ABSTRACT

Title of Dissertation: UTTERANCE-LEVEL PREDICTORS OF STUTTERING-LIKE, STALL, AND REVISION DISFLUENCIES IN THE SPEECH OF YOUNG CHILDREN WHO DO AND DO NOT STUTTER

Julianne Terese Garbarino, Doctor of Philosophy, 2021

Dissertation directed by: Professor Nan Bernstein Ratner, Department of Hearing and Speech Sciences

Disfluencies are generally divided into two types: stuttering-like disfluencies (SLDs), which are characteristic of the speech of people who stutter, and typical disfluencies (TDs), which are produced by nearly all speakers. In several studies, TDs have been further divided into stalls and revisions; stalls (fillers, repetitions) are thought to be prospective, occurring due to glitches in planning upcoming words and structures, while revisions (word and phrase repetitions, word fragments) are thought to be retrospective, occurring when a speaker corrects language produced in error.

This dissertation involved the analysis of 15,782 utterances produced by 32 preschool-age children who stutter (CWS) and 32 matched children who do not stutter (CWNS). The first portion of this dissertation focused on how syntactic factors relate to disfluency. Disfluencies (of all three types) were more likely to occur when utterances were ungrammatical. The disfluency types thought *a priori* to relate to

planning (SLDs and stalls) occurred significantly more often before errors, which is consistent with these disfluencies occurring, in part, due to difficulty planning the error-containing portion of the utterance. Previous findings of a distributional dichotomy between stalls and revisions were not replicated. Both stalls and revisions increased in likelihood in ungrammatical utterances, as the length of the utterance increased, and as the language level of the child who produced the utterance increased. This unexpected result suggests that both stalls and revisions are more likely to occur in utterances that are harder to plan (those that are ungrammatical and/or longer), and that as children's language develops, so do the skills they need to produce both stalls and revisions.

The second part of this dissertation assessed the evidence base for the widespread recommendation that caregivers of young CWS should avoid asking them questions, as CWS have been thought to stutter more often when answering questions. CWS were, in fact, less likely to stutter when answering questions than in other utterance types. Given this finding, the absence of previous evidence connecting question-answering to stuttering, and the potential benefits of asking children questions, clinicians should reconsider the recommendation for caregivers of CWS to reduce their question-asking.

UTTERANCE-LEVEL PREDICTORS OF STUTTERING-LIKE, STALL, AND REVISION DISFLUENCIES IN THE SPEECH OF YOUNG CHILDREN WHO DO AND DO NOT STUTTER

by

Julianne Terese Garbarino

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Nan Bernstein Ratner, Chair
Professor Natasha Cabrera
Professor Rochelle Newman
Associate Professor Jared Novick
Associate Professor Stacy Wagovich

Dedication

For Robbie.

Acknowledgements

I would like to thank my mentor, Nan Bernstein Ratner. Thank you for your guidance, starting with our very first conversation during my PhD program search. You started mentoring me that day in 2015 and have not taken a moment off from being my mentor since. I would also like to thank members of my committee. Rochelle Newman and Jared Novick: Thank you for your support during the dissertation process and also during my entire doctoral program. Stacy Wagovich and Natasha Cabrera: I am grateful for your joining my committee this past year and for your insights about this project. I have also learned great deal over the past five years from Yi Ting Huang and Jan Edwards, and am thankful for their investment in my development as a doctoral student.

My experience as a PhD student would not have been the same without the members of UMD's Language Science Center. I am thankful to all of the students, faculty, and staff who have made the LSC what it is, and especially to Shevaun Lewis, Tess Woods, Caitlin Eaves, and Colin Phillips.

This project could not have been completed without the help of a wonderful group of research assistants who spent many hours carefully coding language sample transcripts: Maddie Burns, Valerie Hsieh, Elizabeth Kim, Kaitlin McCloskey, Rachel Raqueno, Allie Vance, and Gabrielle Woska. I am also grateful to the many members of the UMD Language Fluency Lab who have contributed their time, over the past 20+ years, to collecting and transcribing the language samples used in this project, especially project managers Ning Hsu, Veronica Builes and Mark Baer. And of course, I am thankful for all of the families who have participated in these research projects. I also want to thank Allie Johnson, Anna Tinnemore, and Einat Korman for your friendship. I could not have asked to have get to know more wonderful people during my time in the PhD program.

Lastly, I want to thank my family: John, Irma, and Christina Garbarino for being my cheerleaders over the past five years, but more importantly, long before I decided to start a PhD program. And most of all, Robbie Yeager: Thank you for your endless support.

This work was supported in part by the National Institute on Deafness and Other Communication Disorders (NIDCD) Grant 1 R01 DC015494-01 to Nan Bernstein Ratner and Brian MacWhinney, and by the National Science Foundation (NSF) Grant CS-1626300/1626294 to Nan Bernstein Ratner PI and Brian MacWhinney, Co-I. This work was also supported in part by NSF Grant 1449815 to PI Colin Phillips; I was supported as a trainee on this grant.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Disfluencies, broadly defined, are disruptions to the forward flow of speech (Ambrose & Yairi, 1999). They are typically broken down into two major subtypes: *stuttering-like disfluencies* (*SLDs*) and *typical disfluencies* (*TDs*). SLDs are usually defined as part-word or monosyllabic whole-word repetitions, blocks, prolongations, and broken words (e.g., Ambrose & Yairi, 1999; Logan & Conture, 1995). These disfluencies, particularly part-word and monosyllabic whole-word repetitions, can and do occur in the speech of people who do not stutter (PWNS), but they are more frequent and often involve more iterations (e.g., “a-a-a-a-a-and” vs. “a-and”) in the speech of people who stutter (PWS) (Ambrose & Yairi, 1999). SLDs are often accompanied by muscle tension and awareness when they are produced by PWS (Ambrose & Yairi, 1994; Tichenor et al., 2018). TDs include multisyllabic whole-word repetitions, phrase repetitions, word and phrase revisions, word fragments, and fillers such as “um” and “uh” (Ambrose & Yairi, 1999; Logan & Conture, 1995; Yaruss et al., 1999), and they occur in the speech of PWS and PWNS (e.g., Ambrose & Yairi, 1999; Buhr & Zebrowski, 2009). See Table 1 for examples of SLDs and TDs.

Table 1

Examples of Stuttering-Like Disfluencies (SLDs) and Typical Disfluencies (TDs)

Disfluency Type	Example
SLD Type	
Part-word repetition	Wh-wh-wh-what is that?
Monosyllabic whole-word repetition	I I I I I want that one.
Block	#My doll.
Prolongation	Mmmmom help me.
Broken word	The ba#by.
TD Type	
Filler	I um don't like those.
Multisyllabic whole-word repetition	Baby- baby wants to eat.
Phrase repetition	I need- I need a car.
Word revision	Mom- dad helped me.
Phrase revision	My car- my truck is blue.
Word fragment	The fi- shark is swimming.

Note. # indicates a tense pause.

Stuttering as a Developmental Speech-Language Disorder

Stuttering is unusual as a developmental speech-language disorder in its emergence after a period of typical development; children typically begin to stutter between the ages of 2;0 and 4;0 (Yairi & Ambrose, 2005, 2013). This onset is followed by a period of stuttering, and then by spontaneous recovery in approximately 80% of cases (Bloodstein et al., 2021; Yairi & Ambrose, 2013). At the preschool age, there are small gender differences, with 1.34 to 2.2 boys who stutter for every one girl who stutters. Because girls recover from stuttering at higher rates than boys, the male-to-female ratio for adults is approximately 4:1 (see Yairi & Ambrose, 2013, for a review).

Over the past century, many accounts of the causes of stuttering have been proposed (see Bloodstein et al., 2021, for a review). One of the clinically and theoretically influential perspectives about what causes stuttering is the Demands and

Capacities Model (DCM) (Starkweather, 1987). Under the DCM, limitations in a child's ability to quickly respond to stimuli, coordinate movements of the speech mechanism, and/or plan and execute sequences of movements (i.e., capacity limitations) interact with internal and external demands to cause SLD production. Despite criticism over its lack of specificity and circular arguments (e.g., Bernstein Ratner, 2000; Siegel, 2000), the DCM is important to understand because it forms the theoretical basis for commonly applied indirect stuttering therapies in which modifications are made to decrease demands on a child's speech-language production system and/or to increase capacities for fluent speech (e.g., de Sonneville-Koedoot et al., 2015; Millard et al., 2008; Richels & Conture, 2007).

The model currently favored by many experts, due to the body of research supporting its major tenets, is the Multifactorial Dynamic Pathways Theory (MDP) (Smith & Weber, 2017). The MDP explains that PWS have underlying speech-motor vulnerabilities (e.g., Kleinow & Smith, 2000). Their speech-motor systems are then destabilized by high motoric, linguistic, or emotional processing demands, and overt stuttering behaviors such as prolongations result (Smith & Weber, 2017). The MDP (Smith & Weber, 2017) is by its authors' description consistent with, but more detailed than, the DCM. They write "we believe the MDP is consistent with the basic tenets of demands and capacities. The two accounts 'explain' stuttering at very different levels" (Smith & Weber, 2017, p. 2499). The MDP advances our understanding of why linguistic planning demands, which will be explored in the next section, relate to SLD production by incorporating a series of research findings on motor stability. Findings of (a) increased motor instability for PWS, (b) significant increases in motor instability with increasing utterance length and complexity, and (c) some evidence of interactions between whether a person stutters and degree of increase in motor instability provide a mechanism by which utterances with higher linguistic planning demands can elicit stuttering in a PWS (Kleinow & Smith, 2000; MacPherson & Smith, 2013; Maner et al., 2000; Usler & Walsh, 2018).

Disfluency-Language Relationships

SLDs and Utterance-Level Language Factors

Decades of research have demonstrated that SLDs are more likely to occur in longer and/or more syntactically complex utterances. The relationship between SLD production and longer utterances has been shown with length measured in syllables (Logan & Conture, 1995; Yaruss, 1999), morphemes (Richels et al., 2010; Yaruss, 1999; Zackheim & Conture, 2003), words (Buhr & Zebrowski, 2009; Gaines et al., 1991; Hollister et al., 2017; Wagovich et al., 2009; Yaruss, 1999) and syntactic constituents (Melnick & Conture, 2000; Yaruss, 1999). While length in syllables does relate to stuttering, there is evidence that length in words or morphemes may be a better measure of length to use when studying disfluency-language relationships. Brundage and Bernstein Ratner (1989) assessed three measures of utterance length—number of syllables, morphemes, and words—and how these related to measures of SLDs per utterance. They found that increasing the number of morphemes in an utterance was more strongly positively correlated with stuttering incidents than increasing the number of words or syllables.

To assess stuttering-utterance complexity relationships, several studies have split language samples produced by young CWS into fluent and SLD-containing utterances and compared Developmental Sentence Score (DSS) values obtained for the two sets of utterances. DSS is computed by assigning each utterance in a sample with a score based on its sentence components, and giving an additional “sentence point” if it contains no grammatical errors (Lee, 1974). Significantly higher DSS scores are reported for SLD-containing utterances than for fluent utterances (Buhr & Zebrowski, 2009; Gaines et al., 1991; Weiss & Zebrowski, 1992). Wagovich and Hall (2018) reported mixed results regarding whether samples in which children produced more SLDs had higher complexity scores using another measure, the Index of Productive Syntax (IPSyn) (Scarborough, 1990). These authors looked at samples produced by CWS close to stuttering onset over a series of 10 monthly visits, and divided the samples into the first five collected (early samples) and the last five collected (later samples). The authors found that the most disfluent of the late samples (i.e., the ones with the highest rates of SLDs production) had higher IPSyn scores than the most fluent of the late samples (i.e., the ones with the lowest rates of SLD production). However, this was not true of the most and least disfluent early samples. Overall, findings were mixed.

If disfluencies result in part from problems in the utterance planning process, then they should not be randomly distributed within the utterance; they would be expected to occur at locations where psycholinguists believe that planning loads are high. Studies looking at locations of fluency breakdown have found that SLDs tend to occur at the beginnings of syntactic constituents (Bernstein, 1981), clauses (Wall et al., 1981), and utterances (Buhr & Zebrowski, 2009; Choi et al., 2020; Gaines et al., 1991). For example, in their analysis of active declarative sentences produced by 3;0-

to 5;11-year-old CWS, Buhr and Zebrowski (2009) found that while 17% of all words in their corpus appeared in the sentence-initial position, 51% of all stuttered words were in the sentence-initial position. This type of finding is taken as indicative of a linkage between SLD production and linguistic planning (Buhr & Zebrowski, 2009; Choi et al., 2020).

Regularities about SLD positioning within the utterance are also critical to understanding why the fact that SLDs are more likely to occur in longer and more complex utterances is interpreted as evidence that SLD production is influenced by planning load. Because SLDs are much more likely to occur in the utterance-initial position (Buhr & Zebrowski, 2009; Choi et al., 2020; Gaines et al., 1991), the explanation that SLDs occur more often in longer utterances because these provide more opportunities for stuttering is too simplistic. Considered altogether, both length and complexity studies and loci studies have found SLDs to occur where planning loads are highest—in longer and more complex utterances, and early within the utterance or other planning unit (e.g., Bernstein Ratner & Sih, 1987; Buhr & Zebrowski, 2009; Choi et al., 2020; Wagovich et al., 2009; Yaruss, 1999).

TD-Language Relationships

SLDs emerge in the speech of CWS between ages 2;0 and 4;0, and TDs tend to emerge during or before this period as well. Yairi's (1981) documentation of TD production by 2;0- through 2;9-year-olds indicates that even at early stages of multi-word output, multisyllabic whole-word repetitions, phrase repetitions, fillers, and revisions are already present.

Studies that have analyzed TDs as a single group have shown that they are similar to SLDs in their relationship with utterance-level variables. Like SLDs, TDs as a group tend to occur in longer (Buhr & Zebrowski, 2009; Yaruss et al., 1999; Zackheim & Conture, 2003) and more complex (Buhr & Zebrowski, 2009; Gordon & Luper, 1989; Haynes & Hood, 1978; Yaruss et al., 1999; Zackheim & Conture, 2003) utterances.

Despite the tendency for TDs to be analyzed as one type of disfluency, it is not clear that all types of TDs actually do behave similarly to SLDs in their relationships with utterance length and complexity. Rispoli and colleagues have differentiated between *stalls* and *revisions* (Rispoli, 2003, 2018; Rispoli et al., 2008). Under the definitions provided by Rispoli and colleagues, stalls simply delay and do not change any words or morphemes. These include fillers, and word and phrase repetitions. Revisions, in contrast, involve changes to previously produced words or morphemes and include word and phrase revisions and fragments. Regarding their proposed functions, stalls are considered "prospective" and occur due to glitches in planning language that has not yet been articulated. Revisions, in contrast, are "retrospective" and occur when the speaker corrects already-produced words or parts of words that were produced even though they did not match the speaker's intention (Rispoli, 2003). To be consistent with previous studies (Rispoli, 2003, 2018; Rispoli et al., 2008; Wagovich et al., 2009), I will refer to the group of typical disfluencies including word and phrase revisions and fragments as "revisions." The more specific subtypes will be described as "word revisions," or "phrase revisions." I will refer to fillers, word repetitions, and phrase repetitions as a group as "stalls."

An informative discrepancy has emerged in studies that have separated stalls and revisions. Stalls do not have an apparent relationship with a child's overall language abilities; a longitudinal study tracking children from ages 1;9 to 2;9 found that for different children, stalls increased, decreased, or remained constant as mean length of utterance in morphemes (MLU-m) and IPSyn scores increased (Rispoli et al., 2008). These results converged with evidence of an earlier cross-sectional study that enrolled 1;10- to 4;0-year-olds (Rispoli, 2003) and found that revision rate correlated with MLU-m and IPSyn but that stall rate did not correlate with either of these. While they do not appear to relate to language development, stalls do have a relationship with utterance planning load; stalls are more likely to occur in longer utterances (Rispoli, 2003; Rispoli et al., 2008). In contrast, revisions have not been found to have a significant relationship with utterance length but do increase with language skills, as measured by MLU-m and IPSyn (Rispoli, 2003; Rispoli et al., 2008).

Rispoli (2003, 2018) has hypothesized about why stalls and revisions have different relationships with utterance length and complexity and language development. He points to differing prerequisite skills for production of these two types of disfluencies. For a stall to occur, all that is required is that a glitch occurs in sentence planning. In contrast, revisions require that the child is using a syntactic frame in which they insert lexical items, and that the child has a repertoire of other words that can be substituted. Support for this explanation of revision prerequisites was provided by Rispoli's (2018) finding that the number of different subject noun phrases that children produced in their language samples (i.e., subject diversity) was a significant predictor of subject revision rate. Children who had more available substitutes for that specific sentence role made more revisions in that position. Rispoli (2003, 2018) also suggests that attention to one's own speech, language comprehension, and sufficient monitoring skills are prerequisites of the ability to revise, which further explains why revisions increase as language develops. Importantly, Rispoli (2003) acknowledges a potential alternative explanation for this relationship—that the higher revision rate might occur when a child's language abilities increase because they are revising grammatical errors more often. However, this explanation seems unlikely because there were only 10 successful grammatical revisions produced by children in that study, contrasting with 419 nongrammatical revisions.

Re-analysis of the numbers of different types of TDs reported across studies shows that at least half of TDs produced by young children are stalls. For example, 55.2% of the TDs reported by Haynes and Hood (1977) were stalls and 59.3% of those reported by Wexler and Mysak (1982) were stalls. The results reported by Rispoli and colleagues suggest that it may in fact only be stalls that pattern like SLDs in occurring more often in longer and more complex utterances (Rispoli, 2003; Rispoli et al., 2008). In other words, it may be that only stalls occur in longer and more complex utterances and that revisions are more consistent across utterances of varying length and complexity levels and do not contribute to the finding that TDs overall occur in longer and more complex utterances. In a longitudinal study following nine CWS for approximately 10 months, Wagovich et al. (2009) used the stall-revision framework. They found that SLDs patterned similarly to stalls in their occurrence in longer utterances and variability over time, with revisions not relating to utterance length but increasing over time. The results of this study, along with those of Rispoli (2003) and Rispoli et al. (2008) suggest that the best way to look at TDs and their relationship to utterance-level and developmental language factors may not be to consider them all together as one group.

Finally, like SLDs, TDs tend to occur towards utterance beginnings (Buhr & Zebrowski, 2009) and early in syntactic constituents (Bernstein, 1981). Buhr and Zebrowski (2009) reported that while only 17% of all words occurred in the utterance-initial position, 71% of TDs occurred in this position. The tendency for TDs to occur in specific locations in the utterance suggests that the relationship between longer utterances and the presence of TDs is not due solely to the fact that they contain more words that can be delayed or revised. Because of the patterning of TDs into specific locations in utterances, the relationship is likely due at least in part to longer utterances carrying greater planning loads (Rispoli & Hadley, 2001).

Answers to Questions and Disfluency

The ways in which disfluency production relates to answering questions is somewhat unclear. Most studies looking at the relationship between utterance length and/or complexity and disfluency have grouped answers together with other declarative and imperative utterances (Buhr & Zebrowski, 2009; Gaines et al., 1991; Logan & Conture, 1995; Richels et al., 2010; Zackheim & Conture, 2003).

Several studies have assessed fluency properties of questions and answers to questions. Weiss and Zebrowski (1992) reported that children stuttered more often on assertive utterances (including requests for information, actions, clarifications, or attention, utterances that label or describe, and performatives such as jokes) than on responses to requests (including answers to questions). Weiss and Zebrowski provided two possible explanations for this discrepancy. Assertive utterances were longer and more complex than responsive utterances (many of which were one-word responses to questions), so length or complexity factors could have accounted for this finding. Another possible explanation offered by Weiss and Zebrowski is that assertions have higher levels of "communicative responsibility." This means that they may introduce new information to the discourse or have higher information loads. When a child's utterance has higher communicative responsibility, the demands of carrying the conversation forward and ensuring communicative success are more heavily placed on the child (Stocker & Usprich, 1976). Linguistic factors (length and complexity), and communicative responsibility may both have increased the likelihood that an assertive utterance would contain SLDs, or one of these may have driven the higher rate of stuttering in assertive utterances. Weiss and Zebrowski's analysis was not designed to differentiate between these possibilities. Similarly, Yaruss (1999) found that questions, a type of assertive utterance, were more likely to contain SLDs than declarative utterances, but did not assess whether answers to questions had particular fluency properties.

Byrd et al. (2011) provided additional clarity about the relationship between an utterance's assertiveness or responsiveness and stuttering. They found that while assertive utterances were more likely to contain SLDs than responsive utterances were, this difference was no longer significant after accounting for length and complexity. They broke utterances down into more specific types, but their coding system did not specify whether each responsive utterance was an answer to a question (in comparison to a response to a statement). The distinction between responses to statements and responses to questions is particularly clinically important. The recommendation for caregivers to reduce their own use of questions so as to help children stutter less by requiring fewer responses to questions, is widespread (Bernstein Ratner, 2004; Franken & Putker-de-Bruijn, 2007; Kelman & Nicholas, 2020; Wilkenfeld & Curlee, 1997), despite evidence that parents of young CWS do not ask questions more often than parents of CWNS (Meyers & Freeman, 1985).

These issues are particularly important in the period close to onset of stuttering because (a) this age range is included among those in which environmental modifications can involve reductions in caregiver use of questions (Franken & Putker-de-Bruijn, 2007; Kelman & Nicholas, 2020), and (b) question formulation is a developing skill during this period (Santelmann et al., 2002). Thus, withholding models of questions from CWS may have an unintended impact on language development.

Children's Sentence Planning

Disfluency-Planning Relationships

The study of disfluency in children's speech and children's sentence planning has been closely linked with and has built on some of the early work in adult sentence planning (e.g., Boomer, 1965; Goldman-Eisler, 1968; Maclay & Osgood, 1959). Clause junctures appear to be important planning locations for children, as they are for adults (McKee et al., 2017). While children's sentence planning shares many features with adult planning, children may not be able to plan as much of an upcoming utterance at once. McDaniel et al. (2010) found that linguistic factors that stress adults' and children's sentence planning are similar, even for children as young as 3;5 to 5;11, but that children are likely to have to stall to plan during utterance production more often than adults (McDaniel et al., 2010).

One theoretical perspective on how disfluencies, particularly TDs, relate to sentence planning is the *leading edge* theory (Rispoli, 2003; Rispoli et al., 2008; Rispoli & Hadley, 2001). Under this theory, children have a *comfort zone* of utterances that they can produce fluently, and a leading edge of utterances that are still within their reach, but that are more likely to be disfluent. Under this theory, disfluent utterances are likely to be the most syntactically advanced utterances within a child's output.

Grammatical Errors and Sentence Planning

I will use the term *grammatical error* in the sense of Lee (1974) and Eisenberg et al. (2012), which is deviation from adult grammar, specifically, from the grammar of adult speakers of the same dialect. This use of "error" is only intended to indicate that a production is non-adultlike, and not to suggest that a child's utterance diverges from what is expected of a child of their age. Errors, under this definition, are in fact common in the speech of young children with typical language development. A study of 2;6- to 6;6-year-olds reported that young typically-developing children produced errors in 11% of utterances on average, with a range across children of 0 to 35% (M. Dunn et al., 1996). Another study specifically looking at 3-year-olds found that an average of 29% of utterances contained errors, with an error rate range across children of 11 to 53% (Eisenberg et al., 2012).

McKee et al. (2017) describe the following apparent irony in the study of children's expressive language development: Changes in the structure of utterances that children produce are assumed to reflect language learning (i.e., improvements in competence), while there is in fact evidence that children's processing abilities (i.e., performance factors) impact what children's utterances look like. What a child produces does not always reflect their highest level of language knowledge. Much of the evidence to support this is outlined by McKee et al. (2017). To start, children understand grammatical morphemes before they consistently produce them. In addition, utterance factors of length and complexity, which impact fluency (e.g., Buhr & Zebrowski, 2009), affect whether an inconsistently produced morpheme is produced in a particular utterance (see Charest & Johnston, 2011, for a review). Similarly, prosodic processing can impact whether a morpheme inconsistent across contexts is produced within a particular utterance. For example, stressed, utterance-initial lexical nouns are more likely to be produced than unstressed, utterance-initial subject pronouns (Gerken, 1991; McGregor & Leonard, 1994).

Morphosyntactic errors are useful in assessing how language relates to disfluency production on this basis: Disfluency is more likely to be produced when language planning demands are high. Whether an utterance contains an error is one way of measuring which structures are difficult for a particular child (Watson et al., 2011).

Utterance Grammaticality and Disfluency

To my knowledge, there are three published reports directly exploring the relationship between grammaticality and disfluency, and these all focus on SLDs and

CWS, without including TDs or CWNS. Bernstein Ratner (1997) split utterances from language samples produced by 15 CWS into subsamples that contained and did not contain grammatical errors. A higher proportion of utterances with grammatical errors than error-free utterances contained SLDs. This finding suggests some relationship between grammaticality and SLD production. However, several questions were unanswered by this analysis, as length and complexity were not controlled. It is possible that error-containing utterances were longer and/or more complex than utterances without errors, and that this accounted for the grammaticality effect.

A more recent study did control for length and complexity of utterances and looked at relationships between grammaticality and disfluency in 11 young monolingual Spanish-speaking CWS (Watson et al., 2011). Watson et al. (2011) separated utterances into subsamples that were fluent or SLD-containing. They found that the odds of an utterance containing an SLD was higher for ungrammatical utterances, even when controlling for age, utterance length in syllables, and utterance complexity. The authors interpreted these findings as consistent with prior research reporting that syntactic complexity is an important contributor to stuttering, and as consistent with hypotheses that both stuttering and grammatical errors result from attempting to produce difficult structures. In contrast to Watson et al., and Bernstein Ratner (1997), Yaruss (1999) did not find a significant relationship between utterance grammaticality and the presence of stuttering in the language samples of 12 young CWS.

One unaddressed question here is whether grammatical errors temporally follow disfluencies that are related to planning (stalls and SLDs). If a stall or SLD is caused by difficulty planning the part of an utterance containing a grammatical error, the stall or SLD would have to be produced before the error. To be clear, however, this does not mean that the stall or SLD would have to immediately precede the error. MacWhinney and Osser (1977) differentiated between disfluencies reflecting different types of planning. Coplanning occurs when later words in an utterance are planned while the utterance is already being produced and many of these disfluencies would be classified as revisions. In contrast, preplanning occurs when words are planned farther in advance, particularly towards utterance or clause beginnings. Disfluencies associated with preplanning, particularly stalls, will occur earlier than the error but do not necessarily have to occur immediately before the difficult element. Within this framework, it is possible that preplanning of an ungrammatical element would increase planning demand to a high enough level to cause a stall or SLD. In this case, the stall or SLD would occur earlier in the utterance but not necessarily immediately before the error. Because stalls and SLDs are associated with planning (e.g., MacWhinney & Osser, 1977; McDaniel et al., 2010; Wagovich et al., 2009), they would have to precede an omission or other error, or occur at the beginning of a word containing an error to potentially be caused by difficulty planning that portion of the utterance. Of the studies measuring both grammaticality and disfluency (Bernstein Ratner, 1997; Watson et al., 2011; Yaruss, 1999), none considered whether the disfluency occurred before the grammatical error. I expect that stalls and SLDs will tend to occur before, or on the same word as errors, rather than after them. This prediction is based on the expectation that stalls and SLDs will occur where a child is planning the error-containing portion of the utterance (i.e., on or before the error).

Summary, Motivation, and Hypotheses

Summary of Theories of SLD and TD Production

Stuttering is a neurodevelopmental disorder that affects speakers who have underlying speech-motor vulnerabilities; linguistic factors can interact with this underlying motor instability to result in SLD production (Smith & Weber, 2017). The conceptualization of language factors within this model is possible due to decades of research documenting robust relationships between utterance length and complexity and SLD production (e.g., Bernstein Ratner & Sih, 1987; Buhr & Zebrowski, 2009). Recent reports have also indicated that language growth is a predictor of recovery, with CWS who exhibit greater improvements in their language skills after onset being more likely to experience recovery (Hollister et al., 2017; Leech et al., 2017, 2019). One theoretical approach to understanding how TD production interacts with language development is that these disfluencies, particularly the ones that delay language production, occur in utterances that are at the leading edge of what a child is able to produce; under this theory, children have the linguistic knowledge to produce these utterances but do not have the processing capacity to do this entirely fluently (Rispoli, 2003; Rispoli et al., 2008; Rispoli & Hadley, 2001). Further, in revisions, a speaker produces a word or part of a word that does not match their intention, realizes this mismatch, and then goes back to fix it; this involves coordinating at many levels of the language production process. The skills required to revise parts of an utterance are thought to increase with language development, enabling increased revision production (Rispoli, 2003; Rispoli et al., 2008).

Stall-Revision Dichotomy and SLDs

Findings that SLDs tend to occur early in utterances and in longer and more complex utterances have been taken to indicate that SLDs are related to planning (e.g., Buhr & Zebrowski, 2009; Yaruss, 1999). They have also been incorporated into the MDP model of stuttering, which proposes that SLDs occur when high loads in other domains (i.e., language or emotional) destabilize the already vulnerable speech-motor systems of PWS (Smith & Weber, 2017). Some prior studies on TD production by young children have considered all TDs together and found that TDs have similar relationships to language variables as SLDs (Buhr & Zebrowski, 2009; Gordon & Luper, 1989; Haynes & Hood, 1978; Yaruss et al., 1999; Zackheim & Conture, 2003). Other studies have separated stalls from revisions and found that stalls occur more often in longer and more complex utterances but that revisions do not (Rispoli, 2003; Rispoli et al., 2008; Wagovich et al., 2009). This background motivates the first question I will ask:

1. Do SLDs and stalls differ from revisions in their relationships with length, complexity, grammaticality, and language level in the speech of young children who do and do not stutter?

I expect that SLDs and stalls will occur more often in longer and more complex utterances, and in utterances containing grammatical errors, and that they will be variable across different children but not show a clear relationship with a child's overall language level. In contrast, I predict that revisions will increase with a child's overall language level, but will not show a relationship with length, complexity, or grammaticality. More broadly, SLDs and stalls will pattern similarly to each other in their relationships with length, complexity, grammaticality, and a child's language level, but revisions will show a different pattern. This finding would be consistent with those reported by Rispoli (2003), Rispoli et al. (2008), and Wagovich et al. (2009), but it would be the first time that revisions were compared to SLDs and stalls in matched groups of CWS and CWNS and assessed all in the same (large) study. If SLDs and stalls, which are both thought to serve a prospective purpose in allowing the child more utterance planning time, relate to length, complexity, grammaticality, and language level in similar ways, but revisions, which are thought to serve a retrospective function do not (Rispoli & Hadley, 2001), this would lend further support to the belief that SLDs reflect some aspect of sentence planning.

Grammaticality of Disfluent Utterances

Prior research with relatively small samples of 11 CWS (Watson et al., 2011) and 15 CWS (Bernstein Ratner, 1997) has suggested that CWS do produce more SLDs in ungrammatical utterances, but another study with 12 CWS did not find this relationship to be significant (Yaruss, 1999). Only one of these studies controlled for utterance length and complexity (Watson et al., 2011), and no study has looked at matched CWS and CWNS, and both SLDs and TDs or looked at stalls and revisions separately. Considering these two groups of children and three types of disfluencies (SLDs, stalls, and revisions) together will provide a broader picture of how SLDs, stalls, and revisions, fit into the leading edge theory (Rispoli & Hadley, 2001). The leading edge theory is certainly consistent with the findings that length and complexity are related to disfluency, particularly stall, production. However, length and complexity are clearly not the only indicators, and arguably, not the best structural indicators, of what a child is able to produce easily. Whether the child can produce an utterance accurately, or whether the child is attempting structures they have not yet mastered and thus produce ungrammatically (at least in that particular utterance) may be a better indicator of which utterances are in the individual child's comfort zone or leading edge. Measuring grammaticality allows for more child-specific, rather than population-wide, characterization of which utterances will be difficult for a child to produce. If SLDs and stalls both occur more frequently in the utterances containing grammatical errors, this would allow for an expansion of the leading edge theory. It would suggest that grammatical utterances generally occur within the comfort zone, and grammatical errors generally occur at the leading edge, sometimes preceded by stalls and SLDs. Rispoli (2003) and Rispoli et al. (2008) measured TDs in the speech of CWNS, and Wagovich et al. (2009) measured SLDs and TDs in the speech of CWS. The proposed study will expand upon prior findings by including CWS and CWNS, and measurement of SLDs and TDs, all in the same study.

Regarding models of SLDs, the development of the MDP model (Smith & Weber, 2017) was enabled by previous research into the ways that language, and speech-motor skills relate to stuttering, as well as research assessing how emotional processing, genetics, and neurobiology impact stuttering. This model does not currently account for utterance grammaticality as a linguistic factor that may destabilize the motor system and lead to SLD production. Finding that grammaticality relates to SLD production would contribute to our theoretical understanding of stuttering by showing that structures that we know to be hard for a particular child (indicated by an error) are where SLDs occur. This would be a more individualized way of understanding where the child is putting in additional effort. If grammaticality is in fact related to SLD production, then there would be a need for future research to investigate whether grammaticality, like length and complexity, is related to motor instability in PWS. And further, if SLDs occur before the error, this would be additional support for the proposal that planning of difficult structures contributes to SLD production (Watson et al., 2011).

Studying the ways in which grammaticality relates to disfluency in the speech of young CWS is also clinically important, particularly in the construction of treatment hierarchies. As one example, in DCM-based treatment focused on decreasing internal and external demands for fluent speech and/or increasing a child's capacity for fluent speech, linguistic hierarchies are often used (e.g., Richels & Conture, 2007). For instance, the specific group-based treatment described by Richels and Conture (2007) takes the following factors into consideration, as they relate to the child's capacities and the internal and environmental demands for fluent speech: (a) utterance length and complexity, (b) conversational time demands, (c) speech rate, (d) emotional tendencies, and (e) the family's role. Utterance length and complexity are the primary linguistic factors considered in hierarchy development for this type of fluency treatment. There is no reason to doubt that length and complexity should continue to be considered in treatment planning for young CWS. However, there is reason to believe that utterance grammaticality may also be important to consider. Grammaticality could be controlled by a clinician who knows the child's language profile (including the types of structures they can produce without errors) modifying the types of utterances that are prompted in therapy. This background leads to the next two questions that I plan to address:

2. How does the grammaticality of an utterance produced by a young CWS or CWNS relate to the likelihood that it contains an SLD or stall?

3. Do SLDs and stalls tend to precede grammatical errors in utterances produced by young CWS and CWNS?

I expect that utterances containing grammatical errors will be more likely than grammatically correct utterances to contain stalls and SLDs. I make this prediction in part because of the two sets of findings reported by Bernstein Ratner (1997) and Watson et al. (2011) with SLDs (though these contrast with Yaruss's (1999) results). I am extending this prediction to look at stalls, and expect this to be true for both SLD- and stall-containing utterances. I also make this prediction because this finding would be consistent with the general principle that utterances with higher planning loads—typically measured by length and complexity—are more likely to contain SLDs and TDs (when TDs are measured as a whole) or stalls only when revision and stalls are measured separately. As the analysis section will make clear, while this question has a different theoretical basis than question one, questions one and two will be linked by sharing statistical models.

Further, because stalls and SLDs are thought to occur where children are planning challenging speech structures (e.g., Buhr & Zebrowski, 2009; McKee et al., 2017; Rispoli et al., 2008), I hypothesize that stalls and SLDs in utterances containing grammatical errors will tend to occur before and not after the error. Because children can plan more than one word at a time (MacWhinney & Osser, 1977), "before" does not necessarily mean "immediately preceding." If disfluencies tend to occur before errors, this would strengthen any conclusions drawn about the role of planning difficult (i.e., error-containing) upcoming portions of the utterance in SLD and stall production.

Utterance Function and Disfluency

There are two primary reasons why the study of the relationship between an utterance's function[1] and its fluency is important. First, the frequency with which caregivers are advised to reduce their own use of questions and to use more comments with young CWS makes the study of questions and responses to questions and their relationship with SLDs, and disfluency more broadly, of high clinical relevance (Bernstein Ratner, 2004; Franken & Putker-de-Bruijn, 2007; Kelman & Nicholas, 2020; Wilkenfeld & Curlee, 1997). Second, little attention has been paid to the role of communicative responsibility in early childhood disfluency. It would be clinically and theoretically helpful to know whether a child adding new information to the discourse is something that seems to contribute to the likelihood that the child will be disfluent. This background forms the motivation for the fourth question that I will ask:

[1] The "function" of an utterance can be defined in many ways and will be used here to describe whether the utterance is a question, other assertive utterance, answer to a question, or other responsive utterance.

4. How does utterance function relate to disfluency, in the speech of young CWS and CWNS?

The question about how utterance function relates to SLD production for CWS is particularly motivated by the clinical background literature. The primary hypotheses is that questions and other assertive utterances produced by CWS will be more likely to contain SLDs than answers to questions and other responsive utterances, but that this relationship may no longer be significant after controlling for syntactic features of the utterance. This hypothesis is based on findings from previous studies that have looked at questions and stuttering in CWS just older than those in which the onset of stuttering typically occurs (Byrd et al., 2011; Weiss & Zebrowski, 1992; Yaruss, 1999). One of these previous studies enrolled 12 CWS and found that questions were more likely than declarative utterances to contain SLDs (Yaruss, 1999). The second previous study enrolled eight CWS and found that assertive utterances (a broad category including requests for information, actions, clarifications, or attention, utterances that label or describe, and performatives such as jokes) contained SLDs more often than responsive utterances (including responses to questions) (Weiss & Zebrowski, 1992). The third study enrolled 15 CWS and found that assertive utterances were more disfluent than responsive utterances, but that this difference was no longer significant after controlling for utterance length and complexity (Byrd et al., 2011).

Chapter 2: General Methods

The studies addressing syntactic predictors of three disfluency types, disfluency-error order, and how utterance function relates to disfluency had some shared methods as they used the same language samples from the same participants. Shared methods will be described in the current chapter, and methods specific to each study will be described in subsequent chapters.

Participants

Corpora and Matching

This study involved the analysis of language samples collected from 32 CWS and 32 matched CWNS. These participants were originally enrolled in one of three previous or ongoing studies on stuttering and/or language development. Seventeen of the CWS came from the Ratner corpus, available at Fluency.Talkbank.org (Bernstein Ratner & Silverman, 2000; Miles & Bernstein Ratner, 2001; Silverman & Bernstein Ratner, 2002; Wagovich & Bernstein Ratner, 2007), which was collected in the mid-to-late 1990s. Some of the 17 Ratner corpus CWS enrolled in the current study were part of the Bernstein Ratner (1997) study. The other 15 CWS came from the Ratner-MacWhinney corpus (Bernstein Ratner & MacWhinney, 2018). The data from these children was collected between 2016 and 2020 at the University of Maryland. This study is still ongoing, and thus data are not yet posted in FluencyBank (at Fluency.Talkbank.org).

Each CWS was matched to a CWNS whose language samples were collected in similar play-based contexts. CWS and CWNS were matched on age (within 4 months as the upper limit, and with 75% of pairs matched within 1 month) and gender, and as closely as possible on maternal education. For Ratner-MacWhinney corpus participants, matching was completed during the recruitment process. Ratner corpus CWS were matched by Luckman et al. (2020). Of the 17 Ratner corpus CWS, 10 were matched to CWNS from the Ratner corpus, and seven were matched to CWNS from the Weismer corpus available at Talkbank.org (Ellis Weismer et al., 2013; Heilmann et al., 2005; Moyle et al., 2007). The 15 CWS from the Ratner-MacWhinney corpus were matched to CWNS from the same corpus.

Participants were between the ages of 28 and 50 months. Twelve participants were female and 52 were male.[2] All participants were monolingual English speakers. Maternal education levels were available for the 15 CWS and 15 CWNS from the Ratner-MacWhinney corpus. For these CWS, 14 mothers had college degrees or higher and one parent had some college courses, and for these CWNS, all 15 mothers had college degrees or higher.

Original Participant Recruitment

CWS and CWNS in the Ratner corpus were recruited through flyers posted in pediatricians' offices in the Maryland-Washington, D.C. metropolitan area. CWS and CWNS in the Ratner-MacWhinney corpus were recruited from flyers posted in the community, recruitment emails sent through community organizations, referrals from

---

[2] This 4.3:1 male-to-female ratio is higher than what is reported in epidemiological studies of the gender ratio for young children (see Yairi & Ambrose, 2013). It is, however, within the range of ratios reported for other studies with CWS assessing similar questions. These studies range from their CWS participants being all boys (Gaines et al., 1991; Logan & Conture, 1995; Melnick & Conture, 2000; Yaruss, 1999; Zackheim & Conture, 2003) to gender ratios that more closely approximated those reported at the period near onset, including a 2.75:1 ratio in Byrd et al. (2011), a 2:1 ratio in Wagovich et al. (2009), and a 1.75:1 ratio in Watson et el. (2011).

local speech-language pathologists, and a research participant database. CWNS in the Ellis Weismer corpus were recruited through a research participant database.

Diagnosis of Stuttering

The diagnosis of stuttering for all CWS was made by a speech-language-pathologist who was one of the original cadre of American Speech-Language-Hearing Association Board Recognized Specialists in Fluency Disorders (ASHA BRS-FD). CWS in the Ratner corpus were tested within four months of the onset of stuttering and CWS in the Fluency Bank corpus were tested within one year of onset.

Screening for Typical Language Development

All participants were determined, by language sample analysis and standardized testing, to have language skills falling within normal developmental ranges. Available standardized test scores plus MLU-m were used to construct a language development profile for each child. Each participant was required to have a majority of their scores at 1.5 standard deviations below the mean or higher. These guidelines were generally based on Tomblin et al.'s (1996) recommended criteria for diagnosis of language disorders in kindergarteners,[3] but were slightly relaxed in order to allow for more variation in language skills within the sample. The aim of these criteria was to exclude any children with clear evidence of language impairment, but to still allow for variation in language abilities.

The three corpora used different sets of measures to construct this profile of participants' language skills. In the Ratner-MacWhinney corpus, available tests were

---

[3] According to Tomblin et al.'s (1996) recommended criteria for diagnosing language impairment, children must score at least 1.25 standard deviations below the mean on two out of five composite scores.

the Peabody Picture Vocabulary Test-4 (PPVT-4) (L. M. Dunn & Dunn, 2007), the Clinical Evaluation of Language Fundamentals Preschool-2 (CELF-P2) Sentence Structure subtest, and the CELF-P2 Concepts and Following Directions subtest (Semel et al., 2004). In the Ratner corpus, available test scores for each child were some combination of the Peabody Picture Vocabulary Test-Revised (PPVT-R) (L. M. Dunn & Dunn, 1981), the Expressive One-Word Picture Vocabulary Test-Revised (EOWPVT-R) (Gardner, 1990), the Clinical Evaluation of Language Fundamentals-Preschool (CELF-P) Linguistic Concepts subtest, and CELF-P Word Structure subtest (Wiig et al., 1992). In the Weismer corpus, the MacArthur Communicative Development Inventory-Words and Sentences (M-CDI) (Fenson et al., 1993), Preschool Language Scale-3 (PLS-3) Auditory Comprehension subtest, and the PLS-3 Expressive Language subtest (Zimmerman et al., 1992) were administered. MLU-m was based on all fully intelligible utterances so that it could be compared to norms available through the CHILDES database (Bernstein Ratner et al., 2020).

Language Sample Collection and Preparation

CWS and CWNS from all three corpora participated in play-based language sampling tasks with examiners and/or caregivers, and transcripts document language produced in these play-based interactions. The set of toys provided differed across corpora. While seven of the CWS-CWNS pairs contained children from two different original studies, there is no reason to believe that the toys provided would have impacted the relationships between language and fluency as they will be reported.

All transcripts were compiled in CHAT format, for analysis with CLAN language sample analysis software (MacWhinney, 2000). For the 64 participants combined, there were 98 total initial transcripts. Because transcripts created in CHAT format are linked to the media file, if there were separate video or audio files created by examiners for parent-child play and examiner-child play from the same visit, this resulted in two transcripts. Whether there were two initial transcripts or one was determined by examiners' decisions about when to start and stop their video or audio recorders. Therefore, the data from the 34 participants whose data were initially treated as two separate transcripts are not thought to differ in any systematic way from data from the 30 participants with one initial transcript. Prior to analysis, all samples for each child were combined into a single file so that 64 composite transcripts were analyzed. For transcription and reliability procedures, the 98 separate language sample transcripts were treated separately to permit playback of linked media files. The 98 initial transcripts ranged from 35 to 348 fully intelligible utterances in length (M = 166.6, SD = 67.8), and the 64 composite transcripts ranged from 41 to 514 fully intelligible utterances (M = 255.1, SD = 113.9).

Disfluency Coding

SLDs and TDs were coded using CHAT's fluency coding conventions (MacWhinney, 2000). SLDs coded were part-word repetitions, prolongations, blocks, broken words, and monosyllabic word repetitions. TDs coded were fillers "um" and "uh," word fragments, multisyllabic word repetitions, multisyllabic word revisions, phrase repetitions, and phrase revisions. Fillers were limited to "um" and "uh" as these are the two fillers specifically listed by Rispoli (2003) and Rispoli et al. (2008); lexical items such as "well" or "like" that are sometimes included in the filler (or interjection) category were not coded as fillers in this study. These match the disfluencies presented in Table 1.

In some of the transcripts, there were repetitions of (a) a name (e.g., "mommy mommy I wanna cut a pizza cutter, or "pig pig where are you?") (b) "look," "hey," or "oh" (e.g., "look look he's knocking the bird down," "oh oh space right here," or "hey hey that's the dragon wing," (c) "yes," "no," or another word marking agreement or disagreement (e.g., "no no you're the doctor"), or (d) an adjective said multiple times for emphasis (e.g., "he's a big big cat"). When there was a repetition of one of these types, and coders perceived the repetition as intentional rather than disfluent, they did not include the repetition marker in the transcript. Omitting the repetition marker in a CHAT file results in the utterance being classified as fluent unless it contains another disfluency elsewhere. This decision was made to avoid over-marking disfluencies in cases in which the speaker intended to say a word more than once; this impacted 99 out of 16,328 fully intelligible utterances.

Segmentation Procedures

Samples were segmented into utterances based on the *two out of three criteria* rule. This rule is described by Bernstein Ratner, Brundage and Fromm (2020) and based on Stockman's (2010) finding that use of at least two criteria in boundary placement resulted in more reliable segmentation decisions. According to this rule, an utterance boundary is placed when two of the following three criteria are met: (a) a perceptible pause, (b) terminal intonation contour, and (c) complete grammatical structure. The following rule, similar to Rispoli (2003) was also used: A maximum of two independent clauses connected by a coordinating conjunction could occur in one utterance; after this, another utterance automatically began. Because participants were so young, this additional rule was not often required. Also following Bernstein Ratner et al. (2020), preposed elements such as “yes” were joined to the main utterance unless they met the two out of three criteria.

Consensus Reliability Procedures

For morpheme-by-morpheme transcription, segmentation, and disfluency coding, consensus reliability procedures similar to those used in previous language sample analysis studies assessing disfluency behaviors (Hollister et al., 2017; Rispoli, 2003; Watson et al., 2011) were used. This procedure involved two passes through the transcript, and a consensus meeting with a third coder to discuss any locations where the research assistant completing the second pass disagreed with the first pass transcription.

First Pass. Transcripts were created by undergraduate, postbaccalaureate, or master’s student research assistants. Because of the high number of research assistants involved in transcript creation (at least 26, with the transcriber name unavailable for some transcripts), these transcripts were considered a starting point for the first pass. To increase consistency, I reviewed and corrected all transcripts to complete the first pass.

Second Pass. Second passes were then completed by a different set of undergraduate research assistants who had not worked on the original transcription. The second pass involved checking (a) morpheme-by-morpheme transcription (i.e., whether the words and morphemes transcribed matched what was heard), (b) segmentation (i.e., utterance boundary placement), and (c) disfluency transcription.

Before initiating second passes, these research assistants completed approximately 20 hours of training and were required to obtain high levels of consistency with master transcripts created by the first author.

Specifically, for morpheme-by-morpheme transcription, research assistants were asked to transcribe 50-morpheme-long portions of transcripts. They were required to achieve at least 90% point-by-point agreement with a master transcript on two separate 50-morpheme transcripts. For segmentation, research assistants were presented with transcript sections in which consecutive utterances produced by the child (i.e., the child's entire "turn") were presented on one line. These transcript sections were 30 turns long. They then inserted utterance boundaries. Research assistants were required to achieve 90% point-by-point agreement to pass these tests as well. For disfluency coding, research assistants were required to insert disfluencies into transcript sections containing only bare words segmented into utterances. Word fragments and fillers were removed, as were markings indicating the various disfluency types. Each transcript section used contained 50 child utterances. Research assistants' fluency-coded transcripts were then compared to master transcripts. Utterances were classified into those that were (a) fluent, (b) containing at least one SLD and no TDs, (c) containing at least one TD and no SLDs, or (d) containing at least one SLD and at least one TD. Research assistants were required to achieve a Cohen's kappa value of .80 or higher on two separate transcripts. Once these standards were met, research assistants were permitted to participate in checking transcripts via a second pass and to participate in the consensus process.

During the second pass, research assistants read the transcript while playing the video or audio. They marked any locations where they disagreed with the first pass transcription on (a) morpheme-by-morpheme transcription, (b) segmentation, or (c) disfluency coding. Research assistants were simply told that the first pass was created by a different transcriber to avoid their deferral to my coding decision. They were told that the goal was to check the transcript and ensure that they were as representative as possible of the child's output.

Consensus Procedures. During consensus discussions, the second pass research assistant directed a consensus rater (another research assistant) to the section of the transcript where the second pass research assistant disagreed with the first pass. The second pass and consensus raters discussed the section in question and attempted to reach consensus. The mean number of marked utterances per participant brought to these discussions were 0.83 (*SD* =1.34) for segmentation, 4.77 (*SD* = 4.24) for fluency coding, and 6.64 (*SD* = 6.30) for morpheme-by-morpheme transcription. Including partially unintelligible utterances, which were reviewed during this process to ensure that there was agreement on their intelligibility, and utterances that would be excluded in the data reduction process, there were 17,426 utterances reviewed. This means that 0.3% of utterances were marked and discussed for segmentation, as were 1.8% for fluency coding, and 2.4% for morpheme-by-morpheme transcription. Through this process, consensus was reached for 100% of utterances.

Available Data Summary

After morpheme-by-morpheme, segmentation, and disfluency coding consensus had been achieved, the samples were labeled with exclusion codes so that each research question could be addressed using only the utterances appropriate for answering that question. As is traditional in language sampling studies, fully and partially unintelligible utterances were excluded from all analyses. As previously stated, there were 16,328 fully intelligible utterances, with a mean of 255.1 per participant (range = 41 to 514). See Table 2 for more information about the available utterances produced by each participant. The utterances included in the analyses for each question will be later described.

Table 2

Description of Participants

			CWS		CWNS	
Pair	Corpora	Gender	Age.	Intell. Utts	Age.	Intell. Utts
1	RM-RM	male	28	99	30	242
2	R-W	male	28	183	29	254
3	R-W	male	29	274	29	177
4	R-R	male	29	276	29	167
5	RM-RM	male	31	64	32	258
6	R-W	male	31	248	31	147
7	R-W	male	31	255	31	297
8	R-W	male	32	48	32	196
9	R-R	male	32	120	32	267
10	R-R	female	32	139	31	183
11	RM-RM	female	32	237	33	357
12	RM-RM	male	32	442	31	483
13	R-W	male	33	230	33	107
14	R-R	male	33	252	33	288
15	R-W	female	33	348	32	238
16	RM-RM	male	34	215	34	305
17	RM-RM	female	34	272	30	277
18	RM-RM	male	34	274	33	480
19	RM-RM	female	34	470	35	451
20	R-R	male	35	71	35	174
21	RM-RM	male	35	338	35	323
22	R-R	male	36	460	39	41
23	R-R	male	37	284	37	147
24	RM-RM	male	39	373	36	350
25	R-R	male	40	137	40	279
26	RM-RM	male	40	238	38	415
27	R-R	male	41	65	41	240
28	R-R	male	45	166	46	147
29	RM-RM	male	46	352	48	319
30	RM-RM	male	47	172	47	200
31	RM-RM	female	47	298	45	285
32	RM-RM	male	47	514	50	320

Note. Age is in months. Intell. utts. is the number of fully intelligible utterances produced by the child. R-W indicates that a CWS from the Ratner corpus was matched to a CWNS from the Weismer corpus. R-R indicates that both participants were from the Ratner corpus. RM-RM indicates that both participants were from the Ratner-MacWhinney corpus.

Variable Coding

The following utterance-level factors were automatically computed in CLAN or hand-coded.

Length

Length in morphemes was obtained using CLAN's automatic parser, MOR. This parser has been found to be 94% accurate and is thought to result in utterance lengths that are more accurate and consistent than hand-MLU computation (Bernstein Ratner & MacWhinney, 2016).

DSS Complexity Level

As discussed in the introduction, language samples can be scored for overall DSS level by examining 50 utterances that contain a noun and verb in a subject-predicate relationship and determining which of a set of language structures are included in each utterance (Lee, 1974). For instance, the first person pronoun "I" receives one point, but the reflexive pronoun "myself" receives five points. Additionally, under DSS guidelines, utterances without grammatical errors receive an additional sentence point. Under traditional DSS guidelines, the DSS score is the mean score of the 50 utterances, including points awarded for the structures in the utterance and the sentence points. Because grammaticality was its own predictor in this study, the *DSS complexity* score used in this project was defined as the DSS level for the utterance, without including the additional point for grammatical accuracy. DSS complexity score was determined using CLAN's DSS function. One benefit of coding individual utterances for DSS complexity, rather than using an overall DSS score, is that it was not necessary for a participant to have produced 50 DSS-eligible utterances for their utterances to be included in the analyses. Any individual DSS-eligible utterance could be coded with a DSS complexity score, regardless of how many DSS-eligible utterances a participant produced.

Language Level

The child's overall language level was represented by what I will refer to as *MLU-m-alternative (MLU-m-a).* This is the MLU-m score computed after eliminating utterances not thought to best reflect the child's ability to spontaneously generate language. The utterances used in computation of MLU-m-a were the same utterances that were included in the full analyses for questions 1 and 2. For MLU-m-a and questions 1 and 2, any utterances that were not considered to reflect the child's ability to spontaneously generate language were eliminated. Similar to exclusion criteria used by Rispoli (2003), Hollister et al. (2017), and Watson et al. (2011), the following were excluded: (a) immediate self-repetitions, (b) imitative utterances (i.e., exact repetitions of the immediately preceding adult utterance), (c) isolated filler words or word fragments, (d) utterances that were incomplete (i.e., those that trailed off or were interrupted), and (e) rote utterances or labeling (e.g., singing, counting, or labeling a set of objects).

Another set of utterances were excluded because their inclusion may have inflated or deflated the impression of a child's language ability. These were (a) single word responses to questions, (b) utterances consisting of only "yes", "no" or another word meaning "yes" or "no" (e.g., "sure"), and (c) following Rispoli et al. (2008) and Wagovich et al. (2009), utterances consisting of only "I don't know." "I don't know" was excluded also because it is common in young children's speech and it may not reflect the planning load of a typical four-morpheme utterance (Rispoli et al., 2008). After excluding these utterances, 11,142 utterances remained (for individual participants, $M = 174.1$, range = 41 to 343).

MLU-m-a was different from the child's overall MLU-m that was used to determine eligibility, as the overall MLU-m used for eligibility determination was based on all fully intelligible utterances so that it could be compared to available norms. Because most excluded utterances were one-word responses to questions or one-word productions of a word meaning "yes" or "no" (even if not a response to a question), MLU-m-a was higher than MLU-m; it was 0.8 morphemes higher on average.

Grammaticality

Utterances were coded for grammaticality based on whether they were acceptable in the adult grammar of the dialect spoken by the child. Three CWS and two CWNS were thought to be speakers of both Mainstream American English (MAE) and African American English (AAE) and the other 29 CWS and 30 CWNS were thought to be speakers of MAE only. Grammaticality coding for MAE-speaking participants followed a guide based largely on Eisenberg et al. (2012). For participants who spoke both MAE and AAE, utterances were considered grammatical if they were adultlike in either dialect. Coding for these participants was guided by an internal lab manual based on Oetting and MacDonald (2001) and Oetting and Pruitt (2005) that was developed for a project assessing the utility of language sample analysis across dialects.

Specifically, the “intended utterance” (i.e., the utterance after removing any disfluencies) was coded for grammaticality. See Table 3 for examples of intended utterances.

Table 3

Intended Utterance Examples

Disfluent utterance	Intended utterance
Um that- this- that's a unicorn.	That’s a unicorn.
I- um the the bl-black one mmmakes me s- fall.	The black one makes me fall.
They’re just- uh these guys are just lining up.	These guys are just lining up.
I want- I want do the puzzle on the table.	I want do the puzzle on the table.

Coding Reliability

For length and MLU-m-a, reliability depends primarily on accurate morpheme-by-morpheme transcription and utterance boundary placement, which were checked through the consensus transcription reliability process. Length and DSS complexity level were computed automatically by CLAN. MLU-m-a was also computed automatically by CLAN, and it also depended on the accurate and reliable coding of utterances for exclusion. Of the fully intelligible, excluded utterances, 81.5% were excluded automatically through the use of functions available in CLAN. The other 18.5% were excluded through hand-coding by research assistants. I checked the first five files coded by each research assistant, and all subsequent files were checked for accuracy by a second research assistant.

Grammaticality was hand-coded by research assistants and underwent consensus reliability procedures. Before participating in grammaticality coding for MAE-only speaking participants, research assistants completed training and passed two tests of grammaticality coding for MAE-speakers. Research assistants who worked on grammaticality coding in transcripts where the child spoke AAE and MAE received additional training in AAE features and passed two tests of grammaticality coding for AAE and MAE speakers. Tests involved coding grammaticality in 30-utterance-long transcript sections. To pass each test, research assistants were required to achieve 90%, point-by-point agreement with a master transcript that I coded. Research assistants coded the transcript for grammaticality to complete the first coding pass. Then, I completed the second pass and marked any utterances on which I disagreed with the first pass decision. Finally, a third coder reviewed utterances on which there was disagreement between the two coders (and to avoid their deferral to my coding choice, they were not told which coder chose which code). Through this process, consensus was reached on 100% of utterances.

Chapter 3: Study 1 – Grammaticality and Other Utterance-Level Syntactic Predictors of Stalls, Revisions, and SLDs

Questions 1 and 2 Analysis and Results

Initial Analysis Plan

All statistical models were run in R version 3.6.1 (R Core Team, 2019), using the lme4 package version 1.1-21 (Bates et al., 2015). The analysis plan involved using the same set of logistic mixed effects models to address questions 1 and 2 and running separate models to assess the impact of various predictors on the production of each disfluency type (SLDs, stalls, and revisions).

I predicted that ungrammatical, longer, and more complex utterances would be more likely to contain SLDs and stalls, but that the MLU-m-a of the child who produced the utterance would not relate to the likelihood of a stall or SLD occurring. For revisions, I expected a contrasting pattern in which utterances from children with higher MLU-m-as would be more likely to contain revisions, but that grammaticality, length, and complexity would not relate to the likelihood of a revision occurring.

The initial plan was to run a set of three models including grammaticality, length, complexity, and MLU-m-a as fixed factors, with a random intercept for participant. Continuous level 1 predictors (i.e., those that varied across clusters, or participants), which were length and complexity, were cluster-mean centered. Grammaticality was a binary predictor coded as 0 = grammatical and 1 = ungrammatical. There was one continuous level 2 predictor (i.e., a factor that varied at the level of the clustering variable, participant), MLU-m-a, and this was grand mean centered.

One limitation of this approach is that only about half of the total set of utterances otherwise available for these analyses were DSS-eligible (specifically, 50.6%). In order to determine whether DSS complexity level contributed to model fit, two sets of models were run on the 50.6% of utterances that were DSS-eligible. In the first set of models, DSS complexity was included, and in the second set it was omitted (and "set of models" here means three models with the three disfluency types as outcomes). Likelihood ratio tests were planned to determine whether the inclusion of DSS complexity level contributed to model fit. These likelihood ratio tests would assess models using only DSS-eligible utterances that included or did not include DSS complexity to determine whether DSS complexity improved model fit. If DSS complexity did not significantly contribute to model fit, then models using the entire set of utterances eligible for questions 1 and 2 would be used, without DSS complexity included as a predictor.

The initial analysis plan, which was later modified, involved comparing utterances that contained SLDs only and no other disfluencies to fluent utterances; utterances that contained stalls only and no other disfluencies to fluent utterances; and utterances that contained revisions only and no other disfluencies to fluent utterances. The way in which the outcome variable was coded changed between the initial plan and the final dissertation, but this initial approach was used in pilot analyses that will be outlined in the next section.

Pilot Analyses

Pilot analyses were conducted using data from 10 pilot participants (5 matched pairs of CWS and CWNS). With data from the pilot participants, likelihood ratio tests indicated that DSS complexity did not significantly contribute to model fit for any disfluency type. Therefore, results from pilot analyses using all eligible utterances and omitting the DSS complexity predictor will be described.

Both grammaticality and length were significant predictors of SLD and stall production in the pilot models, so no power analyses were needed to determine whether 64 participants would provide sufficient power. I expected higher MLU-m-a to be associated with a higher likelihood of revisions, and the pilot models indicated a nonsignificant relationship going in the opposite direction. Because of the directionality of the relationship, power analyses could not be run to determine the sample size that would be needed to have sufficient power for detecting the predicted effect of MLU-m-a on revisions. In sum, the planned sample size was expected to have sufficient power to detect most effects. It was unclear, however, whether there was sufficient power to detect a positive association between MLU-m-a and the likelihood of utterances containing revisions because of the direction of this relationship in the pilot data.

Modifications to the Initial Analysis Plan

Under the initial plan for classifying fluency outcomes, utterances with both an SLD and a TD, or with both types of TDs, or with SLDs, stalls, and revisions, were removed from the analyses. This approach may have worked well if the study had assessed CWS only or CWNS only. However, when assessing the final data set, it became apparent that a much higher percentage of utterances from CWS than CWNS would be discarded under this method because a high percentage of TD-containing utterances from CWS also had SLDs. For CWS, 44.2% of their TD-containing utterances also had an SLD, in contrast to 15.2% for CWNS. In order to not bias the sample of TD-containing utterances analyzed for CWS, a different classification system was adopted. With the new approach, utterances were coded for whether they contained at least one SLD or not, at least one stall or not, and at least one revision or not. For example, when determining whether an utterance was SLD-containing, stalls and revisions would be ignored. An utterance with both an SLD and a stall would be coded as SLD-containing for the purpose of models assessing whether certain predictors made SLDs more likely to occur. It would also be coded as stall-containing for the purpose of assessing whether certain predictors made stalls likely to occur. It would be coded as not revision-containing in the models assessing whether certain predictors made revisions more likely to occur. This approach improved on the initial plan by preserving a higher percentage of the utterances produced by CWS. The impact of this approach (and associated concerns) will be explored in detail in *post-hoc* analyses and the discussion section. The results of the primary models that were used to address questions 1 and 2 were also run as they were initially planned (i.e., comparing SLD-only utterances to fluent utterances, stall-only utterances to fluent utterances, and revision-only utterances to fluent utterances) and the output of these models can be seen in Appendix A.

Final Analysis Plan

The final analysis plan made the single modification of coding each disfluency type independently within the utterance. The same procedures outlined in the initial analysis plan were used to determine whether to restrict the utterances to those that were DSS-eligible. Therefore, the final models run would be logistic mixed effects models including the following fixed effects: a binary grammaticality predictor coded as 0 = grammatical, 1 = ungrammatical, cluster-mean centered length in morphemes, a grand-mean centered MLU-m-a predictor, and possibly a cluster-mean centered DSS complexity (depending on the results of the likelihood ratios tests). They also would include a random intercept for participant to account for the nesting of utterances within participants. The outcomes, across the three models run, would be (a) whether the utterance contained at least one SLD or not, (b) whether the utterance contained at least one stall or not, and (c) whether the utterance contained at least one revision or not. Outcomes were coded as 0 = not containing the disfluency of interest, and 1 = containing the disfluency of interest (SLD, stall, or revision). Positive regression coefficients, therefore, would reflect increasing likelihood of disfluency with an increase in the value of the predictor. There were 11,142 fully intelligible utterances that met the inclusion criteria for questions 1 and 2 as outlined in the general methods section. Alpha was set to .017 because of the three models for the three disfluency types. See Table 4 for descriptive information about the utterances and predictors.

Table 4

Length and Grammaticality of Fluent Utterances and Utterances Containing Three Disfluency Types

	n	% of Utts.	Length *M (SD)*	% Ungram.
CWS				
Fluent	3,554	66.8	3.4 (2.0)	16.7
SLD-containing	1,332	25.0	5.0 (2.5)	34.7
Stall-containing	538	10.1	4.8 (2.6)	36.1
Revision-containing	312	5.9	5.5 (2.7)	38.8
Total	5,322	100	3.9 (2.3)	22.2
CWNS				
Fluent	4,852	83.3	4.0 (2.3)	17.8
SLD-containing	410	7.0	5.6 (2.9)	25.4
Stall-containing	365	6.3	5.2 (2.8)	21.4
Revision-containing	346	5.9	5.5 (2.8)	28.0
Total	5,820	100	4.3 (2.4)	18.8
All Participants				
Fluent	8,406	75.4	3.8 (2.2)	17.3
SLD-containing	1,742	15.6	5.1 (2.6)	32.5
Stall-containing	903	8.1	5.0 (2.7)	30.1
Revision-containing	658	5.9	5.5 (2.7)	33.1
Total	11,142	100	4.1 (2.4)	20.4

Note. % of Utts. = percentage of utterances produced by CWS, CWNS, or all participants that were fluent or contained each disfluency type. % Ungram. = % ungrammatical. Some utterances are represented more than once in the above chart because they contained more than one type of disfluency. Therefore, percentages do not add up to 100%.

All stalls were generally considered as one group in the analyses that follow, as were all revisions. For descriptive purposes, breakdowns of the types of stalls and revisions are shown in Table 5 and Table 6.

Table 5

Types of Stalls in Stall-Containing Utterances

Stall Type	*n*	% of Stalls
Filler(s) only	597	66.1
Phrase repetition(s) only	246	27.2
Multisyllabic word repetition(s) only	21	2.3
Phrase repetition(s) & filler	33	3.7
Multisyllabic word repetition(s) & filler	3	0.3
Multisyllabic word repetition(s) & phrase repetition(s)	3	0.3
Total	903	100

Table 6

Types of Revisions in Revision-Containing Utterances

Revision Type	*n*	% of Revisions
Phrase revision(s) only	214	32.5
Word revision(s) only	205	31.2
Fragment(s) only	128	19.5
Phrase revision(s) & fragment(s)	78	11.9
Word revision(s) & phrase revision(s)	24	3.6
Word revision(s) & fragment(s)	5	0.8
Word revision(s), phrase revision(s) & fragment(s)	4	0.6
Total	658	100

Note. Utterances in which the phrase being replaced (the reparandum) contained a fragment, such as, “the chair is too s- the chair is too big” were classified under “phrase revision(s) & fragment(s)” or under “word revision(s), phrase revision(s) & fragment(s)” if they also contained at least one word revision.

Questions 1 and 2 Results

The first step was to determine whether DSS complexity improved model fit (or, in other words, accounted for additional variance when controlling for grammaticality, length, and MLU-m-a). Using the 5,641 DSS-eligible utterances only, models including grammaticality, length, MLU-m-a, and DSS complexity were compared to models including grammaticality, length, and MLU-m-a. Results of likelihood ratio tests indicated that DSS complexity did not significantly improve model fit for the SLD model, ($\chi^2(1) = 0.01$, ns), stall model ($\chi^2(1) = 1.16$, ns), or revision model ($\chi^2(1) = 0.11$, ns). Because inclusion of DSS complexity did not improve model fit for any of the models, and the inclusion of DSS complexity required elimination of about half of the available utterances, models using all utterances and omitting DSS complexity as a predictor will be reported and interpreted.

Utterances with grammatical errors had higher odds of containing all three disfluency types, when holding utterance length and MLU-m-a constant ($z = 8.34$, $p < .001$ for SLDs, $z = 5.06$, $p < .001$ for stalls, and $z = 5.97$, $p < .001$ for revisions). The presence of a grammatical error was associated with a 1.76-times increase in the odds it would contain at least one SLD, a 1.53-times increase in the odds it would contain at least one stall, and a 1.73-times increase in the odds it would contain at least one revision, all when holding length and MLU-m-a constant. Similarly, increasing length increased the odds that utterances would contain all three disfluency types, when holding grammaticality and MLU-m-a constant ($z = 19.88$, $p < .001$ for SLDs, $z = 8.67$, $p < .001$ for stalls, and $z = 12.98$, $p < .001$ for revisions). An increase of one morpheme in the length of an utterance was associated with a 1.29-times increase in the odds that it would contain at least one SLD, a 1.14-times increase in the odds it would contain at least one stall, and a 1.23-times increase in the odds it would contain at least one revision, all when controlling for grammaticality and MLU-m-a. MLU-m-a was unrelated to the odds that an utterance would contain at least one SLD, when holding utterance length and grammaticality constant ($z = 0.46$, ns). Utterances produced by participants with higher MLU-m-as had higher odds of containing stalls and revisions, when holding utterance length and grammaticality constant ($z = 3.41$, $p < .001$ for stalls, and $z = 3.09$, $p = .002$ for revisions). An increase in one morpheme in the MLU-m-a of the child who produced the utterance was associated with a 1.58-times increase in the odds it would contain at least one stall and a 1.29-times increase in the odds it would contain at least one revision, when controlling for utterance grammaticality and length. See Table 7 for model output and *OR*s for the models used to address questions 1 and 2.

Table 7

Results from Mixed Effects Models of Predictors of Disfluency for Three Disfluency Types

Model and Effect	*B*	*OR*	*z*	*p*
SLD				
Fixed effects				
Intercept	-2.35	0.10	-16.06	<.001
Grammaticality	0.57	1.76	8.34	<.001
Length	0.26	1.29	19.88	<.001
MLU-m-a	0.08	1.09	0.46	.646
Random effect	*Var.*			
Participant	1.21			
Stall				
Fixed effects				
Intercept	-2.90	0.06	-26.61	<.001
Grammaticality	0.43	1.53	5.06	<.001
Length	0.13	1.14	8.67	<.001
MLU-m-a	0.46	1.58	3.41	<.001
Random effect	*Var.*			
Participant	0.55			
Revision				
Fixed effects				
Intercept	-3.10	0.04	-42.67	<.001
Grammaticality	0.55	1.73	5.97	<.001
Length	0.21	1.23	12.98	<.001
MLU-m-a	0.26	1.29	3.09	.002
Random effect	*Var.*			
Participant	0.13			

Note. Outcomes coded as 0 = no SLD in utterance, 1 = SLD in utterance, 0 = no stall in utterance, 1 = stall in utterance, and 0 = no revision in utterance, 1 = revision in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Var. = variance. Number of participants = 64. Number of utterances = 11,142.

Summary of Results

All three disfluency types had higher odds of occurring in ungrammatical and longer utterances. Both stalls and revisions had higher odds of occurring in utterances produced by children with higher MLU-m-as, while MLU-m-a was not a significant predictor of whether SLDs would occur. Complexity was removed from the models because it did not significantly contribute to model fit, so no conclusions can be drawn about whether predictions about complexity were supported. In the context of these findings, there are several additional analyses that may be valuable in interpreting the primary findings.

Additional Post-Hoc Analyses

Group. First, the relationship between linguistic predictors and disfluency is of particular interest for CWS because of the potential clinical application of findings for this group. In addition, some prior studies that had results contrasting with some of the findings here have only enrolled CWNS (Rispoli, 2003; Rispoli et al., 2008) or CWS (Wagovich et al., 2009) so assessing these relationships within each group would enable closer comparisons to this previous work. For these reasons, the role of group in the results was assessed.

In order to determine whether the factors that predicted SLD, stall, and revision production for all participants together were also significant predictors of the disfluency types for each group, models were run on utterances from CWS and CWNS separately. It was possible that production of disfluencies, particularly SLDs, may have been influenced by different factors across groups. For these *post-hoc* tests, alpha was set to .017 due to the three models run for each group.

Results of the three models run on utterances produced by CWS looked very similar to those reported for the full set of participants. As with the models run using utterances from all participants, ungrammatical and longer utterances had higher odds of containing SLDs ($z = 7.70$, $p < .001$ for grammaticality and $z = 17.83$, $p < .001$ for length), stalls ($z = 4.82$, $p < .001$ for grammaticality and $z = 6.59$, $p < .001$ for length) and revisions ($z = 5.03$, $p < .001$ for grammaticality and $z = 9.19$, $p < .001$ for length), holding all other factors constant. As with the models run using utterances from all participants, utterances produced by CWS with higher MLU-m-as had higher odds of containing stalls, and revisions, holding utterance length and grammaticality constant ($z = 3.00$, $p = .003$ for stalls, and $z = 3.35$, $p < .001$ for revisions). See Appendix B for CWS-only model output.

Results from the models run only on utterances produced by CWNS differed in a few ways from those run on utterances from both groups. First, utterances produced by CWNS with higher MLU-m-as had higher odds of containing SLDs, holding utterance length and grammaticality constant ($z = 2.78$, $p = .005$). A one-morpheme increase in the MLU-m-a of a CWNS was associated with a 1.47-times increase in the odds that an utterance they produced would contain an SLD, holding grammaticality and length constant. Secondly, grammaticality was not a significant predictor of whether an utterance produced by a CWNS would contain a stall, holding utterance length and MLU-m-a constant ($z = 1.80$, $p = .072$). If an unadjusted alpha of .05 had been used, there would have been a trend towards an effect in which ungrammatical utterances produced by CWNS had higher odds of containing stalls, controlling for grammaticality and MLU-m-a. Ungrammatical utterances produced by CWNS had 1.29-times higher odds of containing stalls compared to grammatical utterances. This contrasts with the 1.68-times increase for CWS. And third, for utterances produced by CWNS, an increase in MLU-m-a was not associated with a significant increase in the odds that it would contain a revision, controlling for grammaticality and length ($z = 1.03$, ns). See Appendix C for CWNS-only model output.

Finally, group was added to the three original models to determine whether the odds of each disfluency type occurring differed by group when controlling for all other predictors. When controlling for grammaticality, length, and MLU-m-a, utterances produced by CWS (unsurprisingly) had greater odds of containing SLDs ($z = 8.59$, $p < .001$) and stalls ($z = 2.71$, $p = .007$), but there was no significant difference between groups in the odds of revisions occurring ($z = 0.06$, ns). If an utterance was produced by a CWS rather than a CWNS, there was a 5.61-times increase in the odds that it would contain an SLD and a 1.73-times increase in the odds that it would contain a stall, holding grammaticality, length, and MLU-m-a constant. See Appendix D for full model output.

Gender. Next, it was potentially useful to explore the possibility that the results identified so far may not be reflective of all children. In the recruiting process, efforts were not made recruit any particular gender balance for the CWS (and CWNS were matched on gender to the CWS). This resulted in a 4.3:1 male-to-female ratio. Therefore, it was of interest to determine whether relationships between linguistic factors and disfluency were different for girls than they were for the boys; boys made up 81.3% of the participants and 77.9% of the utterances included in the models and therefore contributed greater weight to the outcomes. This question was particularly important in the context of a substantial number of studies assessing language-fluency relationships in young CWS that have enrolled only boys (Gaines et al., 1991; Logan & Conture, 1995; Melnick & Conture, 2000; Yaruss, 1999; Zackheim & Conture, 2003). In addition, MacWhinney and Osser (1977) found that boys engaged in coplanning more (typically seen through revisions), while girls engaged in preplanning more (typically seen through stalls), so it was of interest to know whether there were gender differences in the use of stalls or revisions.

The same factors that were significant predictors in models including utterances from all children were significant in the models run on male-produced utterances only, with one exception. While MLU-m-a was a significant predictor of whether an utterance would contain a revision when all utterances were included, it did not reach significance (at the adjusted alpha level of .017) as a predictor of revisions when only utterances produced by males were included ($z = 2.24$, $p = .025$). See Appendix E for full output of the boys-only models.

Because there were only 12 girls enrolled in the study, models run on only their utterances included substantially fewer utterances than models run on all participants' utterances (2,462 utterances, compared to 11,142). In the three models looking at relationships between linguistic predictors and disfluency in utterances produced by girls, all relationships occurred in the same direction as those in the models using utterances produced by all participants. Grammaticality and length had positive coefficients in the SLD, stall, and revision models for girls' utterances. MLU-m-a had a positive coefficient in the stall and revision models, and had a positive coefficient close to zero in the SLD model ($B = 0.01$, $z = 0.02$, ns). However, in these models using utterances from girls only, grammaticality did not reach significance as a predictor of whether a revision would occur, though there would have been a trend towards significance if an unadjusted alpha had been used ($z =$

1.82, $p = .069$), and MLU-m-a did not reach significance as a predictor of whether a stall would occur ($z = 1.04$, ns). See Appendix F for full output of the girls-only models.

Finally, to further assess the relationship between gender and disfluency production, gender was added to the three original models. Controlling for length, accuracy, and MLU-m-a, gender was not a significant predictor of whether an utterance contained at least one SLD, stall, or revision ($z = 0.43$, ns for SLDs, $z = 1.52$, ns for stalls, and $z = 0.13$, ns for revisions). See Appendix G for full model output.

Types of Revisions. As expected, a relationship was identified between revision production and MLU-m-a. This relationship was expected based upon the thinking that, as children's language level increases, skills needed to produce revisions also increase (i.e., monitoring and comprehension skills, and enough lexical or structural alternatives) (Rispoli, 2003, 2018; Rispoli et al., 2008). An alternative explanation for this relationship, that was offered and rejected by Rispoli (2003), was that older children could be better able to detect their grammatical errors. Rispoli (2003) reported that their participants produced only 10 successful grammatical revisions, compared to 419 nongrammatical revisions, and therefore ruled out increased correcting of grammatical errors as an explanation for the increasing revisions with language level. If children in the current study with more advanced language produced more grammatical revisions, this may partially explain the relationship between MLU-m-a and revisions in the current study. To assess this possibility, I examined the types of revisions to better understand whether revisions were occurring in this dataset to correct grammatical errors.

Revision-containing utterances were classified based on how the revision related to the grammaticality of the utterance. It was difficult to tell what a child was beginning to say when they produced word fragments (e.g., “no m- your coffee get out”), so only utterances including at least one word or phrase revision were included in this analysis. Successful grammatical revisions were identified when the portion of the utterance that was replaced (the reparandum) would have made the utterance ungrammatical, but the replacement made that part of the utterance grammatical. For instance, in “put in it- put it in my ear,” “put in it,” was replaced by “put it in.” Without the revision, the utterance would have been “put in it my ear” and been ungrammatical. The revision made the utterance grammatical, so it was coded as a successful grammatical revision. If an utterance had two or more word or phrase revisions, it was labeled as having a successful grammatical revision if at least one of the word or phrase revisions was a successful grammatical revision. There were 658 revision-containing utterances, and 530 of these contained at least one word or phrase revision. (One hundred twenty-eight only had word fragments.) Of the 530 utterances with word or phrase revisions, only 19 of these contained a successful grammatical revision. Standardized MLU-m-as (z-scores) were identified for the 15 participants who produced successful grammatical revisions to determine whether children with more developed expressive language skills produced more successful grammatical revisions. Of these 15 participants, 10 had positive standardized MLU-m-as, indicating MLU-m-as above average. Of the five participants with negative standardized MLU-m-as, two had standardized MLU-m-as of -0.01, which means that their MLU-m-as were approximately at the mean. The mean standardized MLU-m-a for participants who produced at least one successful grammatical revision was 0.65, which indicates that these participants, on average, were nearly two-thirds of a standard deviation above the mean with their MLU-m-as. Therefore, it may have been the case that a very small portion of the effect of MLU-m-a on revision production was due to increasing successful grammatical revisions with language level. However, these 19 utterances with successful grammatical revisions only made up 2.9% of the utterances with revisions. See Table 8 for a description of participants who produced successful grammatical revisions.

Table 8

Participants with Successful Grammatical Revisions

Participant	SGRs	Utts. with word/phrase rev.	Percent	Standardized MLU-m-a
CWS-2	1	8	12.5	0.59
CWS-9	1	6	16.7	0.56
CWS-14	3	14	21.4	-0.46
CWS-16	1	1	100.0	-0.60
CWS-17	1	5	20.0	-0.01
CWS-25	1	6	16. 7	0.71
CWS-30	1	10	10.0	1.76
CWS-31	1	16	6.3	1.10
CWNS-6	1	11	9.1	-0.10
CWNS-12	1	24	4.2	1.45
CWNS-16	2	8	25.0	0.74
CWNS-20	1	13	7.7	0.16
CWNS-21	2	11	18.2	1.45
CWNS-26	1	10	10.0	-0.01
CWNS-32	1	12	8.3	2.35

Note. SGR = Successful grammatical revisions. Utts. = utterances. Rev. = revision. Percent = percent of the participant's utterances with word or phrase revisions that had an SGR.

Question 3 Analysis and Results

Disfluency-Error Order Coding

For any disfluent utterances containing at least one SLD or stall (the disfluencies expected *a priori* to relate to planning), the order of the SLD(s)/stall(s) and grammatical error(s) was coded. Some disfluencies occur in the space between words, and these were considered to occur on the word following the disfluency. Similarly, errors of omission were considered to occur on the word following the omission. This question about order was asked because SLDs or stalls must occur on or before an error for it to be possible that planning the error portion of the utterance is a contributing factor to production of the SLD or stall. Therefore, disfluencies occurring before or on the same word as the error were grouped together in the analysis. Utterances with SLDs or stalls and errors were excluded from this analysis if either (a) the location of the error could not be definitively determined to be before or after the SLD or stall, such as an utterance with incorrect word order or one that was clearly ungrammatical but where the intended utterance was unclear (e.g., “same i-i-ice cream I got” was excluded); (b) there was more than one SLD/stall and at least one occurred before the error and at least one occurred after; or (c) there was more than one error and at least one occurred before the SLD(s)/stalls(s) and at least one occurred after. See Table 9 for examples of order coding.

Table 9

Examples of Disfluency-Error Order Coding for an MAE-Speaking Child

Utterance	Disfluency-error order coding
That that that _ for me.	Disfluency first or on same word as error
Those babies *i-i-is* mine.	Disfluency first or on same word as error
I _ gonna put that *in-inside.*	Error first
Her *d-d-doesn't* want two doll_.	Excluded: One error is before the disfluency and one is after.

Note. MAE = Mainstream American English. Disfluencies are italicized. Error locations are indicated by underlining or an underscore.

Disfluency-Error Order Reliability

Disfluency-error order was hand-coded by research assistants who passed two tests prior to beginning this coding. For each test, they were asked to classify 30 utterances as (a) having the error before all SLDs and stalls, (b) having all SLDs and stalls before/on the same word as the error, or (c) excluded from disfluency-error order analysis due to unclear order. To pass these tests, they were required to achieve a Cohen's kappa value of .80 or higher. Research assistants who conducted order coding for participants thought to be both AAE and MAE speakers had previously passed two tests of AAE and MAE grammaticality coding.

When qualifying utterances (i.e., those eligible for inclusion in the analyses for questions 1 and 2 that also had at least one SLD or stall and an error) in a transcript were coded by a research assistant for disfluency-error order, this completed the first pass. Then I completed a second pass by marking any utterances on which I disagreed with the first pass coding. There were 17 disagreements out of 718 qualifying utterances (2.3%). Finally, a different research assistant who had passed order coding tests reviewed utterances with these disagreements. In cases of disagreement, the disagreement-resolving research assistant was not told which code I chose and which code the first pass coder chose. Consensus was reached for all utterances through this process.

Question 3 Planned Analyses

Question 3 asks whether disfluencies tend to precede grammatical errors. In order to address this question, the 718 utterances that were eligible for analysis in questions 1 and 2 and contained at least one SLD or stall and a grammatical error were identified, as previously outlined. Then, 106 of these were excluded from the analysis because there was no clear error-disfluency order. This left 612 utterances remaining, across 59 participants, 551 having the disfluency first or on the same word as the error, and 61 having the error first. A logistic mixed effects model was run, predicting order (disfluency first or on the same word as the error vs. error first) with a random intercept for participant and no fixed effects.

A pilot analysis was conducted as part of a planned power analysis. This pilot analysis used data from the same 5 CWS and 5 CWNS who were part of the pilot analyses for questions 1 and 2. Results of the pilot analysis found a significant effect consistent with predictions when using data from the 10 participants, so the full dataset was expected to have sufficient power.

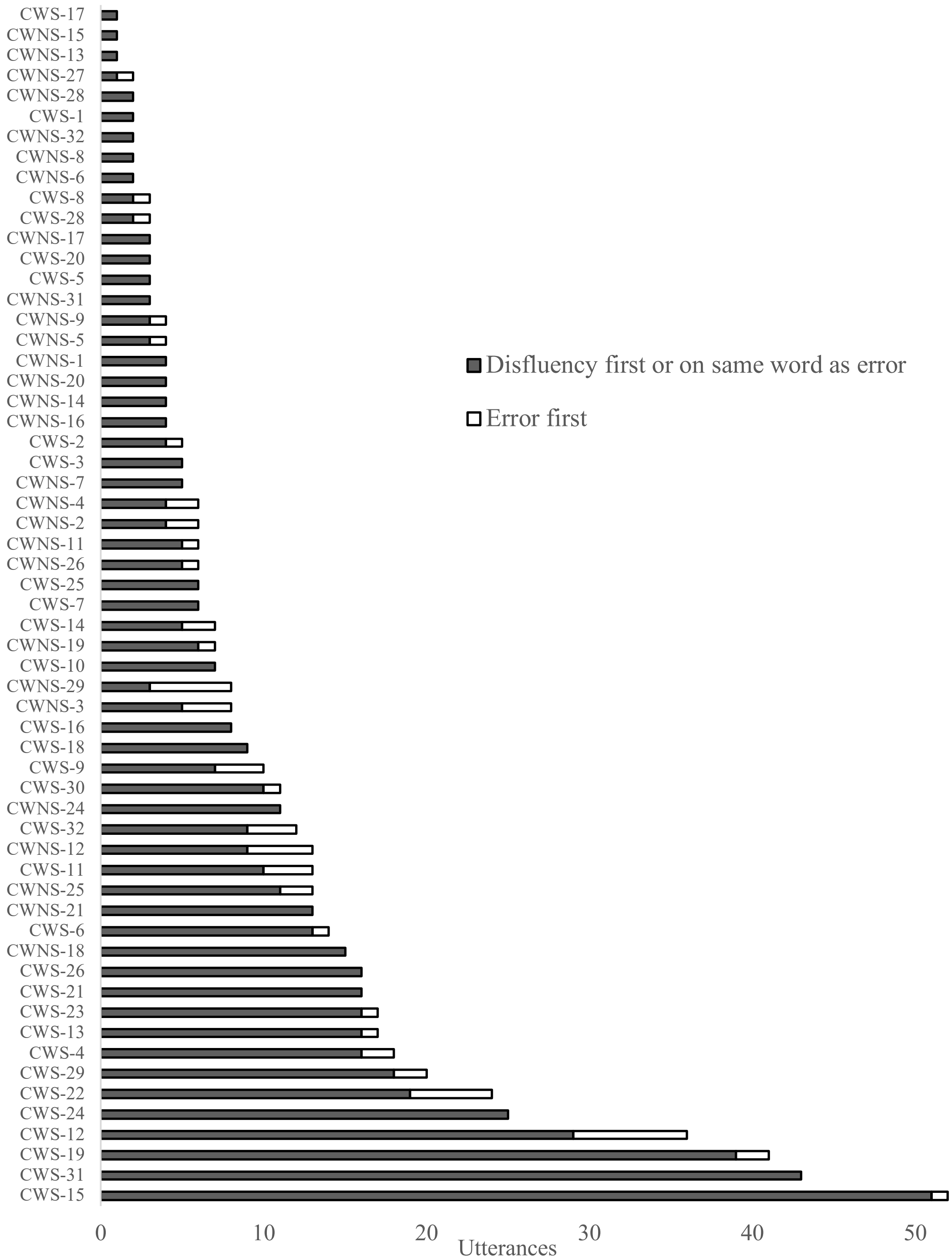
Question 3 Results

There was a significant intercept in the model, reflecting greater odds of the stall(s) and/or SLD(s) occurring before or on the same word as the error than of the error occurring before the stall(s) and/or SLD(s) ($z = 9.92$, $p < .001$). Utterances had

11.57 times greater odds of having the stall(s) and/or SLD first than of having the error first. See Figure 1.

Figure 1

Order of SLDs/Stalls and Errors.



Note. The y-axis shows the group and pair number. CWS = child who stutters. CWNS = child who does not stutter. n = 612 utterances across 59 participants.

Question 3 Additional Post-Hoc Analysis

Disfluencies tend to occur in the utterance-initial position (e.g., Buhr & Zebrowski, 2009), so it was possible that the significant effect was driven by this tendency. Therefore, an additional analysis was run only on utterances that had at least one non-initial stall or SLD. Utterances that had their only stall(s) and/or SLD(s) occurring on or before the first intended word were removed. There were 263 ungrammatical utterances across 51 participants with at least one non-initial stall and/or SLD. The model run on this limited set of utterances also had a significant intercept ($z = 5.05$, $p < .001$), reflecting greater odds of the stall(s) and/or SLD(s) occurring before or on the same word as the error than of the error occurring before the stall(s) and/or SLD(s). With utterances containing only initial-position stalls(s) and/or SLDs removed, the remaining utterances had 3.81 times greater odds of having the stall(s) and/or SLD(s) first or on the same word as the error than having the error first.

Discussion

The primary questions addressed in this study concerned how specific linguistic factors related to the production of SLDs, stalls, and revisions in the speech of young CWS and CWNS. I will first discuss results from the primary models used to address these questions, which used data from all participants, and will discuss the results of the order analysis in this context. I will then discuss results from *post-hoc* analyses that were conducted to more fully understand and interpret the primary findings.

Primary Findings

SLDs. First, I expected that ungrammatical and longer utterances would be more likely to contain SLDs, but I did not expect an association between the language level of the child who produced the utterance and whether it was SLD-containing. These predictions were supported; such results are consistent with a large body of previous research showing that longer utterances are more likely to contain SLDs (e.g., Richels et al., 2010; Yaruss, 1999; Zackheim & Conture, 2003) as well as with two of three previous studies assessing the role of grammaticality in SLD production (Bernstein Ratner, 1997; Watson et al., 2011). These findings are also consistent with the long-held proposal that SLDs occur in part due to language planning demands (e.g., Hall et al., 2007). The absence of an association between language level and SLD production, controlling for length and grammaticality, is not definitive evidence that there is no association. However, this study was well-powered to detect predicted effects, and the *p*-value for the effect of MLU-m-a on SLD production was .646, so it seems fair to say that the absence of an effect was probably not due to insufficient power. The absence of an association between the number of SLDs and language level is consistent with the conception of stuttering as a highly variable disorder (e.g., Yaruss, 2004).

Stalls. Next, I expected that stalls, like SLDs, would be more likely to occur in ungrammatical and longer utterances, because of previous suggestions that stalls are related to planning difficulty (Rispoli, 2003; Rispoli et al., 2008; Wagovich et al., 2009). These predictions were supported. Because of previous suggestions that the prerequisites to stall production are limited—that a “glitch” in sentence planning is all that is needed for a child to produce a stall—and because of previous studies failing to find an association (Rispoli, 2003; Rispoli et al., 2008; Wagovich et al., 2009), I did not expect a relationship between language level and stall production. Contrary to predictions, more advanced expressive language level was associated with a greater likelihood of stall production, when holding grammaticality and length of the utterance constant. It was not simply the case that children with more advanced language development were attempting longer sentences and that this drove their higher likelihood of stalls. Because length and grammaticality were controlled in these models, the result means that, if two children produced utterances of the same length, and the utterances were both grammatical or both ungrammatical, the utterance from the child with higher mean utterance length would be more likely to contain a stall. This result was surprising.

Given this finding, an obvious question raised is whether there are prerequisite skills needed to produce stalls (or possibly, what the prerequisite skills for stall production are). For additional context, 66.1% of the stall-containing utterances only had a filler or fillers, and 27.2% of the stall-containing utterances only had a phrase repetition or phrase repetitions and no other stalls. Therefore, understanding filler development would be the most useful in explaining the current findings.

There are several possible explanations for the increase in stall-containing utterances for children with more advanced language development. The current data cannot differentiate between these, so they may provide directions for additional research. One possible explanation for increasing stalls with language level concerns the child's ability to monitor for upcoming pauses. In an influential study, Clark and

Fox Tree (2002) found that adults tended to produce “um” before longer pauses (where syntactic planning is thought to be occurring more often) and “uh” before shorter pauses (where lexical access is thought to be occurring more often). Hudson Kam and Edwards (2008) discuss the production of “um” and “uh” as being, “deceptively complicated,” (p. 315) because they require a child to anticipate a delay. They found that 3- and 4-year-olds produced “um” and “uh” in ways that were close to adult-like patterns (i.e., before pauses) but that they did not yet differentiate between the two fillers by placing “um” before longer pauses and “uh” before shorter pauses. In the context of these findings, it is possible that children with more advanced language skills are better at monitoring their speech and knowing when a delay is approaching, and therefore when inserting a filler may be useful to the listener. This is testable; researchers could assess whether younger children have silent pauses in their speech, without preceding fillers.

A second possibility concerns the insertion of the filler into the sentence plan. Fraundorf and Watson (2014) asked adults to retell three stories while viewing bulleted lists of major plot points. They hypothesized that fillers would tend to be used at the beginnings of new messages, where the speaker would not be required to re-plan in order to put a filler into the utterance. They found evidence supporting this hypothesis; participants tended to use fillers when encountering message-level, rather than grammatical or phonological-level difficulties. If children with more advanced language skills are better able to insert fillers into their sentence plans, this could partially explain the increase in stalls with language level.

A third possibility is related to the social function of fillers. Recent work has assessed filler production by speakers with autism spectrum disorder (ASD), whose diagnosis of ASD indicates some degree of difficulty with social communication. Speakers with ASD have been found to produce “um” less frequently than speakers with neurotypical development (Gorman et al., 2016; Irvine et al., 2016, 2016; Lake et al., 2011; McGregor & Hadden, 2018). This difference has been attributed to lower social communication skills for speakers with ASD, resulting in their less frequent use of a behavior that is intended to aid the listener (Irvine et al., 2016; Lake et al., 2011). If children with more advanced language levels also have more advanced social skills (or have more cognitive capacity available during language production to use for social processing), they could be better able to use fillers as social signals to aid with holding the floor. These three potential reasons for increases in stalls (specifically, fillers, as these accounted for approximately two-thirds of stall-containing utterances) are not mutually exclusive, nor are they an exhaustive list of possible explanations. Each may account for some portion of the increase in stalls with language level.

SLDs and Stalls and Disfluency-Error Order. SLDs and stalls had much greater odds (11.57 times greater) of preceding errors or occurring on the same word as the error than of following errors. Disfluencies tend to occur on the first word of the utterance (e.g., Buhr & Zebrowski, 2009), so their positional distribution could have fully or partially accounted for this finding. To assess this alternative explanation, utterances with only utterance-initial SLDs or stalls were removed and a second model was run. With these utterances removed, the effect remained, though at a reduced level, with SLDs and stalls now having 3.81 times greater odds of occurring before or on the same word as the error than after it. The overall effect, therefore, appears to be partially, but not fully, accounted for by the tendency of disfluencies to occur in the initial position of the utterance.

These findings help with understanding the effect of errors on the types of disfluencies that were expected, *a priori*, to be associated with planning demands (SLDs and stalls). As far as I know, this was the first analysis of the order of disfluencies and grammatical errors within utterances. It strengthens the interpretation that planning of the error-containing portion of the utterance could be influencing SLD and stall production. However, it should be stated that planning an error-containing portion of an utterance is not always associated with disfluency. Still, in utterances that have SLDs/stalls and errors, the SLD/stall tends to occur first.

Revisions. I expected that revisions would increase with language level, but be unrelated to factors measuring utterance-level demand (grammaticality and length). As predicted, utterances from children with higher language levels were more likely to contain revisions. Rispoli (2003) proposed that the reasons for increases in revisions with increasing utterance lengths are that some of the skills needed for revisions (i.e., attention to one's own speech/monitoring, language comprehension, and availability of alternative words or structures) increase with expressive language skills. Rispoli also offered the possible alternative explanation that children with more advanced language were better able to revise grammatical errors, and then rejected this alternative explanation on the basis that there were only 10 successful grammatical revisions compared to 419 nongrammatical revisions in that study's corpus. In the current study, there were 19 successful grammatical revisions out of 658 total revisions (530 which contained at least one word or phrase revision). The mean standardized MLU-m-a score of the 15 children who produced them was 0.65; children who successfully revised grammatical errors did, on average, have a higher MLU-m-a than the mean level for the total sample. However, since successful grammatical revisions only made up 2.9% of the total revisions in this sample, it is unlikely that more advanced grammatical skills were the primary reason for the increase in revisions as language level increased. The increase in revisions with MLU-m-a seems more likely attributable to one or more of the skills Rispoli discussed: language comprehension or monitoring, the ability to stop oneself from speaking and make a quick change, or having sufficient alternative words or structures to substitute. Future experimental work measuring attention to a child's own output, their comprehension skills, and their error-monitoring skills is needed in order to better understand these relationships.

This study also produced the unexpected results that ungrammatical utterances and longer utterances had higher likelihoods of containing revisions. Revisions were varied across fragments, word revisions, phrase revisions, and combinations of these, with no single category making up more than a third of the revision-containing utterances. The two possible explanations are that more-difficult-to-plan utterances (i.e., those that are longer and/or ungrammatical) are (a) more likely to contain production slips (not necessarily grammatical errors) that the child might want to revise, or (b) production slips that are produced equally across utterances are more likely to be noticed and corrected if the utterance is longer and/or ungrammatical.

The first of these possibilities seems more probable, but it is not possible to differentiate between these based on the current data. One challenge in studying production errors in the speech of young children is that it is difficult to know what utterance the child was aiming for and therefore whether there was a one-time production slip, or whether the utterance contained some kind of error reflecting the child's general language abilities (Hanley et al., 2016). To differentiate between these two possibilities, future experimental studies would have to elicit production slips in paradigms that permit the experimenter to know the intended target. A previous study used a paradigm called the "moving animals" task, in which children were asked to describe the positioning of animals moving on a screen and to correct any slips that they made. Children ages 5 to 8 did produce sufficient semantic slips and revisions to successfully analyze results (Hanley et al., 2016). If, for example, length, were manipulated within this paradigm, experimenters could determine whether production errors are more common in longer utterances and/or whether utterance length was associated with likelihood that a child was able to observe and correct those slips. Alternatively, a simple sentence repetition task, such as the one used by Bernstein Ratner and Sih (1987) may provide sufficient slips and revisions to assess relationships between planning difficulty (e.g., length or complexity), the number of slips, and the likelihood that slips are revised. An elicited production task, like the one used by McDaniel et al. (2010) may also result in enough slips and revisions. Experimenters in the McDaniel et al. study told brief stories and then used these to prompt children to produce sentences with particular structures. A moving animals-type task has the benefit of making it easy to elicit plenty of slips and revisions, while sentence repetition and McDaniel et al.-type elicited production tasks have the benefit of making it easy to manipulate length and complexity of the utterances.

Predictors of Stalls and Revisions. Previous research has suggested a dichotomy between the linguistic factors associated with stall production and those associated with revision production (Rispoli, 2003; Rispoli et al., 2008; Wagovich et al., 2009). However, the current data indicated that both stalls and revisions were associated with planning-related factors (length and complexity) and the child's language level (MLU-m-a). Language level was particularly surprising with regard to how it impacted stalls compared to revisions. While MLU-m-a was a significant predictor of both stalls and revisions, the *OR* was higher for stalls; a one-morpheme increase in MLU-m-a was associated with a 1.58-times increase in the odds of at least one stall occurring, and a 1.29-times increase in the odds of at least one revision occurring. The higher *OR* for stalls than for revisions was surprising, since no relationship was expected for stalls and a relationship was expected for revisions. The current data, therefore, suggest that there should be a reconsideration about whether there is a dichotomy in predictors of stalls and revisions and whether it generalizes to a broad range of utterance types (compared to the active declarative sentences analyzed by Rispoli, 2003, Rispoli et al., 2008, and Wagovich et al., 2009).

Additional Analyses

Group. Factors that predicted significantly increased likelihoods of SLDs, stalls, and revisions when utterances produced by all children were included were also significant predictors in the models only looking at utterances produced by CWS.

This was not true, however, when looking at utterances from CWNS alone. There were three differences for CWNS.

First, for CWNS, MLU-m-a was a significant predictor of whether an utterance would contain an SLD. This relationship can be understood by examining the definition of SLDs. As Wingate (2001) pointed out, "'stutter-like' is not stutter" (p. 381). Wingate expressed concern about the inclusion of monosyllabic whole-word repetitions among behaviors counted when stuttering is quantified and argued that monosyllabic whole-word repetitions do not share some features of other SLD types. He argued that monosyllabic whole-word repetitions do not share two core features of stuttering—iteration of some small unit of speech and perseveration—and therefore monosyllabic whole-word repetitions should be studied with regard to whether they are SLDs but not necessarily included in this category.

Among the 410 SLD-containing utterances produced by CWNS, 205 contained monosyllabic word repetitions and no other SLDs, and 133 contained part-word repetitions and no other SLDs. These part-word repetitions, when produced by CWNS, are also arguably not stuttering; they are qualitatively different, likely lacking muscle tension, awareness, and a feeling of being "stuck," all features of stuttering (Ambrose & Yairi, 1994; Tichenor et al., 2018; Tichenor & Yaruss, 2019). Monosyllabic whole-word repetitions, part-word repetitions (and probably other SLDs) produced by CWNS may be better understood as stall-like. Because stalls increase in likelihood as MLU-m-a increases, it is unsurprising that SLDs produced by CWNS do as well. I am not suggesting that monosyllabic whole-word repetitions should be removed from the SLD category, but I do suggest that researchers studying

SLD and TD production by both CWS and CWNS consider definitions and the impact of such definitions on outcomes.

The second difference for utterances produced by CWNS was that grammaticality did not reach significance as a predictor of whether an utterance would contain a stall. For utterances from CWNS, ungrammatical utterances had 1.29 times greater odds of containing at least one stall, but the *p*-value of this predictor was .072 and not significant. This contrasts with an *OR* of 1.68 and *p*-value of < .001 when only utterances from CWS were used. To better understand why this may have happened, I considered the possibility that stall-containing utterances produced by CWNS differed from those produced by CWS. One possibility is that stalls produced by CWS occur as part of the SLD (if the SLD can be defined broadly).

Of the 365 stall-containing utterances produced by CWNS, 52 (14.2%) also contained an SLD. In contrast, 223 of 538 stall-containing utterances (41.4%) produced by CWS also contained an SLD. This indicates a qualitative difference between groups in their stall-containing utterances. The next step was to determine whether the stalls occurred as part of the stuttering. Utterances were examined to determine whether the stalls immediately preceded the stuttered word or occurred in the middle of a stuttered disfluency. Some examples of utterances having their only stall (or their only string of stalls) before or in the middle of a stuttered disfluency are "uh llllet's dump it," "um uh hi-hi-him will be wo-working too," "um um um um um um um p-put it in in in in in in your face," "the baby- the baby wa-wa-wants to to get a t-temperature," and "her uh her eating it." In the first five of these examples, fillers or phrase repetitions occurred only in positions (or in strings) immediately preceding

SLDs. In the last example listed, the stall "uh" is produced between iterations of the monosyllabic word repetition.

Of the 223 CWS-produced utterances with both SLDs and stalls, only 107 (48.0%) still would have still contained stalls if the stalls immediately preceding stuttered words or occurring in the middle of an SLD (between iterations) were removed. In other words, 116 (52.0%) of the utterances from CWS with both stalls and SLDs had only stalls that immediately preceded stuttered words (or occurred in strings immediately preceding SLDs). In these cases, it is possible (and maybe probable) that the stalls were functioning as postponements or avoidance behaviors. To briefly explain, stuttered disfluencies can be accompanied by secondary behaviors. These behaviors are not stuttered speech, but are physical concomitants that occur alongside the stuttered disfluencies. Through conditioning, they become associated with the stuttered speech. A postponement is a specific type of secondary behavior that develops based on the speaker's attempts to delay or avoid the disfluency (Bloodstein et al., 2021). Secondary stuttering behaviors have been documented close to stuttering onset, in the age range of the participants in this study (Schwartz et al., 1990; Yairi, 1983). Another possibility is that some or all of these fillers had not yet developed into a full-fledged secondary behaviors, but were produced by CWS simply in attempt to avoid or delay the SLD, as anticipation and avoidance are well-documented processes in stuttering (Jackson et al., 2019; Kimmell, 1938; Tichenor & Yaruss, 2019). For example, in the utterance "uh llllet's dump it" the "uh" may be occurring because the speaker has developed a secondary behavior of producing a filler before some of their SLDs (as a postponement), or because in that particular utterance, the child was attempting to delay the prolongation.

Given this co-occurrence of stalls and SLDs in the samples of CWS, the reason for the effect of grammaticality on stall production in the full sample could be that it is primarily the SLDs in the combined dataset and CWS-only dataset that are impacted by grammaticality. The *OR*s would support this possibility; in the models using utterances from all participants, grammaticality had an *OR* of 1.76 in the SLD model, compared to the *OR* of 1.53 in the stall model. In some cases, these stalls and SLDs may both be occurring due to planning difficulty. In other cases, particularly for the 52.0% of stall-containing utterances from CWS where stalls only occurred immediately before or within SLDs, the SLD was likely partially caused by the planning difficulty, and the stall may have occurred secondarily to the SLD. To help with understanding whether this was the case, an additional model was run on the 10,919 utterances remaining after removing utterances produced by CWS that contained both SLDs and stalls. In this model, ungrammatical utterances had 1.30 times greater odds of containing at least one stall than grammatical utterances ($z = 2.63$, $p = .009$), holding utterance length and MLU-m-a constant. The full results of this model are in Appendix H. This contrasts with the *OR* of 1.53 when SLD-stall utterances from CWS are included among those containing stalls. The grammaticality effect was decreased but did not disappear when the utterances from CWS with SLDs and stalls were removed. Based on these findings altogether, it appears that there is a somewhat weaker (though still present) effect of grammaticality on stall production when these stalls appear on their own, apart from SLDs. It also seems that there is a stronger effect of grammaticality on SLD production; this effect seems to impact outcomes when looking at stall-containing utterances, since SLDs produced by CWS can be preceded by stalls that occur as avoidance/delay behaviors related to the SLD.

The third difference is that, for utterances produced by CWNS, the MLU-m-a of the child who produced the utterance was not a significant predictor of whether it contained a revision. For utterances from CWNS, a one-morpheme increase in MLU-m-a was associated with only a non-significant 1.11-times increase in the odds of the utterance containing a revision, compared to an *OR* of 1.59 and significant association for utterances from CWS. It is difficult to explain why there was a weaker association for CWNS. This finding particularly highlights the inconsistency between the current findings and those of Rispoli (2003) and Rispoli et al. (2008), who enrolled only CWNS. Rispoli (2003) and Rispoli et al. (2008) saw relationships between the child's language level and revisions but not stalls, while the opposite was true for CWNS in the current study. (Though it should also be mentioned that an association between language level and revisions but not stalls was also identified in an analysis of longitudinal data from nine CWS by Wagovich et al., 2009.)

There are possible methodological reasons for the differences between these two previous studies and the current one. The current study used a more diverse set of utterances, while these two previous studies with CWNS restricted their analysis to active declarative sentences, and Rispoli et al. (2008) further limited their analysis to utterances that were only seven phonological words or shorter. Therefore, a tighter degree of control was held with regard to which utterances were assessed in the studies by Rispoli and colleagues. It is possible that the relationship they identified between revisions and language development are only present in shorter ADSs, or that the inputs to statistical models must be tightly controlled for a difference between predictors of stalls and revisions to be identified. Rispoli (2003) found increases in revisions as both MLU-m and IPSyn increased in a cross-sectional sample of 22- through 48-month-old children, and Rispoli et al. (2008) found increases in revisions with age in a longitudinal study of 21- through 33-month-old children. The measure of language development, and age of participants (as well as the cross-sectional nature) in the current study are closer to those in Rispoli (2003). If the difference in findings is due to methodological differences, the most likely explanation seems to be the restriction of utterance types to active declarative sentences by Rispoli (2003), compared to the use of a much broader set of utterances analyzed here, as this is the major difference.

One possible reason why CWS showed increased revisions with language level while CWNS did not relates to children's awareness of their own language output. The two groups did not differ in the rate at which their utterances contained revisions. For CWS, 312 of 5,322 total utterances (5.9%) contained at least one revision, and for CWNS, 346 of 5,820 total utterances (also 5.9%) contained at least one revision. Therefore, it does not appear that being a CWS drives children to be so aware of their slips in language production that they revise more often in general. However, it is possible that being a CWS has a differential impact on a child's awareness of their production slips, depending on the child's language level. CWS can be aware of their stuttering even at preschool ages; at least some children show explicit awareness (Ambrose & Yairi, 1994), and others indicate awareness on some level by the development of secondary behaviors (Schwartz et al., 1990; Yairi, 1983). It is possible that CWS who have more advanced language abilities and have greater processing capacities are more capable of being influenced in their revision behavior by increased awareness due to being a CWS.

Finally, group was added to the models to determine whether utterances produced by children from either group were more likely to contain SLDs, stalls, or revisions, when controlling for utterance length, grammaticality, and language level. As expected, utterances produced by CWS were more likely to be SLD-containing, controlling for grammaticality, length, and language level. Utterances produced by CWS were also more likely than utterances produce by CWNS to contain stalls when controlling length, grammaticality, and language level. In the context of CWS potentially producing stalls to avoid SLDs, or as secondary behaviors, it is not especially surprising that CWS produced more stalls. There were no differences between groups on revision production, so there was no evidence that either group overall was more attentive to their production slips.

Gender. The impact of gender was assessed for two reasons. First, it has an impact on many developmental processes, including language development. Secondly, the sample used in this study reflected the fact that more boys than girls stutter (Bloodstein et al., 2021). It was of interest to know whether the general group findings were reflected in both gender groups, or whether results were more indicative of language-fluency relationships as they occur for boys.

The same factors that predicted SLD, stall, and revision occurrence for the entire sample were also significant in models only using utterances from boys. The factors that increased the likelihood of SLD, stall, and revision occurrence for the entire sample had positive coefficients for girls, and most were significant in the models using only girls' utterances. MLU-m-a as a predictor of stalls and grammaticality as a predictor of revisions did not reach significance, however. That these predictors did not reach significance is unsurprising, given that the girls' utterance models only had 2,462 utterances produced by 12 children, compared to 11,142 utterances produced by 64 children in the entire sample. A gender predictor added to the SLD, stall, and revision models was not significant in any of the models. Taken altogether, these results do not suggest any interesting differences related to gender. While future research looking at how SLDs, stalls, and revisions relate to linguistic factors should enroll greater numbers of girls, there is no indication based on the current data that these findings are not generalizable to girls. Current findings are also inconsistent with MacWhinney and Osser's (1977) report of girls producing more preplanning disfluencies (generally stalls) in comparison to boys producing more coplanning disfluencies (generally revisions). In addition, these results suggest that differences between the findings in this study and those reported by Rispoli (2003), Rispoli et al. (2008), and Wagovich et al. (2009), are not due to the lack of gender balance in the current study.

Theoretical Implications

Under the leading edge theory, children are thought to develop a comfort zone of utterances that can be produced without disfluency, while also attempting a set of more recently-acquired and advanced structures that are more likely to exceed a child's ability to produce language easily. This second set of utterances is called the leading edge and are what the theory predicts will be disfluent (Rispoli & Hadley, 2001). Children are thought to have the competence to produce these leading edge utterances, but not the performance capacities to produce them fluently. Rispoli and Hadley (2001) measured utterance length and verb phrase (VP) complexity in the study in which they proposed the theory. The current results do not weaken the suggestion that the longest utterances a child can produce, or those with more complex VPs, are more likely to exceed a child's capacities for uninterrupted speech. The current results do, however, suggest that grammaticality should be added to the list of considerations. Grammaticality is arguably more theoretically in line with the idea of a comfort zone and is a more individualized measure; measuring grammaticality requires examining the utterance and looking inside it to see whether children themselves are able to show via their sentence construction that they can produce it grammatically. The current study shows that grammatical utterances are more likely to be within the comfort zone of utterances that are produced without disruption, while ungrammatical utterances that a child attempts are more likely to be in the disfluent leading edge. Because young children's expressive language skills are growing, this leading edge is constantly moving.

It is also worth noting that the leading edge theory, as it was proposed by Rispoli and Hadley (2001), and as it was studied here, has been focused on syntax. However, it is possible that the leading edge is not only syntactic. Future researchers may wish to assess whether there are semantic contributions to the leading edge. In other words, do utterances containing more recently learned words have greater likelihoods of being disfluent? This could be tested in a study using dynamic assessment. If children are more disfluent in utterances containing pseudowords they are taught in a testing session, then this would support expansion of the leading edge hypothesis to include a semantic/lexical level and would also have clinical ramifications.

In the introduction, I referenced McKee et al.'s (2017) discussion of an irony seen in the study of children's expressive language. They point out that increases in the complexity of the structures that children produce are often assumed to be improvements in competence, even though there is existing evidence that performance factors impact the structure of children's expressive output. For instance, previous studies have shown that prosodic features of an utterance can impact whether utterance-initial subjects are produced (Gerken, 1991; McGregor & Leonard, 1994). The findings of the current study suggest that children have some level of knowledge that their ungrammatical utterances are ungrammatical. If children were completely unaware at all levels of language production that their utterances did not match the adultlike target, then it would be difficult to imagine why ungrammatical utterances would have been more likely to be disfluent. The current findings provide another example of why children's expressive language should not considered to reflect their highest level of competence.

The final set of theoretical implications pertain to the MDP (Smith & Weber, 2017). The MDP synthesizes decades of research examining the factors associated with and contributing to developmental stuttering. It describes stuttering as a neurodevelopmental disorder impacting the motor system that is strongly conditioned by linguistic and emotional factors (Smith & Weber, 2017). The model makes a key connection between linguistic and motoric factors by incorporating research findings showing decreased motor stability in the speech of PWS as length and complexity increase (Kleinow & Smith, 2000; MacPherson & Smith, 2013; Maner et al., 2000; Usler & Walsh, 2018). Based on current findings, grammaticality should be considered for incorporation into the MDP as a factor representing language demand alongside length and complexity; this would be especially true if findings about grammaticality and SLDs continue to be replicated. Current findings that length continues to be a predictor of disfluency, even when controlling for grammaticality and MLU-m-a, support the MDP's current incorporation of linguistic factors as impacting SLD production.

With the current finding of the impact of grammaticality on the likelihood that an utterance contains an SLD, the question arises of what the mechanism is for the connection between grammaticality and SLDs. One possibility is that ungrammatical utterances pose production difficulty for CWS, resulting in lower motor stability for ungrammatical utterances. Because of the magnitude of the association between grammaticality and SLD production (for CWS, a 1.89-times increase in the odds of SLD production for ungrammatical utterances), better understanding the mechanism underlying this relationship seems particularly important to understanding how language factors relate to motor stability. In fact, MacPherson and Smith (2013) asked CWS and CWNS ages 4;0 to 6;11 to complete a sentence repetition task so that their motor stability could be measured. They found that more CWS were unable to repeat the stimuli accurately than CWNS. Inaccurate repetitions were removed from analysis. If this discarded data could be analyzed in an exploratory way, it may begin to help us understand how grammaticality relates to motor stability for CWS.

Clinical Implications

In discussing clinical implications, the most relevant statistical model is the one looking at which factors are related to SLD production for CWS; this model indicated that longer utterances and ungrammatical utterances produced by CWS were more likely to contain SLDs, with *OR*s of 1.89 for grammaticality and 1.36 for length. Because of the different ways that length and grammaticality are measured (i.e., length being measured in morphemes, and grammaticality being a binary predictor), it is not possible to say that grammaticality is necessarily a stronger predictor of whether a CWS will stutter on an utterance than length is. It is, however, possible to say that changing an utterance produced by a CWS from being grammatical to ungrammatical has a greater impact on the likelihood that it will be stuttered than increasing the length of the utterance by one morpheme does. The increase in odds of stuttering with a two-morpheme increase in length is similar to the increase associated with an utterance being ungrammatical rather than grammatical; adding two morphemes increases the odds of an utterance from a CWS being stuttered by 1.85 times.

Clinical intervention is ideally based in research findings that have been replicated or assessed in meta-analyses (e.g., Bernstein Ratner, 2006) so it is important to consider the context of prior findings about grammaticality before discussing clinical recommendations. Current findings are consistent with those from two of three prior studies (Bernstein Ratner, 1997; Watson et al., 2011), one which used some of the same participants as the current study (Bernstein Ratner, 1997). The current study is also larger than the three previous studies (Bernstein Ratner, 1997; Watson et al., 2011; Yaruss, 1999), and clinicians should consider the current larger sample in how they weight previous findings. The two previous studies finding associations between grammaticality and stuttering and the current one have controlled for different factors and used different statistical approaches. Two have enrolled monolingual English-speaking CWS (Bernstein Ratner, 1997, and the current study) while one enrolled monolingual Spanish-speaking CWS (Watson et al., 2011). The diversity in methods and approaches to analysis suggest that there may be a robust relationship between grammaticality and stuttering, though additional research on interactions between language and fluency should continue to measure grammaticality. The clinical applications outlined below may be considered based on the current state of the evidence, and continued replication would increase the strength of these recommendations.

The finding that grammatical errors are related to disfluency production, and most importantly here, SLD production, underscores the need for thorough language assessments in evaluations for young children seen for concerns about stuttering. Comprehensive evaluations have long been recommended, given that we have known about associations between linguistic factors and SLDs for decades, and that the prevalence of language disorders appears to be higher among CWS than in the general population (Hall et al., 2007). With recent findings indicating that language growth is associated with recovery from stuttering (Hollister et al., 2017; Leech et al., 2017, 2019), this recommendation has become even more critical. If language skills are weak for a CWS, then language development must be supported. Further, if a speech-language pathologist is to know which utterances a child is likely to produce with grammatical errors, they have to conduct an in-depth language assessment. Because each child's leading edge, and which utterances are likely to be ungrammatical, is changing as their language develops, frequent language sampling or other monitoring of expressive language skills is required for clinicians to have up-to-date knowledge about the child's leading edge. Finally, dynamic assessment may be a valuable tool as it can provide information about whether structures unattested in language sampling data can be produced grammatically by a CWS.

Another clinical implication of the association between ungrammatical utterances and SLDs is that speech-language pathologists should consider incorporating grammaticality into treatment hierarchies for preschool-age CWS when modeling and teaching new skills. After their thorough language evaluations, clinicians would have information about which structures a child does and does not produce grammatically. Based on previous research findings, clinicians might currently be increasing length and complexity as CWS show increased performance of a new skill such as easy, relaxed speech (e.g., Richels & Conture, 2007). If the grammaticality finding were incorporated, then a hierarchy may involve first working on structures that CWS can produce grammatically before moving to utterances that they are expected to produce with grammatical errors. This can be done by modeling or through a structured elicitation task; with both of these methods, a clinician who has a thorough understanding of the child's expressive language skills can control whether a CWS is attempting an utterance that will likely be grammatical or ungrammatical for that particular child. If the clinician were modeling a structure expected to be produced with a grammatical error, the clinician would of course model the utterance grammatically but know that the child would likely produce an error.

Clinicians should also continue to remember that stuttering is a highly variable disorder. Within a PWS, there is daily and situational variability, particularly in overt stuttering behaviors such as frequency and duration (Constantino et al., 2016; Tichenor & Yaruss, 2021; Yaruss, 1997), and there is variability across speakers in how a particular situation may impact their stuttering (Yaruss, 2004). Because of this, it is important for clinicians to individualize the specific goals of intervention as well as the ways in which these goals will be targeted. Therefore, before incorporating grammaticality into any treatment hierarchy, clinicians should assess whether ungrammatical utterances are more likely to be stuttered by the child.

Further, clinicians should keep in mind that language growth is associated with recovery (Hollister et al., 2017; Leech et al., 2017, 2019), and children produce ungrammatical utterances as they gradually learn new structures (Eisenberg et al., 2012). If specific errored structures have strong associations with stuttering for a particular child, then clinicians should consider adding work on those errored structures to the fluency treatment plan.

The finding that grammatical errors increase the likelihood of stuttering indicates that special care should be taken with CWS who have concomitant language disorders (for whom the amount of clinical guidance is already limited). Severity of overt stuttering behaviors (i.e., SLD production and accompanying behavioral concomitants) will vary across CWS who also have language disorders, just as it varies across all children (e.g., Yaruss, 2004). It is therefore not possible to say that the presence of a concomitant language disorder will necessarily cause overt stuttering behaviors to be more severe for a particular child, even though children with language disorders will produce more grammatical errors (Eisenberg & Guo, 2016, 2018). However, because children with language disorders not only produce more grammatical errors but also persist in producing ungrammatical utterances through later ages (Eisenberg & Guo, 2016), clinicians should consider that CWS who also have language disorders may need additional attention to the development of skills to aid them when they stutter during utterances that are ungrammatical (including learning to tolerate the stuttering). In fact, as percent grammatical utterances (PGU), a language-sampling based measure, has been demonstrated to show acceptable to good diagnostic accuracy for children up to ages 8;11 (Eisenberg & Guo, 2016), associations between grammaticality and stuttering may need to be considered in particular detail for school-age CWS with concomitant language disorders. These children may be working towards more advanced fluency skills while still producing grammatical errors.

Methodological Implications

Another lesson of this study is the complexity of understanding the production of various disfluency types across both CWS and CWNS. An initial plan for this study was to compare utterances containing only SLDs to fluent utterances, utterances containing only stalls to fluent utterances, and utterances containing only revisions to fluent utterances. Under this plan, utterances containing both stalls and revisions or both TDs and SLDs would have been discarded. However, the initial plan would have excluded many more utterances from CWS than from CWNS, as 44.2% of the TD-containing utterances produced by CWS also had an SLD, in comparison to 15.2% of the TD-containing utterances produced by CWNS. Therefore, the approach used in this paper was to assess each utterance with regard to whether an SLD was present or not, whether a stall was present or not, and whether a revision was present or not. This approach was intended to make a more fair assessment of stall and revision-containing utterances for CWS. However, it does have one disadvantage that was described in detail above—that stalls can also serve as avoidance or secondary behaviors for CWS. CWS did not produce higher rates of revisions than CWNS, but it is entirely possible that some revisions produced by CWS were also produced in attempt to avoid words on which they anticipated stuttering, as word substitution is a well-documented behavior among PWS (e.g., Jackson et al., 2015).

Disfluencies occur in clusters (instances in which two or more disfluencies occur in sequence) more often than would be expected by chance, and this is true in the speech of both preschool-age CWS and CWNS (Hubbard & Yairi, 1988; Sawyer & Yairi, 2010). Several explanations have been offered for the occurrence of disfluency clusters, most focusing on ways in which an initial disfluency can increase the likelihood of a second disfluency. These include the possibilities that the first disfluency increases anxiety or tension (Hubbard & Yairi, 1988), or that the speech system takes time to restabilize after the first disfluency. Given what is known about secondary behaviors, as well as anticipation and avoidance being common features of stuttering (e.g., Jackson et al., 2015), the explanation that seems to best explain why stalls precede SLDs is offered by LaSalle and Conture (1995). After discussing explanations in which the first disfluency is assumed to increase the likelihood of the second one, LaSalle and Conture say, "preceding words may *be influenced by* following words. Cognitive, emotional, speech, and/or language adjustments that an individual makes for an upcoming word may affect the words preceding it…in the manner of anticipatory, regressive, or right-to-left assimilation that occurs with phonemes" (emphasis in original, pp. 975-976).

There are no definitive guidelines (or even strong recommendations) regarding how to classify utterances when the research aim is to measure language-fluency relationships for both TDs and SLDs in the speech of both CWS and CWNS. Most previous studies, such as those cited in the introduction, have not enrolled both CWS and CWNS and looked at both SLDs and TDs. Those that did have made varied decisions. For example, Zackheim and Conture (2003) enrolled both CWS and CWNS, and looked only at SLDs in the speech of CWS and only at TDs in the speech of CWNS, decisions made due to the low rate of TDs in the speech of CWS and low rate of SLDs in the speech of CWNS in their sample. This approach avoids the issues encountered in the current study but also restricts the types of questions that can be addressed. Buhr and Zebrowski (2009) discarded utterances containing both a TD and an SLD before conducting their sentence-level analyses because of the known phenomenon of fluency clusters and the possibility that SLDs and TDs were not independent. This approach solves the independence problem, but it adds another problem, which is that a greater proportion of utterances produced by CWS are removed in this case. Future researchers assessing both SLDs and TDs in the speech of both CWS and CWNS should consider disfluency clusters, as well as stuttering avoidance and secondary behaviors, and consider how these issues may impact their particular analyses. To answer certain questions, such as the questions about stalls or TDs generally in CWS versus CWNS, a multi-step approach may be advisable. Researchers may wish to run a primary analysis including utterances containing both stalls (or TDs generally) and SLDs, and then to conduct follow-up analyses omitting utterances containing both stalls (or TDs generally) and SLDs, or omitting utterances with fillers or stalls occurring in utterance positions immediately preceding or within SLDs.

Limitations and Future Directions

This study suggests that whether an utterance will later contain a grammatical error may influence disfluency occurring earlier in the utterance, and it suggests that stalls and revisions may both be caused by planning demand and increases in language skill *in play-based language contexts, for monolingual English-speaking children who do not present with any obvious language delay or disorder*. The generalizability of the findings may be limited to children with similar profiles. Further, while CWS from the Ratner corpus are known to be matched to CWNS (from the Ratner and Weismer corpora) on maternal education (as this matching was done prior to their inclusion in Luckman et al., 2020), the actual maternal education levels were unavailable. As in most studies of parent-child interaction the data are skewed to higher SES and education levels; the maternal education levels for CWS and CWNS from the Ratner-MacWhinney corpus are higher than typical parent education the United States and this limits generalizability as well.

Because this study used cross-sectional data, it is also restricted in the conclusions that it can draw about how disfluencies relate to general profiles of language development. Fortunately, the CWS and CWNS from the Ratner-MacWhinney corpus are being followed for two years (for three total visits), and if researchers wish to know how language development within the same group of children relates to stall and revision production, this data will soon be available. It will eventually be shared at Fluency.Talkbank.org, thus available to any interested researcher. There are also follow-up transcripts available at Fluency.Talkbank.org for eight of the CWS from the Ratner corpus. These children were seen at 3, 6, 9, or 12 months after their initial sessions, or at more than one of these intervals.

It would be of clinical interest to know how the results of a similar study would turn out if it enrolled CWS (or CWNS) with language disorders. It may be the case that ungrammatical utterances are particularly hard for these children to produce fluently, or it may simply be the case that they have more ungrammatical utterances.

It is also currently unknown what happens to the impact of grammaticality as children age and improve their language skills to the point where they no longer make as many grammatical errors. It is possible that older children who only produce a limited number of grammatical errors may have heightened awareness of the ungrammatical nature of their ungrammatical utterances, and be even more impacted by them. Or their more advanced language skills may decrease the extent to which grammatical errors impact their fluency. A related question is whether there is an age so young that children are not aware that their utterances are ungrammatical. It is uncommon for one-word utterances to be ungrammatical (but there are some, particularly if they are two morphemes, such as “sheeps”). The question of how grammaticality relates to disfluency could be asked of children who are just producing two-word combinations.

Chapter 4: Study 2 – Utterance Function and Disfluency

Data Reduction

The analyses for question 4 included all fully intelligible utterances that could be assessed for their status as to whether they were a question or other assertive utterance, or an answer to a question or other responsive utterance. If the same set of utterances used in the analyses for questions 1 and 2 were used here, this would have excluded many of the utterances used in responding to adults, particularly "I don't know" and one-word responses to questions. Therefore, the set of utterances included in the analyses for question 4 was intentionally larger than the set included in analyses for questions 1 and 2 so as to reflect the ways in which children respond to adults. Utterances were only excluded from the analyses for question 4 if it would have been difficult to determine their intent. Excluded utterances were those that were (a) fully or partially unintelligible, (b) incomplete (i.e., interrupted or trailed off), or (c) composed only of fillers or fragments. There were 7,689 eligible utterances produced by CWS. Of these 3 were coded as having "unclear" functions (per the guidelines described below) and so were excluded from the analyses, leaving 7,686 remaining. For CWNS, there were 8,093 eligible utterances initially, 6 of which had "unclear" functions, leaving 8,087 coded utterances.

Function Coding

Utterances were assigned to one of the five categories described below. These guidelines are based on the system used by Weiss and Zebrowski (1992) and Byrd et al. (2011) and originally described by Fey (1986). Questions and non-question assertive utterances were both considered types of assertive utterances. Answers to questions and non-answer responsive utterances were both considered to be types of responsive utterances.

Questions

All of a child's questions were given this code. This included questions where there was subject-auxiliary inversion (e.g., "can you get this on?"), as well questions with declarative sentence structure but in which intonation indicated that the utterance was a question (e.g., "this one's full?").

Other Assertive Utterances

This code was given to any non-question utterance in which the child (a) brought up a new topic, (b) solicited information, an action or attention, (c) labeled, (d) described facts or events, or (e) stated rules or explanations. For example, when not prompted by an adult's utterance, "sheep go on here," "your name is 'cowboy,'" "I can open it," and "I need something else" would all have been coded as other assertive utterances.

Answer to a Question

Any utterance in which a child was answering an adult's question was given this code. In line with Fey (1986) the child's utterance must have been an attempt to answer the question to receive this code, but the child's answer was not required to be correct. For instance, if the child was holding a toy car and an adult asked, "Is that a car?" and the child said "yes," "no," or "yeah but it's not fast" all of these would have been coded as answers. However, if the child said, "cars go fast," this would have been coded as an assertive utterance.

Other Responsive Utterances

This code was used for any utterance, other than an answer to a question, in which a child was responding to an adult. It was also used for utterances in which the child repeated the previous adult utterance or their own immediately preceding utterance. This included any utterances that were statements of agreement with the adult's preceding declarative utterance (e.g., if the adult said, "your Lego tower is big," and the child said, "yeah it is") and responses to requests for information (e.g., if the adult said, "tell me which truck you want," and the child said, "the green one"). If the adult's utterance had been, "which truck do you want?" and the child had responded in the same way, this would have been coded as an answer to a question.

Unclear Function

This code was used if it was unclear what preceded the child's utterance. For instance, it would have been used if the child was the first person to speak after recording began and this made it difficult to determine the function of the utterance. These utterances were excluded from the analysis.

Reliability

Question 4 asks whether questions and other assertive utterances are more disfluent than answers and other responsive utterances. To address this question, utterance function was coded, using similar reliability methods to those used to code other hand-coded variables (i.e., grammaticality and disfluency-error order). Research assistants completed training and passed two tests of function coding before coding this feature. On each test, research assistants were required to classify utterances in 30-utterance-long transcript sections as (a) an answer to a question, (b) a non-answer responsive utterance, (c) a question, (d) a non-question assertive utterance, or (e) having unclear function. They were required to achieve Cohen's kappa values of .80 or higher on two of these tests. Research assistants who had passed two tests completed the first function coding pass. I completed the second coding pass and identified utterances on which I disagreed with the first pass coding. There were 255 utterances (1.6% of the 15,782 total) with disagreements marked. A third coder reviewed these utterances to attempt to resolve disagreements. The third coder did not know which coder had chosen which code so that they would not defer to my coding choice. Consensus was reached on 100% of utterances.

Planned Analyses

As in analyses addressing questions 1, 2, and 3, statistical models reported in this chapter were run in R version 3.6.1 (R Core Team, 2019) with the lme4 package version 1.1-21 (Bates et al., 2015). Question 4 asks whether questions and other assertive utterances contain disfluencies more often than answers to questions and other responsive utterances. Because caregivers of CWS are sometimes advised to avoid asking questions so that the CWS will have to provide fewer answers (and thereby, potentially stutter less), the first set of models assessed the impact of the utterance function on whether CWS would produce SLDs. The outcome for all of these models was whether the utterance contained at least one SLD or not. Because the aim of this recommendation is to reduce SLD production and it does not focus on TDs in any way, TDs were disregarded for these analyses. That is, utterances containing stalls and/or revisions along with at least one SLD were counted as SLD-containing. Utterances containing stalls and/or revisions but without SLDs were counted (along with fluent utterances, of course) as not SLD-containing.

The central prediction was that, without controlling for other utterance-level factors such as length and grammaticality, assertive utterances would be more likely to contain SLDs than responsive utterances, but that this difference may no longer be significant after controlling for syntactic features. The first model run to test this hypothesis included a binary predictor representing whether an utterance was assertive (including questions and other assertive utterances) or responsive (including answers to questions and other responsive utterances), and a random intercept for participant. The second model was the same as the first but it controlled for grammaticality and length by adding a binary grammaticality predictor and cluster-mean centered utterance length (in morphemes) as fixed factors.

The next question was whether answers were more disfluent than all other utterance types combined. To address this question, answers were compared to all utterances that were not answers. Because answers made up 77.3% of responsive utterances produced by CWS, results from the third and fourth models were not expected to differ greatly from those of the first two models, which grouped answers and other responsive utterances together. However, in order to address the belief that caregiver questions should be avoided so as to decrease a child’s production of disfluent utterances, answers as a single group should be compared to (a) all other utterances and (b) other responsive utterances.

The last question concerned the two types of responsive utterances. It aimed to determine: If a child is responding to an adult’s utterance, then does it matter what they are responding to? Answers were compared to non-answer responsive utterances. The fifth model, therefore, included a binary answer vs. other responsive utterance predictor and a random effect of participant. The sixth model added a binary grammaticality predictor and cluster-mean-centered utterance length in morphemes. Because there were six models run, alpha was set to .008.

Pilot Analyses

Pilot analyses were run looking at the impact of function on whether utterances would the two disfluency types thought *a priori* to be more strongly related to planning demands (SLDs and stalls). This pilot analysis was run on data from 5 CWS and 5 matched CWNS, and did not separate participants by group. It found that with the 10 pilot participants, assertive utterances were more likely to contain SLDs or stalls than responsive utterances, but that this difference was no longer significant after controlling for grammaticality and length. It also found that answers were less likely to contain SLDs or stalls than utterances than the other three types when combined into one category, and that this significant association persisted after controlling for grammaticality and length. Last, the pilot analysis found that answers were more likely to contain SLDs or stalls than other responsive utterances, and that this relationship was still significant after controlling for SLDs and stalls. Given that these revealed significant associations between function and SLD/stall production with the 10 pilot participants, no power analyses were warranted.

The question better motivated by the background literature strictly regarded whether CWS, specifically, were more likely to stutter when asked questions.

Therefore, the primary analyses were modified to focus on CWS and SLDs, using models as outlined in the planned analyses section.

Results

Without controlling for other factors, assertive utterances produced by CWS had higher odds of being SLD-containing than responsive utterances ($z = 15.33$, $p < .001$). The odds of an utterance containing an SLD increased by a factor of 2.76 when an utterance was assertive rather than responsive, when other predictors were not controlled. When controlling for grammaticality and length, assertive utterances produced by CWS still had higher odds of being SLD-containing than responsive utterances ($z = 7.56$, $p < .001$). When holding grammaticality and length constant, assertive utterances produced by CWS had 1.74 times higher odds of being SLD-containing than responsive utterances.

Without controlling for grammaticality or length, utterances produced by CWS that were not answers had higher odds of containing at least one SLD than answers ($z = 9.97$, $p < .001$). Utterances that were not answers had 1.98 times the odds of being SLD-containing compared to answers. Put differently, answers had 0.51 times the odds of being stuttered compared to other utterances. Controlling for grammaticality and length, utterances produced by CWS that were not answers still had higher odds of containing at least one SLD compared to answers ($z = 4.72$, $p < .001$). When holding grammaticality and length constant, utterances that were not answers had 1.43 times the odds of being SLD-containing compared to answers.

Finally, the last two models looked within responsive utterances and compared answers to other responsive utterances. Results of the model not controlling for other factors indicated that answers had greater odds of being SLD-containing than other responsive utterances (z = -4.50, p < .001). Without controlling for other factors, for CWS, other responsive utterances had 0.52 times the odds of answers of being stuttered. Stated differently, without controlling for other factors, answers produced by CWS had 1.92 times greater odds of being stuttered than other responsive utterances. When holding grammaticality and length constant, there was a trend towards answers from CWS having greater odds of being SLD-containing than other responsive utterances (z = -0.42 p = .009). When controlling for grammaticality and length, other responsive utterances from CWS had 0.66 times the odds of being stuttered compared to answers. Put differently, when controlling for grammaticality and length, answers produced by CWS had 1.52 times greater odds of being stuttered compared to other responsive utterances. See Table 10 for descriptive statistics and Table 11 for model output from the six models.

Table 10

Characteristics of Assertive and Responsive Utterances Produced by CWS

	n	Length *M (SD)*	% Ungrammatical	% SLD-containing
Assertive utterances				
Questions	898	3.4 (2.0)	22.4	27.3
Other assertive	2,894	3.9 (2.4)	22.3	23.6
Assertive total	3,792	3.8 (2.3)	22.3	24.5
Responsive utterances				
Answers	3,008	2.6 (2.2)	10.0	13.4
Other responsive	886	2.1 (1.7)	6.4	7.0
Responsive total	3,894	2.5 (2.1)	9.2	12.0
Total	7,686	3.1 (2.3)	15.7	18.1

Note. Length is in morphemes.

Table 11

Results from Mixed Effects Models of SLD Production by Utterance Function for CWS

Model and Effect	*B*	*OR*	*z*	*p*	*B*	*OR*	*z*	*p*
Responsive vs. Assertive		Function alone			Function, grammaticality, length			
Fixed effects								
Intercept	-2.31	0.10	-15.59	<.001	-2.45	0.09	-14.59	<.001
Function	1.01	2.76	15.33	<.001	0.55	1.74	7.56	<.001
Grammaticality					0.74	2.10	8.99	<.001
Length					0.38	1.46	23.42	<.001
Random effect	*Var.*				*Var.*			
Participant	0.59				0.76			
Answers vs. All Other Utts.		Function alone			Function, grammaticality, length			
Fixed effects								
Intercept	-2.18	0.11	-14.77	<.001	-2.38	0.09	-14.05	<.001
Function	0.68	1.98	9.97	<.001	0.36	1.43	4.72	<.001
Grammaticality					0.77	2.16	9.33	<.001
Length					0.39	1.48	24.40	<.001
Random effect	*Var.*				*Var.*			
Participant	0.56				0.75			
Answers vs. Oth. Resp.		Function alone			Function, grammaticality, length			
Fixed effects								
Intercept	-2.09	0.12	-13.87	<.001	-2.64	0.07	-14.70	<.001
Function	-0.66	0.52	-4.50	<.001	-0.42	0.66	-2.61	.009
Grammaticality					0.90	2.45	5.91	<.001
Length					0.44	1.56	17.12	<.001
Random effect	*Var.*				*Var.*			
Participant	0.55				0.75			

Note. Var. = variance. Utts. = utterances. Oth. resp. = other responsive utterances. Function coded as 0 = all responsive, 1 = all assertive; 0 = answers, 1 = all other utts.; 0 = answers, 1 = other responsive utterances. Outcomes coded as 0 = not SLD-containing, 1 = SLD-containing. Number of participants = 32. For the responsive vs. assertive and answers vs. all other utterances models, number of utterances = 7,686. For the answers vs. other responsive utterances model, number of utterances = 3,984.

Question 4 Additional Post-Hoc Analyses

CWNS. Results indicated that for CWS, answers are not more likely to contain SLDs than utterances serving other conversational functions. One possibility is that the origin of the recommendation for caregivers to avoid asking questions was early clinicians' and researchers' observations of how disfluency related to question-answering in children with typical levels and types of disfluencies in their speech.

Therefore, I also assessed whether disfluencies in the speech of CWNS were increased in answers to questions. Six models were run with the same inputs as those run to assess SLD production in the speech of CWS. However, in this set of models, only utterances from CWNS were used, and the outcome variable was whether the utterance was fluent or disfluent. Simple fluency or disfluency was used because, as discussed in the interpretation of previous findings in this dissertation, it is unclear that there are qualitative distinctions between TDs and the SLDs produced by CWNS that would make a meaningful difference in the outcomes of these models. Alpha was again set to .008 because six models were run.

Without controlling for other factors, there was a significant increase in the odds of an utterance being disfluent if it was assertive rather than responsive ($z = 3.21$, $p = .001$). For CWNS, assertive utterances had 1.24 times greater odds of being disfluent compared to responsive utterances. Controlling for grammaticality and length, there was a significant increase in the odds of an utterance being disfluent if it was responsive rather than assertive ($z = -3.25$, $p < .001$). For CWNS, assertive utterances had 0.79 times the odds of responsive utterances of being disfluent, when controlling for grammaticality and length. Put differently, for CWNS, responsive utterances had 1.27 times greater odds of being disfluent compared to assertive utterances, controlling for grammaticality and length.

Without controlling for other factors, there was an nonsignificant trend towards an increase in the odds that an utterance from a CWNS would be disfluent if it was an answer compared to an utterance with another function. For CWNS, utterances had 0.87 times the odds of being disfluent if they were not answers compared to answers (a nonsignificant trend). Stated differently, for CWNS, utterances had 1.15 times greater odds of being disfluent if they were answers, and this association approached the corrected alpha level of .008 (z = -1.96, p = .050). When controlling for grammaticality and length, answers from CWNS had significantly greater odds of being disfluent, compared to all of their other utterances (z = -6.52, p < .001). For CWNS, utterances that were not answers had 0.62 times the odds of being disfluent compared to answers, when holding grammaticality and length constant. Put differently, for CWNS, answers had 1.61 times greater odds of being disfluent compared to utterances with other functions, when holding grammaticality and length constant.

Without controlling for other predictors, answers produced by CWNS had greater odds of being disfluent than other responsive utterances (z = -7.74, p < .001). For CWNS, other responsive utterances had 0.24 times the odds of answers of being disfluent. Stated differently, for CWNS, answers had 4.17 times greater odds of being disfluent compared to other responsive utterances. Finally, controlling for grammaticality and length, answers produced by CWNS still had greater odds of being disfluent than other responsive utterances (z = -6.67, p < .001). Controlling for grammaticality and length, other responsive utterances from CWNS had 0.29 times the odds of answers of being disfluent. Put differently, controlling for grammaticality and length, answers from CWNS had 3.45 times greater odds of being disfluent compared to other responsive utterances. See Table 12 for descriptive statistics and Table 13 for model output.

Table 12

Characteristics of Assertive and Responsive Utterances Produced by CWNS

	n	Length *M (SD)*	% Ungrammatical	% Disfluent
Assertive utterances				
Questions	1,041	3.7 (2.1)	18.9	12.6
Other assertive	3,180	4.4 (2.5)	20.3	14.9
Assertive total	4,221	4.2 (2.4)	20.0	14.4
Responsive utterances				
Answers	3,045	2.7 (2.4)	7.9	14.6
Other responsive	821	2.1 (1.8)	4.0	4.3
Responsive total	3,866	2.6 (2.3)	7.1	12.4
Total	8,087	3.4 (2.5)	13.8	13.4

Note. Length is in morphemes.

Table 13

Results from Mixed Effects Models of Disfluency Production by Utterance Function for CWNS

Model and Effect	*B*	*OR*	*z*	*p*	*B*	*OR*	*z*	*p*
Responsive vs. Assertive		Function alone			Function, grammaticality, length			
Fixed effects								
Intercept	-2.15	0.12	-20.69	<.001	-2.11	0.12	-19.46	<.001
Function	0.22	1.24	3.21	.001	-0.24	0.79	-3.25	.001
Grammaticality					0.43	1.54	4.74	<.001
Length					0.25	1.28	18.02	<.001
Random effect	*Var.*				*Var.*			
Participant	0.24				0.26			
Answers vs. All Other Utts,		Function alone			Function, grammaticality, length			
Fixed effects								
Intercept	-1.95	0.14	-18.52	<.001	-1.94	0.14	-17.60	<.001
Function	-0.14	0.87	-1.96	.050	-0.48	0.62	-6.52	<.001
Grammaticality					0.44	1.56	4.83	<.001
Length					0.26	1.29	18.61	<.001
Random effect	*Var.*				*Var.*			
Participant	0.24				0.26			
Answers vs. Oth. Resp.		Function alone			Function, grammaticality, length			
Fixed effects								
Intercept	-2.01	0.10	-14.64	<.001	-3.18	0.04	-21.22	<.001
Function	-1.41	0.24	-7.74	<.001	-1.25	0.29	-6.67	<.001
Grammaticality					0.40	1.50	2.40	.016
Length					0.27	1.31	12.96	<.001
Random effect	*Var.*				*Var.*			
Participant	0.44				0.37			

Note. Var. = variance. Utts. = utterances. Oth. resp. = other responsive utterances. Function coded as 0 = all responsive, 1 = all assertive; 0 = answers, 1 = all other utts.; 0 = answers, 1 = other responsive utterances resp. Outcomes coded as 0 = fluent, 1 = disfluent. Number of participants = 32. For the responsive vs. assertive and answers vs. all other utterances models, number of utterances = 8,087. For the answers vs. other responsive utterances model, number of utterances = 3,866.

Types of Other Responsive Utterances. In the coding system discussed so far, all other responsive utterances were grouped together. The clinical question that this study aims to address is whether the strategy often taught to caregivers of CWS of making comments rather than questions is in fact effective at reducing the likelihood that an utterance will be SLD-containing. To gain a satisfactory answer to this question, additional steps in controlled experiments would have to be taken.

There is one step towards better understanding this clinical issue that can be taken with the data available here: the nature of the utterances that children are responding to in their "other responsive" utterances could be examined. Are children responding to comments that caregivers may use in the place of questions? Or are children responding to utterances that are not specifically intended to prompt the child to respond? To address this question, other responsive utterances were classified further based on the communicative interaction that they were a part of, as outlined in Table 14. The child's other responsive utterance itself was coded, but the preceding adult utterance and conversational interaction as a whole were also considered.

Table 14

Other Responsive Utterance Subtypes

Oth. Resp. Subtype	Definition	Example
Prompted	Child is responding to an utterance in which the adult was clearly attempting to get the child respond.	Adult: I wonder what this is. Child: That'll help us get on there.
Repetition	Child's utterance is a partial or complete repetition of preceding child or adult utterance.	Adult: I think that's actually supposed to be pickles. Child: Pickles.
Spontaneous	Child appears to have decided on their own to say something, including spontaneous agreements, disagreements, and protests.	Adult: I think we should play with the castle. Child: No.

Note. Oth. Resp. = Other responsive.

When the 886 other responsive utterances produced by CWS were coded according to the definitions in Table 14, 20 (2.3%) were prompted, 241 (27.2%) were repetitions and 625 (70.5%) were spontaneous. Prompted other responsive utterances are the subtype of greatest interest because they are the ones that would occur in response to an adult substituting a comment for a question. Given the small number of prompted other responsive utterances, no additional statistical analyses were performed.

Discussion

Utterance Function and SLD Production by CWS

The primary aim of this study was to assess the empirical basis of the recommendation that caregivers should consider avoiding asking questions of CWS, so as to help CWS avoid stuttering. Three sets of models were run. Within each set, utterances were first compared by function without controlling for other syntactic predictors. These models run without controlling for other predictors are arguably the more clinically relevant models; when caregivers are avoiding asking questions in order to reduce a child's SLD production, they are not thinking about controlling for grammaticality and length of a child's response. Then the models run while controlling for grammaticality and length can provide more useful information about the role of function and communicative responsibility in disfluency production.

I hypothesized that assertive utterances (which include questions), would be more likely than responsive utterances (which include answers) to contain SLDs when grammaticality and length were not held constant, and expected that this difference might no longer be significant when controlling for grammaticality and length (as was no longer significant after controlling for length and complexity in a study by Byrd et al., 2011). The first part of the hypothesis was supported; assertive utterances from CWS had 2.76 times greater odds of being stuttered compared to responsive utterances. This difference was still significant when controlling for length and grammaticality (though the *OR* was reduced to 1.74). That this difference remained significant in the current study when controlling for other linguistic factors at first appears to diverge from what Byrd et al. (2011) reported. However, the current study is larger and better-powered than Byrd et al.'s, and Byrd et al. reported a near-significant *p*-value of .097 for their analyses. Therefore, it probably should not be considered a surprising, or even divergent result, that the current study's findings found assertiveness to have an impact when controlling for grammaticality and length. On the basis of their results, Byrd et al. concluded that the length and complexity of assertive utterances, rather than communicative responsibility, accounted for the increase in stuttering in these utterances. Though the results of the current study are not surprising in light of the power difference, the conclusions about the role of communicative responsibility are different. Current results suggest that assertiveness on its own contributes to increased likelihood of stuttering for CWS (and grammaticality and length play a role as well).

On the basis of these results, however, what it is about assertive utterances that makes them more likely to be stuttered cannot be identified. Weiss and Zebrowski (1992) suggested that if it was not the length and complexity of assertive utterances that made them more likely to be stuttered, then it may have been their "communicative responsibility," defined in the sense of Stocker and Usprich (1976). In high communicative responsibility utterances, the burden of communicative success is held more by the child; when the child is introducing a novel topic or in some other way taking control of the conversation, the demands on the child to communicate clearly so that the listener can follow along increase. In assertive utterances, children may have been introducing a new topic; the difficulty of retrieving words that are newly introduced into the conversation may have been part of what makes assertive utterances more likely to be stuttered. Finally, assertive utterances may have imposed greater emotion regulation demands on CWS; moments with high levels of positive emotionality (e.g., when receiving a desirable gift) have been linked to higher rates of SLD (and TD) production for CWS (Johnson et al., 2010). The ways in which a child's emotional reactivity and emotional state impact their stuttering are complex (Jones et al., 2014), and since emotional reactivity and stress levels were not measured or manipulated in the current study, additional research would be needed to assess how assertive and responsive utterances relate to emotion regulation demands.

In order to better assess whether answers are associated with higher levels of stuttering, answers were compared to all utterances that were not answers (assertive utterances plus other responsive utterances). For CWS, utterances that were not answers were 1.98 times (and significantly) more likely than answers to be stuttered. When controlling for length and grammaticality, these other utterances had 1.43 times (and still significantly) greater odds of being stuttered compared to answers. That answers have only about half the odds of being stuttered of other utterances, without controlling for other factors, is the most critical result here in terms of indicating why caregiver questions do not need to be avoided by adults interacting with CWS. That the significant relationship persists after controlling for other factors is likely attributable to one or some of the explanations discussed regarding assertive utterances, as 81.1% of utterances that were not answers were assertive utterances.

Finally, answers were compared to other responsive utterances. When caregivers are instructed to avoid questioning, the proportion of the child's utterances that are responses to non-question comments are expected to increase, so comparing answers to other responsive utterances addressed the question of the child utterances that may replace answers are likely to be more fluent than answers. Answers produced by CWS had 1.92 times (and significantly) higher odds than their other responsive utterances to contain SLDs. Controlling for grammaticality and length, answers produced by CWS had 1.52 times higher odds of being stuttered compared to other responsive utterances, and this difference approached significance. These results might appear to suggest that if caregivers are trying to prompt children to talk, prompting children with statements, that will result in non-answer responsive utterances, may be helpful. However, in order to determine whether these non-question prompts elicited a large portion of the other responsive utterances, additional coding was required. This coding indicated that only 2.3% of these were prompted by adult statements of the type that might be used instead of questions. The majority (70.5%) of other responsive utterances were spontaneous responses to adults, such as agreements, disagreements, and protests, and the remaining 27.2% were full or partial repetitions of preceding utterances. Given this breakdown, the increased likelihood of stuttering in other responsive utterances does not indicate that caregivers should avoid questions and instead make comments. It also raises the question of how successful leading a child to speak using primarily statements actually is; measurement of the rate of children's responses to prompts was not part of this study but may be worth assessing in future studies. Research measuring how often children respond to adults' prompts may be useful; if children rarely respond to them (which may be the case based on the low percentage of other responsive utterances that are children's responses to prompts), then this would be even further reason for clinicians to move away from asking caregivers to comment rather than question.

Clinical recommendations can be made more strongly when they are based on higher levels of evidence, such as randomized controlled trials (RCTs) and meta-analyses (Bernstein Ratner, 2006), so current findings will be contextualized. In 1992, Weiss and Zebrowski found that assertive responses were more likely to contain SLDs than responsive utterances. Assertive utterances were also longer than responsive utterances, and since length was not controlled in their analyses, it was unclear whether this difference was attributable to utterance length or assertiveness. More recently, Byrd et al.'s (2011) findings were similar to the ones from this study, though less well-powered. Therefore, current findings that answers are stuttered less often than other utterances are consistent with past reports.

When the recommendation is made for caregivers to reduce their questioning of young children, an implicit assumption is made that the benefit to CWS from the supposed decrease in SLD production outweighs any benefit they would have gained from being asked questions. However, being asked questions does in fact have benefits in language development. One obvious benefit is that it provides children with models of questions, which are some of the more advanced structures that they will learn (Birbili & Karagiorgou, 2009; Santelmann et al., 2002). In addition, mother's usage of yes/no questions has been linked to the development of their children's auxiliaries (Newport et al., 1977). Questions also help with moving conversations forward; the use of wh- questions maintaining the child's previous topic has been shown to increase the length of conversations between adults and children with varied language abilities (Yoder et al., 1994). Children with language and cognitive impairments have been found to use longer utterances in response to caregivers' questions than in response to other parent utterances (Yoder & Davies, 1990). Finally, questions can be an important part of shared book reading, a practice known to support language and literacy development (Birbili & Karagiorgou, 2009; Crain-Thoreson & Dale, 1999).

Utterance Function and Disfluency Production by CWNS

When answers were found to have lower rates of stuttering than all other utterance types combined, I wondered what the source of the clinical recommendation to help CWS avoid having to answer questions may have been. This recommendation for a reduction in caregiver use of questioning has been made by experts since at least the 1970's. Van Riper (1973) wrote:

> We have observed thousands of verbal interactions between mothers and children and the amount of questioning which takes place just before volleys of stuttering occur is surprisingly great. We have no direct research on this but an examination of daily logs kept by parents in which they reported the characteristics of the communicative situation which yielded the most fluent and the least fluent responses shows clearly that parental questioning preceded

stuttering in an impressive number of instances whereas only rarely did it precede a markedly fluent response. (p. 379)

The overall percentage of answers from CWS that contained SLDs (13.4%) was lower than the percentage of utterances from CWS overall that contained SLDs (18.1%); it was not the case that without controlling for length and grammaticality, answers would have appeared to have been highly disfluent to a casual observer of these samples. One possibility is that this concern that researchers and clinicians (outside of Van Riper, who made this observation with CWS) have had about question-answering was derived from elevated disfluency rates for CWNS when answering questions. The question about whether CWNS produce more disfluencies when they are answering questions was testable.

For CWNS, responsive utterances had 1.24 times (and significantly) higher odds compared to assertive utterances of being disfluent, after controlling for length and complexity. Answers to questions had 1.61 times (and significantly) higher odds of being disfluent compared to other utterances, controlling for grammaticality and length. Finally, answers, when compared to other responsive utterances, were 3.45 times more likely to be disfluent, controlling for grammaticality and length.

These findings suggest that higher communicative responsibility did not increase the likelihood that CWNS would be disfluent in the same way that it increased the likelihood that CWS would stutter. One potential explanation for the decrease in stuttered disfluency for CWS when answering questions is that parents and examiners could have been aware of the impact that demand and time pressure may have on stuttering. If parents and examiners interacting with CWNS were not modifying their speech in other ways to make their questions less demanding, then this might explain the relative increase in disfluency with question-answering in CWNS.

An alternative explanation concerns potential temperamental differences between CWS and CWNS. Conflicting findings have driven continue debate about whether CWS and CWNS have differing temperamental profiles (see Bloodstein, Bernstein Ratner, & Brundage, 2021, for a review) but if these differences do in fact turn out to be present, they may partially explain the differing fluency patterns in the ways that the two groups answer questions. This would provide a child-focused explanation. For instance, CWS have been reported to have higher behavioral inhibition than CWNS (Choi et al., 2013; Ntourou et al., 2020), and higher behavioral inhibition has been related to the production of fewer words in a language sampling context (Tumanova et al., 2020). CWS with higher behavioral inhibition may be more resistant to answering questions, and therefore may choose to answer different types of questions than CWNS. This, in turn, may have impacted the profiles of fluency we observed here, and warrant future investigations that probe question-asking and -answering patterns between CWS and their parents on structural, functional, and temperament-focused levels of analysis.

Limitations and Future Directions

As with results from the previous study, the generalizability of the findings about answers and other utterance functions and their relationships with disfluency may be limited to other children who share characteristics with these participants. This means that generalization may be most appropriate when considering *monolingual, English-speaking children without obvious language delays or disorders*. The maternal education level for this sample, when available, was also not representative of the American population, and this may further limit generalizability.

The current study was restricted to data available from language sample analysis. While there are benefits to using this type of data, particularly ecological validity, there are limits to the types of information that this data can provide. I will first discuss potential research directions that could be taken with language sampling data (and even with the current corpus) and then I will discuss a research direction that would require a different paradigm.

First, future researchers aiming to provide clinical guidance may wish to identify different question types and measure the disfluency of responses to these (controlling for length, as length will likely vary greatly depending on question type). For instance, yes/no questions, and questions presenting choices (e.g., “do you want the car or the truck?”) may elicit more fluent answers than open-ended questions (e.g., “how does baby get ready for bed?”). Questions differ in structural complexity required of the asker, as well as structural complexity typically demanded by the person who answers. Even yes-no questions and certainly Wh-questions differ in the lexical and conceptual demands placed on speaker and answerer. Beyond structural distinctions, questions can vary in the cognitive demands placed on the respondent, as well as the functional load within the interaction (e.g., “Why don’t we play with the blocks”, vs. “Why do you think the train keeps falling off the track?”). To summarize succinctly, one can ask many questions about questions on numerous levels, all of which require coding of specific aspects of the utterance and interchange.

In addition, future researchers also may wish to disentangle the impact of the pace of the conversation and questioning. This may particularly be helpful in understanding why answers to questions were *less* likely to be stuttered for CWS while they were more likely to be disfluent for CWNS. If parents and examiners were sensitive to the pace of the conversation when speaking to CWS, but parents and examiners rapidly asked questions of CWNS, this may partially explain the divergent findings across groups. This could be studied by measuring the amount of time that parents and examiners paused after their questions and also by measuring their speech rate. The likelihood of parents of CWS having encountered information suggesting a slowed pace of conversation may have differed across corpora because of the time periods when they were collected; the Ratner corpus was collected in the late 1990s and these parents were probably less likely to have encountered information on the internet about how to adjust their communication styles. Therefore, this explanation may not be particularly likely to explain this discrepancy for Ratner corpus CWS and CWNS, but it may partially explain the discrepancy in the fluency of answers for Ratner-MacWhinney corpus CWS and CWS.

To gain a better understanding of the impact that asking fewer questions would have on the functions of utterances that CWS produce, researchers may wish to assess the rate at which CWS answer questions they are asked. A cursory look at this can be gained by looking at the overall number of adult questions (i.e., the number of adult utterances ending in a question mark) and the number of child answers. There were 6,443 adult questions in samples with CWS, and only 3,008 fully intelligible utterances coded as answers. (The numbers are similar for CWNS, so higher selectivity about which questions to answer does not explain the relative fluency of CWS in their answers.) A careful analysis would be needed before concluding that CWS respond to fewer than half of adult questions, as some of the answers may have been fully or partially unintelligible, or interrupted or trailed off and therefore not coded for their function. A finding that children only answer about half of the questions they are asked, in a well-controlled study, would further suggest that it is not particularly helpful for caregivers to reduce their use of questions. If children are happy to not respond to questions when they do not want to, caregivers would not need to reduce their questioning to reduce demands.

Further analysis of language samples could also provide more clarity about the role of introducing new information into the discourse in the disfluency of assertive utterances. Researchers could code utterances for simply whether new information was introduced in the utterance. This type of analysis would help us better understand the specific role of introducing new information into the discourse, in comparison to other features of assertive utterances (e.g., potentially higher emotion regulation demands).

In order to confirm the association between answers and lower levels of disfluency for CWNS, elicited probes that control the content of the child's answer could be used. For instance, researchers may wish to pose comments and questions intended to elicit answers/non-answer responses with similar content and syntax. For instance, the parallel "What does baby want to do?" with a clear expectation of an answer, and "Hm, I wonder what baby wants to do," where it is clear from the context that an answer is optional, could be posed to CWS to determine whether the expectation of answering the question has an impact on their fluency.

Chapter 5: General Discussion

The major findings of this dissertation can be summarized as follows. This study is arguably the largest to demonstrate that ungrammatical utterances are more likely to be disfluent (stuttered and TD-containing) than grammatical utterances. To my knowledge, it is also the first study to assess grammaticality as it relates to both SLDs and TDs, and in the speech of both CWS and CWNS. Second, this dissertation found that SLDs and stalls tend to precede grammatical errors when they occur in the same utterance, which is consistent with SLDs and stalls occurring during the planning of potentially more challenging, error-containing portions of utterances. Third, this dissertation has failed to replicate a dichotomy between predictors of two different TD types- stalls and revisions. Both stalls and revisions increased in ungrammatical utterances, longer utterances, and as language level increased. Fourth, this dissertation found that, in the speech of young CWS, answers to questions are less likely to contain SLDs than other utterance types, but that in the speech of CWNS, answers are more likely to be disfluent than other utterance types.

All of these findings can be interpreted in the context of stuttering being a highly variable disorder (Tichenor & Yaruss, 2021; Yaruss, 1997). For any particular CWS at any particular moment, it is not possible to definitively determine whether a word or utterance will be stuttered. There will be long, ungrammatical, assertive utterances produced by CWS that will be fluent. For example, consider participant CWS-32, whose MLU-m-a was 4.90, and who stuttered in 14.3% of his utterances that were eligible for the syntactic analyses in questions 1 and 2 (and was disfluent in some way in 19.2% of them). He fluently produced the nine-morpheme, ungrammatical question, “why are these wheels keep getting off?” According to the output of the statistical models, this utterance would have had a high likelihood of being stuttered due to its grammatical error, length, and assertiveness. However, stuttering is variable, and other factors not measured in this study may have contributed to the non-production of any SLDs in that utterance. Some of these other factors, such as those relating to emotional regulation demands, are known to stuttering researchers (Jones et al., 2014). There also are likely other factors contributing to the occurrence or non-occurrence of an SLD (or TD) in a particular moment that researchers have not yet thought to study.

The association between grammatical errors and disfluency, as well as the absence of an identified relationship between answering questions and stuttering for CWS are group-level findings was based on data from 32 CWS and 32 CWS. While these are large samples in comparison to many previous studies of disfluency or stuttering and language, clinicians should remember analyses included data from only 64 children in total. The findings of this dissertation, therefore, should be considered starting points, particularly given that stuttering is a disorder characterized by great heterogeneity, both across speakers and even within speakers when sampled at multiple time points, even close in time (Constantino et al., 2016; Tichenor & Yaruss, 2021; Yaruss, 1997). For instance, clinicians may wish to assess whether an individual CWS presents with increased stuttering in ungrammatical utterances. If this pattern is not seen for a particular CWS, then it may not be worthwhile to consider grammaticality in developing treatment hierarchies for that child. In contrast, particularly with young children with emergent language profiles, a profile of occasional ungrammaticality in expressive language samples might logically prompt the SLP to inventory those structures that the child appears to have strong productive control over, and those with less consistent evidence of mastery. In planning fluency-related activities, such information would be taken into account in planning expected verbal responses to clinician prompts.

Two of the major findings of this dissertation can serve as examples of how facilitation of language development should be considered in stuttering intervention for young children. First, the finding that ungrammatical utterances are more likely to be stuttered can be applied in specific contexts by clinicians. It can be used when introducing new skills such as easy, relaxed speech. The other finding with direct clinical applicability is that CWS are less likely to stutter when answering questions than in other utterances. This finding is more applicable to the child's general environment, outside of therapy. It indicates that there is little reason for caregivers to generally avoid asking questions of CWS to facilitate their fluency, particularly given the potential benefits of being asked questions (e.g., Birbili & Karagiorou, 2009; Newport et al., 1977). When implementing any fluency intervention with a young CWS, clinicians should consider how the intervention might impact the child's language development. The close links between language and fluency indicate that in order to be a good fluency therapist for young CWS, a clinician must also be a good language therapist.

Appendices

Appendix A

Results from Mixed Effects Models of Predictors of Disfluency Comparing Utterances With Each Disfluency Type Only to Fluent Utterances

Model and Effect	*B*	*OR*	*z*	*p*
SLD Only vs. Fluent				
Fixed effects				
Intercept	-2.51	0.08	-16.84	<.001
Grammaticality	0.49	1.63	6.16	<.001
Length	0.27	1.31	17.94	<.001
MLU-m-a	0.08	1.08	0.46	.648
Random effect	*Var.*			
Participant	1.23			
Stall Only vs. Fluent				
Fixed effects				
Intercept	-3.11	0.04	-27.54	<.001
Grammaticality	0.35	1.42	3.15	<.001
Length	0.14	1.15	7.29	<.001
MLU-m-a	0.51	1.67	3.68	<.001
Random effect	*Var.*			
Participant	0.53			
Revision Only vs. Fluent				
Fixed effects				
Intercept	-3.44	0.03	-39.16	<.001
Grammaticality	0.44	1.55	3.47	<.001
Length	0.22	1.25	9.78	<.001
MLU-m-a	0.24	1.27	2.55	.011
Random effect	*Var.*			
Participant	0.13			

Note. Outcomes coded as 0 = fluent, 1 = SLD(s) only; 0 = fluent, 1 = stall(s) only; and 0 = fluent, 1 = revision(s) only. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Var. = variance. Number of participants = 64. For the SLD only vs. fluent model, number of utterances = 9,702. For the stall only vs. fluent model, number of utterances = 8,961. For the revision only vs. fluent model, number of utterances = 8,772.

Appendix B

Results from Mixed Effects Models of Predictors of Disfluency for Three Disfluency Types for CWS Only

Model and Effect	*B*	*OR*	*z*	*p*
SLD				
Fixed effects				
Intercept	-1.58	0.21	-9.91	<.001
Grammaticality	0.63	1.89	7.70	<.001
Length	0.31	1.36	17.83	<.001
MLU-m-a	0.22	1.24	1.04	.297
Random effect	*Var.*			
Participant	0.71			
Stall				
Fixed effects				
Intercept	-2.75	0.06	-17.48	<.001
Grammaticality	0.52	1.68	4.82	<.001
Length	0.13	1.14	6.59	<.001
MLU-m-a	0.61	1.85	3.00	.003
Random effect	*Var.*			
Participant	0.59			
Revision				
Fixed effects				
Intercept	-3.21	0.04	-28.50	<.001
Grammaticality	0.65	1.91	5.03	<.001
Length	0.21	1.24	9.19	<.001
MLU-m-a	0.47	1.59	3.35	<.001
Random effect	*Var.*			
Participant	0.15			

Note. Outcomes coded as 0 = no SLD in utterance, 1 = SLD in utterance, 0 = no stall in utterance, 1 = stall in utterance, and 0 = no revision in utterance, 1 = revision in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Var. = variance. Number of participants = 64. Number of utterances = 5,322.

Appendix C

Results from Mixed Effects Models of Predictors of Disfluency for Three Disfluency Types for CWNS Only

Model and Effect	*B*	*OR*	*z*	*p*
SLD				
Fixed effects				
Intercept	-2.95	0.05	-24.89	<.001
Grammaticality	0.36	1.43	2.84	.004
Length	0.18	1.20	9.01	<.001
MLU-m-a	0.38	1.47	2.78	.005
Random effect	*Var.*			
Participant	0.26			
Stall				
Fixed effects				
Intercept	-3.01	0.05	-22.83	<.001
Grammaticality	0.25	1.29	1.80	.072
Length	0.12	1.13	5.62	<.001
MLU-m-a	0.43	1.54	2.74	.006
Random effect	*Var.*			
Participant	0.35			
Revision				
Fixed effects				
Intercept	-3.01	0.05	-33.21	<.001
Grammaticality	0.42	1.53	3.23	.001
Length	0.20	1.22	9.13	<.001
MLU-m-a	0.10	1.11	1.03	.305
Random effect	*Var.*			
Participant	0.09			

Note. Outcomes coded as 0 = no SLD in utterance, 1 = SLD in utterance, 0 = no stall in utterance, 1 = stall in utterance, and 0 = no revision in utterance, 1 = revision in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Var. = variance. Number of participants = 64. Number of utterances = 5,820.

Appendix D

Results from Mixed Effects Models of Predictors of Disfluency for Three Disfluency Types With Group as a Predictor

Model and Effect	*B*	*OR*	*z*	*p*
<u>SLD</u>				
Fixed effects				
Intercept	-3.18	0.04	-21.75	<.001
Grammaticality	0.56	1.76	8.29	<.001
Length	0.26	1.29	19.86	<.001
MLU-m-a	0.31	1.36	2.43	.015
Group	1.72	5.61	8.59	<.001
Random effect	*Var.*			
Participant	0.50			
<u>Stall</u>				
Fixed effects				
Intercept	-3.16	0.04	-21.77	<.001
Grammaticality	0.43	1.53	5.03	<.001
Length	0.13	1.14	8.67	<.001
MLU-m-a	0.52	1.69	4.03	<.001
Group	0.55	1.73	2.71	.007
Random effect	*Var.*			
Participant	0.48			
<u>Revision</u>				
Fixed effects				
Intercept	-3.11	0.04	-32.81	<.001
Grammaticality	0.55	1.73	5.97	<.001
Length	0.21	1.23	12.98	<.001
MLU-m-a	0.26	1.29	3.06	.002
Group	0.01	1.01	0.06	.951
Random effect	*Var.*			
Participant	0.13			

Note. Outcomes coded as 0 = no SLD in utterance, 1 = SLD in utterance, 0 = no stall in utterance, 1 = stall in utterance, and 0 = no revision in utterance, 1 = revision in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Group coded as 0 = CWNS, 1 = CWS. Var. = variance. Number of participants = 64. Number of utterances = 11,142.

Appendix E

Results from Mixed Effects Models of Predictors of Disfluency for Three Disfluency Types for Boys Only

Model and Effect	*B*	*OR*	*z*	*p*
SLD				
Fixed effects				
Intercept	-2.39	0.09	-15.00	<.001
Grammaticality	0.64	1.89	8.16	<.001
Length	0.25	1.29	17.37	<.001
MLU-m-a	0.11	1.11	0.52	.601
Random effect	*Var.*			
Participant	1.15			
Stall				
Fixed effects				
Intercept	-2.95	0.05	-26.91	<.001
Grammaticality	0.42	1.52	4.05	<.001
Length	0.13	1.13	7.37	<.001
MLU-m-a	0.50	1.65	5.52	<.001
Random effect	*Var.*			
Participant	0.41			
Revision				
Fixed effects				
Intercept	-3.14	0.04	-37.05	<.001
Grammaticality	0.61	1.83	5.83	<.001
Length	0.21	1.23	11.54	<.001
MLU-m-a	0.23	1.26	2.24	.025
Random effect	*Var.*			
Participant	0.15			

Note. Outcomes coded as 0 = no SLD in utterance, 1 = SLD in utterance, 0 = no stall in utterance, 1 = stall in utterance, and 0 = no revision in utterance, 1 = revision in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Var. = variance. Number of participants = 52. Number of utterances = 8,680.

Appendix F

Results from Mixed Effects Models of Predictors of Disfluency for Three Disfluency Types for Girls Only

Model and Effect	*B*	*OR*	*z*	*p*
SLD				
Fixed effects				
Intercept	-2.18	0.11	-6.06	<.001
Grammaticality	0.35	1.42	2.51	.012
Length	0.27	1.30	9.73	<.001
MLU-m-a	0.01	1.01	0.02	.988
Random effect	*Var.*			
Participant	1.44			
Stall				
Fixed effects				
Intercept	-2.63	0.07	-8.65	<.001
Grammaticality	0.42	1.52	2.82	.005
Length	0.13	1.14	4.60	<.001
MLU-m-a	0.33	1.39	1.04	0.30
Random effect	*Var.*			
Participant	0.96			
Revision				
Fixed effects				
Intercept	-2.97	0.05	-22.57	<.001
Grammaticality	0.35	1.42	1.82	.069
Length	0.20	1.22	6.01	<.001
MLU-m-a	0.31	1.36	2.47	.013
Random effect	*Var.*			
Participant	0.04			

Note. Outcomes coded as 0 = no SLD in utterance, 1 = SLD in utterance, 0 = no stall in utterance, 1 = stall in utterance, and 0 = no revision in utterance, 1 = revision in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Var. = variance. Number of participants = 12. Number of utterances = 2,462.

Appendix G

Results from Mixed Effects Models of Predictors of Disfluency for Three Disfluency Types With Gender as a Predictor

Model and Effect	*B*	*OR*	*z*	*p*
SLD				
Fixed effects				
Intercept	-2.38	0.09	-14.67	<.001
Grammaticality	0.57	1.76	8.33	<.001
Length	0.26	1.29	19.88	<.001
MLU-m-a	0.08	1.09	0.46	.645
Gender	0.15	1.17	0.43	.671
Random effect	*Var.*			
Participant	1.20			
Stall				
Fixed effects				
Intercept	-2.97	0.05	-25.04	<.001
Grammaticality	0.43	1.53	5.04	<.001
Length	0.13	1.14	8.67	<.001
MLU-m-a	0.46	1.58	3.47	<.001
Gender	0.39	1.48	1.52	.128
Random effect	*Var.*			
Participant	0.53			
Revision				
Fixed effects				
Intercept	-3.11	0.04	-39.13	<.001
Grammaticality	0.55	1.73	5.96	<.001
Length	0.21	1.23	12.98	<.001
MLU-m-a	0.26	1.29	3.09	.002
Gender	0.02	1.02	0.13	.898
Random effect	*Var.*			
Participant	0.13			

Note. Outcomes coded as 0 = no SLD in utterance, 1 = SLD in utterance, 0 = no stall in utterance, 1 = stall in utterance, and 0 = no revision in utterance, 1 = revision in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Gender coded as 0 = male, 1 = female. Var. = variance. Number of participants = 64. Number of utterances = 11,142.

Appendix H

Results from Mixed Effects Models of Predictors of Disfluency for Stall-Containing Utterances, Without SLD-Stall Utterances From CWS

Model and Effect	*B*	*OR*	*z*	*p*
Stall, without SLDs				
Fixed effects				
Intercept	-3.08	0.05	-29.40	<.001
Grammaticality	0.26	1.30	2.63	.009
Length	0.10	1.11	6.05	<.001
MLU-m-a	0.52	1.69	4.07	<.001
Random effect	*Var.*			
Participant	0.46			

Note. Outcomes coded as 0 = no stall in utterance, 1 = stall in utterance. Grammaticality coded as 0 = grammatical, 1 = ungrammatical. Var. = variance. Number of participants = 64. Number of utterances = 10,919

Bibliography

Ambrose, N. G., & Yairi, E. (1994). The development of awareness of stuttering in preschool children. *Journal of Fluency Disorders*, *19*, 229–245. https://doi.org/10.1016/0094-730X(94)90002-7

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research*, *42*(4), 895–909. https://doi.org/10.1044/jslhr.4204.895

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

Bernstein, N. (1981). Are there constraints on childhood disfluency? *Journal of Fluency Disorders*, *6*(4), 341–350. https://doi.org/10.1016/0094-730X(81)90021-8

Bernstein Ratner, N. (2000). Performance or capacity, the model still requires definitions and boundaries it doesn't have. *Journal of Fluency Disorders*, *25*(4), 337–346. https://doi.org/10.1016/S0094-730X(00)00083-8

Bernstein Ratner, N. (2004). Caregiver-child interactions and their impact on children's fluency: Implications for treatment. *Language, Speech, and Hearing Services in Schools*, *35*(1), 46–56. https://doi.org/10.1044/0161-1461(2004/006)

Bernstein Ratner, N. (2006). Evidence-based practice: An examination of its ramifications for the practice of speech-language pathology. *Language,* *Speech, and Hearing Services in Schools*, *37*(4), 257–267. https://doi.org/10.1044/0161-1461(2006/029)

Bernstein Ratner, N. (1997). Linguistic and perceptual characteristics of children at stuttering onset. In E. C. Healey & H. F. M. Peters (Eds.), *Proceedings of the Second World Congress on Fluency Disorders* (pp. 3–6). Nijmegen University Press.

Bernstein Ratner, N., Brundage, S. B., & Fromm, D. (2020). *A clinician's complete guide to CLAN and PRAAT*. https://talkbank.org/manuals/Clin-CLAN.pdf

Bernstein Ratner, N., & MacWhinney, B. (2016). Your laptop to the rescue: Using the Child Language Data Exchange System Archive and CLAN utilities to improve child language sample analysis. *Seminars in Speech and Language*, *37*(02), 074–084. https://doi.org/10.1055/s-0036-1580742

Bernstein Ratner, N., & MacWhinney, B. (2018). Fluency Bank: A new resource for fluency research and practice. *Journal of Fluency Disorders*, *56*, 69–80. https://doi.org/10.1016/j.jfludis.2018.03.002

Bernstein Ratner, N., & Sih, C. C. (1987). Effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders*, *52*(3), 278–287. https://doi.org/10.1044/jshd.5203.278

Bernstein Ratner, N., & Silverman, S. (2000). Parental perceptions of children's communicative development at stuttering onset. *Journal of Speech, Language, and Hearing Research*, *43*(5), 1252–1263. https://doi.org/10.1044/jslhr.4305.1252

Birbili, M., & Karagiorgou, I. (2009). Helping children and their parents ask better questions: An intervention study. *Journal of Research in Childhood Education*, *24*(1), 18–31. https://doi.org/10.1080/02568540903439359

Bloodstein, O., Bernstein Ratner, N., & Brundage, S. B. (2021). *A handbook on stuttering* (7th ed.). Plural Publishing, Inc.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*(3), 148–158. https://doi.org/10.1177/002383096500800302

Brundage, S. B., & Bernstein Ratner, N. (1989). Measurement of stuttering frequency in children's speech. *Journal of Fluency Disorders*, *14*(5), 351–358. https://doi.org/10.1016/0094-730X(89)90015-6

Buhr, A., & Zebrowski, P. (2009). Sentence position and syntactic complexity of stuttering in early childhood: A longitudinal study. *Journal of Fluency Disorders*, *34*(3), 155–172. https://doi.org/10.1016/j.jfludis.2009.08.001

Byrd, C. T., Coalson, G., & Bush, C. (2011). The communicative intent of stuttered utterances. *Journal of Interactional Research in Communication Disorders*, *1*(2), 253–275. https://doi.org/10.1558/jircd.v1i2.253

Charest, M., & Johnston, J. R. (2011). Processing load in children's language production: A clinically oriented review of research. *Canadian Journal of Speech-Language Pathology and Audiology*, *35*(1), 18–31.

Choi, D., Conture, E. G., Walden, T. A., Lambert, W. E., & Tumanova, V. (2013). Behavioral inhibition and childhood stuttering. *Journal of Fluency Disorders*, *38*(2), 171–183. https://doi.org/10.1016/j.jfludis.2013.03.001

Choi, D., Sim, H., Park, H., Clark, C. E., & Kim, H. (2020). Loci of stuttering of English- and Korean-speaking children who stutter: Preliminary findings. *Journal of Fluency Disorders*, *64*, 1–14. https://doi.org/10.1016/j.jfludis.2020.105762

Clark, H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Constantino, C. D., Leslie, P., Quesal, R. W., & Yaruss, J. S. (2016). A preliminary investigation of daily variability of stuttering in adults. *Journal of Communication Disorders*, *60*, 39–50. https://doi.org/10.1016/j.jcomdis.2016.02.001

Crain-Thoreson, C., & Dale, P. S. (1999). Enhancing linguistic performance: Parents and teachers as book reading partners for children with language delays. *Topics in Early Childhood Special Education*, *19*(1), 28–39. https://doi.org/10.1177/027112149901900103

de Sonneville-Koedoot, C., Stolk, E., Rietveld, T., & Franken, M.-C. (2015). Direct versus indirect treatment for preschool children who stutter: The RESTART randomized trial. *PLOS ONE*, *10*(7), e0133758. https://doi.org/10.1371/journal.pone.0133758

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Pearson Assessments.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. American Guidance Service.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech, Language, and Hearing Research*, *39*(3), 643–654. https://doi.org/10.1044/jshr.3903.643

Eisenberg, S., & Guo, L.-Y. (2016). Using language sample analysis in clinical practice: Measures of grammatical accuracy for identifying language impairment in preschool and school-aged children. *Seminars in Speech and Language*, *37*(02), 106–116. https://doi.org/10.1055/s-0036-1580740

Eisenberg, S., & Guo, L.-Y. (2018). Percent grammatical responses as a general outcome measure: Initial validity. *Language, Speech, and Hearing Services in Schools*, *49*(1), 98–107. https://doi.org/10.1044/2017_LSHSS-16-0070

Eisenberg, S., Guo, L.-Y., & Germezia, M. (2012). How grammatical are 3-year-olds? *Language, Speech, and Hearing Services in Schools*, *43*(1), 36–52. https://doi.org/10.1044/0161-1461(2011/10-0093)

Ellis Weismer, S., Venker, C. E., Evans, J. L., & Moyle, M. J. (2013). Fast mapping in late-talking toddlers. *Applied Psycholinguistics*, *34*(01), 69–89. https://doi.org/10.1017/S0142716411000610

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur Communicative Development Inventory*. Singular Publishing Group.

Fey, M. (1986). *Language intervention with young children*. Allyn & Bacon.

Franken, M.-C., & Putker-de-Bruijn, D. (2007). *Restart-DCM Method. Treatment protocol developed within the scope of the ZonMW project Cost-effectiveness of the Demands and Capacities Model based treatment compared to the Lidcombe programme of early stuttering intervention: Randomised trial.* https://restartdcm.nl

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, *29*(9), 1083–1096. https://doi.org/10.1080/01690965.2013.832785

Gaines, N. D., Runyan, C. M., & Meyers, S. C. (1991). A comparison of young stutterers' fluent versus stuttered utterances on measures of length and complexity. *Journal of Speech and Hearing Research*, *34*, 37–42. https://doi.org/10.1044/jshr.3401.37

Gardner, M. F. (1990). *Expressive One-Word Picture Vocabulary Test—Revised*. Academic Therapy Publications.

Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language*, *30*(4), 431–451. https://doi.org/10.1016/0749-596X(91)90015-C

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press, Inc.

Gordon, P. A., & Luper, H. L. (1989). Speech disfluencies in nonstutterers. *Journal of Fluency Disorders*, *14*(6), 429–445. https://doi.org/10.1016/0094-730X(89)90029-6

Gorman, K., Olson, L., Hill, A. P., Lunsford, R., Heeman, P. A., & van Santen, J. P. H. (2016). Uh and um in children with autism spectrum disorders or language impairment. *Autism Research*, *9*(8), 854–865. https://doi.org/10.1002/aur.1578

Hall, N. E., Wagovich, S. A., & Bernstein Ratner, N. (2007). Language considerations in childhood stuttering. In E. G. Conture & R. F. Curlee (Eds.), *Stuttering and related disorders of fluency* (3rd ed., pp. 153–167). Thieme Medical Publishers, Inc.

Hanley, J. R., Cortis, C., Budd, M.-J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology*, *142*, 36–47. https://doi.org/10.1016/j.jecp.2015.09.008

Haynes, W. O., & Hood, S. B. (1978). Disfluency changes in children as a function of the systematic modification of linguistic complexity. *Journal of Communication Disorders*, *11*(1), 79–93. https://doi.org/10.1016/0021-9924(78)90055-2

Heilmann, J., Weismer, S. E., Evans, J., & Hollar, C. (2005). Utility of the MacArthur—Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology*, *14*(1), 40–51. https://doi.org/10.1044/1058-0360(2005/006)

Hollister, J., Van Horne, A. O., & Zebrowski, P. (2017). The relationship between grammatical development and disfluencies in preschool children who stutter and those who recover. *American Journal of Speech-Language Pathology*, *26*(1), 44–56. https://doi.org/10.1044/2016_AJSLP-15-0022

Hudson Kam, C. L., & Edwards, N. A. (2008). The use of uh and um by 3- and 4-year-old native English-speaking children: Not quite right but not completely wrong. *First Language*, *28*(3), 313–327. https://doi.org/10.1177/0142723708091149

Irvine, C. A., Eigsti, I. M., & Fein, D. A. (2016). Uh, um, and autism: Filler disfluencies as pragmatic markers in adolescents with optimal outcomes from autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(3), 1061–1070. https://doi.org/10.1007/s10803-015-2651-y

Jackson, E. S., Rodgers, N. H., & Rodgers, D. B. (2019). An exploratory factor analysis of action responses to stuttering anticipation. *Journal of Fluency Disorders*, *60*, 1–10. https://doi.org/10.1016/j.jfludis.2019.03.001

Jackson, E. S., Yaruss, J. S., Quesal, R. W., Terranova, V., & Whalen, D. H. (2015). Responses of adults who stutter to the anticipation of stuttering. *Journal of Fluency Disorders*, *45*, 38–51. https://doi.org/10.1016/j.jfludis.2015.05.002

Johnson, K. N., Walden, T. A., Conture, E. G., & Karrass, J. (2010). Spontaneous regulation of emotions in preschool children who stutter: Preliminary findings. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1478–1495. https://doi.org/10.1044/1092-4388(2010/08-0150)

Jones, R. M., Conture, E. G., & Walden, T. A. (2014). Emotional reactivity and regulation associated with fluent and stuttered utterances of preschool-age children who stutter. *Journal of Communication Disorders*, *48*, 38–51. https://doi.org/10.1016/j.jcomdis.2014.02.001

Kelman, E., & Nicholas, A. (2020). *Palin parent-child interaction therapy for early childhood stammering* (2nd ed.). Routledge.

Kimmell, M. (1938). Studies in the psychology of stuttering, IX: The nature and effect of stutterers' avoidance reaction. *Journal of Speech Disorders*, *3*(2), 95–100. https://doi.org/10.1044/jshd.0302.95

Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, *43*(2), 548–559. https://doi.org/10.1044/jslhr.4302.548

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review*, *18*(1), 135–140. https://doi.org/10.3758/s13423-010-0037-x

LaSalle, L. R., & Conture, E. G. (1995). Disfluency clusters of children who stutter: Relation of stutterings to self-repairs. *Journal of Speech, Language, and Hearing Research*, *38*(5), 965–977. https://doi.org/10.1044/jshr.3805.965

Lee, L. L. (1974). *Developmental sentence analysis*. Northwestern University.

Leech, K. A., Bernstein Ratner, N., Brown, B., & Weber, C. M. (2017). Preliminary evidence that growth in productive language differentiates childhood stuttering persistence and recovery. *Journal of Speech, Language, and*

Hearing Research, *60*(11), 3097–3109. https://doi.org/10.1044/2017_JSLHR-S-16-0371

Leech, K. A., Bernstein Ratner, N., Brown, B., & Weber, C. M. (2019). Language growth predicts stuttering persistence over and above family history and treatment experience: Response to Marcotte. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1371–1372. https://doi.org/10.1044/2019_JSLHR-S-18-0318

Logan, K. J., & Conture, E. G. (1995). Length, grammatical complexity, and rate differences in stuttered and fluent conversational utterances of children who stutter. *Journal of Fluency Disorders*, *20*(1), 35–61. https://doi.org/10.1016/0094-730X(94)00008-H

Luckman, C., Wagovich, S. A., Weber, C., Brown, B., Chang, S.-E., Hall, N. E., & Bernstein Ratner, N. (2020). Lexical diversity and lexical skills in children who stutter. *Journal of Fluency Disorders*, *63*, 1–14. https://doi.org/10.1016/j.jfludis.2020.105747

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *WORD*, *15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

MacPherson, M. K., & Smith, A. (2013). Influences of sentence length and syntactic complexity on the speech motor control of children who stutter. *Journal of Speech, Language, and Hearing Research*, *56*(1), 89–102. https://doi.org/10.1044/1092-4388(2012/11-0184)

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

MacWhinney, B., & Osser, H. (1977). Verbal planning functions in children's speech. *Child Development*, *48*, 978–985. https://doi.org/10.2307/1128349

Maner, K. J., Smith, A., & Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research*, *43*(2), 560–573. https://doi.org/10.1044/jslhr.4302.560

McDaniel, D., McKee, C., & Garrett, M. F. (2010). Children's sentence planning: Syntactic correlates of fluency variations. *Journal of Child Language*, *37*(1), 59–94. https://doi.org/10.1017/S0305000909009507

McGregor, K. K., & Hadden, R. R. (2018). Brief report: "Um" fillers distinguish children with and without ASD. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3736-1

McGregor, K. K., & Leonard, L. B. (1994). Subject pronoun and article omissions in the speech of children with specific language impairment: A phonological interpretation. *Journal of Speech, Language, and Hearing Research*, *37*(1), 171–181. https://doi.org/10.1044/jshr.3701.171

McKee, C., McDaniel, D., & Garrett, M. F. (2017). Children's performance abilities: Language production. In E. M. Fernandez & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 459–480). Wiley-Blackwell.

Melnick, K. S., & Conture, E. G. (2000). Relationship of length and grammatical complexity to the systematic and nonsystematic speech errors and stuttering of children who stutter. *Journal of Fluency Disorders*, *25*(1), 21–45. https://doi.org/10.1016/S0094-730X(99)00028-5

Meyers, S. C., & Freeman, F. J. (1985). Are mothers of stutterers different? An investigation of social-communicative interaction. *Journal of Fluency Disorders*, *10*, 193–209. https://doi.org/10.1016/0094-730x(85)90010-5

Miles, S., & Bernstein Ratner, N. (2001). Parental language input to children at stuttering onset. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1116–1130. https://doi.org/10.1044/1092-4388(2001/088)

Millard, S. K., Nicholas, A., & Cook, F. M. (2008). Is Parent–Child Interaction Therapy effective in reducing stuttering? *Journal of Speech, Language, and Hearing Research*, *51*(3), 636–650. https://doi.org/10.1044/1092-4388(2008/046)

Moyle, M. J., Ellis Weismer, S., Evans, J. L., & Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 508–528. https://doi.org/10.1044/1092-4388(2007/035)

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 109–149). Cambridge University Press.

Ntourou, K., DeFranco, E. O., Conture, E. G., Walden, T. A., & Mushtaq, N. (2020). A parent-report scale of behavioral inhibition: Validation and application to preschool-age children who do and do not stutter. *Journal of Fluency Disorders*, *63*, 105748. https://doi.org/10.1016/j.jfludis.2020.105748

Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research*, *44*(1), 207–223. https://doi.org/10.1044/1092-4388(2001/018)

Oetting, J. B., & Pruitt, S. (2005). Southern African-American English use across groups. *Journal of Multilingual Communication Disorders*, *3*(2), 136–144. https://doi.org/10.1080/14769670400027324

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Richels, C. G., Buhr, A., Conture, E. G., & Ntourou, K. (2010). Utterance complexity and stuttering on function words in preschool-age children who stutter. *Journal of Fluency Disorders*, *35*(3), 314–331. https://doi.org/10.1016/j.jfludis.2010.06.001

Richels, C. G., & Conture, E. G. (2007). An indirect treatment approach for early intervention for childhood stuttering. In E. G. Conture & R. F. Curlee (Eds.), *Stuttering and related disorders of fluency* (3rd ed., pp. 77–99). Thieme Medical Publishers, Inc.

Rispoli, M. (2003). Changes in the nature of sentence production during the period of grammatical development. *Journal of Speech, Language, and Hearing Research*, *46*(4), 818–830. https://doi.org/10.1044/1092-4388(2003/064)

Rispoli, M. (2018). Changing the subject: The place of revisions in grammatical development. *Journal of Speech, Language, and Hearing Research*, *61*(2), 360–372. https://doi.org/10.1044/2017_JSLHR-L-17-0216

Rispoli, M., & Hadley, P. (2001). The leading-edge: The significance of sentence disruptions in the development of grammar. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1131–1143. https://doi.org/10.1044/1092-4388(2001/089)

Rispoli, M., Hadley, P., & Holt, J. (2008). Stalls and revisions: A developmental perspective on sentence production. *Journal of Speech, Language, and Hearing Research*, *51*(4), 953–966. https://doi.org/10.1044/1092-4388(2008/070)

Santelmann, L., Berk, S., Austin, J., Somashekar, S., & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language*, *29*(4), 813–842. https://doi.org/10.1017/S0305000902005299

Scarborough, H. (1990). Index of productive syntax. *Applied Psycholinguistics*, *11*(1), 1–22. https://doi.org/10.1017/S0142716400008262

Schwartz, H. D., Zebrowski, P. M., & Conture, E. G. (1990). Behaviors at the onset of stuttering. *Journal of Fluency Disorders*, *15*(2), 77–86. https://doi.org/10.1016/0094-730X(90)90034-P

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals Preschool* (2nd ed.). Pearson Assessments.

Siegel, G. M. (2000). Demands and capacities or demands and performance? *Journal of Fluency Disorders*, *25*(4), 321–327. https://doi.org/10.1016/S0094-730X(00)00081-4

Silverman, S., & Bernstein Ratner, N. (2002). Measuring lexical diversity in children who stutter: Application of vocd. *Journal of Fluency Disorders*, *27*(4), 289–304. https://doi.org/10.1016/S0094-730X(02)00162-6

Smith, A., & Weber, C. (2017). How stuttering develops: The multifactorial dynamic pathways theory. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2483–2505. https://doi.org/10.1044/2017_JSLHR-S-16-0343

Starkweather, C. W. (1987). *Fluency and Stuttering*. Prentice-Hall, Inc.

Stocker, B., & Usprich, C. (1976). Stuttering in young children and level of demand. *Journal of Childhood Communication Disorders*, *1*(2), 116–131. https://doi.org/10.1177/152574017600100205

Stockman, I. J. (2010). Listener reliability in assigning utterance boundaries in children's spontaneous speech. *Applied Psycholinguistics*, *31*(3), 363–395. https://doi.org/10.1017/S0142716410000032

Tichenor, S., Leslie, P., Shaiman, S., & Yaruss, J. S. (2018). Speaker and observer perceptions of physical tension during stuttering. *Folia Phoniatrica et Logopaedica*, *69*(4), 180–189. https://doi.org/10.1159/000486032

Tichenor, S., & Yaruss, J. S. (2019). Stuttering as defined by adults who stutter. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4356–4369. https://doi.org/10.1044/2019_JSLHR-19-00137

Tichenor, S., & Yaruss, J. S. (2021). Variability of stuttering: Behavior and impact. *American Journal of Speech-Language Pathology*, *30*(1), 75–88. https://doi.org/10.1044/2020_AJSLP-20-00112

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1284–1294. https://doi.org/10.1044/jshr.3906.1284

Tumanova, V., Woods, C., & Razza, R. (2020). The role of behavioral inhibition for conversational speech and language characteristics of preschool-age children who stutter. *American Journal of Speech-Language Pathology*, *29*(2), 638–651. https://doi.org/10.1044/2019_AJSLP-19-00026

Usler, E. R., & Walsh, B. (2018). The effects of syntactic complexity and sentence length on the speech motor control of school-age children who stutter. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2157–2167. https://doi.org/10.1044/2018_JSLHR-S-17-0435

Van Riper, C. (1973). *The treatment of stuttering*. Prentice-Hall.

Wagovich, S. A., & Bernstein Ratner, N. (2007). Frequency of verb use in young children who stutter. *Journal of Fluency Disorders*, *32*(2), 79–94. https://doi.org/10.1016/j.jfludis.2007.02.003

Wagovich, S. A., & Hall, N. E. (2018). Stuttering frequency in relation to lexical diversity, syntactic complexity, and utterance length. *Communication Disorders Quarterly*, *39*(2), 335–345. https://doi.org/10.1177/1525740117702454

Wagovich, S. A., Hall, N. E., & Clifford, B. A. (2009). Speech disruptions in relation to language growth in children who stutter: An exploratory study. *Journal of Fluency Disorders*, *34*(4), 242–256. https://doi.org/10.1016/j.jfludis.2009.09.004

Wall, M. J., Starkweather, C. W., & Cairns, H. S. (1981). Syntactic influences on stuttering in young child stutterers. *Journal of Fluency Disorders*, *6*(4), 283–298. https://doi.org/10.1016/0094-730X(81)90017-6

Watson, J. B., Byrd, C. T., & Carlo, E. J. (2011). Effects of length, complexity, and grammatical correctness on stuttering in Spanish-speaking preschool children. *American Journal of Speech-Language Pathology*, *20*(3), 209–220. https://doi.org/10.1044/1058-0360(2011/10-0019)

Weiss, A. L., & Zebrowski, P. M. (1992). Disfluencies in the conversations of young children who stutter: Some answers about questions. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1230–1238. https://doi.org/10.1044/jshr.3506.1230

Wiig, E. H., Secord, W. A., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals- Preschool*. Psychological Corporation.

Wilkenfeld, J. R., & Curlee, R. F. (1997). The relative effects of questions and comments on children's stuttering. *American Journal of Speech-Language Pathology*, *6*(3), 79–89. https://doi.org/10.1044/1058-0360.0603.79

Wingate, M. E. (2001). SLD Is not stuttering. *Journal of Speech, Language, and Hearing Research*, *44*(2), 381–383. https://doi.org/10.1044/1092-4388(2001/031)

Yairi, E. (1981). Disfluencies of normally speaking two-year-old children. *Journal of Speech and Hearing Research*, *24*(4), 490–495. https://doi.org/10.1044/jshr.2404.490

Yairi, E. (1983). The onset of stuttering in two- and three-year-old children: A preliminary report. *Journal of Speech and Hearing Disorders*, *48*(2), 171–177. https://doi.org/10.1044/jshd.4802.171

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, *38*(2), 66–87. https://doi.org/10.1016/j.jfludis.2012.11.002

Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering: For clinicians by clinicians*. Pro-ed, Inc.

Yaruss, J. S. (1997). Clinical implications of situational variability in preschool children who stutter. *Journal of Fluency Disorders*, *22*(3), 187–203. https://doi.org/10.1016/S0094-730X(97)00009-0

Yaruss, J. S. (1999). Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language, and Hearing Research*, *42*(2), 329–344. https://doi.org/10.1044/jslhr.4202.329

Yaruss, J. S. (2004). Documenting individual treatment outcomes in stuttering therapy. *Contemporary Issues in Communication Science and Disorders*, *31*(Spring), 49–57. https://doi.org/10.1044/cicsd_31_S_49

Yaruss, J. S., Newman, R. M., & Flora, T. (1999). Language and disfluency in nonstuttering children's conversational speech. *Journal of Fluency Disorders*, *24*(3), 185–207. https://doi.org/10.1016/S0094-730X(99)00009-1

Yoder, P. J., & Davies, B. (1990). Do parental questions and topic continuations elicit replies from developmentally delayed children?: A sequential analysis. *Journal of Speech, Language, and Hearing Research*, *33*(3), 563–573. https://doi.org/10.1044/jshr.3303.563

Yoder, P. J., Davies, B., Bishop, K., & Munson, L. (1994). Effect of adult continuing wh-questions on conversational participation in children with developmental disabilities. *Journal of Speech, Language, and Hearing Research*, *37*(1), 193–204. https://doi.org/10.1044/jshr.3701.193

Zackheim, C. T., & Conture, E. G. (2003). Childhood stuttering and speech disfluencies in relation to children's mean length of utterance: A preliminary study. *Journal of Fluency Disorders*, *28*(2), 115–142. https://doi.org/10.1016/S0094-730X(03)00007-X

Zimmerman, I., Steiner, V., & Pond, R. (1992). *Preschool Language Scale* (3rd ed.). Psychological Corporation.