentences by monolingual and bilingual children when controlling for English syntactic knowledge?

b. To what extent does complex working memory account for the time spent looking at pictures representing the agent, the patient, and the location of the action in canonical and noncanonical sentences by monolingual and bilingual children when controlling for English syntactic knowledge?

2. Among bilinguals, what is the additional role of Spanish syntactic knowledge in long-term memory on comprehension of English canonical and noncanonical sentences?

   a. To what extent does complex working memory account for response accuracy (selecting the correct agent) and response-time measures of canonical and noncanonical comprehension of English sentences by bilingual children when controlling for English and Spanish syntactic knowledge?

   b. To what extent does complex working memory account for the time spent looking at pictures representing the agent, the patient, and the location of the action in canonical and noncanonical sentences by bilingual children when controlling for English and Spanish syntactic knowledge?

METHOD

This study employs a quasi-experimental design that makes use of data collected in a larger multimodal study of language and literacy. The larger study involved four sessions of data collection in which children completed cognitive, language, and literacy tasks. While these tasks were completed, participants had their neural activity indirectly measured with functional near infrared spectroscopy (fNIRS) and their eye gaze recorded with an eye tracker. This dissertation focuses on the English sentence comprehension portion of the larger study.

Participants

The sample of children used for this analysis consisted of 19 bilingual children (12 girls, 7 boys) and 19 monolingual children (11 girls, 8 boys) who were equivalent in age. The mean age of the monolingual group was 11;8 (years; months) and the mean age of the bilingual group was 11;5 (years; months).

Recruitment flyers were distributed at several schools surrounding the university area and at community events. To be eligible for this proposed study, children had to be between the ages of 9-14, have no history of language impairment, and be either a monolingual English speaker or a bilingual Spanish-English speaker. Participants' guardians completed an extensive demographic form to delineate aspects of their child's language development, such as: age at onset of second-language exposure, language of formal education, country of birth, current usage of both languages on a daily basis, and reported proficiency (speaking, listening, reading, and writing). Additional demographic information was collected, such as: age, sex, each parent's highest level of education completed, income, ethnicity/race, and vision/hearing information.

See Table 1 for the participants' demographic information. Though the groups did not differ by age, sex, or family income, there was a statistically significant difference in maternal educational level ($p < .001$) in favor of the monolingual English group.

Table 1. *Summary of Demographic Information*

	Monolingual	Bilingual	P-Value
	n = 19	n = 19	
Sex*			
Male	8 (42.1%)	7 (36.8%)	
Female	11 (57.9%)	12 (63.2%)	
Age (in years)			0.508
	11.7 (1.7)	11.4 (1.5)	
Income (in dollars)*			0.564
8000-12000	0 (0%)	0 (0%)	
13000-15000	3 (15.8%)	3 (15.8%)	
16000-19000	1 (5.3%)	1 (5.3%)	
20000-22000	0 (0%)	2 (10.5%)	
23000-25000	0 (0%)	0 (0%)	
26000-29000	0 (0%)	1 (5.3%)	
30000-36000	2 (10.5%)	4 (21.1%)	
37000-50000	2 (10.5%)	3 (15.8%)	
51000-75000	4 (21.1%)	2 (10.5%)	
76000+	5 (26.3%)	2 (10.5%)	
NR	2 (10.5%)	1 (5.3%)	
Mother's Education*			<.001
High School	1 (5.3%)	11 (57.9%)	
2-year college	7 (36.8%)	0 (0%)	
4-year college	6 (31.6%)	3 (15.8%)	
Graduate or Professional	3 (15.8%)	1 (5.3%)	
NR	2 (10.5%)	4 (21.1%)	

Note: NR = not reported, * = chi-square test

Standardized Assessments

All children received a battery of measures to assess their cognitive abilities and their linguistic abilities in English and Spanish (see Appendix).

Cognitive tasks. Visual working memory was measured with the Symbolic Memory subtest of the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998). This subtest was designed to measure short-term visual memory and complex sequential memory for meaningful material. The administrator provided the child with a total of 10 tiles, each containing a symbol for *baby, girl, boy, woman,* and *man* depicted in green and black. The child looked at a sequence of these symbols for five seconds. Once the sequence was removed, the child recreated the sequence using the tiles. The reliability of this assessment was adequate with an internal consistency of .85 and test-retest reliability of .72 (corrected).

Complex auditory working memory was assessed with the Auditory Working Memory subtest from the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock, McGrew & Mather, 2001). This subtest was designed to measure the recoding of verbalizable acoustic information. The child listened to a series of mixed up numbers and words. The child repeated first, the series of words in the sequence heard, and second, the series of numbers in the sequence heard. The reliability of this assessment was good with a median reliability .88.

Phonological short-term memory was assessed with the Non-word Repetition subtest from the Comprehensive Test of Phonological Processing-Second Edition (CTOPP-2; Wagner, Torgesen, Rashotte, & Pearson, 2013). This subtest measured the child's ability to repeat nonwords, ranging from 3-15 sounds. The child listened to audio-recorded nonwords and was told to repeat them exactly as they heard them. The reliability of this assessment was good with an internal consistency of .77, test-retest reliability of .77, and rater reliability of .99.

Linguistic tasks. Vocabulary knowledge in English was assessed with the Antonyms subtest from the Comprehensive Assessment of Spoken Language-Second Edition (CASL-2; Carrow-Woolfolk, 2017). The Antonyms subtest was designed to measure word knowledge, retrieval, and oral expression in a decontextualized format. The administrator orally presented a word and the child was expected to respond with one word that was opposite in meaning to the stimulus given. The reliability of this assessment was good with an internal consistency of .92-.99, test-retest reliability of .94 (corrected), and rater reliability of .92.

Syntactic knowledge in English was assessed with the Grammaticality Judgment subtest of CASL-2. This subtest was designed to measure syntactic judgment and construction, was used as the procedural LTM measure. For early items, the administrator orally presented an incorrect sentence and the child was expected to correct the sentence by adding, changing, or removing only one word without changing the meaning of the sentence. For later items, the administrator orally presented a sentence and the child was expected to say “yes” if the sentence was grammatically correct or “no” if the sentence was not grammatically correct. If the sentence was incorrect, the child was expected to correct the sentence as in the earlier items. The reliability of this assessment was good with an internal consistency of .98-.99, test-retest reliability of .87 (corrected), and rater reliability of .86.

Global language comprehension was assessed with the Narrative Comprehension subtest of the Test of Narrative Language-Second Edition (TNL-2; Gillam & Pearson, 2017). This subtest measured comprehension of narratives. The administrator read three narratives supported by a single scene or sequenced scenes and then the child answered open-ended comprehension questions read by the test administrator. The reliability of this assessment was good with an internal consistency of .81, test-retest reliability of .85, and rater reliability of .99.

The bilingual children were given additional assessments to measure their Spanish language ability. These assessments included the following subtests from the Clinical Evaluation of Language Fundamentals-Spanish Edition (CELF-4 Spanish Edition; Semel, Wiig, and Secord, 2006): Conceptos y Siguiendo Oraciones (Concepts and Following Directions), Recordando Oraciones (Recalling Sentences), Formulación de Oraciones (Formulating Sentences), Clases de Palabras-Receptivo (Word Classes-Receptive), Clases de Palabras-Expresivo (Word Classes-Expressive), and Definiciones de Palabras (Word Definitions).

The Conceptos y Siguiendo Oraciones (Concepts and Following Directions) subtest measured the child's ability to comprehend oral directions of increasing length and complexity, as well as relational terms, while also identifying the objects described. The administrator read a direction aloud and the child followed the directions by pointing to the correct item(s) (in the correct order) pictured on the stimulus book. The reliability of this assessment was good with an internal consistency of .88 and test-retest reliability of .82 (corrected).

The Recordando Oraciones (Recalling Sentences) subtest measured the child's ability to recall and reproduce sentences of varying lengths. The test administrator orally read aloud a sentence and then the child was asked to repeat the sentence back verbatim.

The reliability of this assessment was good with an internal consistency of .95, and test-retest reliability of .89 (corrected).

The Formulación de Oraciones (Formulating Sentences) subtest is a measure of expressive language. It measures a child's ability to produce semantically and grammatically correct sentences. The administrator presented a visual scene and read aloud a word. The child was asked to produce a complete sentence about the scene that contained the word given. The reliability of this assessment was good with an internal consistency of .85, test-retest reliability of .77 (corrected), and rater reliability of .81.

The Clases de Palabaras (Word Classes) subtest was divided into two parts, Clases de Palabras - Receptivo (Word Classes-Receptive) and Clases de Palabras-Expresivo (Word Classes-Expressive). This subtest measured the child's ability to understand and explain the logical relationships between the meaning of related words. For this task, the administrator read aloud four words and then the child was required to select the two words that were related to each other. After selecting the related words, the child was expected to explain how these words related to each other. Correct selection of the related words was scored as Clases de Palabras - Receptivo (Word Classes-Receptive). The reliability of this assessment was good with an internal consistency of .84, test-retest reliability of .76 (corrected), and rater reliability of .99. Correct explanation of how the words were related was scored as Clases de Palabras – Expresivo (Word Classes – Expressive). The reliability of this assessment was good with an internal consistency of .88, test-retest reliability of .76 (corrected), and rater reliability of .99.

The Definiciones de Palabras (Word Definitions) subtest is a measure of vocabulary. It measured the child's ability to define words by describing meaning features, class relationships and shared meanings. For each item, the test administrator read aloud the target word and then read aloud a sentence containing the target word. The child then defined the word. If the child was generally correct, but gave an incomplete answer, the administrator was permitted to prompt the child by saying, “Dime más./Tell me more”. The reliability of this assessment was good with an internal consistency of .89, test-retest reliability of .92 (corrected), and rater reliability of .89.

As shown in Table 2, the two groups did not differ in English measures of nonword repetition and verbal working memory. However, the bilingual children obtained lower scores on measures of grammatical judgment, narrative comprehension, and antonyms. The bilinguals had a wider range of scores, especially for grammatical judgment, as seen in Figure 1. Additionally, they also had a wide range of scores on the measures of Spanish language (Figure 2). Raw and scaled scores for the Spanish language measures are shown in Table 3. There was one bilingual participant who scored lower in Spanish than English, whereas most other children had scores within 1.5 SD. None of the bilingual children obtained scores lower than -1.5 SD on more than two measures for both the Spanish and English measures.

Table 2. *Summary of English Measures-Raw Scores*

	Monolingual	Bilingual	P-Value	Cohen’s *d*
	n = 19	n = 19		
Nonword Repetition (CTOPP-2)	16.8 (1.9)	15.9 (3.1)	0.284	0.35
Narrative Comprehension (TNL2)	37.8 (3.8)	32.6 (7.3)	0.01*	0.89

Narrative Production (TNL2)	53.7 (9.9)	52.5 (11.8)	0.734	0.11
Antonyms (CASL-2)	37.9 (6.0)	28.7 (7.7)	0.001**	1.33
Grammaticality Judgment (CASL-2)	55.8 (5.9)	40.5 (17.5)	0.002**	1.17
Auditory Working Memory	22.3 (8.2)	19.7 (4.8)	0.235	0.39

Note: * indicates *p*-value < 0.01, ** indicates *p*-value < 0.001

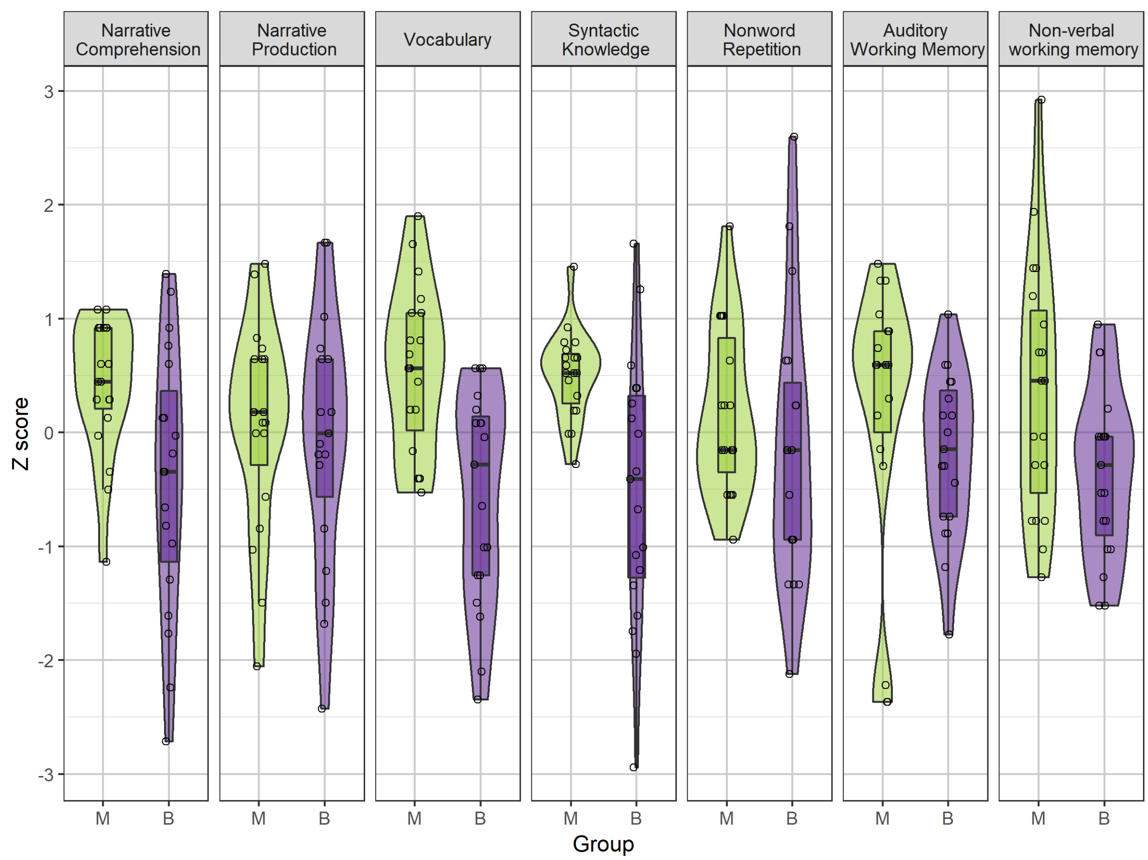

Figure 1. Z scores of English measures for both groups

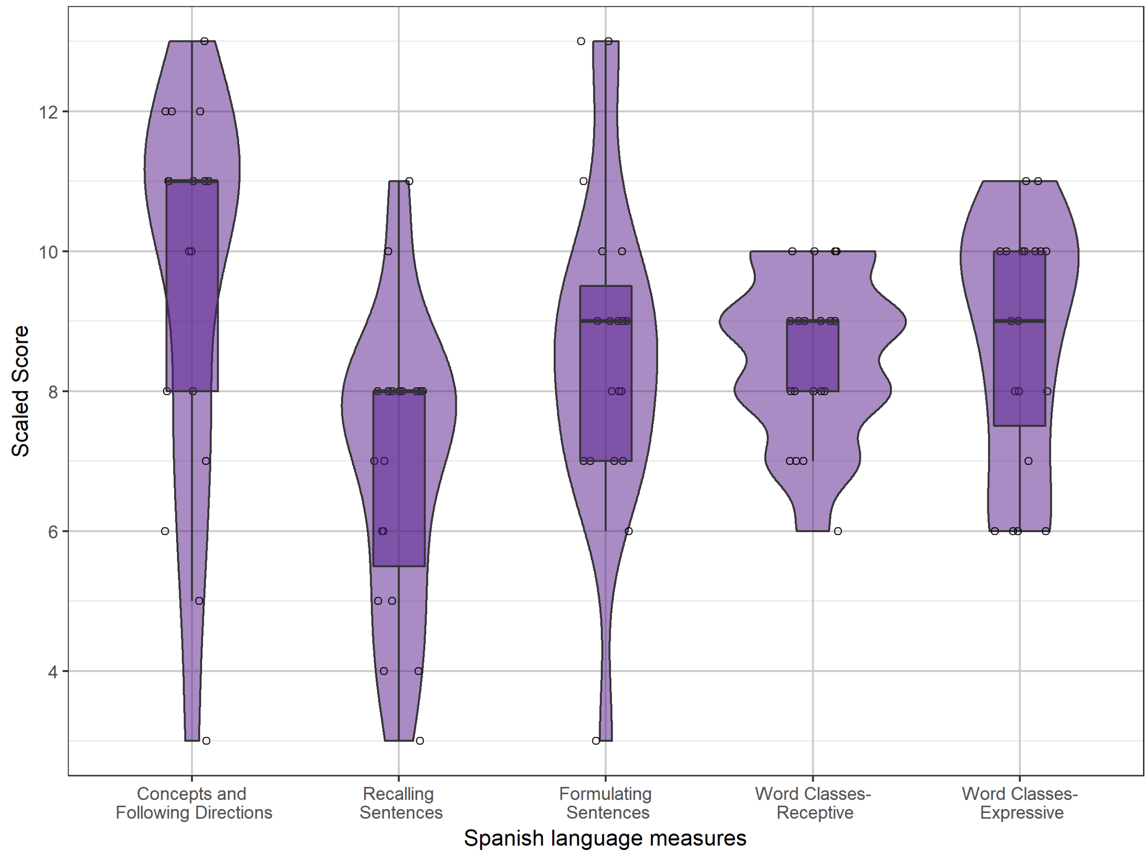

Figure 2. Scaled scores of Spanish measures

Table 3. *Summary of Spanish Measures*

	Raw Score Mean (SD)	Scaled Score Mean (SD)
	n = 19	n = 19
Concepts and Following Directions (CELF-4)	40.5 (6.3)	9.5 (2.8)
Recalling Sentences (CELF-4)	49.2 (17.0)	6.9 (2.1)
Formulating Sentences (CELF-4)	31.9 (7.5)	8.6 (2.3)
Word Class - Receptive (CELF-4)	17.8 (2.8)	8.5 (1.2)
Word Class - Expressive (CELF-4)	13.9 (3.5)	8.7 (1.8)

Core Language Score (CELF-4)	33.7 (6.7)	90.3 (11.0)

Experimental Sentence Interpretation Task

Participants' sentence interpretation was assessed using the "whatdunit?" agent-selection task (Montgomery et al., 2016, 2017). These sentences feature inanimate objects doing something to another inanimate object, for example, "*The hat was kissed by the clock under the cold boot*." The participant was instructed to select the agent in the sentence, which for this example was *the hat*.

Stimuli. Twelve sentences were presented in each of four conditions. Canonical sentences consisted of subject verb object (SVO; *The ring moved the square behind the very bright cold bed*.) and subject relative (SR; *The fork that wiped the boot near the shirt was bright*.). Noncanonical sentences consisted of passive (PAS; *The ring was bathed by the key under the hot bread*.) and object relative (OR; *The hat that the car fixed under the fork was hot*.).

All sentences were derived from Montgomery et al. (2016, 2017) with slight modifications (explained below). These sentences were originally constructed to contain 33 inanimate objects to decrease semantic plausibility, which decreases the reliance on semantic knowledge and increases the reliance on syntactic information. The 33 nouns chosen for these sentences were specifically chosen to be accessible to children with language impairments. These words are typically acquired by four years of age, have good imageability, and have high familiarity and frequency of usage.

The reported internal construct validity of the canonical sentences was very high, .84, p = .0001 (Montgomery et al., 2016). The reported internal construct validity of the noncanonical sentences was also very high, .89, p = .0001 (Montgomery et al., 2016). The noncanonical sentences had lower correlations (.31-.35) with the canonical sentences. The internal reliability was calculated using Cronbach's alpha, which was .97 overall, .88 for SVO, .86 for SR, .95 for PAS, and .94 for OR (Montgomery et al., 2016).

For this study, the English sentences were modified to include only 11 words per sentence compared to 12 words in the original task. An additional 12 control sentences were created using the same nouns and images to mirror the agent selection task. Each control sentence took the form of, "Click on the Noun." This task was designed to control for motor speed and visualization. All sentences were recorded at a normal speaking rate by an adult female speaker of Midwestern American English. The audio files were all low-pass filtered at 20kHz and normalized for intensity.

Sentences were pseudorandomized into two blocks of six sentences per sentence type and control task (see Figure 3). The first block consisted of six tasks each lasting 72s: cross-rest, a control task, SR, SVO, Pas, OR. The second block consisted of another set of these six tasks followed by a third cross-rest task. Each task, including rests, was preceded by a 15s interstimulus rest. Stimuli were presented on an Eizo ColorEdge CS230 screen and through speakers on each side of the monitor directed toward the participant. Visually, one picture representing each noun in the sentence appeared on the screen for 2ms followed by a colorful square in the middle of the screen to center the participant's eyes to the center of the screen for another 2ms. Finally, the center square was removed, and the sentence was auditorily presented so that the participant could select the agent using the mouse.

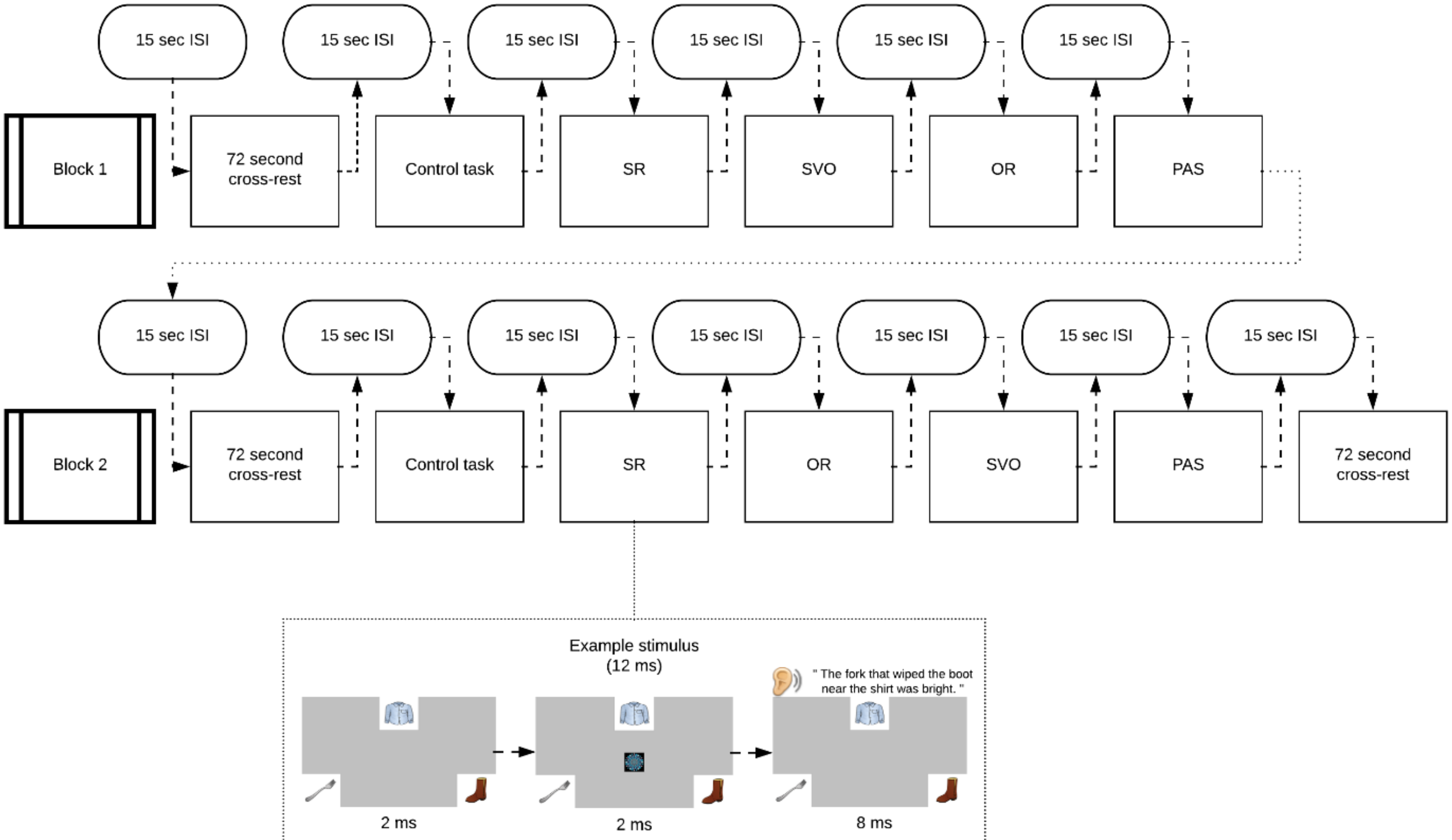

Figure 3. Experimental task design and example stimulus item

Procedures

Children attended three to four testing sessions on separate days. Informed consent was signed on the first day by both the guardian and child. Children were seen individually in a quiet testing room for assessments and in another quiet room for the experimental task. Each session lasted about 1 hour and 15 minutes, with a break between the experimental task and any administration of the assessments.

For the experimental task, children sat in front of the Eizo ColorEdge CS230 monitor with their chins on a chin rest. All children completed a 9-point calibration task in which the children had to follow a dot on the screen. The eye-tracking software was designed so that the next dot appeared when the eye's fixation on the dot was detected. This was followed by a 4-point validation in which the child, again, fixated on the dots presented. During validation, the eye-tracking software calculated and provided the average deviation from the dot and gaze of the eye. Participants were required to obtain a score less than 1° in order to continue with the experimental task.

Once the calibration and validation checks were completed, children performed a demonstration task in which they saw two examples of each sentence type. There were pauses between each type to allow for questions and to check for understanding of the task. After the demonstration, the children were capped with functional near infrared spectroscopy (fNIRS) optodes and they completed a second 9-point calibration and 4-point validation check.

Stimuli were presented via an SMI Red250m eye-tracking system running Experiment Center software. The software automatically detected distance from eyes, and research assistants adjusted the screen so that participants' eyes were centered and were at an adequate distance (approximately 58-62cm) for eye detection and recording. Eye movements were recorded at 250Hz. The SMI Experiment Center software automatically classified eye movements as fixations, saccades, and blinks. Because children were also wearing an fNIRS cap, which emits infrared light, all children wore a blackout cap over the optodes to remove interference from both light sources (i.e., eye-tracker and NIRS). To further control for interference from other light sources, the windows in the room were covered with blackout curtains and the lighting was controlled at a brightness of 28 to 35 Fc using dimmable LED lights with LED drivers.

Eye-tracking Measures

Eye-tracking technology can be used as one of several methodological techniques to measure online processing or the processing of sentences as it occurs. Eye tracking also provides multiple options to analyze data from the numerous measures that can be obtained from recording the eye.

One approach to eye tracking is to take advantage of the fact that we generally tend to look at things as they are mentioned (Cooper, 1974). In this study, participants passively listened to stories while presented with a visual grid of pictures. Listeners' looks to the pictures were time-locked to when those objects were mentioned in the story, suggesting a simple linking hypothesis: *The probability of looking at an object increases when the object is mentioned* (Boland, 2004). Using this basic tenet, researchers use looks to images as insight into how and when sentences are processed.

Carpenter and Just (1980) described their eye-mind hypothesis, in which a reader simultaneously looks at a word and engages in cognitive processing for the full length of the fixation. Whereas in reading we know to look for fixations and saccades as representations of processing, no comparable approach exists in the listening literature (Boland, 2004). In listening tasks, the dependent measures are usually limited to fixation duration and probability of looking at an object within some temporal interval as in the Cooper (1974) and Altmann and Kamide (1999) studies.

Altmann and Kamide (1999) had 24 college students engage in a listening task while looking at a visual scene with a referent (e.g., boy), target object (e.g., cake), and three or four distractor objects (e.g., train, car, ball). The listeners heard sentences such as, "The boy will move the cake" or "The boy will eat the cake" while looking at the visual scene. For each verb (e.g., *eat* or *move*), the researchers calculated the cumulative probability across trials of fixating either on the target (*cake*) or one of the distractor objects for each 50ms interval from the verb onset. There were significantly more looks to the target than distractors before hearing the noun. In the *move* condition, the first saccade occurred 127ms after the onset of the target noun. In the *eat* condition, the first saccade occurred 95ms before the onset of the target noun. Importantly, this study provides evidence that listeners begin to establish anaphoric dependencies at the verb. These results are consistent with the *chunk-and-pass* model of sentence processing, which posits that when listening to sentences, we tend to anticipate the forthcoming words. Similarly, Sussman, (2006) demonstrated that listeners make anticipatory looks to objects using verb knowledge. For example, participants looked at a pencil when hearing "poke the dolphin," but, not when they heard "touch the dolphin".

Other eye-tracking approaches have attempted to determine effort or cognitive load during sentence processing. Measures believed to indicate cognitive effort are thought to include pupil dilation and fixation time or the amount of time spent looking at one area. Schluroff (1982) collected pupil dilation data while participants listened to sentences of varying complexity. After listening to each sentence, participants rated the level of difficulty of the sentence using a 7-point scale. One result of this study was that pupil dilation was more strongly correlated with grammatical complexity than were the participants' ratings, demonstrating how pupil size is utile as an online measure of cognitive effort in relation to varying levels of grammatical complexity. Scheepers and Crocker, (2004) used both gaze duration and pupil size to determine the effects of written sentences classified as subject-object, object-subject, and neutral, as primes for two possible syntactic interpretations of ambiguous sentences presented orally. They used a visual scene, in which one person could be either the patient or the agent of an orally presented ambiguous sentence. Both measures of pupil size and gaze duration or fixation time were used to determine when during an ambiguous sentence stream disambiguation occurred and how difficult it was to disambiguate.

Fixation time as a measure of processing. Holmqvist et al., (2011) summarized research in which the interpretation of fixation time varies across tasks and stimuli. Generally, longer fixations are associated with deeper and more effortful cognitive processing during reading, scene perception, and usability research. In usability research, longer fixations may be an indication of how much difficulty a participant has in extracting information from a display. However, longer fixations could also mean shallow processing as in cases where participants begin to experience low arousal. Expertise in a field such as art or chess leads to longer fixations and fewer fixations because more information is extracted around the fixation. The authors also summarize that neurological impairments may be associated with longer fixations, not as an interpretation of deeper processing, rather as an interpretation of disturbed processing.

Data Analyses

The first research question concerned the nature of the relationship between complex auditory working memory and canonical and noncanonical sentence processing of English sentences in monolingual and bilingual children. Specifically, I wanted to know the extent to which complex working memory accounts for response accuracy (selecting the correct agent) and response-time measures of canonical and noncanonical comprehension of English sentences by monolingual and bilingual children while controlling for English syntactic knowledge. I also wanted to know the extent to which complex working memory accounts for the time spent looking at the picture representing the agent of the action in canonical and noncanonical sentences by monolingual and bilingual children when controlling for English syntactic knowledge.

To answer the three parts of Research Question 1, I used multilevel modeling (MLM) in three separate analyses to explore how working memory moderates the accuracy of agent selection, response time, and fixation time on the agent. Analyses were conducted using the *lme4* (Bates, Mächler, Bolker, & Walker, 2015) packages in R (R Core Team, 2018). Instead of aggregating data for each stimulus, subjects were treated as a cluster with observations at the stimulus level nested under participant. This prevented loss of information and associated loss of statistical power, while avoiding spurious results (e.g., ecological fallacy, Simpson’s paradox; Hox, 2010). This also allowed for the inclusion of participants with missing data at one or more stimulus levels. While repeated measures ANOVA can include correlated observations, it must exclude participants with less than complete data. Also, repeated measures ANOVA is limited to only two levels of observational nesting and assumes a strict pattern of variance and correlation (homogeneity of variance and sphericity). MLM, which is the umbrella under which ANOVA and regression fall, offers a more flexible framework for inclusion of correlation observations on two or more levels without assuming homogeneity of variance and sphericity (Hox, 2010).

To assess the accuracy of agent selection, a binomial logistic linear mixed effects regression was used. A two-level model with correctness of selection of the agent as the outcome was proposed. The two-level model included two random intercepts, for participant and stimulus item (Figure 4). Items were analyzed as crossed effects rather than being nested under each participant because each participant received the same set of stimulus items. Fixed effects at the participant level included *group* as a two-level factor (monolingual, bilingual), *working memory* (WM), and *English syntactic knowledge* (GJ). One fixed effect at the stimulus level was included, *sentence type* as a four-level factor (SVO, SR, PAS, OR). Because the outcome measure, *accuracy*, was binomial, I used the *glme*r() function in the *lme4* package (Bates et al., 2015).

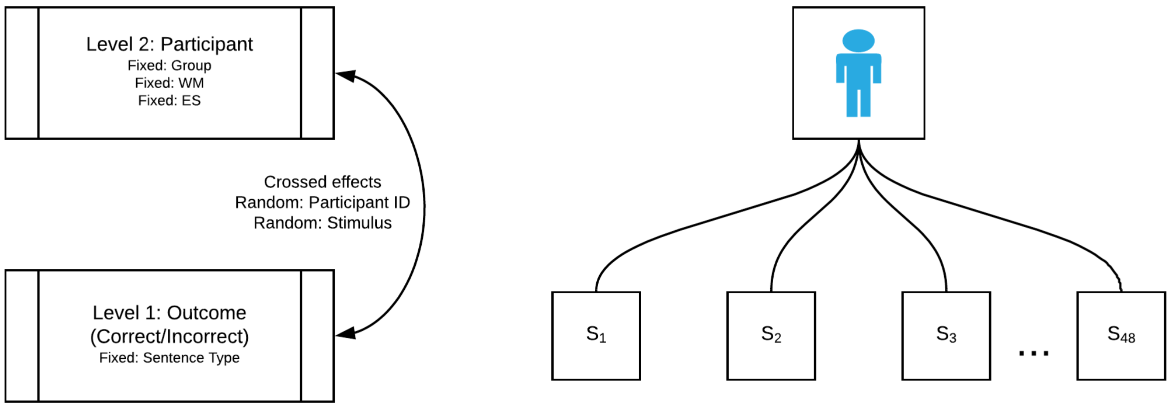

Figure 4. Two-level model of accuracy with participant and stimulus as crossed effects. WM = working memory, ES = English syntactic knowledge

The analysis strategy was theoretically driven, as a strictly top-down or bottom-up exploratory approach may have resulted in overfitting of models and decisions being made by chance (Hox, 2010). The initial model contained a three-way interaction between *group*, *sentence type*, and *working memory*. *English syntactic knowledge* was treated as a non-interacting predictor or covariate. If the three-way interaction was not significant based on a Type III sum of squares F test, a secondary model containing theoretically relevant two-way interactions were fit and compared to the initial model using a Likelihood Ratio Test. Further simplification followed in a similar manner, if needed.

To assess the latency of the agent selection, a linear mixed effects regression was used. A two-level model was proposed with response time as the outcome. The two-level model included random intercepts for both participants and the stimulus items (Figure 5). Items were analyzed as crossed effects rather than being nested under each participant because each participant received the same set of stimulus items. Fixed effects at the participant level included *group* as a two-level factor (monolingual, bilingual), *working memory* (WM), and *English syntactic knowledge* (GJ). Two fixed effects at the stimulus level were included, *sentence type* as a four-level factor (SVO, SR, PAS, PR) and *accuracy* as a two-level factor (correct, incorrect). Because the outcome variable, *response* time, was a continuous variable, this analysis was conducted employing the *lmer()* function in the *lme4* package (Bates et al., 2015).

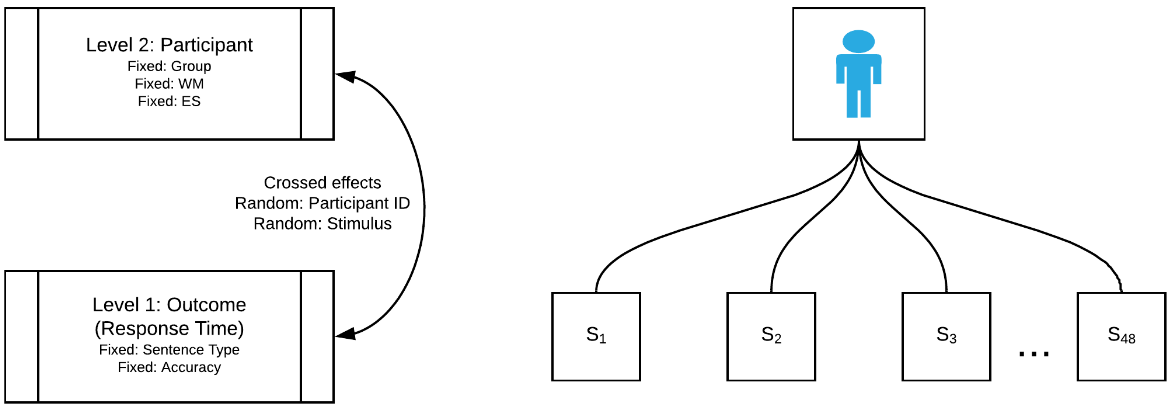

Figure 5. Two-level model of response time with participant and stimulus as crossed effects. WM = working memory, ES = English syntactic knowledge

The analysis strategy was similar to the previous analysis. The initial model contained a four-way interaction between *group*, *sentence type*, *working memory*, and *accuracy*. *English syntactic knowledge* was treated as a non-interacting predictor or covariate. If the three-way interaction was not significant based on a Type III sum of squares F test, a secondary model containing theoretically relevant two-way interactions were fit and compared to the initial model using a Likelihood Ratio Test. Further simplification followed in a similar manner, if needed.

Analysis of the fixation time on the agent was also conducted using linear mixed effects modeling. Because fixation time was a continuous variable, the *lmer()* function in the *lme4* package (Bates et al., 2015) was used. This entailed using a two-level model (Figure 6) with two random effects: random intercepts for the participants (level 2) and random intercepts for the stimulus items (level 1). Again, the random effects were crossed because every participant received every item. Level 1 consisted of fixation time measurements for the picture of the agent as the outcome variable. As in the previous analysis, fixed effects at the participant level included *group* as a two-level factor (monolingual, bilingual), *working memory* (WM), and *English syntactic knowledge* (ES). Two fixed effects at the stimulus level were included: *sentence type* as a four-level factor (SVO, SR, PAS, OR) and *accuracy* as a two-level factor (correct, incorrect). Similar to the previous analysis, model-building started with a model including *English syntactic knowledge* and a four-way interaction of *group*, *sentence type*, *working memory*, and *accuracy*, which was compared to a simpler model using a Likelihood Ratio Test to determine the best fit model.

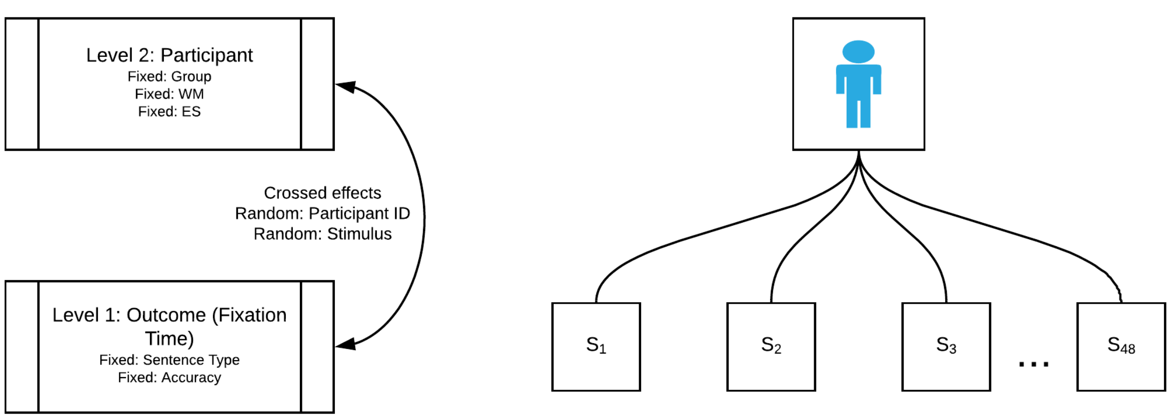

Figure 6. Two-level model of fixation time on the agent with participant and stimulus as crossed effects. WM = working memory, ES = English syntactic knowledge

The second research question concerned the additional role of Spanish syntactic knowledge in long-term memory on comprehension of English canonical and noncanonical sentences. Specifically, we wanted to know the extent to which complex working memory accounts for response accuracy (selecting the correct agent), response time, and fixation measures of canonical and noncanonical comprehension of English sentences by bilingual children when controlling for English and Spanish syntactic knowledge. Additionally, we were interested in the extent to which complex working memory accounts for the time spent looking at the pictures representing the agent of the action in canonical and noncanonical sentences by bilingual children when controlling for English and Spanish syntactic knowledge. For research question number 2, assessing the additional role of Spanish syntactic knowledge (SS), the correlation between English syntactic knowledge and Spanish syntactic knowledge was investigated. Providing that the correlation was not moderately high, analysis followed the same manner as described previously for research question 1. The analysis was restricted to the subset of bilingual children and included the addition of the Spanish syntactic knowledge measure.

Power Analysis

Power analysis is typically conducted to determine the appropriate sample size needed to detect an effect of a given size with a specific test at a desired significance level. Because the proposed study used an existing database for analysis, power analysis would typically be conducted to determine the power to detect an effect given the effect size, sample size, and significance level. However, power analysis of MLM is complicated by having sample sizes at more than one level, fixed effects, and random effects. A common approach to determine sample sizes in MLM, is to conduct simulation studies, in which a statistical model with population models of all parameters and sample sizes at all levels are given to create thousands of datasets.

Because power analysis is complicated, various programs have been devised for the purpose of helping researchers calculate power for a given sample size. Power Analysis in Two-level Designs (PINT; Bosker, Snidjers, & Guldemond, 2003), Optimal Design (Spybrook et al., 2011), and Powerlmm (Magnusson, 2018), all programs designed for power analysis in MLM, are not suitable for the quasi-experimental design of this study that seeks to explore cross-level interactions. Furthermore, the information needed to conduct these analyses, in addition to effect sizes, includes population values of all other parameters, including correlations and variance components (Hox, Moerbeek, & van de Schoot, 2018).

Given the difficulty in obtaining plausible values for all model parameters, various rules of thumbs have been suggested (Hox, Moerbeek, Schoot, Moerbeek, & Schoot, 2017), such as the ‘30/30 rule,’ (Kreft, 1996). This rule suggests a minimum of 30 participants and 30 items per participant. The proposed study exceeded this with 38 participants and 48 items per participant.

Expected Results

For research question 1, I expected to see faster time to selection in canonical sentences and noncanonical sentences for both groups. I expected that accuracy would be similar for each group, but that the time to selection would be slower for the bilingual group than for the monolingual group. I also expected that selection would be slower for the noncanonical sentences than the canonical sentences. I also expected that increases in working memory would relate to faster selection times.

For research question 1c, I expected to see an interaction between sentence type, working memory, and accuracy, such that for increases in working memory ability, there would be increased fixation time on the agent for noncanonical sentences, but not canonical sentences. In terms of group differences, I expected that bilinguals would have greater fixation time on the agent than the monolingual group, especially for noncanonical sentences, and that working memory would moderate performance for the monolingual group, but not as much for the bilingual group.

RESULTS

The purpose of this study was to examine the relationship between complex auditory working memory, syntactic knowledge, and canonical and noncanonical sentence comprehension in bilingual and monolingual children using both offline (behavioral) and online (eye-tracking) measures. There were two main research questions. The first question focused on the extent to which working memory accounted for three outcomes (accuracy, response time, and fixation time) of canonical and noncanonical sentence comprehension in the monolingual and bilingual groups when controlling for English syntactic knowledge. The second question focused on the additional role of Spanish syntactic knowledge.

Agent Selection Task Accuracy

Descriptive data. Raw data for the accuracy of agent selection for each group is presented in Figure 7. Recall that children were asked to listen to four types of sentences (canonical: SVO and SR; and noncanonical: PAS and OR) and to click on one picture from a 3-picture display that best represented the agent of the sentence. Responses were coded as correct or incorrect based on the first click. The fine lines in the figure represent individual participants' accuracy on the four sentence types and the heavy lines represent the group average. Note that there was greater variance for the noncanonical (PAS and OR) than the canonical sentence types (SVO and SR) for both the monolingual and bilingual groups.

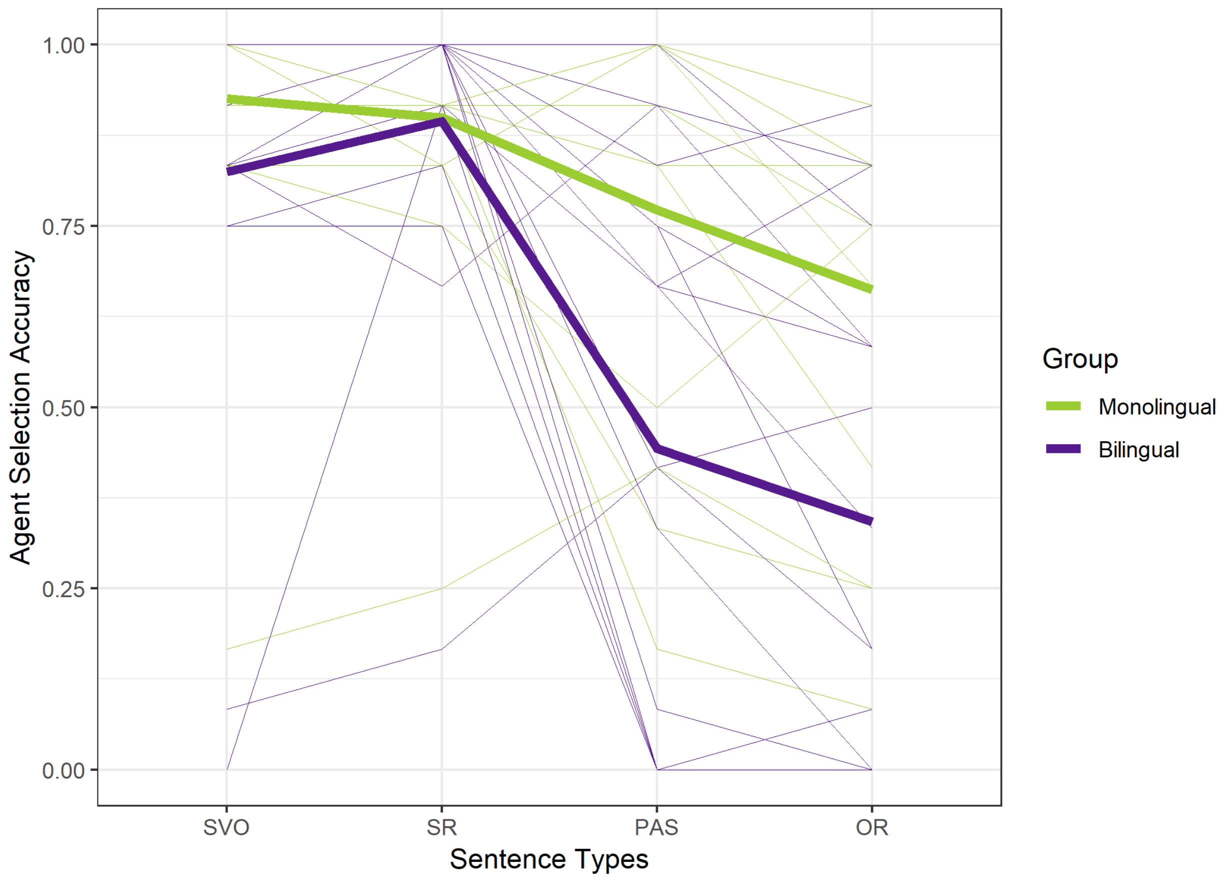

Figure 7. Raw data for accuracy summarized by participant and group. Fine lines represent participant means and heavy lines represent group means. Both groups had greater variance inaccuracy scores for the noncanonical sentences compared to the canonical sentences.

Figure 8 depicts the range of accuracy across the four sentence types. The width of the shape represents the frequency of participants who obtained a particular score. For the canonical sentences, we see that in both groups a majority of participants obtained scores above 75% accuracy and very few participants obtained scores below 25% accuracy. For noncanonical sentences, we see that the monolingual group had more participants score below 75% than in the canonical sentences and that distribution was fairly even across these scores. The bilingual group also had more children obtaining a lower score on the noncanonical sentences as compared to the canonical sentences. There are some differences between the two noncanonical sentence types. Notice that their distribution of scores is fairly even across the full range of scores for the PAS sentence type. For the OR sentence type, the bilingual group's distribution is almost inverted compared to the monolingual group; that is more children obtained lower accuracy scores than the number of children who obtained high accuracy scores. A descriptive summary of the accuracy for each sentence type for each group follows in Table 4.

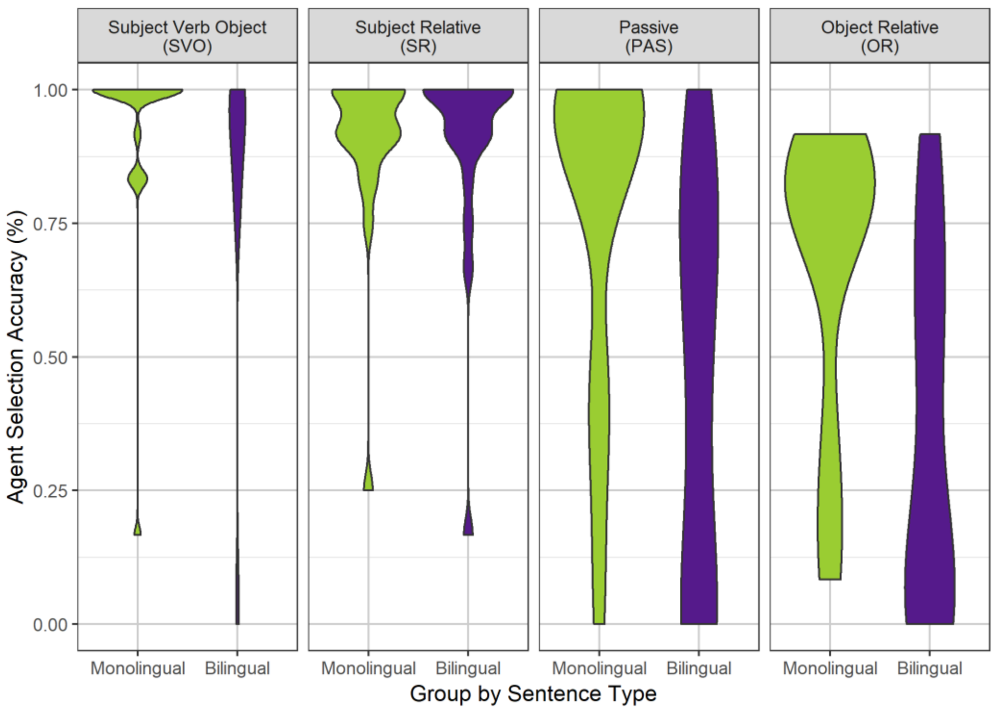

Figure 8. Distribution of agent selection accuracy across sentence types by groups.

Table 4. Summary of Accuracy by Group and Sentence Type

Sentence Type	Monolingual		Bilingual		Cohen's d
	Mean	SD	Mean	SD	
SVO	0.8792	0.3006	0.8246	0.3076	0.179534
SR	0.9	0.3266	0.8947	0.3812	0.014932
PAS	0.7333	0.4431	0.443	0.4755	0.631656
OR	0.6292	0.484	0.3421	0.4978	0.584786
Control	0.9	0.3006	0.9079	0.2898	-0.02676

LME analysis. The first part of question 1 asked to what extent does complex working memory account for response accuracy (selecting the correct agent) of canonical and noncanonical comprehension of English sentences by bilingual children when controlling for English syntactic knowledge. Prior to analysis of the data using linear mixed effects modeling, I checked for multicollinearity using correlation matrices (Figures 9 and 10). The measures are identified on the diagonal. The values to the left of the diagonal indicate the Pearson correlation coefficients and their associated *p*-values. Circles to the right of the diagonal visually represent the correlation with color representing the direction of the relationship (blue = positive, red = negative). Both color and the ellipses represent the strength of the relationship (darker/thinner = stronger, lighter/wider = weaker).

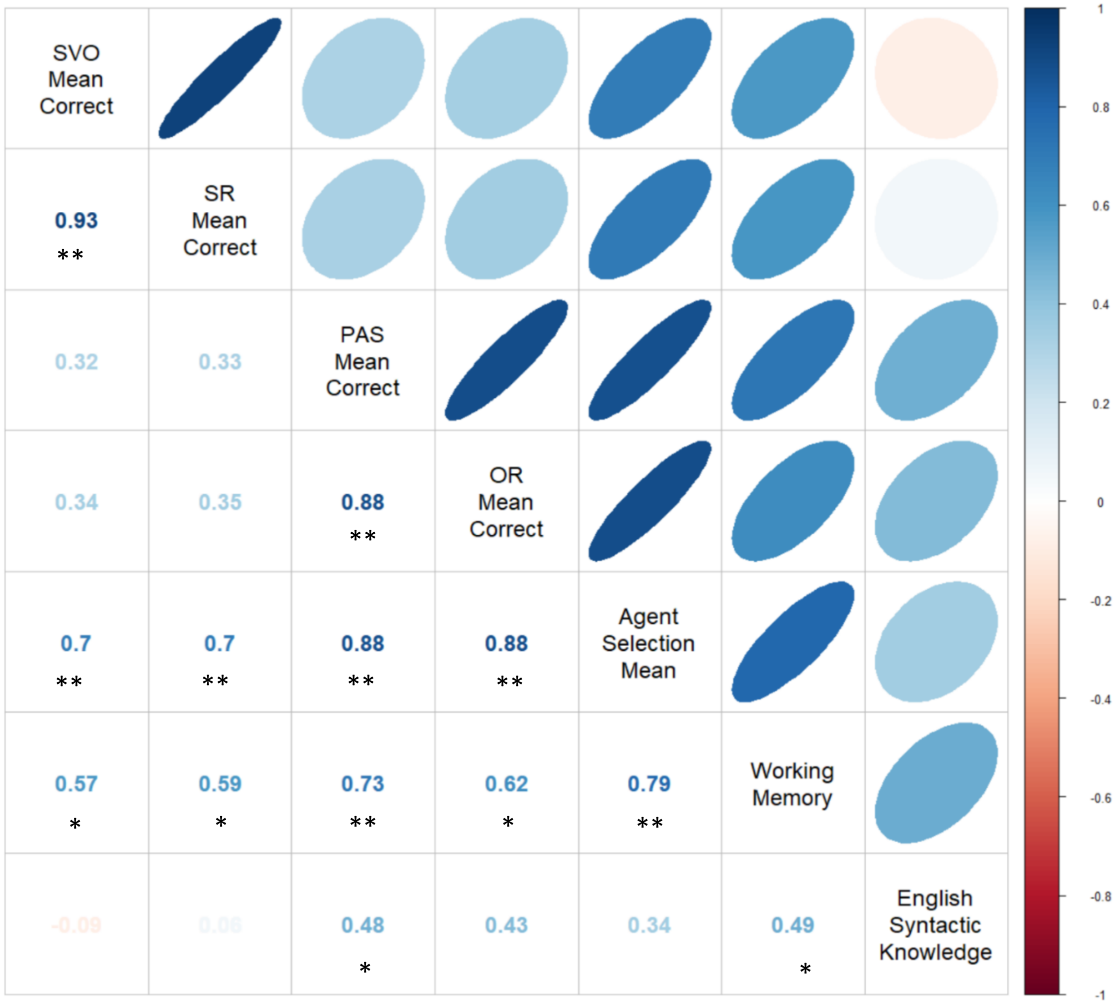

Figure 9. Correlation plot for monolingual group. * p < .05, ** p < .01, *** p < .001. SVO = subject verb object. SR = subject relative. PAS = passive. OR = object relative.

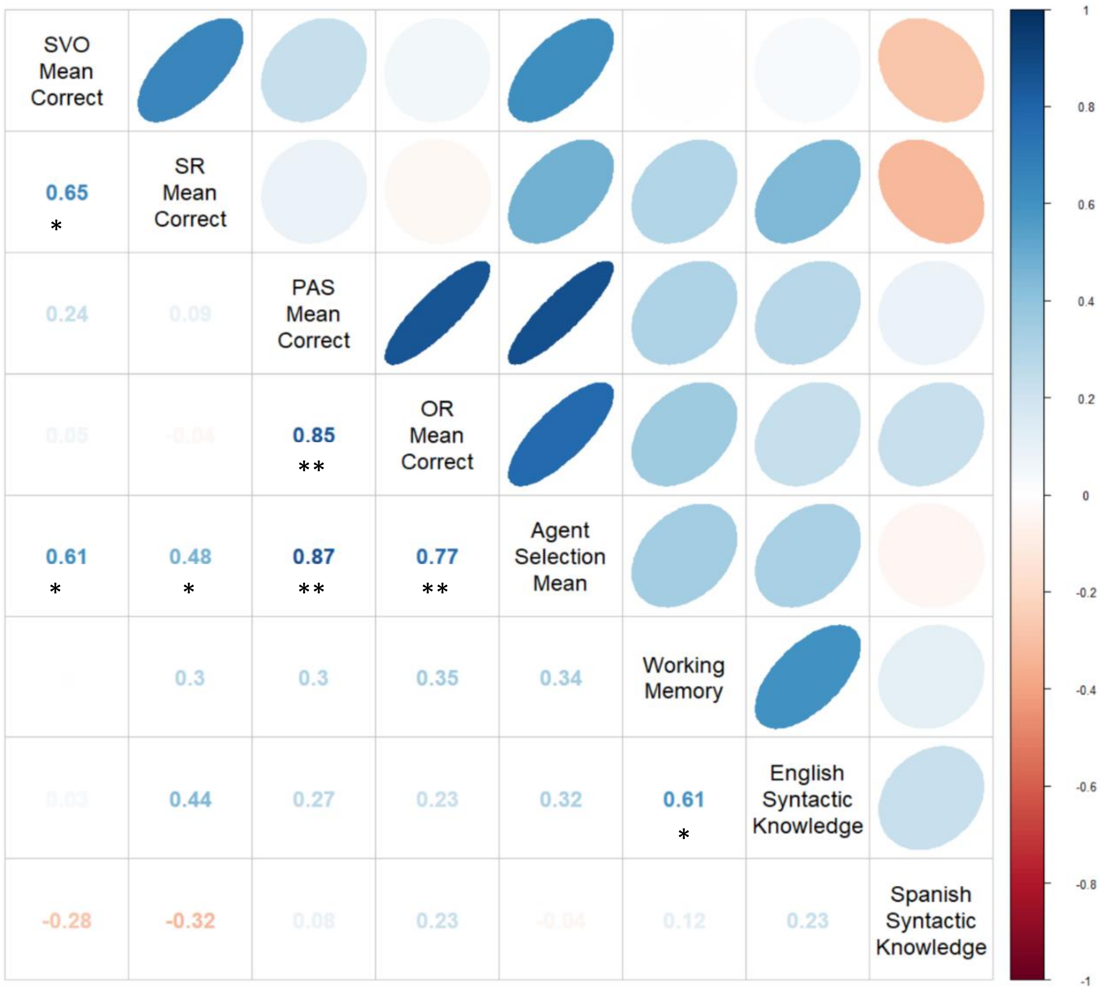

Figure 10. Correlation plot for bilingual group. * $p < .05$, ** $p < .01$, *** $p < .001$. SVO = subject verb object. SR = subject relative. PAS = passive. OR = object relative.

For monolinguals, the correlation between SVO and SR and the correlation between PAS and OR were extremely large. For bilinguals, the correlation between SVO and SR was moderately large and the correlation between PAS and OR was extremely large. For monolinguals, there was a moderate relationship between English syntactic knowledge and working memory. There was a moderately small relationship between English syntactic knowledge and overall sentence comprehension and a moderately large relationship between working memory and overall sentence comprehension. For bilinguals, there was a moderate relationship between English syntactic knowledge and working memory. The relationship between Spanish and English syntactic knowledge was not significant. Bilinguals' overall sentence comprehension had no significant relationship between any of the cognitive/linguistic measures. I also calculated the generalized variance inflation factor (VIF) using *vif*() in the *car* package (Fox & Weisberg, 2019). VIF was less than 2 for all predictor variables, indicating that multicollinearity was not a concern, supporting my decision to utilize LME.

I used linear mixed effects modeling to assess the accuracy of the sentence types. Model building followed a top-down approach in which a complete model (three-way interaction and all subsuming interactions, Model A1) was tested against progressively simpler models (Table 5). Model fitting began with response accuracy as the outcome variable. The fixed effects included group (monolingual and bilingual), sentence type (SVO, SR, PAS, OR), and working memory. These models also included two random intercepts for participants and stimuli. Performance on the grammatical judgment task was modeled as a fixed effect to control for English syntactic knowledge. Using the likelihood ratio test (LRT), Model A 1 was compared to a simpler model that removed the three-way interaction and left the three two-way interactions. The first model [first_click_correct ~ Group*SenType*AWM_raw + CASL_GJ_raw + (1|SenID) + (1|Participant)] with one three-way interaction was the best fit model (χ^2 (19) = 10.486, p = 0.014853). The best fit model (Table 6) was used to create Figure 11, which depicts the three-way interaction between group, sentence type, and working memory. There are two critical contributions to the three-way interaction, one relating to group and working memory differences for canonical sentence comprehension and the other relating to group and working memory difference for noncanonical sentences. Generally, across all sentence types, as working memory increased so did accuracy. The exception to this related to comprehension of SVO sentences by children in the bilingual group. For SVOs, the bilinguals with low working memory had the same sentence comprehension accuracy as bilinguals with high working memory. This differed for the monolingual group, in which low working memory scores were associated with lower performance on the comprehension items. The groups had nearly identical performance on the SR sentence type, with poorer accuracy in children with low working memory.

There were clear group and working memory differences for both noncanonical sentence types (PAS and OR). For noncanonical sentences, monolinguals with high working memory obtained sentence comprehension scores approximating the scores they obtained on the canonical sentences. However, this was not the case for bilinguals with high working memory performance. These children performed more poorly on comprehension of noncanonical sentences than canonical sentences. The children with lower working memory scores had much poorer performance on the noncanonical items than the canonical items with children in both the monolingual and bilingual groups performing at similarly low levels on the PAS sentences. However, for the OR sentences, the low working memory bilingual performed more poorly than the low working memory monolinguals. It appears that working memory relates more strongly to the comprehension of noncanonical sentences than the comprehension of canonical sentence types for monolingual children. The bilinguals had lower accuracy for all levels of working memory for the noncanonical sentence types, especially OR, even while controlling for English syntactic knowledge.

Table 5. *Model Comparisons for Accuracy*

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model A 4	3	1482	1499	-738.2	1476	NA	NA	NA
Model A 3	9	1402	1452	-692.1	1384	92.26	6	1.03E-17
Model A 2	16	1394	1482	-680.9	1362	22.35	7	0.002208
Model A 1	19	1389	1494	-675.7	1351	10.49	3	0.01485

Table 6. *Fixed Effects Estimates (Top, with Standard Error of Estimates in Parentheses) and Variance-Covariance Estimates (Bottom) for the Best fit Model of Accuracy*

	Best Fit Model
(Intercept)	-1.27 (1.44)
GroupB	2.98 (2.09)
SenTypeSR	0.42 (0.79)
SenTypePAS	-2.03 (0.80)*
SenTypeOR	-1.68 (0.79)*
AWM_raw	0.19 (0.05)***
CASL_GJ_raw	0.02 (0.02)
GroupB:SenTypeSR	-2.57 (1.41)
GroupB:SenTypePAS	-2.55 (1.38)
GroupB:SenTypeOR	-4.14 (1.45)**
GroupB:AWM_raw	-0.20 (0.10)*
SenTypeSR:AWM_raw	-0.06 (0.04)
SenTypePAS:AWM_raw	-0.00 (0.04)
SenTypeOR:AWM_raw	-0.06 (0.04)
GroupB:SenTypeSR:AWM_raw	0.22 (0.08)**
GroupB:SenTypePAS:AWM_raw	0.11 (0.07)
GroupB:SenTypeOR:AWM_raw	0.19 (0.07)**
AIC	1389.40
BIC	1494.07
Log Likelihood	-675.70
Num. obs.	1824
Num. groups: SenID	48
Num. groups: Participant	38
Var: SenID (Intercept)	0.49
Var: Participant (Intercept)	1.52

*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$.

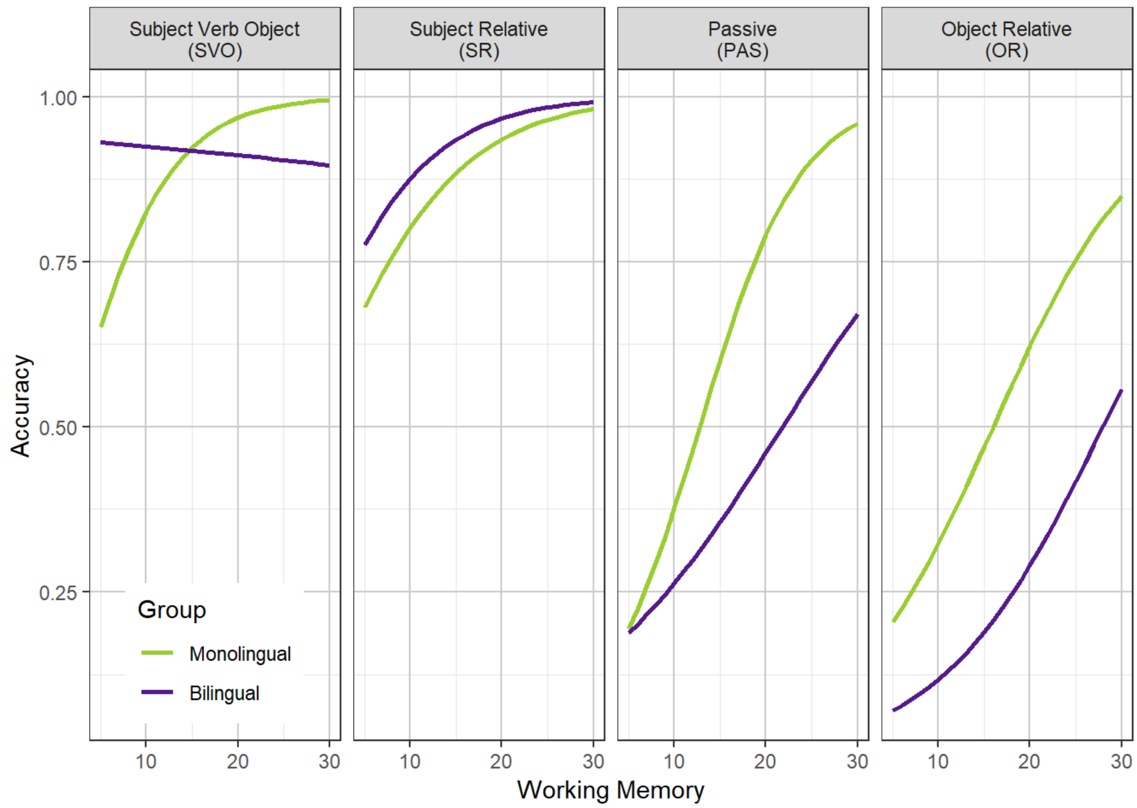

Figure 11. Model fit for accuracy by sentence type, group, and working memory.

Response Time

Descriptive data. Figure 12 displays the raw data for the response time for correct and incorrect agent selections for each sentence type. The data were plotted for each group separately, with purple representing the bilingual group and green representing the monolingual group. The size of each circle represents the frequency of correct or incorrect agent selection. In general, both groups responded more slowly when they were incorrect than when they were correct. Looking at the incorrect items (the dotted lines), we can see that the size of the circles (representing the frequency of incorrect responses) is larger for PAS and OR items for both groups. Consistent with the previous section, accuracy decreases for both groups for the noncanonical sentence types when compared to the canonical sentence types. The bilingual group had more frequent incorrect responses in the noncanonical sentences than the monolingual group. Monolinguals presented very similar response times for incorrect responses across all four sentence types. The response-time pattern for incorrect items was different for the bilinguals. Children in the bilingual group had faster response times for the incorrect noncanonical items as compared to the incorrect canonical sentences.

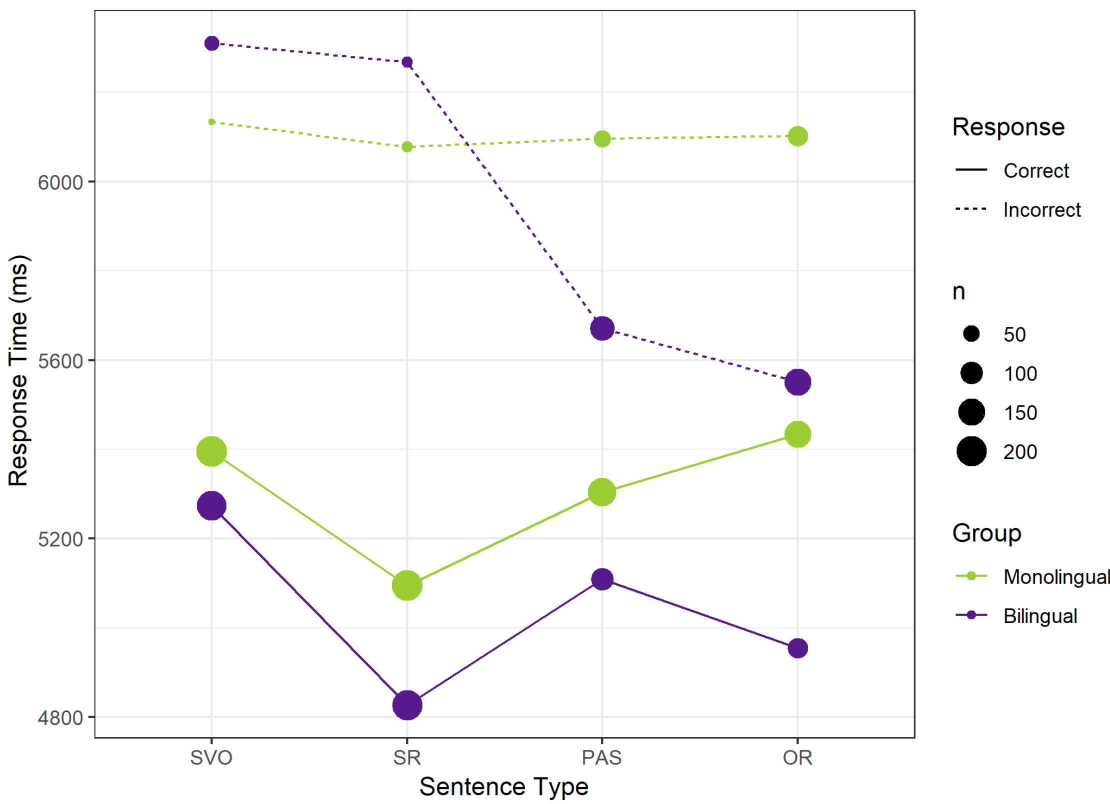

Figure 12. Raw data summarized for the response time by sentence type, group, and response accuracy. The size of the dots represents the frequency of the response.

Looking at the correct items (the solid lines), we can see that the size of the circles (representing the frequency of correct responses) is the inverse of the incorrect responses. That is, circle size is larger for PAS and OR items for monolinguals compared to bilinguals. Accuracy decreases for both groups for the noncanonical sentence types when compared to the canonical sentence types, with the bilingual group having less frequent correct responses in the noncanonical sentences than the monolingual group. Focusing on response time, we see that bilinguals responded faster than monolinguals across all sentence types, with the greatest group differences appearing to occur for the OR items.

LME analysis. The second part of Question 1 relates to online processing of canonical and noncanonical sentences. I asked, to what extent does complex working memory account for the response time of the interpretation of canonical and noncanonical English sentences by bilingual children when controlling for English syntactic knowledge. I answered this question by examining the response time of the children as a function of sentence comprehension accuracy. For the statistical analysis of the response time data, I employed linear mixed effects modeling that controlled for sentence and participant variance by including a random intercept for each sentence and for each participant. The dependent variable was response time. The fixed effects included group, sentence type, and response accuracy. Performance on the grammatical judgment task was modeled as a fixed effect to control for English syntactic knowledge. Nonresponses were recoded as 8000, the maximum time limit. Model building followed a top-down approach in which a complete model (four-way interaction and all subsuming interactions, Model RT 1) was tested against progressively simpler models (Table 7). The best fit model, Model RT 3 [ttfc ~ SenType * Group + SenType * Response + AWM_raw * Response + CASL_GJ_raw + (1 | SenID) + (1 | Participant)], (χ^2 (18) = 10.84, p = 0.0009958) included 3 two-way interactions: Sentence Type x Group, Sentence Type x Response Accuracy, and Working Memory x Response Accuracy. I refit this model using restricted maximum likelihood (REML, Table 8) and used it to create Figure 13. In this figure, the x-axis represents working memory performance; the y-axis represents response time; the colors represent group (green = monolingual, purple = bilingual); and line type represents response accuracy (solid = correct, dotted = incorrect).

Table 7. *Model Comparisons for Response Time*

	Df	AIC	BIC	logLik	Deviance	Chisq	Chi Df	Pr(>Chisq)
RT Model 4	17	30353	30447	-15160	30319	NA	NA	NA
RT Model 3	18	30345	30444	-15154	30309	10.84	1	0.0009958

Table 8. *Fixed Effects Estimates (Top, with Standard Error of Estimates in Parentheses) and Variance-Covariance Estimates (Bottom) for the Best fit Model of Response Time*

	Best fit model
(Intercept)	5396.21 (344.92)***
SenTypeSR	-308.86 (160.35)
SenTypePAS	-57.29 (161.41)
SenTypeOR	106.64 (163.53)
GroupB	-180.29 (170.63)
ResponseIncorrect	714.60 (212.11)***
AWM_raw	9.91 (11.35)
CASL_GJ_raw	-4.36 (5.64)
SenTypeSR:GroupB	-87.65 (125.85)
SenTypePAS:GroupB	-231.36 (130.27)
SenTypeOR:GroupB	-353.42 (130.22)**
SenTypeSR:ResponseIncorrect	150.41 (202.69)
SenTypePAS:ResponseIncorrect	-379.94 (175.32)*
SenTypeOR:ResponseIncorrect	-691.55 (177.15)***
ResponseIncorrect:AWM_raw	27.77 (8.48)**
AIC	30190.37
BIC	30289.52
Log Likelihood	-15077.18
Num. obs.	1824
Num. groups: SenID	48
Num. groups: Participant	38
Var: SenID (Intercept)	105466.98
Var: Participant (Intercept)	140143.03
Var: Residual	889233.19

$p < 0.001$, $p < 0.01$, $p < 0.05$

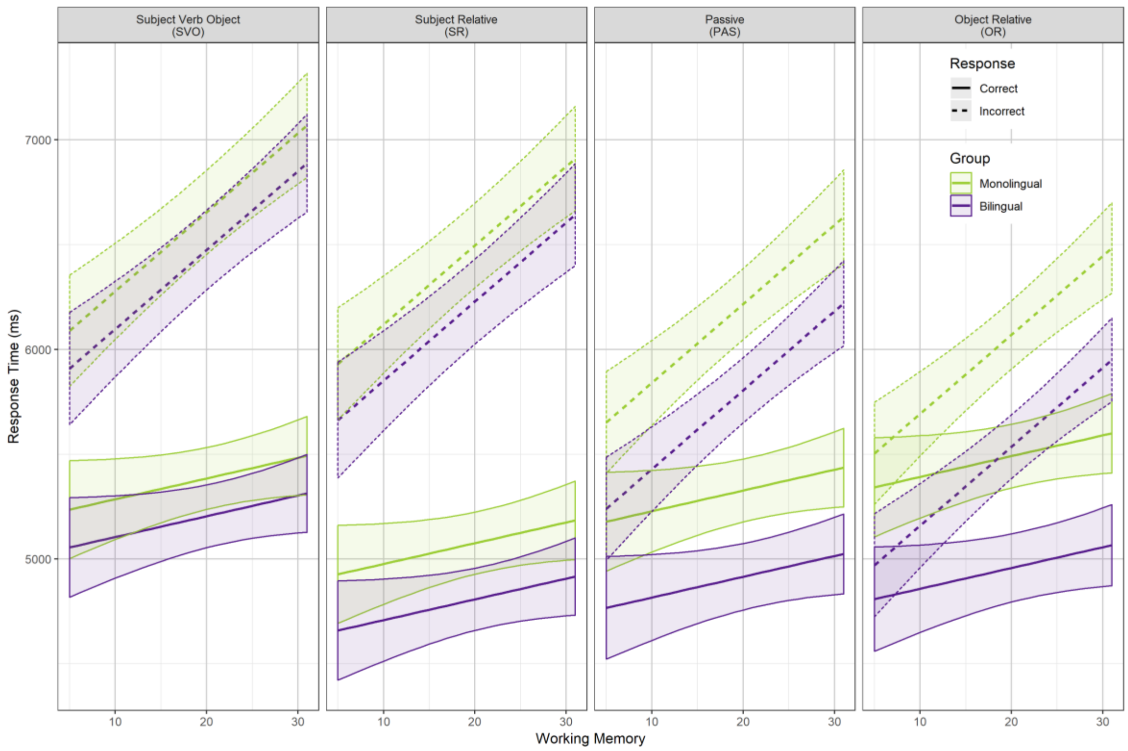

Figure 13. Best fit model for response time by working memory, group, response accuracy, and sentence type

The statistically significant interaction (F (3, 1775.59) = 10.7245, $p < 0.01$) between working memory and response accuracy is best represented by the slope of the lines and line types (dotted vs solid). It appears that, in general, children with low working memory had similar response times regardless of accuracy. However, as performance on the working memory measure increased, the difference between response time on correct or incorrect items also increased. The children with the highest working memory scores had the largest response time differences between correct and incorrect items. This pattern of results suggests that children with higher working memory were taking longer to decide when they were incorrect, whereas children with lower working memory were deciding more quickly.

The statistically significant interaction (F (3, 1778.03) = 8.7960, $p < 0.001$) between sentence type and response accuracy is best represented by the panels and line types (dotted vs solid) as shown in Figure 13. For incorrect items, response time decreased from canonical to noncanonical sentence types. But, for correct responses, response times for SVO and OR were slower than those for SR. This pattern of results suggests that when incorrect, children were slower to respond to canonical sentences than noncanonical sentences.

The statistically significant interaction (F (3, 1737.43) = 2.8625, $p < 0.05$) between sentence type and group is best represented by the panels and the line colors (green = monolinguals, purple = bilinguals) in Figure 13. Response time for the monolingual group did not differ across sentence types, but bilinguals had faster response times for SR, PAS, and OR than for SVO. When compared to monolinguals, bilinguals also had faster response times for the noncanonical sentences than the canonical sentences. The two groups did not differ in response times for the canonical sentences. Thus, the bilingual group processed noncanonical sentences differently than the monolingual group.

Fixation Time

LME analysis. The third part of Question 1 examined attentional focus during sentence comprehension. The question was to what extent does complex working memory account for the time spent looking at pictures representing the agent of the action in canonical and noncanonical sentences by monolingual and bilingual children when controlling for English syntactic knowledge. I used eye-tracking data for both correct and incorrect responses to assess the fixation time on the picture of the agent. The dependent variable was the total fixation time. The fixed effects were group (monolingual, bilingual), sentence type (SVO, SR, PAS, OR), accuracy (correct vs incorrect), and working memory ability (WJ AWM score). Performance on the grammatical judgment task was modeled as a fixed effect to control for English syntactic knowledge. Model fitting followed a top-down approach, with the most complex model compared to progressively simpler models using the LRT (Table 9). The best fit model (χ^2 (36) = 8.981, p = 0.02954) was Model FT 1 [TotFixTime ~ SenType*Group*AWM_raw*Response + CASL_GJ_raw + (1 | SenID)+ (1 | Participant)], which included a four-way interaction for sentence type, group, accuracy, and working memory. I refit the best fit model using REML (Table 10) and used the refit model to create Figure 14. In this figure, color represents the group (green = monolinguals, purple = bilinguals) and the line type represents the accuracy of the response (dotted = incorrect response, solid = correct response). The shaded areas represent standard errors. Each panel represents one of the four sentence types. Generally speaking, fixation time on the agent was higher for correct responses than for incorrect responses across all sentence types. Within the monolingual group, a pattern arose for increased fixation time on sentences including relative clauses (SR and OR) compared to sentences without a relative clause (SVO and PAS) regardless of word order (canonical vs noncanonical) or working memory.

Table 9. *Fixation Time – Comparison of Models*

	Df	**AIC**	**BIC**	**logLik**	**deviance**	**Chisq**	**Chi Df**	**Pr(>Chisq)**
FT Model 2	33	31137	31319	-15535	31071	NA	NA	NA

FT Model 1	36	31134	31332	-15531	31062	8.981	3	0.02954

Table 10. *Fixed Effects Estimates (Top, with Standard Error of Estimates in Parentheses) and Variance-Covariance Estimates (Bottom) for the Best fit Model of Fixation Time*

	Model 1
(Intercept)	2649.61 (696.66)***
SenTypeSR	395.38 (432.08)
SenTypePAS	-328.57 (517.75)
SenTypeOR	313.10 (504.29)
GroupB	-230.34 (986.92)
AWM_raw	-22.80 (21.04)
ResponseIncorrect	-2032.53 (560.34)***
CASL_GJ_raw	4.54 (9.90)
SenTypeSR:GroupB	-247.96 (649.45)
SenTypePAS:GroupB	1872.82 (855.99)*
SenTypeOR:GroupB	-1359.62 (937.22)
SenTypeSR:AWM_raw	1.06 (15.93)
SenTypePAS:AWM_raw	11.09 (19.17)
SenTypeOR:AWM_raw	2.46 (18.61)
GroupB:AWM_raw	22.38 (44.06)
SenTypeSR:ResponseIncorrect	91.22 (724.59)
SenTypePAS:ResponseIncorrect	209.41 (761.55)
SenTypeOR:ResponseIncorrect	-414.09 (739.94)
GroupB:ResponseIncorrect	353.72 (1015.56)
AWM_raw:ResponseIncorrect	51.46 (34.94)
SenTypeSR:GroupB:AWM_raw	8.28 (30.22)
SenTypePAS:GroupB:AWM_raw	-88.52 (38.73)*
SenTypeOR:GroupB:AWM_raw	45.84 (42.29)
SenTypeSR:GroupB:ResponseIncorrect	511.62 (1439.76)
SenTypePAS:GroupB:ResponseIncorrect	-2911.71 (1353.85)*
SenTypeOR:GroupB:ResponseIncorrect	426.95 (1403.63)
SenTypeSR:AWM_raw:ResponseIncorrect	-14.21 (43.03)
SenTypePAS:AWM_raw:ResponseIncorrect	-29.49 (42.51)
SenTypeOR:AWM_raw:ResponseIncorrect	-14.51 (40.76)
GroupB:AWM_raw:ResponseIncorrect	-29.09 (54.28)
SenTypeSR:GroupB:AWM_raw:ResponseIncorrect	-9.64 (78.78)
SenTypePAS:GroupB:AWM_raw:ResponseIncorrect	149.01 (68.83)*
SenTypeOR:GroupB:AWM_raw:ResponseIncorrect	-6.20 (69.96)
AIC	30799.51
BIC	30997.83

Log Likelihood	-15363.76
Num. obs.	1824
Num. groups: SenID	48
Num. groups: Participant	38
Var: SenID (Intercept)	222400.56
Var: Participant (Intercept)	357410.65
Var: Residual	1334191.08

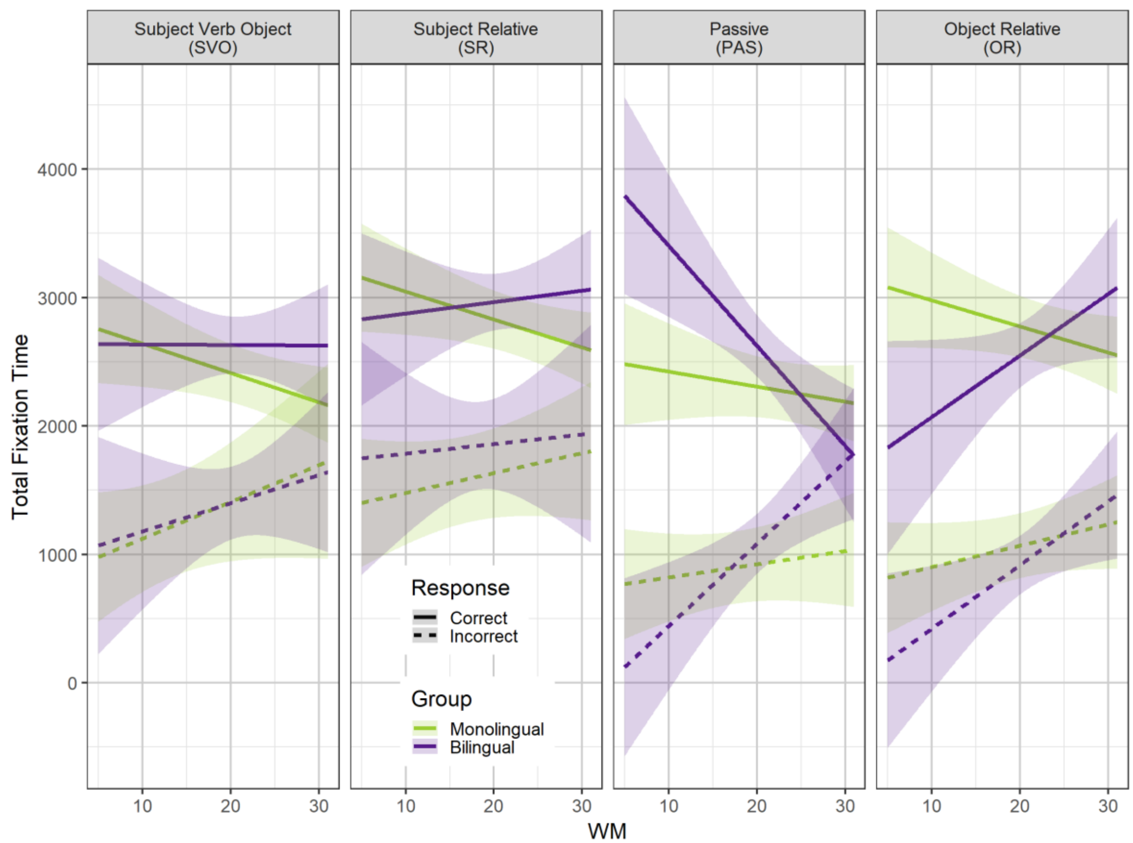

Figure 14. Best fit model for total fixation time.

The statistically significant four-way interaction (F (3, 1731.89) = 2.9558, $p <$ 0.05) is visible by focusing on the PAS and OR sentence type panels. Here we can see that the bilingual group's performance on the working memory measure has a stronger relationship with fixation time on the agent than that of the monolingual group. When responding correctly (solid line), bilinguals with low working memory attended to the agent for significantly more time than monolinguals with low working memory. For children with high working memory, fixation time on the agent was the same in both the monolingual and bilingual groups. Within the bilingual group, there was a significant difference between PAS and OR sentences. For correct responses to PAS sentences, among bilinguals, fixation time decreased as a function of increases in working memory. However, for incorrect responses, the bilinguals' fixation times increased as a function of increases in working memory. This increase in the bilingual group's fixation time on the agent was similar for incorrect responses in both OR and PAS sentence types. However, for correct responses to the OR sentences, fixation time increased instead of decreasing as in the PAS sentences when children obtained higher scores of working memory. This pattern of results suggests that for the bilingual group, the processing of OR sentences was significantly different than for the PAS sentence type. This pattern is unique to the bilingual group. The monolingual group seems to have responded differently to the presence of a relative clause in the sentences more so than to other word orders.

Syntactic Knowledge in Spanish and English Sentence Processing

The second question addressed the relationship between Spanish syntactic knowledge and the bilingual group's processing of sentences. The main research question was what is the additional role of Spanish syntactic knowledge in long-term memory on the comprehension of English canonical and noncanonical sentences? This question about the role of Spanish syntactic knowledge was analyzed in three parts similar to the previous analysis by exploring the same three outcome variables as in Question 1. For these analyses, only the bilingual group's data were used. Spanish syntactic knowledge was represented by children's responses to the Formulated Sentences subtest of the CELF-4 Spanish Edition (Semel, Wiig, & Secord, 2006). As in the previous analyses, I checked for multicollinearity by creating a correlation matrix (Figure 10) of the measures used in all the analyses and by calculating the VIF for each outcome variable (accuracy, response time, fixation time). The VIF was less than 2 for each, indicating that multicollinearity was not a concern. For all the following analyses, model building began using a model similar to that of the best fit model found in the group comparisons in the previous analyses. Spanish syntactic knowledge was added to each of the best fit models and compared using the LRT.

Recall that in the first question, which included both groups in the analysis, the best fit model for accuracy consisted of a three-way interaction between group, sentence type, and working memory. For this second question, I divided the data to include only the bilinguals and created a similar model, Accuracy (Table 11), by removing the group term. I compared this model to a second model, to which I added Spanish syntactic knowledge as a fixed effect, using a LRT. Adding Spanish syntactic knowledge did not improve the model fit (χ^2 (12) = 0.2206, p = 0.6386). This suggests that Spanish syntactic knowledge did not further explain the variance for accuracy when accounting for working memory and sentence type.

I followed a similar procedure for response time by subsetting the data to include the bilingual group only and removing the group term from the best fit model from the previous analysis. This new model, Response Time (Table 11), included 2 two-way interactions, Sentence Type x Accuracy, and Working Memory x Accuracy. I used a LRT to compare this model to a second model that added Spanish syntactic knowledge as a fixed effect. Adding Spanish syntactic knowledge did not improve the model fit model (χ^2 (15) = 0.022, p = 0.882). This suggests that adding Spanish syntactic knowledge did not further explain the variance for response times when accounting for working memory, accuracy of the response, and sentence type.

For the third part of question two, I analyzed the fixation time data for only the bilingual group. The procedure was the same as for response time. I created the model, Fixation Time, which included a three-way interaction between working memory, sentence type, and accuracy of the response. Using a LRT, I compared this model to a second model that added Spanish syntactic knowledge as a fixed effect. Adding Spanish syntactic knowledge did not improve the model fit model (χ^2 (21) = 0.8755, p = 0.3494). This suggests that Spanish syntactic knowledge did not further explain the variance for fixation time when accounting for working memory, accuracy of the response, and sentence type.

The pattern of results was consistent across the three outcome measures (accuracy, response time, fixation time). Adding Spanish syntactic knowledge to the models did not improve the model fits, indicating that Spanish syntactic knowledge did not provide additional information to the model beyond what was provided by the other variables.

Table 11. *Bilingual Model Comparisons*

Model	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Accuracy	11	792.5	845.4	-385.2	770.5	NA	NA	NA
+ Spanish syntactic knowledge	12	794.2	852	-385.1	770.2	0.2206	1	0.6386
Response time	14	15320	15388	-7646	15292	NA	NA	NA
+ Spanish syntactic knowledge	15	15322	15394	-7646	15292	0.02202	1	0.882
Fixation time	20	15556	15653	-7758.1	15516	NA	NA	NA
+ Spanish syntactic knowledge	21	15557	15658	-7757.7	15515	0.8755	1	0.3494

DISCUSSION

The ability to comprehend complex sentences has been shown to involve a combination of cognitive and linguistic processes in monolingual children. The purpose of this study was to examine the role of complex auditory working memory on sentence comprehension (canonical and noncanonical) while controlling for syntactic knowledge in children with varying language experience (monolingual and bilingual children) using both offline (behavioral) and online (eye-tracking) measures. Monolingual and bilingual children between the ages of 9 and 14 completed the "Whatdunit" task (Montgomery et al., 2015) while having their eye movements recorded. For this task children listened to four types of sentences: two canonical sentences (SVO, SR) and two noncanonical sentences (PAS, OR). The task was developed to measure sentence comprehension in a manner that maximized reliance on syntactic structure (word order) and minimized reliance on semantics for interpretation. The sentences, which were all semantically ambiguous and syntactically reversible, featured an inanimate object in the agent position acting upon another inanimate object in the patient position. The children were instructed to select the agent by clicking on one of three pictures on a screen.

I collected a combination of offline and online measures of performance. The offline measure of sentence processing was the children's accuracy on the agent selection task. The online measures were response time of the selection and the total fixation time on the agent as measured by eye-tracking. Linear mixed effect modeling was used for each analysis to account for sentence variance and for participant variance.

Accuracy

Language experience differentially influenced comprehension of canonical and noncanonical sentences. The mean accuracy scores for the monolingual children in this study were quite similar to those in the Montgomery et al. (2016) study (SVO - 87.9% vs. 88.9%, SR - 89%, vs. 84.5%, PAS - 73.3% vs. 66.1%, and OR - 62.9% vs. 58.1%, respectively). The bilingual children in this study had accuracy scores for the canonical sentences (SVO - 82.5% and SR - 89.5%) that were similar to the monolinguals, but their mean accuracy for the noncanonical sentences (PAS - 44.3% and OR - 34.2%) was somewhat lower than their monolingual peers. The fact that the bilingual children performed much like their monolingual peers on the canonical sentences indicates that they understood the task and were able to perform it correctly. Their weaker performance on the noncanonical sentences suggests that the bilingual children had less knowledge of and/or less experience with passive and object relative sentences than their monolingual peers.

For descriptive purposes, we were interested in the relationships among the four tasks. The monolingual group obtained high correlations between performance on the canonical (SVO and SR) sentences (r = .93) and the noncanonical (PAS and OR) sentences (r = .88). These findings were consistent with Montgomery et al. (2016), who reported correlations of .84 between the two types of canonical sentences and .89 between the two types of noncanonical sentences. The sentence correlations for the bilingual group were somewhat different than those for the monolingual group in this study. The correlation between canonical sentences in this study was somewhat lower for the bilinguals (r = .65), while the correlation between the two noncanonical sentences was similar ($r = .85$) to the correlations for the monolingual children. The lower correlation between the canonical sentences could have occurred because one bilingual participant had 0% accuracy on SVO, but 92% accuracy on SR. If we remove that participant, the correlation between the canonical sentences for the bilingual children would be 0.93. Thus, we see no reason to suspect that the canonical sentences represented different categories of linguistic operations for the bilingual group.

The present study uniquely contributed to the literature by exploring the role of working memory in sentence comprehension while also controlling for syntactic knowledge. As expected, both groups were statistically more accurate on the canonical sentences than the noncanonical sentences. Generally, children's interpretation of canonical and noncanonical sentences benefited from having higher working memory. However, this effect was stronger for interpretation of noncanonical sentences. The significant three-way interaction indicated that working memory played a stronger role in interpreting noncanonical sentences compared to canonical sentences. These results suggest that the role of working memory in comprehending canonical sentences was similar for monolingual and bilingual children

However, there were important group differences regarding the role of working memory in monolinguals' and bilinguals' sentence comprehension of noncanonical sentences. Though we controlled for syntactic knowledge, the bilingual children in this study appeared to have poorer comprehension of noncanonical sentences than the monolingual children. The results for the two noncanonical sentences (PAS and OR) were somewhat different. For the passive sentences, individuals in either group with low working memory had similarly poor comprehension accuracy. For both the monolingual and bilingual groups, as working memory scores increased, there was a concomitant increase in comprehension accuracy scores. However, the curve was significantly steeper for the monolinguals, suggesting that better working memory corresponds with greater improvements in PAS sentence comprehension for the monolingual group than the bilingual group. The pattern for the OR sentences was different. The bilingual group had poorer accuracy on the OR sentences for all levels of working memory when compared to the monolingual group. In general, when controlling for English syntactic knowledge, performance on the PAS and OR sentences was affected more by working memory ability than performance on the SVO and SR sentences. These results suggest that working memory plays a stronger role in comprehending noncanonical sentences than canonical sentences. Additionally, stronger working memory seems to have a stronger faciliatory effect on accuracy in monolinguals than bilinguals.

These analyses explain the role of working memory on sentence comprehension as an offline measure. However, offline measures do not fully explain the processing demands during sentence comprehension. Therefore, I analyzed the online processing of sentences using two measures, response time and fixation time. For both of these analyses, I decided to analyze the data for both correct and incorrect responses because omitting incorrect responses would result in the omission of about 70% of the noncanonical data. Keeping these data, I was able to analyze processing differences between correct and incorrect responses. Considering that bilingual children had lower accuracy on the noncanonical sentences, I expected higher fixation times and longer response times in the bilingual group for the noncanonical sentences, but similar times for the canonical sentences.

Response Times

The analysis of response times yielded three significant two-way-interactions: Group x Sentence Type, Working Memory x Accuracy, and Sentence Type x Accuracy. As expected, analysis of the response times for canonical sentences yielded similar results for both groups. However, for the noncanonical sentences, bilinguals had faster response times than monolinguals. Generally, children responded more slowly to incorrect responses than correct responses. Response times also increased as a function of increases in performance on the working memory task. The interaction between accuracy and working memory on response times indicated that children in both groups who had lower working memory scores had response times that were about 600ms slower for incorrect responses than correct answers, whereas children with higher working memory were about 1600ms slower for incorrect responses than for correct responses.

Recall that when no selection occurred, the lack of response was coded as the maximal time limit. It is likely that these slower response times in children with higher working memory were due to both slow responses and no responses, signifying increased processing time needed to decide on the correct answer with noncanonical sentences. Children with lower working memory, as noted in the previous analysis, had very poor accuracy, which, paired with their quicker response times indicates little processing effort with a poor outcome.

Response times did vary according to sentence type and accuracy. Correct responses for SR sentences were faster than both SVO and OR. That response times were slower for SVO than for SR was an unexpected result. This may be due to an order effect. After randomizing the sentence order, SR sentences occurred first in both blocks of sentences presented. In Rosenberg et al.'s study (Rosenberg, Noonan, DeGutis, & Esterman, 2013) of adults, response times increased with greater sustained attentional demands. A similar effect may have occurred in this study with sustained attention over time affecting later tasks and resulting in an advantage for the SR stimuli. When collapsing across the two groups, it is clear that the children generally responded more quickly to the noncanonical sentences than canonical sentences when they selected the agent incorrectly. In other words, the children's incorrect responses were faster with increasing sentence complexity. An opposite trend occurred for correct responses, for which response times tended to be slower with increasing complexity. Contrary to expectation, the bilingual children did not demonstrate increased processing effort as measured by response time with increasing sentence complexity.

I speculate that children with higher working memory were able to make use of this resource to deliberate the correct answer for greater amounts of time in a similar process to the adults with high working memory in the Traxler (2007) study, who had longer reading times than the adults with low working memory. If the bilingual children in this study were not able to chunk the information into higher levels of abstractions due to their limited knowledge of noncanonical structures, perhaps they were selecting their answers based on "good-enough" representations of the sentences and spending less time deliberating their choices. The second online processing measure was the length of time the pupil was fixated on the agent as recorded by the eye-tracker. This resulted in a significant four-way interaction between working memory, group, sentence type, and accuracy. As expected, the children attended more to the agent when they were correct than when they were incorrect.

Fixation Time

There were some significant group differences in the fixation time data. For monolinguals, there was a slight downward trend in fixation time from low memory performance to high memory performance across all sentence types. These findings suggest that children with higher working memory held the image of the agent less in their focus of attention than children with lower working memory by about 500ms. This suggests that the children with lower working memory had to maintain the agent within their focus of attention for a longer duration due to their limited resources. Consistent with Cowan's model of memory (Cowan, 1999, 2001, 2014), it seems that monolingual children with higher working memory were able to more rapidly chunk into meaningful multi-word units and offload the information into long-term memory. That is, they spent less of their focus of attention (as measured by time) on the image of the agent because they were able to move on to a higher level of abstraction beyond the concept of the item that represented the agent. Monolinguals also attended to the agent more for the relative clause sentences than the SVO and PAS sentences. This suggests that the sentences with relative clauses required more processing effort than their counterparts.

For the bilingual children, there were significant differences in the relationship between working memory and fixation time across sentence types. For the SVO sentences, bilingual children fixated on the agent for about 2600ms regardless of their working memory ability. For the SR sentences, the bilingual children had a slight increase in fixation time on the agent as a function of increases in working memory, which was not significant. Therefore, this trend was indistinguishable from their SVO pattern. However, for the PAS sentences, bilinguals with low working memory had the highest duration (approximately 3750ms) of attention on the agent for correct responses, indicating higher levels of cognitive effort were needed to respond correctly. Note also that the bilingual children with the highest working memory had the shortest fixation duration (approximately 1750ms) on the agent for PAS sentences and that this fixation overlapped for both correct and incorrect responses. The trajectories for the other noncanonical sentence type, OR, depict parallel lines for correct and incorrect responses, with distinctly greater fixation time on the agent for correct responses than incorrect responses. It is likely that the high fixation time in PAS of the bilinguals with low working memory is indicative of maintaining the agent in the focus of attention and processing effort. Notice also that for these children, when they were incorrect, they had near 0ms of fixation time on the agent. Recall that the accuracy was also very poor for PAS sentence type. It seems possible that the bilingual children with low working memory, fixated on one image due to the lack of sufficient resources. Perhaps a separate analysis of the other nouns would provide an answer as to the possibility of this.

For the OR sentences, bilingual children with low working memory fixated on the agent for the same amount of time as the bilingual children with high working memory fixated on the agent in the PAS sentences. The finding that fixation time on the agent increased as a function of increases in working memory suggests that children with high working memory attended to the agent more and expended more effort to process the agent in the OR sentences than children with low working memory. Accuracy was poorest for this sentence type and approaching chance levels for some of the children. Based on the accuracy when accounting for working memory, it is more likely that the children with higher working memory were the children selecting at above chance levels.

It is possible that given the difficulty of the sentence and the limited experience with this type of sentence, even the children with high working memory had fewer resources to allocate to eye movements and needed to sustain more of their attention on the image of the agent.

One issue with the analysis of the fixation time data is that the correct responses for bilinguals consisted of a mean of 34% for the OR sentences. Though this may be at chance levels of accuracy, Figure 8 showed that the bilingual children obtained scores ranging from 0% to 92%. However, that does mean that the data for correct responses are of a much smaller sample size. Because language experience can be quite variable for bilingual children, it would be beneficial to explore the data of bilingual children and see which factors may be related for the good comprehenders and poor comprehenders of the noncanonical sentences. It is possible that the inconsistency of the results between PAS and OR may be due to unique factors at the individual level, such amount of time spent listening to English, age of acquisition, and the type of environments in which they hear each language.

Spanish Syntactic Knowledge

Finally, I explored whether the Spanish syntactic knowledge of the bilingual children played a role in their sentence comprehension performance. I thought it was possible that the relationships among syntactic knowledge, WM, and sentence comprehension might differ with respect to Spanish as compared to English knowledge. The results indicate that Spanish syntactic knowledge did not significantly account for changes in accuracy, response time, or fixation time when we controlled for working memory and English syntactic knowledge. This result is consistent with the findings from the previous analyses showing that working memory accounted more strongly for the variance in sentence comprehension performance than English syntactic knowledge. If English syntactic knowledge was not as important as WM in sentence comprehension of English sentences, it should not be surprising that Spanish syntactic knowledge was of minimal importance too.

Models of Sentence Comprehension

These behavioral results of sentence comprehension accuracy are consistent with the GEM model (Montgomery et al. 2018) and *chunk-and-pass*. Recall that in the GEM model (Montgomery et al., 2018), working memory mediates the relationship between long-term language knowledge and sentence comprehension. The results for the monolingual children in the current study were similar to the results for the monolingual children in the Montgomery et al. (2018) study. I believe the monolingual children in this study were able to chunk information contained in the sentence stimuli into relevant units using their long-term memory of language knowledge for both canonical and noncanonical sentences. However, it appears that the bilingual children engaged in a similar process only for the canonical sentences. It is possible that the bilingual group did not have as much familiarity and thus long-term memory knowledge of the noncanonical structure necessary to facilitate the creation of these chunks. In other words, lack of familiarity reduced their ability to use the top-down predictive nature of the *chunk-and-pass* model. If that were the case, then working memory would be taxed by having to hold a greater number of smaller chunks. This may explain why the bilingual children with similar levels of working memory as the monolingual children demonstrated less accuracy while still having similar rates of accuracy increases as a function of working memory increases.

Implications

Previous studies of language comprehension in bilinguals have reported slower lexical access, as shown through slower response times in lexical decision tasks (DeAnda, Poulin-Dubois, Zesiger, & Friend, 2016; Ivanova & Hallowell, 2012; Shook, Goldrick, Engstler, & Marian, 2014). This is thought to result from cross-linguistic interference. However, we know of no other studies comparing the response time in comprehending orally presented sentences by monolingual and bilingual children. That response times were faster for noncanonical sentences in the bilingual group compared to the monolingual group in this study could be a result of the bilinguals' experiences with a flexible word-order language. Though the bilingual group's accuracy was lower than that of the monolingual group, across both correct and incorrect answers, the bilingual group maintained a significantly faster response time than the monolingual group. It is important to view this information keeping in mind that some children in both groups were performing at the full range of accuracy, such that some children were performing at chance levels, but this would have been a minority of these children. This means that for most children the processing information is relevant.

Working memory was an important factor in noncanonical sentence comprehension for both groups. However, the monolingual group's sentence comprehension performance apparently benefited from greater working memory capacity than the bilingual group even when controlling for English syntactic knowledge. This indicates that for bilinguals in this age group, other factors are still hindering their sentence comprehension. The bilingual group is not a homogenous group and it is possible that other factors relating to their language experiences, such as English proficiency, age of onset, amount of time spent listening and speaking English, and the quality of their English experiences may be affecting their knowledge specific to noncanonical sentences. Additionally, we used only one measure of syntactic knowledge, which perhaps did not best capture the syntactic skills needed for the sentence comprehension used in this study. This limitation will be further explained below.

Limitations

One limitation of this study includes the smaller numbers of participants who obtained scores at the upper and lower ends of the working memory measure, which decreases the generalizability to other children with high or low working memory. This problem can be addressed with a larger sample size. Additionally, working memory was measured using only one assessment, which assessed the ability to hold words in mind, to categorize them, and then to repeat them in order within two categories. It could be more advantageous to use multiple measures of working memory (Waters & Caplan, 2003). Specifically, measures like the reading span task, which require participants to remember words within sentences, may have yielded different results.

Another limitation of this study relates to the presentation order. The dataset was obtained from a larger study using fNIRS. Analysis of hemoglobin concentration levels is facilitated by blocking items by sentence type for presentation. Therefore, the sentences were not fully randomized, which may have influenced some of the results. In our pseudorandomization, the SR sentences were presented first in both presentation blocks. The significantly faster response times to the SR items than the SVO items may not have occurred if the items were fully randomized for presentation. We thought something about the relative clause would make it more difficult, and thus slower, to process. However, we see that for the fixation time data indicated that the monolinguals fixated on the agent longer for the SR items than the SVO items, which is consistent with the expectation that SR sentences are slightly more challenging to process than SVO sentences. In regard to accuracy, I found small effect sizes for the accuracy in this study (SVO - .04, SR - .21, PAS - .18, OR - .13) compared to the original Montgomery et al. (2018) study, though SR had the largest effect size of the four sentence types. The experimental task may benefit from randomization such that sentences of the same type are not blocked together.

Additionally, our measure of syntactic knowledge (the grammatical judgment task) may not have been the best index of the level of grammatical knowledge that affects sentence comprehension. The CASL grammatical judgment raw score yielded a nonsignificant beta value of 0.02. We built this measure into the model as a control because the monolingual children scored significantly better. However, this measure did not provide statistically significant levels of explanatory information. The problem is that the CASL grammatical judgment task contained many more items that focus on grammatical morphology than items that address complex syntax. Only 12% of the items were noncanonical in nature, and many of the children reached ceiling before encountering most of those items. Perhaps a measure that better assessed knowledge of complex syntax would have accounted for more of the variance in sentence comprehension. Unfortunately, we know of no formal measures of syntactic knowledge. What is needed is a standardized measure of grammatical judgment of various types of complex sentences. A related point is that the English and Spanish measures of syntactic knowledge were different. We used the Formulated Sentences measure of CELF Spanish because there were no grammatical judgment measures in Spanish that we were aware of. In future investigations, it would be better to construct similar measures of grammatical judgment in English and Spanish.

Future Directions

Because of the inconsistency of results between the two online processing measures (response time and fixation time), analyzing other eye-tracking measures (number of fixations, revisits, and saccades) of online processing could inform the results of the present study. Additionally, the data can be analyzed using exact timestamps of each word in the sentence. Using these timestamps, specific portions of the sentences can be analyzed by fixation times, fixation counts, and saccades to provide a more discriminating measure than the currently used gross measure of total fixation time across the entire sentence presentation. This will also be more similar to how reading research using eye-tracking measures analyze individual words or clauses within a sentence. Furthermore, I plan to explore differences within the bilingual children as they relate to language experience to see if any patterns arise between good and poor comprehenders.

Conclusion

The key findings of this study were that the relationships between working memory and canonical sentence comprehension were similar for the monolingual and bilingual children. However, the processes underlying noncanonical sentence comprehension differed for the monolingual and bilingual children, even though we controlled for their English syntactic knowledge. The bilinguals comprehended noncanonical sentences with less accuracy even in cases in which children in the two groups had similar levels of working memory. Despite the fact that we controlled for English syntactic knowledge in the statistical model, it did not play an important role in sentence comprehension. This could be a reflection of our measure of syntactic knowledge, which focused more on grammatical morphology than complex syntactic structures. It is likely that the bilingual children had less experience with the noncanonical sentence structures. Because of this, greater working memory abilities were insufficient for obtaining high comprehension accuracy scores.

Another important finding relates to our use of two measures of online processing. The eye-tracking measures revealed information about the focus of attention in working memory. In addition, response time measures enabled us to examine children's processing time and cognitive effort. We found that monolingual children with better working memory had lower fixation times on the agent, together with slower response times. It appears that these children more quickly encoded the agent (indicated by shorter fixation time) and then spent more time thinking about the multi-word chunks (especially in the noncanonical sentences) before selecting the agent. This informs our understanding of the relationship between chunking, focusing attention, and sentence processing. Unfortunately, there were inconsistencies in the bilingual data eye-tracking data that are difficult to explain. The bilinguals were poorer at identifying the agent in noncanonical sentences, but they had widely varying fixation times on the agents and faster response times. Thus, the eye-tracking measures did not provide clear evidence of the cognitive processing abilities supporting sentence comprehension in bilinguals, especially those related to the focus of attention.

REFERENCES

Adams, C. (1990). Syntactic comprehension in children with expressive language impairment. *British Journal of Disorders of Communication*, *25*(2), 149–171. https://doi.org/10.3109/13682829009011971

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Bishop, D. V. M., Bright, P., James, C., Bishop, S. J., & Van Der Lely, H. K. J. (2000). Grammatical SLI: A distinct subtype of developmental language impairment? *Applied Psycholinguistics*, *21*(2), 159–181. https://doi.org/10.1017/S0142716400002010

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton Jr. (Eds.), *The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond* (pp. 51–76). Brighton, England: Psychology Press.

Bosker, R. J., Snijders, T. A. B., & Guldemond, H. (2003). User’s manual PINT. Program and manual available at: http://www.gamma.rug.nl/

Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test*. Chicago, IL: The Riverside Publishing Co.

Carrow-Woolfolk, E. (2017). *Comprehensive Assessment of Spoken Language, Second Edition*. Torrence, CA: WPS.

Christiansen, M.H., & Chater, N. (2016). *Creating Language: Integrating Evolution, Acquisition, and Processing*. MIT Press

Christiansen, Morten H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39*. https://doi.org/10.1017/S0140525X1500031X

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Cowan, N. (1999). An Embedded-Processes Model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2014). Working memory underpins cognitive development, dearning, and education. *Educational Psychology Review*, *26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

DeAnda, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). Lexical processing and organization in bilingual first language acquisition: Guiding future research. *Psychological Bulletin*, *142*(6), 655–667. https://doi.org/10.1037/bul0000042

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1077–1088. https://doi.org/10.1044/1092-4388(2010/10-0093)

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, *25*(4), 529–557. https://doi.org/10.1017/S0272263103000238

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*(01), 101. https://doi.org/10.1017/S1366728906002847

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551. https://doi.org/10.1016/S0022-5371(81)90165-1

Fernández, E. M. (2003). *Bilingual sentence processing*. John Benjamins. https://doi.org/10.1075/lald.29

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Fox, J., Weisberg, S. (2019). *An R companion to applied regression*, Third edition. Sage, Thousand Oaks CA. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gaulin, C. A., & Campbell, T. F. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills*, *79*(1), 55–64. https://doi.org/10.2466/pms.1994.79.1.55

Gillam, R. B., Montgomery, J. W., Evans, J. L., & Gillam, S. L. (2019). Cognitive predictors of sentence comprehension in children with and without developmental language disorder: Implications for assessment and treatment. *International Journal of Speech-Language Pathology*, *21*(3), 1–12. https://doi.org/10.1080/17549507.2018.1559883

Gillam, R. B., & Pearson, N. A., (2004). *TNL: Test of Narrative Language*. Austin, Tex.: Pro-Ed.

Gillam, R. B., & Pearson, N. A., (2017). *TNL-2: Test of Narrative Language – Second Edition. A*ustin, TX: Pro-Ed.

Gutiérrez-Clellen, V. F., Calderón, J., & Weismer, S. E. (2004). Verbal working memory in bilingual children. *Journal of Speech, Language, and Hearing Research*, *47*(4), 863–876. https://doi.org/10.1044/1092-4388(2004/064)

Gutiérrez-Clellen, V. F., Restrepo, M. A., & Simón-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1209–1223.

Hernandez, A. E., Bates, E. A., & Avila, L. X. (1994). On-line sentence interpretation in Spanish–English bilinguals: What does it mean to be "in between"? *Applied Psycholinguistics*, *15*(4), 417–446. https://doi.org/10.1017/S014271640000686X

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hox, J. J., Moerbeek, M., Schoot, R. van de, Moerbeek, M., & Schoot, R. van de. (2017). *Multilevel analysis : Techniques and applications, Third Edition*. Routledge. https://doi.org/10.4324/9781315650982

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288. https://doi.org/10.1016/j.actpsy.2007.06.003

Jasinska, K. K., & Petitto, L. A. (2013). How age of bilingual exposure can change the neural systems for language in the developing brain: A functional near infrared spectroscopy investigation of syntactic processing in monolingual and bilingual children. *Developmental Cognitive Neuroscience*, *6*, 87–101. https://doi.org/10.1016/j.dcn.2013.06.005

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. http://dx.doi.org.dist.lib.usu.edu/10.1037/0033-295X.87.4.329

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech*, *32*(1), 1–23. https://doi.org/10.1177/002383098903200101

Kreft, I. G. G. (1996). Are multilevel techniques necessary? An overview, including simulation studies. Unpublished Report, California State University, Los Angeles, CA.

Lahousse, K., & Lamiroy, B. (2012). Word order in French, Spanish and Italian: A grammaticalization account. *Folia Linguistica*, *46*(2). https://doi.org/10.1515/flin.2012.014

Magnusson, K. (2018). powerlmm: Power analysis for longitudinal multilevel models. R package version 0.4.0. https//CRAN.R-project.org/package=powerlmm

Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 269–288. https://doi.org/10.1044/1092-4388(2008/07-0116)

Montgomery, J. W., Evans, J. L., Gillam, R. B., Sergeev, A. V., & Finney, M. C. (2016). “Whatdunit?” Developmental changes in children’s syntactically based sentence interpretation abilities and sensitivity to word order. *Applied Psycholinguistics*, *37*(06), 1281–1309. https://doi.org/10.1017/S0142716415000570

Montgomery, J. W., Gillam, R. B., Evans, J. L., & Sergeev, A. V. (2017). “Whatdunit?” sentence comprehension abilities of children with SLI: Sensitivity to word order in canonical and noncanonical structures. *Journal of Speech Language and Hearing Research*, 1. https://doi.org/10.1044/2017_JSLHR-L-17-0025

Montgomery, J. W., Evans, J. L., Fargo, J. D., Schwartz, S., & Gillam, R. B. (2018). Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *61*(12), 2950–2976. https://doi.org/10.1044/2018_JSLHR-L-17-0421

Morett, L. M., & Macwhinney, B. (2013). Syntactic transfer in English-speaking Spanish learners*. *Bilingualism: Language and Cognition*, *16*(1), 132–151. https://doi.org/10.1017/S1366728912000107

Norbury, C., Bishop, D., & Briscoe, J. (2002). Does impaired grammatical comprehension provide evidence for an innate grammar module? *Applied Psycholinguistics*, *23*(2), 247-268. doi:10.1017/S0142716402002059

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 280–285. http://dx.doi.org.dist.lib.usu.edu/10.1037/a0014276

Qian, Z., Garnsey, S., & Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language,* *Cognition and Neuroscience*, *33*(2), 227–254. https://doi.org/10.1080/23273798.2017.1379606

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosenberg, M., Noonan, S., DeGutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: A novel gradual-onset continuous performance task. *Attention, Perception, & Psychophysics*, *75*(3), 426–439. https://doi.org/10.3758/s13414-012-0413-x

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437–442. https://doi.org/10.3758/BF03208784

Scheepers, C., & Crocker, M. W. (n.d.). Constituent order priming from reading to listening: A visual-world study. In M. Carreiras & C., Clifton Jr. (Eds.), *The on-line study of sentence comprehension: eyetracking, ERPs, and beyond* (pp. 167–185). Brighton, England: Psychology Press.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, *17*(1), 133–145. https://doi.org/10.1016/0093-934X(82)90010-4

Shook, A., Goldrick, M., Engstler, C., & Marian, V. (2014). Bilinguals show weaker lexical access during spoken sentence comprehension. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-014-9322-6

Spybrook, J., Bloom, H., Congdon, R., Hill, C. Martinez, A., & Raudenbush, S. W. (2011). Optimal design plus empirical evidence: Documentation for the "Optimal

Design" software (version 3.0). Retrieved from http://hlmsoft.net/od/od-manual-20111016-v300.pdf

Sussman, R. S. (2006). *Verb -instrument information during on -line processing* (Ph.D.). University of Rochester, United States -- New York. Retrieved from http://search.proquest.com/pqdtglobal/docview/305272148/abstract/E1B774DABE084EA1PQ/2

U.S. Census Bureau (2017). *Characteristics of people by language spoken at home, 2017 American Community Survey 1-year estimates*. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

van der Lely, H. K. J. (1996). Specifically language impaired and normally developing children: Verbal passive vs. adjectival passive sentence interpretation. *Lingua*, *98*(4), 243–272. https://doi.org/10.1016/0024-3841(95)00044-5

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *CTOPP-2: Comprehensive Test of Phonological Processing – Second Edition.* Austin, TX: Pro-Ed.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 550–564. https://doi.org/10.3758/BF03195534

Wiig, E., Secord, W. A., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals–Fourth Edition–Spanish*. San Antonio, TX: The Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition, 35*(5), 1107–1121.

APPENDIX

Summary of assessments

Measure	Standardization population	Target Age	Type of Measures	Reliability		
				Internal Consistency (content)	Test-retest (time)	Rater (scorer)
Symbolic Memory (UNIT)	Representative of 1995 U.S. Census	5-17	Non-verbal, symbolic memory	.85	.72(.68*)	NA
Auditory Working Memory (WJ-III)		2-19	Verbal working memory		.88	
Non-word Repetition (CTOPP-2)	Representative of 2010 U.S. Census	4-24	Non-word repetition	77	77	96
Narrative Comprehension (TNL-2)	Representative of US population reported in *ProQuest Statistical*	4;0 to 15;11	English narrative comprehension	.81	.85	99

	Abstract of the United States 2015 (ProQuest, 2015) and the *Digest of Education Statistics 2014* (Synder, deBrey, & Dillow, 2016)					
Grammaticality Judgment (CASL-2)	Representative of 2012 U.S. Census	7-21	Syntactic judgment and construction	.98-.99	.87(.84*)	.86
Antonyms (CASL-2)	Representative of 2012 U.S. Census	5-21	Word knowledge, retrieval, and oral expression (decontextualized)	.92-.98	.94(.90*)	.92
Conceptos y Siguiendo Oraciones (Concepts and Following Directions, CELF-4 Spanish Edition)	Representative of Hispanic population in the US reported in the *Current population survey, October*	5-12	Comprehension of increasingly complex spoken directions	.88	.82(.81)	NA

	2002: School Enrollment Supplemental File					
Recordando Oraciones (Recalling Sentences, CELF-4 Spanish Edition)	Representative of Hispanic population in the US reported in the *Current population survey*, *October 2002: School Enrollment Supplemental File*	5-21	Recall and repeat spoken sentences with increasing length and complexity	.95	.89(.85)	NA
Formulación de Oraciones (Formulating Sentences, CELF-4 Spanish Edition)	Representative of Hispanic population in the US reported in the *Current population survey*, *October 2002: School Enrollment Supplemental File*	5-21	Formulation of complete, grammatically correct spoken sentences	.85	.77(.75)	.81
Clases de Palabras – Receptivo (Word Classes -	Representative of Hispanic population in the US reported in the *Current*	9-21	Understand	.84	.76(.72)	.99

Receptive, CELF-4 Spanish Edition)	*population survey, October 2002: School Enrollment Supplemental File*		logical relationships in the meanings of associated words			
Clases de Palabras - Expresivo (Word Classes - Expressive, CELF-4 Spanish Edition)	Representative of Hispanic population in the US reported in the *Current population survey, October 2002: School Enrollment Supplemental File*	9-21	Explain logical relationships in the meanings of associated words	.88	.76(.75)	.99
Definiciones de Palabras (Word Definitions, CELF-4 Spanish Edition)	Representative of Hispanic population in the US reported in the *Current population survey, October 2002: School Enrollment Supplemental File*	13-21	Define and describe meanings of words	.89	.92(.91)	.89

Note: *Observed coefficient is in parentheses and corrected coefficients are the values given.

CURRICULUM VITAE

Carla I. Orellana
407-433-0468
carla.orellana@aggiemail.usu.edu

EDUCATION

Utah State University, Logan, UT. Student in the Disability Disciplines Doctoral Program (2015-2019).
Anticipated Graduation: Spring 2020

Dissertation: Working memory and syntactic processing in bilingual and monolingual children
Advisor: Ronald B. Gillam, Ph.D.

University of South Florida, Tampa, FL. Master of Science in Speech-Language Pathology (2009).

University of South Florida, Tampa, FL. Bachelor of Arts in Communication Sciences and Disorders (2007).

Undergraduate Thesis: Exploring the relationship between vocabulary knowledge and non-word processing in bilingual Spanish-English speakers.
Advisor: Stefan A. Frisch, Ph.D.

PROFESSIONAL EXPERIENCE

Bilingual Speech-Language Pathologist, Orange County Public Schools, Orlando, FL, August 2012-June 2015

Speech-Language Pathologist, HCR-ManorCare, West Palm Beach, FL, June 2010-January 2012

Speech-Language Pathologist, Orange County Public Schools, Orlando, FL, August 2009-June 2010

TEACHING EXPERIENCE

Spring 2018 (*Co-teacher)* Cultural Linguistic Diversity in Communicative Disorders

Spring 2017 (*Co-teacher)* Cultural Linguistic Diversity in Communicative Disorders

Spring 2016 (*TA)* Cultural Linguistic Diversity in Communicative Disorders

Summer 2018 (*Guest lecturer*) Topic: Speech Sound Disorders in Multilinguals
Course: Phonological Assessments & Intervention

Summer 2017 (*Guest lecturer*) Topic: Speech Sound Disorders in Multilinguals
Course: Phonological Assessments & Intervention

Fall 2015 (*Guest lecturer*) Topic: Neural System
Course: Speech Science

PUBLICATIONS

Orellana, C. I., Wada, R., Gillam, R.B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology, 28,* 1298-1317; https//doi.org/10.1044/2019_AJSLP-18-0202.

Hartzheim, D. U., Studenka, B. E**., Orellana, C. I**., Hancock, A., Gillam, R. B. (2018) *Patterns of neural activity in children with and without autism spectrum disorders during a pragmatic language task as measured by fNIRS: A proof of concept study*. Manuscript submitted for publication.

Orellana, C., Skaria, J., & Gillam, R.B. (2019). Neural Imaging. In M. Ball & J. Damico (Eds.), *SAGE Encyclopedia of Communication Sciences and Disorders*. Thousand Oaks, CA: Sage Publishing.

Gillam, R.B., Gillam, S.L., Holbrook, S., & **Orellana, C.** (2017). Language Disorder in Children. In S. Goldstein & M. DeVries (Eds.), *Handbook of DSM-5 Disorders in Children and Adolescents* (pp. 57-76), NY: Springer https://doi.org/10.1007/978-3-319-57196-6

INVITED PRESENTATIONS

Orellana, C. I. (2018). Working with bilinguals as a school-based speech-language pathologist. Seminar presented to the Logan City School District speech-language pathologists. Logan, UT, April.

Orellana, C. I. & Risueño, R. (2017). El desarrollo del lenguaje y la lectura: El papel de los padres y maestros [The development of language and literacy: The role of parents and teachers]. Seminar presented to Head Start - Centro de La Familia, Providence, UT, April.

PRESENTATIONS

Alphonsa, S., **Orellana, C.I.**, Schwartz, S., Gillam, R.B. (2019, March). *A multilevel modelling approach to quantify channel-based neural variability during postural working memory dual-tasking in young and old adults using fNIRS*. Poster presentation at the Cognitive Neuroscience Society meeting, San Francisco, CA.

Ding, G., Mohr, K.A.J., **Orellana, C.I**., Hancock, A., Gillam, R.B. (2019, March). *Syntactic processing in bilinguals and monolinguals: Evidence from functional near-infrared spectroscopy (fNIRS)*. Poster presentation at the Cognitive Neuroscience Society meeting, San Francisco, CA.

Orellana, C. I., Juth, S. M., Mohr, K., Gillam, R.B. (2018, November). *How bilinguals and children with language impairment process complex academic language during listening and reading*. Seminar presented at the American Speech Language Hearing Association Annual Convention, Boston, MA.

Orellana, C. I., Wada, R., Gillam, R.B. (2018, November)**.** *The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis*. Poster presentation at the American Speech Language Hearing Association Annual Convention, Boston, MA.

Holbrook, S., **Orellana, C. I.**, Schwartz, S., Gillam, S. (2018, November). *Developmental trajectory of narrative comprehension in literal and inferential questions: Effects of demographics and context*. ePoster presented at the American Speech Language Hearing Association Annual Convention, Boston, MA.

Orellana, C. I., Schwartz, S., Juth, S. M., Ding, G., Wada, R., Gillam, R.B. (2018, June). *Sentence comprehension in children: Multilevel modeling of behavioral and eye-tracking evidence*. Poster presented at the 39th annual Symposium on Research in Child Language Disorders, Madison, WI.

Holbrook, S., **Orellana, C**.I., Schwartz, S., & Gillam, S. (2018, June). *Developmental changes in the response to literal and inferential comprehension questions: Demographic and contextual effects.* Poster presented at the 39th annual Symposium on Research in Child Language Disorders, Madison, WI.

Orellana, C. I., Juth, S. M., Ding, G., Gillam, R.B. (2018, April). *Sentence comprehension in children developing typically, children with language impairment, and Spanish-English bilingual children: Behavioral and eye-tracking evidence*.

Poster presentation at the Utah State University Student Research Symposium, Logan, UT.

Orellana, C. I., Wada, R., Hancock, A., Gillam, R.B. (2017, November). *Syntactic processing in children using fNIRS & eyetracking*. Flash oral presentation at the American Speech Language Hearing Association Annual Convention, Los Angeles, CA.

Juth, S. M., Pickel, M., **Orellana, C. I**., Mohr, K., Gillam, R.B. (2017, November). *Through the eyes of the beholder: An overview of eye tracking physiology, methodology, and research*. Panel presentation at the Association of Literacy Educators and Researchers Annual Conference, St. Petersburg, FL.

Orellana, C. I., Gillam, R.B., Wan, N., Hancock, A. (2016, November). *Syntactic processing in bilingual adults using fNIRS and eyetracking*. Poster presentation at the American Speech Language Hearing Association Annual Convention, Philadelphia, PA.

Orellana, C. I., Frisch, S. A., & Brea-Spahn, M. R. (2007, November). *Metalinguistic phonological judgments in bilingual Spanish-English speakers.* Poster presentation at the American Speech Language Hearing Association Annual Convention, Boston, MA.

AWARDS

Presidential Doctoral Research Fellowship, Utah State University (2015-2019)

NIH Student Travel Award for the 39th Annual Symposium on Research in Child Language Disorders (SRCLD) in Madison, Wisconsin (2018)

RGS Student Travel Award, Utah State University (2018)

RGS Student Travel Award, Utah State University (2017)

RGS Student Travel Award, Utah State University (2016)

Diverse Student Success Fellowship, University of South Florida (2007-2009)

Successful Latino Student Award, University of South Florida (2006)

PROFESSIONAL AFFILIATIONS

Member of the American Speech-Language-Hearing Association (ASHA) (2009-current)

Member of the Hispanic Caucus of the American Speech-Language-Hearing Association (ASHA) (2016-current)

CERTIFICATION AND LICENSURE

Certificate of Clinical Competence in Speech Pathology, American Speech-Language-Hearing Association (2010-current)

Licensed Speech-Language Pathologist, Florida Board of Speech-Language Pathology & Audiology (2010-2015)

SERVICE

Grant Reviewer - Undergraduate Research and Creative Opportunities (URCO) Grant Program at Utah State University

Presentation Reviewer - Student Research Symposium (SRS) at Utah State University

Ad hoc Reviewer – Language, Speech, and Hearing Services (2015)

MENTORING

Language, Education and Auditory Processing (LEAP) Brain Imaging Lab

- Trained and supervised undergraduate and graduate students on fNIRS and eye-tracking operation

SKILLS

- Languages: Spanish and English
- Technology: SMI (eye-tracking), fNIRS, SPSS, Mplus, R
- Vital-Stim Certified
- Picture Exchange Communication System (PECS) Level 1 Certified Implementor