10-1-2014

When Less Can Be More: Dual Task Effects in Stuttering and Fluent Adults

Naomi Nechama Eichorn

Graduate Center, City University of New York

Recommended Citation
http://academicworks.cuny.edu/gc_etds/484

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Graduate Works by Year: Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
WHEN LESS CAN BE MORE: DUAL TASK EFFECTS IN STUTTERING AND FLUENT ADULTS

by

Naomi (Nechama) Eichorn

A dissertation submitted to the Graduate Faculty in Speech-Language-Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2014
This manuscript has been read and accepted for the Graduate Faculty in Speech-Language-Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Klara Marton, Ph.D.
Chair of Examining Committee

Date

Klara Marton, Ph.D.
Executive Officer

Richard G. Schwartz, Ph.D.
Supervision Committee Member

Robert D. Melara, Ph.D.
Supervision Committee Member

J. Scott Yaruss, Ph.D.
External Reviewer

THE CITY UNIVERSITY OF NEW YORK
Abstract

WHEN LESS CAN BE MORE: DUAL TASK EFFECTS IN STUTTERING AND FLUENT ADULTS

by

Naomi Eichorn

Advisor: Klara Marton

The present study tested the counterintuitive hypothesis that engaging cognitive resources in a secondary task while speaking could benefit aspects of speech production. Effects of dual task conditions on speech fluency, rate, and error patterns were examined in stuttering and fluent speakers based on specific predictions derived from three related theoretical frameworks. Twenty fluent adults and 19 adults with confirmed diagnoses of stuttering participated in the study. All participants completed two baseline tasks: (1) a continuous speaking task in which spontaneous speech was produced in response to given prompts; and (2) a working memory (WM) task involving manipulations of WM domain, WM load, and inter-stimulus interval (ISI). In the dual task portion of the experiment, participants simultaneously performed the speaking task with each unique combination of WM conditions. Resulting performance patterns were examined based on speech-related measures (fluency, rate, errors) and WM accuracy in each speaker group. Contrary to predicted outcomes, both groups showed comparable decrements in secondary task performance as well as comparable fluency benefits as a result of dual task conditions. This effect was specific to atypical forms of disfluency and was similar across all manipulations of the WM task. Changes in fluency were accompanied by reductions in speaking rate, but not by corresponding changes in overt speech errors. Overall, findings suggest that WM contributes to disfluencies regardless of stuttering status and that suppressing these resources
enhances speech fluency, possibly by inducing more implicit or automatic modes of movement during speech production. Further research is needed to more precisely identify the cognitive mechanism involved in this effect, clarify the nature of this association, and determine whether and how these findings can inform clinical intervention.
Dedication

This work is dedicated to Gilad, Eyal, Naftali, and to the 64 selfless soldiers of the Israeli Defense Force whose lives and dreams were cut so short, all during the course of this writing.

May their families find comfort and their memories be blessed.
Acknowledgements

I am incredibly grateful to so many individuals whose unique contributions helped me reach this milestone. I am indebted, first and foremost, to the chair of my committee, Dr. Klara Marton. Thank you, Klara, for your expert guidance and endless insights, for the many, varied opportunities you provided for me to learn about research, and for your consistent encouragement. Your approach to science and learning is an inspiration to me and I strive to emulate your openness to new ideas and perspectives, your persistence and positive outlook, and your ability to balance personal and professional obligations with such grace. It was my incredible fortune to work with such a knowledgeable, creative, and supportive advisor and mentor.

I am also grateful to the members of my committee, Dr. Richard Schwartz and Dr. Robert Melara, for helping me develop the concepts explored in this study and for taking the time to review my dissertation proposal and draft. Your nuanced questions helped me focus on aspects of the data I had not considered and your unique perspectives added tremendously to my understanding and interpretation of the results. Thank you for your guidance and support.

I am thankful to Dr. Scott Yaruss for so graciously agreeing to serve as the outside reader. It was a privilege to discuss and analyze my work with someone who has contributed so prominently to the field of stuttering and I deeply appreciate the time and thought put into your questions, comments, and feedback.

I am immensely grateful for the generous dissertation fellowship provided by the CUNY Graduate Center to support this research over the past year and to all the faculty at the Graduate Center for the interest you have each shown in my work and progress. I thank Dr. Martin
Gitterman, in particular, for his ongoing positive feedback, which has always meant a lot to me. I also appreciate the consistent efforts made by Dr. Loraine Obler to facilitate my continued academic growth and achievement.

I owe special gratitude to Dr. Gail Gurland for inspiring me to pursue doctoral studies and for demonstrating how clinical and academic interests can be successfully balanced. I also acknowledge the influence of Dr. Natalie Schaeffer, who provided me with my very first research experience and in so doing, helped shape my professional trajectory.

This accomplishment reflects the value and emphasis consistently placed by my parents and grandparents on hard work, educational achievement, and the pursuit of knowledge. To my sister, Goldie, thank you for paving the way.

Special acknowledgements go to my lab members and friends, within and outside of the Graduate Center, for their constant encouragement and for injecting camaraderie into a process that could sometimes feel frustrating. In particular, I thank Luca Campanelli, for being someone I am always able to learn from and for sharing his time and skills so generously.

I am grateful to all the individuals who participated in this study and helped recruit others – this could have not been done without you.

To my son, Amitai, thank you for infusing unique joy and meaning into this and all my life’s pursuits. Last, to my husband, Yossi… words do not suffice. This accomplishment is yours as much as mine; it could never have happened without the incredible emotional and practical support you provided every step of the way. Thank you for believing in me more than I will ever believe in myself and for your endless, wholehearted encouragement of my goals and dreams.
# TABLE OF CONTENTS

Abstract .......................................................................................................................iv  
Dedication ..................................................................................................................vi  
Acknowledgements ........................................................................................................vii  

Chapter I. Introduction ........................................................................................................1  
1.1 Literature Review .......................................................................................................1  
1.2 Disfluencies as a result of covert repairs ......................................................................2  
1.3 Disfluencies as a result of hypervigilance ...................................................................7  
1.4 Disfluencies due to explicit attention .............................................................................11  
1.5 Dual task paradigm .....................................................................................................18  
1.6 Hypotheses ..................................................................................................................20  

Chapter II. Methods ...........................................................................................................24  
2.1 Participants ..................................................................................................................24  
2.2 General Procedures ......................................................................................................27  
2.3 Stimuli ..........................................................................................................................28  
2.3.1 Speaking tasks ........................................................................................................28  
2.3.2 WM tasks ...............................................................................................................29  
2.4 Procedure ....................................................................................................................30  
2.4.1 Baseline Tasks .......................................................................................................30  
2.4.2 Vigilance Task .......................................................................................................34  
2.4.3 Dual Tasks ............................................................................................................35  
2.4.4 Subjective Measures ............................................................................................37  
2.5 Data processing ..........................................................................................................37
2.6 Data analysis ...........................................................................................................40

Chapter III. Results ........................................................................................................41

3.1 Preliminary analyses ...............................................................................................41
3.2 Subjective ratings ...................................................................................................42
3.3 Secondary task performance ................................................................................46
3.4 Speech rate .............................................................................................................50
3.5 Fluency ..................................................................................................................53
    3.5.1 Between-subject effects ..................................................................................53
    3.5.2 Subject level variables ....................................................................................58
3.6 Speech errors .........................................................................................................62

Chapter IV. Discussion ...................................................................................................65

4.1 Perceived dual task effects ....................................................................................65
4.2 Secondary task performance patterns ....................................................................67
4.3 Dual task effects on speaking rate .........................................................................70
4.4 Dual task effects on fluency ..................................................................................72
4.5 Dual task effects on overt errors ..........................................................................76
4.6 Does a specific cognitive mechanism contribute to speech disfluency? .................77
4.7 Resource allocation in individuals who stutter .......................................................79
4.8 Optimal levels of cognitive control: When less can be more and more can be less ....81
4.9 Is stuttering associated with underlying deficits in WM? .....................................83
4.10 Limitations ..........................................................................................................84
4.11 Future research ....................................................................................................87
4.12 Clinical implications ............................................................................................88
4.13 Summary and conclusion

Appendices

References
List of Tables

Table 1. Participant characteristics. ................................................................. 25
Table 2. Demographic and diagnostic details for stuttering participants (N=19). .......... 26
Table 3. Experimental conditions for WM task. ..................................................... 33
Table 4. Number of speaking and WM trials per condition in dual tasks. ..................... 36
Table 5. Subjective ratings by group and task type. ............................................. 42
Table 6. Proportion of correct responses on secondary task by group and condition. ....... 47
Table 7. Model fit of sequential multilevel logistic regression model of secondary task performance. .............................................................................. 48
Table 8. Results of final model of secondary task performance (M11). ......................... 48
Table 9. Syllable counts (raw and adjusted) by group and condition. Mean (SD). .......... 51
Table 10. Model fit of sequential multilevel regression models of speech rate. ............. 52
Table 11. Results of final model of speech rate (M4). ........................................... 52
Table 12. Atypical disfluencies (raw counts and as a proportion of syllables) by group and condition. ............................................................................. 54
Table 13. Model fit of sequential multilevel negative binomial regression models of disfluencies. .............................................................................. 56
Table 14. Results of final model of disfluencies (M5). ............................................. 57
Table 15. Model fit of sequential multilevel negative binomial regression model of disfluencies including subject-level variables. ........................................... 59
Table 16. Results of final model of disfluencies including subject-level variables (M13). ... 60
Table 17. Model fit of sequential multilevel negative binomial regression models of disfluencies including subject-level severity for AWS only. .................................................................61

Table 18. Results of final model of disfluencies including subject-level stuttering severity for AWS only (M15). .........................................................................................................................62

Table 19. Errors (raw counts and proportion of syllables) by group and condition..................63

Table 20. Model fit of sequential multilevel negative binomial regression models of speech errors. ..............................................................................................................................64
List of Figures

Figure 1. Representation of a single trial for the verbal (left) and spatial (right) WM tasks ........32
Figure 2a. Subjective measures by task type and group: Cognitive effort..................................44
Figure 2b. Subjective measures by task type and group: Speech anxiety ..................................45
Figure 2c. Subjective measures by task type and group: Stuttering severity ...............................45
Figure 3a. Secondary task performance by task type and group: WM domain ..............................49
Figure 3b. Secondary task performance by task type and group: WM load ..................................49
Figure 3c. Secondary task performance by task type and group: ISI ............................................50
Figure 4. Speech rate (syllables/second) by task type and group ..................................................52
Figure 5. Proportion of disfluent syllables by task type and group: Atypical disfluencies .............57
Figure 6. Proportion of disfluent syllables by task type and group: Typical disfluencies ..........58
Figure 7. Proportion of syllables containing errors by task type and group .................................64
Devoting more mental effort to a task, whether by increasing attention or allocating more working memory (WM) resources, is generally assumed to benefit task performance. A corresponding assumption is that performing two activities simultaneously will result in performance decrement on one or both tasks (e.g., Bourke, 1996; Pashler, 1994). The present study explored whether these premises hold true in the case of speech production, particularly for speakers who vary in their level of speech fluency. Within the stuttering literature, specific hypotheses (Kolk, 1991; Postma, Kolk, & Povel, 1991; Postma & Kolk, 1993; Vasić & Wijnen, 2005) have proposed that stuttering arises from excessive speech monitoring activity, and that limiting attentional or working memory (WM) resources available for such monitoring can improve fluency. Evidence from movement sciences and sport psychology (e.g., Beilock, Carr, Macmahon, & Starkes, 2002; DeCaro & Beilock, 2010; Wulf & Lewthwaite, 2010) further suggest that conscious attempts to control learned motor skills interfere with efficient performance and result in performance breakdown. Based on combined ideas of these lines of research, the aim of the present study was to examine the role of WM in speech or movement monitoring, suppress WM resources in a systematic manner, and examine whether resulting spoken output shows appreciable fluency gains in both speaker groups.

1.1 Literature Review

Chronic stuttering affects 1% of the adult population (Bloodstein & Bernstein Ratner, 2008; Craig, Hancock, Tran, Craig, & Peters, 2002; Gordon, 2002; Porfert & Rosenfield, 1978), and is highly resistant to treatment once it persists beyond early childhood. Despite the profound impact of this disorder on communication, self-confidence, and quality of life, the stuttering
literature still lacks a unifying theoretical model to explain the cause of stuttering and to support a specific approach to treatment. Although the importance of managing stuttering through a variety of strategies addressing areas other than overt behaviors is generally recognized, traditional therapies for stuttering (particularly those emphasizing speech restructuring) essentially involve training speakers to focus on specific aspects of speech production, such as respiration, articulation, or prosody (e.g., see Blomgren, 2013 for a recent review). In contrast to this approach, literature reviewed below suggests that enhanced fluency may be achieved by minimizing rather than maximizing conscious control over these component processes. Theoretical and empirical support for this notion is presented below, based on three different accounts for speech disfluencies, two from within the stuttering literature and one based on related research on attention and motor performance. These perspectives on disfluency are followed by the specific aims and predictions motivating the present study.

1.2 Disfluencies as a result of covert repairs

One account of stuttering that implies potential benefit to speech fluency when central resources are limited is the Covert Repair Hypothesis (CRH; Postma et al., 1991; Postma & Kolk, 1993), which is based on Levelt’s influential model of speech production and related studies on speech monitoring and self-repair (Hockett, 1967; Levelt, 1989; Levelt, 1983). As described by Levelt (1989), spoken output is generated through a series of three independently functioning stages or modules: (1) the conceptualizer, which forms preverbal messages based on a speaker’s intentions and ideas; (2) the formulator, which converts preverbal messages into a phonetic plan via grammatical and phonological encoding processes; and (3) the articulatory system, which translates the phonetic plan to specific motor movements so that it can be realized as overt speech. Levelt’s Perceptual Loop Model (1989; 1983) further specifies three feedback
loops by which preverbal messages (ideas), the phonetic plan (inner speech), and spoken output (executed motor movements) can be monitored and corrected at their respective stages of speech production. A total of eleven feedback loops have been identified (see Postma, 2000 for details); however, the one most relevant to the CRH and stuttering is the inner loop, a channel through which the phonetic plan is parsed and monitored approximately 150-200 milliseconds after it is generated in order to detect and correct prearticulatory errors (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999).

The question of whether speech monitoring requires conscious awareness and which cognitive mechanisms are involved in this function remains unclear. According to Levelt (1989), “self-corrections are hardly ever made without a touch of awareness” (p. 21) and occur within the conceptualizer module, possibly via working memory, which receives alarm signals generated by the monitor when errors are detected. Crosson (1985) suggested that prearticulatory monitoring occurs in neurological areas involved in language comprehension, particularly the temporoparietal cortex. Involvement of this region is consistent with Levelt’s assertion that self-produced speech and speech produced by others are monitored by identical comprehension-based processes. Data from more recent neurophysiological studies link speech monitoring processes to medial frontal areas, such as the anterior cingulate cortex (Riès, Janssen, Dufau, Alario, & Burle, 2011), in addition to bilateral regions within the temporal cortex (McGuire, Silbersweig, & Frith, 1996), suggesting that the detection of speech errors may at least partially involve a general-purpose monitoring mechanism, as well as language-specific processes. A shared assumption of these studies is that speech monitoring is under some degree of conscious control, mediated by attention, and dependent upon resources that are subject to capacity limitations (Brocklehurst, 2008; Levelt, 1989; Oomen & Postma, 2002; Postma, 2000).
The term *covert repairs*, as defined by Levelt (1989; 1983) and earlier, by Hockett (Hockett, 1967), refers specifically to the repair of errors in the prearticulatory speech plan via inner loop feedback. This editing can be completed before overt articulation begins, resulting in no evidence of either an error or a repair. When articulation has already been initiated, it is interrupted and either restarted or postponed until the corrected phonetic plan is ready for execution, thus disrupting progress of the utterance and resulting in observable disfluencies. Specific disfluency types will vary depending on the specific type of error detected, the strategy utilized for correction (restart or postponement), and the point at which the utterance is interrupted (see Kolk, 1991; Postma et al., 1991; Postma & Kolk, 1993 for a complete list of disfluency types and corresponding internal errors and repair strategies). This link between covert editing and speech disfluency was first described by Hockett in his 1967 writing, where he suggested that covert editing could account for “everyday stuttering” but only partially explained the articulatory spasms characteristic of stuttering disorders (Hockett, 1967, p. 115).

Postma and Kolk extended Hockett’s proposal to stuttering by attributing the excessive disfluencies observed in stuttering speakers to an underlying deficit in phonological encoding that ultimately caused a large number of errors to arise during speech planning. According to the CRH and based on concepts described in spreading activation models (e.g., Dell, 1986), this deficit results in slow activation of phonemic elements during preparation of the articulatory plan and a correspondingly long period of uncertainty during which many competing phonemic responses are simultaneously activated. Unless the speaker slows his rate of speech accordingly, inappropriate phonological nodes are selected and errors occur. These errors are subsequently detected by an internal speech monitor, which generates an alarm signal. Based on the CRH
account, this alarm signal triggers repair action, causing frequent disfluencies as an unintended byproduct.

Empirical support for the CRH is broadly derived from two categories of studies: (1) studies examining phonological encoding in individuals who stutter, and (2) dual task studies examining effects of suppressing the speech monitoring mechanism. Overall, evidence for phonological encoding deficits in stuttering speakers is mixed. Comparisons of stuttering and fluent children have shown that the two groups produce similar rates of overt errors in spontaneous speech (LaSalle & Conture, 1995; Yaruss & Conture, 1996), show equal articulatory proficiency (Ryan, 1992), benefit from phonological priming to the same extent (Melnick, Conture, & Ohde, 2003), and demonstrate similar patterns of phonological development overall (see Nippold, 2002 for review). A recent study of phonological priming benefits in stuttering and fluent adults likewise showed no differences between speaker groups (Vincent, Grela, & Gilbert, 2012). On the other hand, review of the literature reveals considerable evidence of at least subtle weaknesses in phonological encoding among children and adults who stutter, based on analyses of spoken output (Anderson & Byrd, 2008; Howell, Au-Yeung, & Sackin, 2000) and experimental studies of priming effects (Byrd, Conture, & Ohde, 2007), rhyme judgments (Jones, Fox, & Jacewicz, 2012; Weber-Fox, Spencer, Spruill, & Smith, 2004; Weber-Fox, Spruill, Spencer, & Smith, 2008), phoneme monitoring (Sasisekaran, Brady, & Stein, 2013; Sasisekaran, De Nil, Smyth, & Johnson, 2006), and nonword repetition (Hakim & Ratner, 2004; Sasisekaran & Weisberg, 2014).

Dual task studies examining predictions of the CRH are similarly inconclusive. Postma and colleagues (Postma et al., 1991) found that stuttering and fluent speakers and controls
exhibited fewer errors and more disfluencies when speech accuracy was emphasized. These findings are predicted by the CRH and provide what is probably the clearest evidence for its key claims. Several early dual task studies also showed a reduction in disfluencies for stuttering individuals exposed to various forms of distraction, such as masking noise (Maraist & Hutton, 1957) or continuous visual tracking (Arends, Povel, & Kolk, 1988), suggesting that limiting attentional resources available for monitoring can bring about a reduction in disfluency. However, this finding has not been replicated consistently in similarly structured dual task studies (e.g., Oomen & Postma, 2002) and several studies have reported the opposite effect (i.e., increased disfluencies under dual task conditions) (Bosshardt, 1999; Bosshardt, 2002; Greiner, Fitzgerald, & Cooke, 1986; Oomen & Postma, 2001). To some degree, inconsistent findings can be explained by methodological differences across studies, such as: (1) the nature of the speaking task used (e.g., repetition of single words, repetition of tongue twisters, narratives, reading); (2) features of the secondary task, such as complexity (e.g., tapping vs. mathematical calculations), stimulus and processing domains (e.g., visual/auditory, verbal/spatial, cognitive/motor); and failure of most studies to distinguish between specific types of disfluency (typical vs. atypical) which may be differentially influenced by dual tasking demands. An important theoretical limitation in previous literature on this topic is that the majority of studies do not identify a specific cognitive resource being taxed by the secondary tasks selected or explain how these tasks may suppress resources normally involved in speech monitoring (for exceptions, see Bosshardt, 1999; De Nil & Bosshardt, 2000).

The present study focused on two key predictions of the CRH. First, although the specific mechanism responsible for speech monitoring is not clearly understood, Levelt’s original model
assumes that working memory (WM) receives “alarm signals” generated by the monitor when errors are detected (Levelt, 1983, p. 50) and that control of self-repair is therefore subject to expected WM limitations. Accordingly, secondary tasks that reliably tax WM should produce a clear effect on monitoring functions and speech fluency. Second, the CRH asserts that differences between stuttering and fluent speakers are quantitative rather than qualitative (Kolk, 1991). Thus, stuttering and fluent speakers should be similarly affected by the dual-task condition and comparable changes should be observed across different types of disfluency. In sum, expected results based on the CRH include: (1) reduced disfluencies when WM resources are suppressed; and (2) similar effects for fluent and stuttering speakers and across categories of disfluency.

1.3 Disfluencies as a result of hypervigilance

Like the CRH, the Vicious Circle Hypothesis (VCH; Vasić & Wijnen, 2005) links stuttering to speech monitoring functions but discards the notion of underlying phonological encoding deficits. Instead, the VCH maintains that the speech monitor of individuals who stutter is a faulty, hypervigilant mechanism that emits alarm signals in the absence of true encoding errors. This notion is largely based on earlier research by Martin (1970), who first attributed stuttering to excessively conservative monitoring criteria, and Sherrard (1975), who described stuttering as “false alarm responses” to monitoring feedback. Drawing on these principles, the VCH claims that individuals who stutter: (1) devote more effort to speech monitoring than needed; (2) pay excessive attention to temporal discontinuities in speech that are often normal interruptions; and (3) set exceedingly low thresholds during error monitoring, all of which result in frequent detection and repair of errors. Vasić and Wijnen (2005) further speculate that early stuttering may arise from a child’s heightened awareness (internally generated or triggered by
others) to his own disfluencies, causing maladaptive adjustments of monitoring parameters. This hypervigilance, in turn, results in a *vicious circle* of increasing discontinuity and counterproductive forms of compensation.

Initial support for the VCH was based on a series of dual task experiments conducted by Vasić and Wijnen (2005) which showed a significant reduction in atypical forms of disfluency (particularly blocks) for stuttering speakers engaged in attention demanding tasks (easy and difficult versions of the PONG computer game, linguistic distractor task) while speaking. As previously found in certain dual task studies (Arends, Povel, & Kolk, 1988; Postma et al., 1991), disfluency levels in all dual conditions differed significantly from the baseline condition; however, the largest effect occurred during the linguistic distractor task, in which participants continuously monitored their spoken output for the presence of a particular word (*that*) while they were speaking. Combined findings were interpreted as evidence that overt disfluencies are reduced when: (1) attentional resources available for speech monitoring are suppressed; or (2) monitoring focus is shifted away from its habitual focus on temporal fluctuations. This outcome was inconsistent with several other dual task studies in the stuttering literature (e.g., Bosshardt, 1999; De Nil & Bosshardt, 2002; Oomen & Postma, 2002; Oomen & Postma, 2001b), likely due to methodological considerations, as previously described.

Although the experiments initially described by Vasić and Wijnen have not been replicated, other studies have examined the general role of evaluative criteria in stuttering and reported findings that support some central claims of the VCH. Lickley and colleagues (Lickley, Hartsuiker, Corley, Russell, & Nelson, 2005) found that stuttering adults used more stringent criteria than fluent speakers when rating the fluency of pre-recorded utterances (regardless of whether the utterances were actually produced by a stuttering speaker). Others have suggested
links between stuttering persistence and general perfectionistic tendencies, based on results of self-report scales measuring attitudes and beliefs related to perfectionism (Amster & Klein, 2007; Brocklehurst 2008b). Notably, however, Postma & Kolk (1992) found comparable error detection performance (based on rate, speed, and false alarms) for stuttering and fluent participants monitoring self-produced speech, which contradicts the VCH and other studies on this topic.

Predictions of the VCH have also been examined through the use of neurophysiological tools, with overall results suggesting atypical neural monitoring activity in individuals who stutter. Specifically, several neuroimaging studies have identified higher levels of activity in the basal ganglia (e.g., Giraud et al., 2008) and anterior cingulate cortex (ACC; Brown, Ingham, Ingham, Laird, & Fox, 2005; De Nil, Kroll, Kapur, & Houle, 2000) in stuttering individuals compared to controls. Although its role in stuttering is often understood to involve motor control and planning (e.g., Alm, 2004), the basal ganglia has also been identified as an important component of the neural monitoring circuit (Holroyd & Coles, 2002) based on its anatomical connection and dopaminergic input to the ACC. The ACC has long been associated with cognitive control and its specific contribution to performance monitoring includes the detection of conflicts in information processing (such as response competition) and modulation of cognitive control in response to conflict (Botvinick, Cohen, & Carter, 2004; Carter et al., 1998). Research examining electrophysiological components associated with the ACC (error-related negativity [ERN] and error positivity [Pe]) in stuttering and fluent controls demonstrated larger ERN amplitudes among adults who stutter (Arnstein, Lakey, Compton, & Kleinow, 2011). Group differences in this study were observed across linguistic (rhyme judgments) as well as nonlinguistic (flanker) tasks, and for both correct and incorrect responses, suggesting that error
detection processes are heightened in individuals who stutter even when tasks do not involve covert monitoring of phonemes and even in the absence of errors. Overall, neurophysiological data point to ACC involvement in speech monitoring and, consistent with the VCH, provide neural evidence for hyperactive monitoring processes in individuals who stutter. WM capacity, particularly as conceptualized by Engle and others (e.g., Broadway, Redick, & Engle, 2010; Engle & Kane, 2004; Kane, Conway, Habrick, & Engle, 2007), is defined as the executive control of attention. Performance monitoring mechanisms, in turn, are believed to optimize performance by deploying increased control (see Botvinick & Cohen, 2014 for recent review). Based on this perspective, neuroimaging data related to stuttering suggest that excessive reliance on control processes associated with performance monitoring may contribute to disfluency. Accordingly, conditions in which WM resources are engaged should limit the extent to which this control can be deployed, and thereby promote increased fluency in people who stutter.

From a theoretical standpoint, the VCH and CRH are similar in their shared focus on speech monitoring processes; however, the VCH makes several testable assumptions that contradict the CRH and that are explored in the present study. First, unlike the CRH, the VCH does not imply a tradeoff between speech errors and overt disfluencies. Accordingly, conditions that minimize hypervigilant monitoring in individuals who stutter should not be accompanied by increases in the number of speech errors. Second, whereas the CRH views disfluencies in stuttering speakers as only quantitatively different from those in fluent speakers and provides the same account for disfluencies in both speaker groups, the VCH applies specifically to individuals who stutter. Monitoring integrity is assumed to be intact in fluent speakers and thus, control participants should not show any fluency benefit from conditions that alter the amount of monitoring resources or the monitor’s focus. It is difficult to discern, however, whether results
obtained by Vasić & Wijnen (2005) supported either prediction, since error data were not reported at all and complete details regarding fluency effects were provided for the stuttering group but not for controls. The present study addressed these limitations and examined two specific questions based on the VCH: (1) If dual task conditions reduce the frequency of disfluencies, is this reduction accompanied by an increase in speech errors? (2) If dual task conditions enhance speech fluency, is this effect specific to stuttering speakers?

1.4 Disfluencies due to explicit attention

A final approach to stuttering, which has not been considered in existing literature, is to view disfluent speech production as a type of performance breakdown caused by heightened self-awareness or excessive attention to the process of speaking. Although conceptually similar to both the CRH and VCH, this view does not assume involvement of a speech monitoring mechanism at all and suggests disruption further downstream, at the level of motor movements involved in articulation rather than the level of speech planning. A key principle in this literature is the differentiation between implicit and explicit forms of information processing (defined below), with the former being well-suited for skilled motor performance but the latter being counterproductive (Liao & Masters, 2001; Masters, 1992; Poolton, Maxwell, Masters, & Raab, 2006).

The explicit processing system is a rule-based system linked to conscious awareness and WM, whereas the implicit system is skill- or experience-based and involves content that cannot be verbalized and is not available for representation in WM (e.g., Maddox & Ashby, 2004). This distinction is derived more generally from dual-process theories across various disciplines, which use varied terminology but ultimately posit the same two forms of cognitive processing: (1) analytic, declarative, effortful, rule-based, System 2 processes, which are analogous to explicit
processing; and (2) automatic, procedural intuitive, associative, heuristic, System 1 processes, which resemble implicit processing (e.g., Kahneman, 2003; Maxwell, Masters, & Eves, 2003; Shiffrin & Schneider, 1977). Critically, implicit or procedural processing is exceedingly efficient, resilient, and less vulnerable to stress or distraction (e.g., Beilock et al., 2002), making it the ideal information system for managing skilled motor performance. The concept of autotelic experience has been used by researchers in this area to describe the optimal, effortless performance patterns achieved when implicit forms of processing are emphasized (see Dietrich & Stoll, 2010 for review). In contrast, the terms choking under pressure or freezing (Baumeister, 1984; Beilock & Gray, 2007; Lewis & Linder, 1997) metaphorically describe the phenomenon in which well-learned skills are performed sub-optimally under situations of pressure.

A theoretical account of choking first described by Baumeister (Baumeister, 1984; Baumeister & Showers, 1986) proposed that situational demands for excellence or perfection result in increased conscious attention to the internal process of performance (i.e., coordination and precision of muscle movements), disrupting automaticity. In Baumeister’s classic example, a typist’s attempt to ensure accuracy by consciously monitoring finger movements predictably results in less accurate and efficient typing performance. As interpreted by Baumeister and others (Baumeister, 1984; Baumeister & Showers, 1986; Masters, 1992; see Beilock & Gray, 2007 for review), increased pressure induces an inward focus of attention and results in explicit processing of proceduralized knowledge that normally functions without WM involvement. The outcome is step-by-step control, or dechunking of movement sequences into independent units (Masters, 1992) in a way that resembles the irregular and effortful movement patterns observed
during early phases of motor learning and that is ultimately counterproductive for skilled performers.

In support of this account, several studies have found that experienced athletes (golfers, soccer players, baseball batters) perform more poorly during experimental conditions that require skill-focused attention (e.g., attending to timing of golf swing) (Beilock & Gray, 2012; Beilock et al., 2002; Gray, 2004). Skilled typists similarly became slower and less accurate under conditions requiring them to attend to details of their performance (Snyder & Logan, 2013). A comparison of rock climbing performance in low and high anxiety conditions (low and high height) likewise found that high levels of anxiety resulted in more rigid movements, slower overall climbing speed, and longer grasped holds (Pijpers, Oudejans, & Bakker, 2005). Collectively, these results indicate that situations involving pressure or that inherently call attention to processes underlying motor performance negatively influence movement fluidity and efficiency.

Drawing upon the same concept, other studies have examined whether suppressing resources used for explicit processing could enhance motor learning and performance by forcing participants to rely primarily on implicit control systems. Many of these studies utilized the dual task paradigm and included secondary tasks such as random letter generation (Masters, 1992), tone detection (Beilock, et al., 2002), or cumulative tone counting (Poolton, et al., 2006) in order to tax WM, which is the presumed source of explicit performance control. It is not clear from these studies whether WM is viewed as a unitary construct or a multiple-components system (which would imply that the strength of the effect may vary based on the degree of similarity between tasks); however, results generally indicate an advantage for skilled participants forced to
use implicit rather than explicit forms of knowledge (Beilock, et al., 2002; Liao & Masters, 2001; Masters, 1992).

Similar outcomes have been demonstrated through experimental manipulations of the focus of attention during motor task performance. Accumulating evidence suggests that instructing participants to focus externally, on the effects of movements, rather than internally, on the movements themselves, promotes greater automaticity and enhances overall performance. This result has been replicated across a wide variety of motor tasks, including golf-putting (Wulf & Su, 2007), jumping (Wulf & Dufek, 2009), balancing (McNevin, Shea, & Wulf, 2003), dart-throwing (McKay & Wulf, 2012), and notably, oral motor performance (Freedman, Maas, Caligiuri, Wulf, & Robin, 2007). Based on their findings, Wulf and colleagues proposed the Constraining Action Hypothesis (CAH), which posits that conscious control of movements through an internal attentional focus constrains the movement system and disrupts automatization; whereas external focusing induces reflexive control processes and results in more effortless and optimal movement. In corroboration of this account, studies incorporating direct measures of neuromuscular effort via electromyography (EMG) demonstrate that efficient movement patterns (induced by an external focus) are accompanied by less rather than more EMG activity (Vance, Wulf, Tollner, McNevin, & Mercer, 2004; Wulf, Dufek, Lozano, & Pettigrew, 2010; Zachry, Wulf, Mercer, & Bezodis, 2005). Overall, Wulf’s explanation is closely in line with explicit monitoring models for choking under pressure, as reviewed above. Specifically, both approaches share two central principles: (1) motor performance is enhanced by minimizing the expenditure of task-related effort; and (2) deconstructing movement sequences is detrimental to task performance. It is important to note that the majority of these studies focus on complex skills involving gross motor movements (golf, soccer, simulated skiing), which are
inherently different from the fine motor sequences underlying speech production. A study of attentional focus effects on isolated oral and manual pressure bursts (Freedman et al., 2007), however, suggests that similar outcomes occur in simple, fine motor movements and specifically, in the oral facial system.

Although this theoretical framework has never been applied to stuttering disorders, it is remarkably similar to Bloodstein’s well-known Anticipatory Struggle Hypothesis, which proposed that speech-related anxiety and excessive attention to the process of speech production play a key role in the development of stuttering (1972, 1975, 1984). Specifically, Bloodstein suggested that the perception or anticipation of difficulty results in “exaggerated speech consciousness,” which causes the speaker to control speech movements by assuming more rigid articulatory postures and producing speech in a “piece by piece” fashion (Bloodstein, 1975, p. 5). This description of the maladaptive fragmentation underlying tense blocks closely resembles the performance breakdown described by the CAH.

Fragmentation of motor sequences also impedes automaticity, which may be critical to fluent speech production. Levelt (2001) suggested that the well-practiced motor patterns required for articulatory syllables represent a form of overlearning and are stored as a mental syllabary in premotor cortex. This description implies that motor movements involved in producing speech are highly automatized and generally rely upon implicit rather than explicit processing systems. Although this may be the case for fluent speakers, evidence suggests that speech as well as nonspeech processes are less automatized for people who stutter (Saltuklaroglu, Teulings, & Robbins, 2009; Smits-Bandstra, De Nil, & Saint-Cyr, 2006; Smits-Bandstra, De Nil, & Rochon, 2006; Smits-Bandstra & Gracco, 2013). A dual task study involving a repetitive manual task (circle tracing) and simultaneous production of choral speech (a fluency enhancing condition)
showed that stuttering adults exhibited more manual disfluency on the tracing task than controls, even when their stuttering was virtually eliminated (Saltuklaroglu et al., 2009). As interpreted by the authors, results indicate that stuttering speakers may expend greater amounts of effort when speaking, both in response to, or in anticipation of, stuttering. Studies of explicit sequence learning, in which practice effects are measured for trained target sequences, also show poorer performance (typically, slower reaction time) in people who stutter relative to controls during production of both nonverbal (finger-tapping) (Smits-Bandstra, De Nil, & Saint-Cyr, 2006; Smits-Bandstra & De Nil, 2013; Smits-Bandstra, De Nil, & Rochon, 2006) as well as verbal sequences (Bauerly & De Nil, 2011; Smits-Bandstra & De Nil, 2009). These differences tend to be most evident on earlier stages of practice, suggesting that stuttering is associated with difficulty achieving automaticity.

Finally, the distinction between implicit and explicit forms of processing has important clinical relevance. The process of speech production is subjectively perceived as effortful by individuals who stutter (Ingham, Warner, Byrd, & Cotton, 2006) and to a large extent, can remain effortful even after intensive treatment (Boberg & Kully, 1994). Available research related to attentional effort suggests that fluency treatment may be enhanced by incorporating methods to induce a more proceduralized approach to speech production; however, exploration of this possibility are still generally absent in the literature.

The only available data pertaining to this question are described by Metten and her colleagues (Metten et al., 2011, Experiment 3, p.940), who incorporated dual tasks during the final stage of stuttering treatment to facilitate automatization of recently learned speaking patterns and prepare participants for real life speaking demands. Procedures involved
spontaneous speaking and a 1-back categorization task (“Does this word belong to the same category as previous word?”). Results of the intervention trial are not very reliable, given the small number of participants (N=3) and lack of objective data for this portion of the study; however, feedback revealed that the three participants all found the dual task procedure beneficial. Specifically, participants indicated that the task helped them experience the potential ease with which they could speak when not simultaneously engaged in a secondary task. This subjective report indicates that participants did not actually experience any change in the automaticity of their speech production; thus, the nature of the perceived benefit is not entirely clear. It is possible that objective measures of disfluency during or following the dual task condition might have reflected increased automaticity, but no such data is reported.

Although the present study draws upon concepts derived from the CAH and related theoretical approaches to effort, attention, and movement, predicted outcomes for stuttering speakers are somewhat conjectural, since this literature makes no specific claims about stuttering. Nonetheless, several reasonably direct deductions can be drawn. Like the CRH and VCH, this literature predicts fluency benefits when specific cognitive resources are suppressed, although targeted resources would be those specifically associated with explicit processing. Further, this effect should theoretically benefit all speakers, although stuttering participants may benefit to a greater degree if excessive disfluency reflects general dependence on declarative rather than proceduralized representation of motor speech patterns. With respect to anticipated error effects, attention-based explanations for stuttering assume no involvement of a speech monitor in speech disfluencies; thus, conditions that increase fluency should not be associated with increased errors. A question that remains open relates to the specific cognitive mechanism associated with explicit control. Several researchers assume WM involvement in explicit action
monitoring and develop secondary tasks that clearly draw upon WM resources during dual task conditions (e.g., Masters, 1992; Poolton et al., 2006); however, specific domains within WM (e.g., verbal/spatial) are not differentiated. Moreover, many studies show performance effects as a result of generally distracting tasks (e.g., tone detection task in Beilock & Gray, 2012) or conditions that involve changes in attentional focus (e.g., external vs. internal) (Freedman et al., 2007; McKay & Wulf, 2012; Vance et al., 2004; Wulf & Lewthwaite, 2010) but do not tax WM at all. Thus, it is unclear whether explicit processing necessarily involves WM or whether it reflects more general attentional resources. In summary, previous hypotheses and findings in motor performance literature suggest that: (1) the benefit of implicit processing may not be unique to stuttering participants; (2) occurrence of speech errors should not increase in conjunction with fluency benefits; and (3) fluency effects, if observed, may or may not be dependent on resources specifically associated with WM.

1.5 Dual task paradigm

The performance of two tasks simultaneously can provide important insights into the way cognitive resources are shared between activities and is a frequently utilized paradigm in stuttering research (e.g., Bosshardt, 2002; Kamhi & McOsker, 1982; Mallard & Webb, 1980; Metten et al., 2011; Smits-Bandstra & De Nil, 2009) as well as in various domains related to cognitive science. Three important considerations related to experimental applications of the dual task paradigm are reviewed below, followed by an overview of the general design and specific hypotheses of the present study.

One critical factor influencing dual task performance is the degree to which concurrently performed tasks compete for the same limited resources. Although tasks that rely on similar processes typically result in interference (Leclercq, 2002), several studies demonstrate that
healthy adults are able to perform certain combinations of demanding tasks without significantly compromising performance on either task when tasks do not involve similar processing domains or output modalities (Allport, Antonis, & Reynolds, 1972; Cocchini, Logie, Sala, MacPherson, & Baddeley, 2002; Duff & Logie, 2001). In many studies of this nature (e.g., see Cocchini et al., 2002) concurrent tasks are specifically performed at a level representing individual capacity, based on performance of each task in isolation. This design ensures that each processing domain is being taxed to an extent that can potentially produce interference under concurrent task conditions, if the tasks interfere with each other. These findings are consistent with a multicomponent view of working memory, as proposed by Baddeley (Baddeley & Logie, 1999; Baddeley & Hitch, 1974) as well as the multichannel hypothesis described by Allport and colleagues (Allport et al., 1972), which proposed “special-purpose” independent processors that are able to operate in parallel. In line with this perspective, neuroimaging studies suggest that recruitment of overlapping areas of the cortex, particularly in frontal and parietal regions form the neurophysiological basis of dual task interference (Klingberg, 1998; Rémy, Wenderoth, Lipkens, & Swinnen, 2010; Watanabe & Funahashi, 2014).

A second consideration in dual task design is the extent to which tasks overlap in time. Based on the bottleneck explanation for dual task interference, operations simultaneously vying for dedicated use of central resources will result in either delayed or impaired performance on one or both tasks (Pashler, 1994; Pashler, 1992). This effect is classically demonstrated by the psychological refractory effect, in which a response to the second of two discrete stimuli is increasingly delayed as the interval between stimuli decreases (i.e., as temporal overlap increases). Similar competition occurs between two continuous tasks being performed simultaneously, although in this case, the tasks intermittently compete for access to the central
mechanism, and the extent of interference will depend on how frequently this access is required by each task (Dux, Ivanoff, Asplund, & Marois, 2006; Pashler, 1992).

Last, the degree of dual task interference is highly influenced by the degree of automaticity associated with each task being performed (Pashler, 1999; Poldrack et al., 2005). Performance is considered automatic when it requires no capacity demands and is not affected by other ongoing mental activity. When a task is well-practiced, larger pieces of information are chunked and continuous access to central resources is no longer necessary. Thus, secondary tasks can clarify the degree to which a primary task relies on conscious control. Several studies within the stuttering literature have demonstrated that compared to fluent speakers, individuals who stutter perform more poorly on secondary tasks executed while speaking (Saltuklaroglu et al., 2009; Smits-Bandstra & De Nil, 2009), indicating that speaking is less automatized in this group and continues to require regular access to central resources.

In accordance with these general considerations, the present dual task study combined a spontaneous speaking (primary) task and secondary WM task that incorporated manipulations of domain (verbal vs. spatial), WM load, and inter-stimulus interval (ISI). The overarching goal of the study was to determine how specific patterns of dual task interference caused by these manipulations affected aspects of speech production in adult speakers. The degree to which dual task performance was affected by automaticity was further examined through the use of speaker groups that differed in their fluency status, as described in more detail below.

1.6 Hypotheses

Based on literature reviewed above, the present study examined the following central hypotheses:
1. Adults who stutter (AWS) will demonstrate poorer accuracy on WM tasks performed in combination with spontaneous speaking, particularly when conditions involve greater dual task interference (verbal domain, short ISI) and higher WM load. Although the secondary (WM) task was specifically designed so that participants could maintain high levels of accuracy under dual task conditions (in order to reliably measure effects of the secondary task on speech production), previous dual task studies suggest that secondary task performance of AWS may be more vulnerable to dual task decrement due to incomplete automatization of speech production (Saltuklaroglu et al., 2009; Smits-Bandstra & De Nil, 2009) and greater overall susceptibility to interference under dual task conditions (Bosshardt, 1999, 2002, 2006).

2. All speakers will exhibit a reduction in speaking rate under dual compared to baseline speaking conditions. Most dual task studies within the stuttering literature focus on fluency effects without considering speech rate changes (Bosshardt, 2002; Bosshardt, 1999; Kamhi & McOsker, 1982; Oomen & Postma, 2001a; Postma et al., 1991; Vasić & Wijnen, 2005); however, slower speech rate under dual task conditions is a logical outcome of imposing simultaneous demands while speaking and has been reported in at least one previous study of fluent speakers (Oomen & Postma, 2002). A smaller reduction in speech rate is predicted for AWS, based on their slower speech rate at baseline (due to frequent disfluency) and anticipated fluency benefit under dual conditions (which could increase their rate).

3. Speaking under dual task conditions (while performing concurrent WM task) will affect fluency across participants; however, effects will differ based on speaker group
(stuttering vs. fluent speaker), disfluency type (typical vs. atypical form of disfluency), and dual task condition (WM domain, WM load, ISI).

a. The occurrence of atypical forms of disfluency, which are generally associated with stuttering (see Methods section and Appendix X for details regarding disfluency types), will decrease under dual task conditions; however, this effect will be specific to AWS. Fluent speakers are not expected to produce a significant number of this disfluency type and therefore, not expected to demonstrate the same dual task effect predicted for stuttering participants.

b. The occurrence of typical forms of disfluency will increase under dual task conditions for all speakers. Concurrent demands of the speaking and WM tasks place an obvious strain on the language system, which is expected to manifest across speakers as an increase in typical disfluencies (e.g., fillers, phrase repetitions, revisions).

c. Dual task effects will be specific to secondary tasks involving verbal (but not spatial) WM. This prediction is based on the consideration that the speaking and verbal WM tasks draw upon similar resources and are therefore more likely to result in interference (Cocchini et al., 2002; Leclercq, 2002) than concurrent performance of the speaking and spatial WM task.

d. Increasing the load of WM and decreasing ISI in the secondary WM task will strengthen dual task effects. These manipulations place greater demands on cognitive resources by requiring larger amounts of information to be processed and more frequent access to the central bottleneck. Predicted speech effects include an increased fluency benefit for AWS (fewer atypical disfluencies) but
greater interference in language formulation (more typical disfluencies) for both
groups.

4. Overt speech errors will not increase under dual task conditions for either group.

Although a reverse outcome is predicted by the CRH, most previous studies examining
its claim have not found any appreciable rise in speech errors as a result of diverting
central resources (e.g., Oomen & Postma, 2001b; Vasić & Wijnen, 2005).
Chapter II. Methods

2.1 Participants

Twenty self-identified adults who stuttered (AWS) and 20 adults who did not stutter (AWNS), all between the ages of 18 and 35, served as participants in the study. Participants were recruited via flyers posted or distributed at college campuses, speech-language clinics, and stuttering support group meetings in and around New York City. Informed consent was obtained before beginning experimental procedures. Participants were each compensated $40 for their time. Data from one AWS was excluded due to poor speech intelligibility in recorded samples.

All participants met the following inclusionary criteria: (1) nonverbal intelligence within at least the average range (minimum standard score 85) on the Test of Nonverbal Intelligence - 4th Edition (Brown, Sherbenou, & Johnsen, 2010); (2) expressive vocabulary within at least the average range (minimum standard score 85) on the Expressive One-Word Picture Vocabulary Test - 4th Edition (Martin & Brownell, 2011); and (3) absence of significant medical history, learning disability, hearing loss, head injury or cognitive impairment, as determined by a screening questionnaire and interview. Four potential participants (2 AWS and 2 AWNS) could not be included in the experimental sample based on failure to meet TONI-4 and/or EOWPVT-4 performance criteria. All participants additionally completed computer-based operation span (OSpan) and symmetry span (SSpan) tasks (Unsworth, Heitz, Schrock, & Engle, 2005; also see Redick et al., 2012 for reliability, validity and scoring details). The purpose of these tasks was to verify that AWS and AWNS did not differ in working memory capacity for verbal and spatial stimuli. All participants spoke English as their primary language. See Table 1 for a summary of participant demographics and scores on standardized cognitive and language measures.
Table 1. Participant characteristics. Mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>AWNS</th>
<th>AWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>25.60 (4.58)</td>
<td>27.21 (4.18)</td>
</tr>
<tr>
<td>Males/Females</td>
<td>11/9</td>
<td>12/7</td>
</tr>
<tr>
<td>Right-/Left-Handedness</td>
<td>17/3</td>
<td>16/3</td>
</tr>
<tr>
<td>SSI-4 (raw score)</td>
<td>4.11 (0.32)</td>
<td>15.78 (8.64)</td>
</tr>
<tr>
<td>TONI-4 (standard score)</td>
<td>107.85 (12.90)</td>
<td>107.56 (12.54)</td>
</tr>
<tr>
<td>EOWPVT-4 (standard score)</td>
<td>104.60 (9.81)</td>
<td>102.11 (12.45)</td>
</tr>
<tr>
<td>SSpan: Absolute score</td>
<td>20.65 (7.04)</td>
<td>19.61 (12.52)</td>
</tr>
<tr>
<td>OSpan: Absolute score</td>
<td>45.85 (16.16)</td>
<td>42.56 (17.39)</td>
</tr>
<tr>
<td>Vigilance Task: mean accuracy</td>
<td>96.67 (2.42)</td>
<td>96.75 (3.49)</td>
</tr>
<tr>
<td>Vigilance Task: mean RT</td>
<td>476.74 (60.21)</td>
<td>510.35 (52.99)</td>
</tr>
</tbody>
</table>


Participants in the stuttering group met additional inclusion criteria including: (1) a rating of at least very mild stuttering on the Stuttering Severity Instrument – 4th Edition (Riley, 2009); and (2) at least mild total impact score on the Overall Assessment of Stuttering Experience Survey, a subjective measure of the impact of stuttering on various aspects of a speaker’s life (Yaruss & Quesal, 2006). Scoring of the SSI was based on three speech samples (monologue, conversation, reading), all of which were audio and video recorded, transcribed and coded for disfluencies (see Data Processing section for methodological and technical details), and scored according to standardized procedures recommended by Riley (2009). All AWS reported childhood onset of stuttering and all but one had received prior speech therapy for their stuttering. Among stuttering participants with a history of treatment, most (14/19) reported fairly recent speech therapy (ranging from 6 months to approximately 5 years prior to study). One participant reported more than 10 years since last treatment; three AWS were receiving therapy.
at the time of the study. Demographic and diagnostic details for stuttering participants are summarized in Table 2.

Table 2. Demographic and diagnostic details for stuttering participants (N=19).

<table>
<thead>
<tr>
<th></th>
<th>Handedness</th>
<th>Gender</th>
<th>Age</th>
<th>Most recent therapy</th>
<th>OASES: Impact Rating</th>
<th>SSI-4: Severity Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>right</td>
<td>male</td>
<td>23</td>
<td>3 years ago</td>
<td>Moderate/Severe</td>
<td>Very mild</td>
</tr>
<tr>
<td>2</td>
<td>right</td>
<td>female</td>
<td>27</td>
<td>Present</td>
<td>Moderate</td>
<td>Severe</td>
</tr>
<tr>
<td>3</td>
<td>right</td>
<td>female</td>
<td>21</td>
<td>2 years ago</td>
<td>Mild/Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>4</td>
<td>right</td>
<td>male</td>
<td>35</td>
<td>5 years ago</td>
<td>Mild/Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>5</td>
<td>left</td>
<td>male</td>
<td>31</td>
<td>2 years ago</td>
<td>Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>6</td>
<td>right</td>
<td>male</td>
<td>32</td>
<td>1 year ago</td>
<td>Moderate</td>
<td>Mild</td>
</tr>
<tr>
<td>7</td>
<td>left</td>
<td>female</td>
<td>28</td>
<td>1 year ago</td>
<td>Moderate</td>
<td>Severe</td>
</tr>
<tr>
<td>8</td>
<td>right</td>
<td>male</td>
<td>27</td>
<td>5 years ago</td>
<td>Mild/ Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>9</td>
<td>right</td>
<td>male</td>
<td>34</td>
<td>3 years ago</td>
<td>Moderate /Severe</td>
<td>Very mild</td>
</tr>
<tr>
<td>10</td>
<td>right</td>
<td>female</td>
<td>30</td>
<td>5 years ago</td>
<td>Mild/ Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>11</td>
<td>right</td>
<td>male</td>
<td>30</td>
<td>10 years ago</td>
<td>Moderate</td>
<td>Mild</td>
</tr>
<tr>
<td>12</td>
<td>right</td>
<td>male</td>
<td>22</td>
<td>Present</td>
<td>Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>13</td>
<td>right</td>
<td>male</td>
<td>26</td>
<td>Present</td>
<td>Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>14</td>
<td>right</td>
<td>female</td>
<td>26</td>
<td>6 months ago</td>
<td>Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>15</td>
<td>right</td>
<td>male</td>
<td>29</td>
<td>1 year ago</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>left</td>
<td>female</td>
<td>26</td>
<td>1 year ago</td>
<td>Moderate</td>
<td>Mild</td>
</tr>
<tr>
<td>17</td>
<td>right</td>
<td>female</td>
<td>20</td>
<td>n/a</td>
<td>Mild/ Moderate</td>
<td>Very mild</td>
</tr>
<tr>
<td>18</td>
<td>right</td>
<td>male</td>
<td>24</td>
<td>1 year ago</td>
<td>Mild/ Moderate</td>
<td>Mild</td>
</tr>
<tr>
<td>19</td>
<td>right</td>
<td>male</td>
<td>26</td>
<td>6 months ago</td>
<td>Mild/ Moderate</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Note: OASES=Overall Assessment of Subjective Experience of Stuttering; SSI=Stuttering Severity Instrument – 4th Edition. Diagnostic details unavailable for Participant #15 due to missing data (participant was unable to complete scheduled session and did not respond to attempted follow-up).

Groups were well-matched on characteristics and abilities potentially related to key dependent variables. There was no between group difference in age, $t(37) = -1.14, p = .26$, and the proportion of males to females was similarly distributed in both groups, $X^2(1, N = 39) = 0.27, p = .61$. The proportion of right- to left-handedness was heavily weighted toward right handedness but did not differ between groups, $X^2(1, N = 39) = 0.005, p = .95$. There were also no
between-group differences on standardized measures of nonverbal IQ (TONI-4), \( t(36) = 0.07, p = .94 \), and expressive vocabulary (EOWPVT-4), \( t(36) = 0.69, p = .50 \). Group performance patterns on span measures (based on an overall absolute scoring procedure, in which perfectly recalled sets are summed; see Unsworth et al., 2005) revealed no between-group differences on either span task, \( t(36) = 0.61, p = .55 \) and \( t(36) = 0.32, p = .75 \) for OSpan and SSpan tasks, respectively.

More detailed analysis of span task performance using mixed-design analysis of variance to examine response accuracy at each set size (3-7 for OSpan, 2-5 for SSpan) likewise revealed no between-group differences on either task, \( F(1,36) = 0.48, p = .50 \), partial \( \eta^2 = .01 \) and \( F(1,36) = 0.46, p = .50 \), partial \( \eta^2 = .01 \) for OSpan and SSpan respectively, and no Group x Set Size interaction, \( F(4,144) = 0.56, p = .69 \), partial \( \eta^2 = .02 \) and \( F(3,108) = 2.43, p = .07 \), partial \( \eta^2 = .06 \) for OSpan and SSpan, respectively, indicating similar WM capacity in stuttering and fluent adults, based on span length measures.

2.2 General Procedures

Following screening procedures (described above), all participants performed three baseline tasks, which were administered in a random order and included a spontaneous speaking task and two WM tasks, one primarily verbal (adding numbers), the other primarily nonverbal (tracking spatial positions). Both WM tasks also incorporated manipulations of WM load and ISI, which were randomly sequenced and are described in more detail below. Following the three baseline tasks, participants completed a preliminary dual task condition, in which they engaged in spontaneous speaking while simultaneously performing a simple vigilance task (see Vigilance Task in Procedure section for details). Participants that met pre-specified performance criteria (80%) on baseline and vigilance tasks proceeded to the two full dual task conditions in which
spontaneous speaking was combined with the verbal and spatial WM tasks completed in the baseline condition. Presentation order was randomized for the two full dual tasks and for all experimental conditions within each dual task. A final baseline speaking task, identical to the initial one, was administered following completion of the dual task conditions. All tasks were administered via a desktop computer using E-Prime 2.0 software (Psychology Software Tools, Inc.) to present stimuli and record button press responses from a standard keyboard. Spoken output was recorded via an adjustable headworn unidirectional microphone (Shure SM10A) connected to a preamplifier (Switchcraft 308TR). A digital video recorder was positioned beside the computer monitor and set to record continuously throughout speech production tasks. All conditions of the experimental task were completed in a single session. Remaining procedures (screening tasks, span measures) were generally administered in a separate session but were occasionally completed on the same day, based on participant preference. Breaks were provided between tasks as needed.

2.3 Stimuli

2.3.1 Speaking tasks

Stimuli for all speaking tasks consisted of simple, highly engaging prompts designed to elicit continuous speaking for the duration of 60 seconds. Prompts were presented as questions (e.g., Do smart phones make us smarter?), open statements (e.g., A teacher that made a difference) or direct instructions (Describe a favorite restaurant) and were largely adapted from the Student Opinion question list regularly maintained by the Learning Network blog section of the New York Times (Slotnik & Schulten, 2012). See Appendix A for a complete list of topic prompts presented. Feedback on a list of prospective topics during the piloting phase of the study
revealed that individuals (even those of similar ages) varied significantly in the topics they felt comfortable speaking about and for which they had a sufficient amount of material to discuss. Based on these responses, the task was modified so that prompts were always presented in sets of three throughout the experiment, allowing participants to select topics based on personal preference and experiences. Repetition of a previously selected prompt was possible on practice trials but was prevented from occurring on recorded speaking trials.

2.3.2 WM tasks

Stimuli for the verbal WM task consisted of a start number, followed by a sequence of 2, 3, or 4 single-digit operations (e.g., +3), each presented individually. Original stimuli developed for the experiment consisted of both addition and subtraction operations, consistent with procedures described by Salthouse, Babcock, & Shaw (1991); however, pilot data suggested that alternating between two operations was too difficult and subtraction trials were therefore discarded. Additional manipulations of the WM stimuli, including the use of degraded stimuli and/or distractors were considered during piloting phases of the study but ultimately rejected. This decision was based on the lack of a clear effect for the former (degraded versions of the verbal and spatial WM stimuli did not produce reliable changes in either accuracy or RT) and because the latter would have required recruitment of executive processes other than WM. In the final stimulus set, all start numbers were between 0 and 4; operation numbers were between 1-4; and outcomes were all below 10. The purpose of these criteria was to ensure that the task was simple enough to be performed at high levels of accuracy (at or above 80%), even while engaging in continuous speaking. Individual operations were never repeated more than once during a single trial. Each WM load condition resulted in four potential numeric outcomes (2-5 for WM load of 2; 4-7 for load of 3; and 6-9 for load of 4), with each outcome having equal
probability (25%) in each experimental condition. Pseudo trials (on which outcome responses were not entered; see below for details) in the dual task condition were similarly constrained and met identical criteria.

In the spatial WM task, a single colored circle (approximately 1-inch diameter) was presented in one of four cells within a 2 x 2 grid, similar to Salthouse and colleagues (1991). Presentation of the circle was followed by a sequence of 2, 3, or 4 arrows, each presented individually and pointing in one of the four cardinal directions. As in the verbal WM task, individual operations (directions) were never repeated more than once during a single trial. Each WM load condition resulted in four potential position outcomes, each having equal probability (25%) within any experimental condition. Pseudo trials in the dual task condition were developed with identical criteria and constraints.

2.4 Procedure

2.4.1 Baseline Tasks

In baseline speaking task, participants were instructed to produce continuous spontaneous speech for a period of 60 seconds for each of 4 topic prompts (e.g., “Describe a recent vacation”). Prompts were always presented in a set of three, as previously mentioned. The first speaking trial was considered practice and omitted from all analyses. Following topic selection, prompts disappeared and were replaced by the word “GO” on the screen, which remained visible for the duration of the speaking trial. During this time, E-Prime automatically recorded all spoken output and saved each trial (except practice trials) in individual audio files. These were subjected to extensive off-line coding and analysis, as described in the Data Processing section below. The baseline speaking task was administered twice: once at the beginning of the
experiment (pretest baseline) and once at the end (posttest baseline) in order to help account for possible order and practice effects during the experiment.

WM tasks were modeled after procedures described by Salthouse, Babcock, & Shaw (1991) for measuring operational capacity of WM but were adapted to examine effects of three experimental manipulations: WM domain (verbal vs. spatial); WM load (2, 3, and 4); and ISI (long vs. short). Trials within both verbal and spatial domains were identically structured and began with a centrally positioned stimulus, followed by a sequence of operations to be performed on the stimulus, and ending with a keyboard response from the participant. Although the task was originally designed with individual operations appearing in the same (center) screen location as the start stimulus, consistent with Salthouse and colleagues (1991), pilot data suggested that this format made recall more difficult and increased the likelihood of below criterion performance (<80%) during dual task conditions. Presentation of stimuli was therefore modified so that the starting stimulus always appeared in the center of the screen and individual operations appeared in designated positions from the far left to the far right of the lower third of the screen. All stimuli (start numbers/circles and operations) remained visible for 3000 ms. See Figure 1 for a visual representation of the WM task and organization of stimuli on individual trials.

In the verbal domain (verbal WM task), participants viewed a start number, followed by a series of addition operations and a prompt to enter the correct numeric outcome (2-9). In the spatial WM task, participants viewed a 2x2 grid with a colored circle in one cell, followed by a series of directional arrows and a prompt to enter the number representing the final location of the circle, based on a numbered grid provided on the screen. Feedback regarding performance accuracy was provided after every set of (2) WM trials. The purpose of the feedback was to
facilitate consistent engagement in the secondary task and help participants monitor their performance so that they met criteria (80%) for proceeding to dual tasks.

Figure 1. Representation of a single trial for the verbal (left) and spatial (right) WM tasks, each shown with WM load of 2. Note: WM=working memory; ISI=inter-stimulus interval; ms=milliseconds.

WM load was manipulated by varying the number of sequential operations (2, 3 or 4) to be performed. ISI varied between individual operations, with intervals of 3000 milliseconds (ms) and 1000 ms for the long and short ISI conditions, respectively. The purpose of ISI manipulations was to vary the relative frequency with which WM resources needed to be accessed during performance of the secondary task. The shorter ISI was expected to increase competition between simultaneous tasks for dedicated use of the same resource. The decision to use a 1000 ms interval was largely based on practical considerations, as feedback from pilot testing suggested that shorter ISIs would compromise participants’ ability to produce continuous speech output. One-second intervals (between secondary task stimuli) have also been used in several previous dual task studies involving continuous speech production (Declerck & Kormos,
The long ISI (3000 ms) was based on data demonstrating clear performance differences with stimulus onset asynchronies varying by an approximate multiple of three (Dux et al., 2006). Within each domain, two practice items were presented, both representing a WM load of 3 and ISI of 1000 ms, followed by a total of 60 test items, with 10 items for each unique combination of WM load and ISI. All trials after practice items were presented in randomized order. Accuracy of keyboard responses was automatically recorded and used for subsequent analysis. See Table 3 for a representation of the experimental paradigm with timing details for all conditions, and Figure 1 for a sample series of screens representing the baseline verbal and spatial WM tasks.

Table 3. Experimental conditions for WM task.

<table>
<thead>
<tr>
<th>Domain</th>
<th>ISI</th>
<th>WM load</th>
<th>Sample trial</th>
<th>Total time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>Long (ISI=3000 ms)</td>
<td>4 3 +1 +2 +1 +1</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 1 +1 +2 +3</td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2 2 +3 +2</td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Short (ISI=1000 ms)</td>
<td>4 3 +2 +1 +1 +2</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 1 +1 +2 +3</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2 2 +3 +1</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Long (ISI=3000 ms)</td>
<td>4 - - - -</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 - - - -</td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2 - - - -</td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Short (ISI=1000 ms)</td>
<td>4 - - - -</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 - - - -</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2 - - - -</td>
<td></td>
<td></td>
<td>12</td>
</tr>
</tbody>
</table>

Note: WM=working memory; ISI=inter-stimulus interval.
2.4.2 Vigilance Task

The vigilance task was a preliminary dual task condition in which participants engaged in spontaneous speech while simultaneously responding to simple stimuli (colored circles) appearing at varying locations on the right or left side of the screen. The purpose of the vigilance task was to obtain a basic measure of participants’ ability to manage dual task demands and was therefore designed with minimal attention demands and no working memory load (other than task instructions). As in baseline speaking tasks, participants selected one of three topic options, based on personal preference, and produced a 60-second sample of spoken output for each of four speaking trials. The first trial was considered practice and omitted from all analyses. Participants were instructed to respond as quickly as possible to all circle stimuli appearing on the screen by pressing designated buttons on the keyboard (‘J’ or ‘F’) when stimuli appeared on the right or left side of the screen, respectively. Twenty circle stimuli were presented during each speaking trial. Individual circles remained visible for 1000 ms; ISI randomly varied between 1000, 1500, 2000, 2500, and 3000 ms, with each ISI occurring four times per speaking trial. Multiple variations of ISI were included to minimize the predictability of target responses. Responses were considered accurate if they occurred during the 1000 ms period of stimulus presentation and corresponded to the side on which the stimulus was presented. Performance measures for the task consisted of accuracy and RT, which were recorded automatically via E-Prime. Spoken output was recorded for the duration of each speaking trial, primarily to ensure that speech production remained continuous. One potential AWS could not be included in the study sample based on failure to meet performance criterion (80%) on this task; all 39 participants in the final sample performed to criterion.
2.4.3 Dual Tasks

In the two full dual-task conditions, participants concurrently performed the speaking task and each WM task (verbal and spatial). The speaking task was identical to that performed in the baseline condition. The structure of each WM task likewise matched its baseline counterpart, with three manipulations of WM load (2, 3, and 4) and two manipulations of ISI (long and short ISI of 3000 and 1000 ms, respectively). The total number of WM trials per speaking trial varied, depending on WM load and ISI. For example, trials with an ISI of 3000 ms and WM load of 4 consisted of two WM trials per speaking trial, with each WM trial lasting 30 seconds, but trials with ISI of 1000 ms and WM load of 2 consisted of five WM trials per speaking trial, with each WM trial lasting 12 seconds.

Certain conditions required an incomplete WM trial, due to timing constraints imposed by the uniformly 60-second speaking trial (e.g., for ISI of 1000 ms and WM load of 3, speaking trials included 2½ WM trials, with complete WM trials lasting 24 seconds and the partial trial lasting 12 seconds). During incomplete WM trials, an initial stimulus was presented, followed by one or more operations, followed by a large X. The X signaled participants to disregard previous stimuli for that trial and begin again. Regardless of specific WM load and ISI conditions, keyboarded responses (to indicate numeric or spatial outcome) were only performed twice over the course of each speaking trial. The purpose of collecting only two outcome responses per trial was to keep the number of WM trials consistent across load and ISI conditions (highest load + long ISI condition only allowed for two WM trials; see Table 4). Participants were informed that some secondary task trials would be followed by outcome responses but some would not. The order of WM trials with and without outcome responses was randomized during each speaking
trial. Spoken output was saved whenever keyboard responses to WM trials were entered, resulting in two audio files per speaking trial.

Table 4. Number of speaking and WM trials per condition in dual tasks.

<table>
<thead>
<tr>
<th>Domain</th>
<th>ISI</th>
<th>WM load</th>
<th>Number of speaking trials</th>
<th>WM trials per speaking trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Long (ISI=3000 ms)</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td>2 ½</td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
<td>2</td>
<td>5</td>
<td>3 ⅓</td>
</tr>
<tr>
<td></td>
<td>Short (ISI=1000 ms)</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td>3 ⅓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Long (ISI=3000 ms)</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td>2 ½</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>5</td>
<td>3 ⅓</td>
</tr>
<tr>
<td>Spatial</td>
<td>Short (ISI=1000 ms)</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td>3 ⅓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Note: WM=working memory; ISI=inter-stimulus interval.

Participants were instructed to pay close attention to the secondary task but to speak continuously while the “GO” signal remained visible. A practice speaking trial with concurrent WM trials (2 WM trials per speaking trial) consisting of a WM load of 3 and 1000 ms ISI was provided within each domain. Participants proceeded to the full task only when both WM responses for the practice item were accurate; the practice trial was repeated if either response was incorrect. Participants were also observed to ensure that their speech production remained continuous when attention was divided. Feedback and reiteration of task instructions were
provided when indicated. Practice items were followed by five speaking trials for each combination of WM load and ISI, resulting in a total of 30 speaking trials and 60 WM trials within each WM domain. See Table 4 for a complete list of experimental conditions, number of speaking trials per condition, and number of WM trials per speaking trial.

2.4.4 Subjective Measures

Following completion of each baseline speaking task (pretest, posttest) and each dual task (dual verbal, dual spatial), participants rated two subjective measures: (1) mental effort expended during task, and (2) speech-related anxiety. Stuttering participants additionally rated their perceived stuttering severity for each task type. Individual 9-point Likert scales were used to measure ratings, with a rating of 1 indicating no mental effort, no stuttering, or no anxiety; and rating of 9 representing significant mental effort, extremely severe stuttering, or extreme anxiety. The use of a nine-point scale was based on similar measures commonly used for subjective assessment of speech-related variables in stuttering research (e.g., Block, Onslow, Packman, Gray, & Dacakis, 2005; Ingham, Warner, Byrd, & Cotton, 2006; Karimi, Jones, O’Brian, & Onslow, 2013; Riley, Riley, & Maguire, 2004). Subjective rating questions were administered via E-Prime, which automatically presented rating items and recorded keyboard responses (1-9).

2.5 Data processing

Dependent variables for WM tasks consisted of accuracy scores (proportion correct responses). Dependent measures for the vigilance task consisted of accuracy scores (proportion correct) and mean RT. Audio files from speaking tasks were orthographically transcribed and coded for disfluencies and speech errors. Specific behaviors coded off-line included blocks,
prolongations, repetition of sounds or syllables, repetition of monosyllabic words, repetition of multisyllabic words, repetition of phrases, interjections/fillers, revisions, sound errors, and word errors. Speech interruptions were subsequently categorized as one of three interruption types: (1) typical disfluencies, which included fillers, revisions, repetitions of phrases, and repetitions of multisyllabic words; (2) atypical disfluencies; which consisted of repetitions of monosyllabic words, repetitions of sounds or syllables, prolongations, blocks, and broken words; and (3) speech errors, which consisted of corrected or uncorrected sound and word errors. This system of categorization was based on typologies and widely accepted distinctions between disfluency types within the stuttering literature (Ambrose & Yairi, 1999; Ratner, Rooney, & MacWhinney, 1996; Vasić & Wijnen, 2005; Wingate, 1962; Yairi & Lewis, 1984; Yairi, 1996; Yaruss, 1997; Yaruss, 1998; however, see Cordes & Ingham, 1995; Einarsdottir & Ingham, 2005 for other perspectives) as well as speech monitoring research (Levelt, 1983; Oomen & Postma, 2001; Postma, 2000).

Transcription and coding were completed by the investigator and a second experienced transcriber who had clinical training as a speech pathologist, but no specialized experience in the area of stuttering. Using speech samples from stuttering speakers, the second transcriber was trained to 100% agreement with the investigator before beginning independent transcription. Praat software (Boersma & Weenink, 2014) was utilized to automate repetitive procedures (preparation of annotation grids, file operations) through the use of scripts and to automatically detect silent intervals (based on a consistent default silence threshold) in speech samples. The majority of transcription was completed based on audio recordings alone; however, supplementary video recordings were consulted as needed to ensure reliable quantification of
disfluencies, particularly for AWS who demonstrated silent blocks (as recommended by Rousseau, Onslow, Packman, & Jones, 2008).

A customized script utilizing the qdap package (Rinker, 2013) in R (R Core Team, 2013) was used to automatically generate frequency counts of all spoken syllables as well as counts of each interruption type for individual speaking trials. Output was used to compute values for four speech-related variables: (1) total number of syllables spoken; (2) total number of typical disfluencies; (3) total number of atypical disfluencies; and (4) total number of speech errors. Syllable counts were used (together with measures of total trial duration) to derive measures of speaking rate (disfluent syllables were excluded from syllable counts, as recommended by Guitar, 2006, pp. 193-194).

Counts of disfluencies and errors were used (with syllable counts) to calculate rates of disfluency and errors as a proportion of total syllables spoken. Thirty-nine speech samples (one per participant) were randomly selected and transcribed a second time by the transcriber who had not originally coded that sample. Inter-rater reliability for transcription and coding of key dependent variables was assessed via Pearson’s product-moment correlations, with results indicating high levels of reliability across outcome measures: \( r(37) = 1.00 \) for syllable counts, \( r(37) = .96 \) for typical disfluencies, \( r(37) = .92 \) for atypical disfluencies, and \( r(37) = .88 \) for errors.

To prepare data for statistical analyses, a coding scheme was developed with six dummy variables representing specific contrasts of interest, based on a priori hypotheses. Contrasts included: (1) Disfluency type, coded with typical disfluencies as the reference level and atypical as the comparison; (2) Task type, coded with baseline tasks as reference level and dual tasks as comparison; (3) Domain, coded with spatial as reference level and verbal as comparison; (4) ISI,
coded with long ISI as reference level and short ISI as comparison; (5) WM Load1, coded with load of 2 as reference level and load of 4 as comparison; and (6) WM Load2, coded with load of 3 as reference level and loads of 2 and 4 as the comparison. Subject level predictors, including measures of nonverbal IQ (TONI-4), vocabulary (EOWPVT-4), working memory span (absolute score), and stuttering severity (SSI raw score, OASES total impact) were centered at their grand mean to reduce collinearity and facilitate interpretation of final models.

2.6 Data analysis

Data were analyzed to examine effects of experimental manipulations and speaker types on five sets of outcome measures: (1) subjective ratings of cognitive effort, speech anxiety, and stuttering severity; (2) secondary task performance; (3) speech rate; (4) speech fluency; and (5) speech errors.

Subjective rating data were analyzed using mixed-design analysis of variance (ANOVA) in SPSS (version 21.0). For the remaining four sets of outcome measures, individual multilevel models were estimated, using the Lme4 package in R (Bates, Maechler, Bolker, & Walker, 2014). Successive models were compared using log-likelihood ratio, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) statistics.
Chapter III. Results

3.1 Preliminary analyses

Comparison of pretest and posttest baseline measures for both typical and atypical forms of disfluency within each group (calculated as a proportion of syllables) indicated no significant difference between the two baseline measures: for typical disfluencies, $t(19) = -1.44, p = .17$ and $t(18) = 0.77, p = .45$ within AWNS and AWS, respectively; for atypical disfluencies, $t(19) = -0.16, p = .86$ and $t(18) = 0.93, p = .36$ within AWNS and AWS, respectively. Disfluency measures for both tasks were therefore included in models as baseline measures without further differentiation.

Performance patterns on the vigilance task were examined to determine whether groups differed in their ability to respond quickly and accurately to on-screen stimuli while they were engaged in a continuous speaking task. Results indicated no between-group differences in accuracy, $t(37) = -0.09, p = .93$, or RT, $t(37) = -1.85, p = .07$, suggesting that the two groups were similar in their general ability to manage dual task demands.

Statistical methods varied based on the outcome measure. Performance on the WM task was scored as either correct (1) or incorrect (0) for each trial and was therefore analyzed using a multilevel logistic regression model. Speech rate was measured via syllable counts, which followed a normal distribution and was therefore analyzed using multilevel models. Preliminary analyses of fluency data indicated that a nonzero inflated negative binomial distribution best fit disfluency counts; therefore, all multilevel generalized linear models for disfluency counts utilized a negative binomial link function (Hardin & Hilbe, 2007). Theoretically, counts of event occurrence are best represented by the negative binomial distribution rather than a normal
distribution (Hardin & Hilbe, 2007) and empirical data revealed that the distribution was truncated at zero, inflated at lower counts, and therefore poorly matched to the normal distribution (AIC = 47800, BIC = 47815, -2LL = -23898). The distribution of disfluencies did appear well matched to the negative binomial distribution and inferential testing confirmed that the negative binomial fit the data reasonably well and provided significantly better fit compared to the normal distribution (AIC = 42980, BIC = 42994, -2LL = -21488, $\chi^2(2) = 2410, p < .0001$).

Results below first consider subjective ratings and secondary task performance, as these data provided evidence relating to the integrity of the task design. These data are followed by analyses of specific dual task effects on three aspects of speech production: rate, fluency, and speech errors.

3.2 Subjective ratings

Subjective ratings of cognitive effort and speech anxiety were analyzed via a 4 x 2 mixed-design analysis of variance (ANOVA), with independent variables consisting of Task (pretest baseline, dual verbal, dual spatial, posttest baseline) and Speaker type (AWS and AWNS). A separate one-way analysis of variance was performed on ratings of stuttering severity, which were only considered for stuttering participants. See Table 5 for a summary of descriptive statistics.

Table 5. Subjective ratings by group and task type. Mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>Pretest Baseline</th>
<th>Dual Spatial</th>
<th>Dual Verbal</th>
<th>Posttest Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive Effort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWNS</td>
<td>4.75 (1.55)</td>
<td>5.85 (2.62)</td>
<td>6.35 (2.11)</td>
<td>3.75 (2.29)</td>
</tr>
<tr>
<td>AWS</td>
<td>4.84 (1.80)</td>
<td>5.79 (1.84)</td>
<td>6.32 (2.03)</td>
<td>3.16 (1.83)</td>
</tr>
<tr>
<td><strong>Speech Anxiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWNS</td>
<td>5.00 (1.89)</td>
<td>3.80 (2.59)</td>
<td>4.35 (2.60)</td>
<td>3.05 (2.63)</td>
</tr>
<tr>
<td>AWS</td>
<td>3.63 (2.19)</td>
<td>3.00 (2.13)</td>
<td>3.11 (2.13)</td>
<td>2.32 (1.49)</td>
</tr>
<tr>
<td><strong>Stuttering Severity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWS</td>
<td>3.32 (1.57)</td>
<td>3.63 (2.06)</td>
<td>3.32 (1.80)</td>
<td>2.68 (1.64)</td>
</tr>
</tbody>
</table>

Note: AWNS=Adults who do not stutter; AWS=Adults who stutter.
Mauchly’s test indicated violation of the sphericity assumption for the Task factor across all three analyses, $\chi^2(5) = 16.38, p = .006$, $\chi^2(5) = 21.13, p = .001$, $\chi^2(5) = 11.87, p = .04$ for cognitive effort, speech anxiety, and stuttering severity, respectively. Degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .81, \varepsilon = .73, \varepsilon = .68$ for the three analyses, respectively) for this factor. All post hoc pairwise comparisons were performed with Bonferroni adjustments.

For subjective ratings of cognitive effort, analysis revealed a statistically significant main effect of Task, $F(2.43, 89.94) = 34.91, p < .001$, partial $\eta^2 = .49$ (Figure 2a). The main effect of Group and Group x Task interaction were not significant, $F(1, 37) = .08, p = .78$, partial $\eta^2 = .002$ and $F(2.43, 89.94) = .40, p = .65$, partial $\eta^2 = .01$, respectively.

Pairwise task comparisons revealed that the two dual tasks resulted in significantly higher ratings of cognitive effort compared to either baseline task (pretest baseline – dual nonverbal, $p = .02$, pretest baseline – dual verbal, $p < .001$, posttest baseline – dual nonverbal, $p < .001$, posttest baseline – dual verbal, $p < .001$) but that effort ratings for the two dual tasks did not differ from each other, $p = .08$. Effort ratings for the posttest baseline were also significantly lower than ratings for the pretest baseline, $p = .001$.

For subjective ratings of speech anxiety, analysis again revealed a statistically significant main effect of Task, $F(2.18, 80.55) = 9.74, p < .001$, partial $\eta^2 = .21$ (Figure 2b). The main effect of Group did not reach significance, $F(1, 37) = .28, p = .61$, partial $\eta^2 = .07$, and the interaction between Group and Task was likewise nonsignificant, $F(2.18, 80.55) = .53, p = .61$, partial $\eta^2 = .01$. 
Pairwise comparisons indicated that the pretest baseline, dual verbal, and dual nonverbal were all characterized by comparable levels of reported anxiety but that reported anxiety was significantly lower on the posttest baseline compared to the pretest baseline task, $p < .001$, the dual nonverbal task, $p = .005$, and the dual verbal task, $p = .002$.

Analysis of stuttering severity ratings (within AWS only) revealed no significant main effect of Task, $F(2.04, 36.75) = .2.7$, $p = .08$, partial $\eta^2 = .13$, indicating no overall change in perceived stuttering severity across the four general task types (Figure 2c).

Figure 2. Subjective measures by task type and group: (a) cognitive effort; (b) speech anxiety; (c) stuttering severity. Note: AWS=Adults who stutter; AWNS=Adults who do not stutter; Baseline 1=Pretest baseline; Baseline 2= posttest baseline.
Speech anxiety ratings by task type and group

(2b)

Stuttering severity ratings (AWS only) by task type

(2c)
3.3 Secondary task performance

Secondary task performance was measured based on response accuracy to each WM trial and was analyzed via multilevel logistic regression modeling. See Table 6 for a summary of descriptive data by group and task conditions. A series of successively complex models was used to examine effects of stuttering status and experimental conditions on secondary task performance (Table 7). Initial results suggested that a model including Speaker (AWNS vs. AWS), Task (baseline vs. dual), WM Domain (spatial vs. verbal), ISI (long vs. short), interactions between these factors, and WM Load (2 vs. 4 and 3 vs. 2 /4) (without interactions) provided the best model fit (Model 9). Examination of model coefficients, however, indicated that the Speaker variable and its interactions were only marginally significant. Two subsequent models were compared to Model 9: (1) A full model (Model 10) without Speaker and its interactions but with all remaining factors and interactions; and (2) a final more parsimonious model (Model 11) that included only Task, Domain, ISI, and Load without interactions. Results indicated that AIC and BIC increased for the full model (Model 10), suggesting that this model was overparamaterized. The difference in log likelihood for the comparison between Model 9 and 11 was only marginally significant (p =.04) and AIC/BIC decreased, suggesting that the more parsimonious model fit the data just as well as the more complex one.

Examination of coefficients for the final model (Model 11, see Table 8) indicates that each experimental manipulation showed an overall effect on secondary task performance (Figures 3a-3c). Accuracy on the secondary task was 18% lower under dual vs. baseline conditions ($\beta = -1.69, SE = .08, t = 20.46, p < .00001$); 72% lower in verbal compared to spatial WM tasks ($\beta = -0.33, SE = .04, t = 7.71, p < .00001$), 113% higher under short compared to long ISI conditions ($\beta = .12, SE = .04, t = 2.92, p = .004$); and 75% lower under
the highest WM load (load of 4) compared to the lowest load (load 2) condition ($\beta = .28, SE = .05, t = 5.36, p < .00001$). However, there were no interactions between these predictors and no differences between AWS and AWNS. These findings did not support Hypothesis 1, which predicted poorer secondary task performance among AWS compared to fluent controls.

Table 6. Proportion of correct responses on secondary task by group and condition. Mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>AWNS</th>
<th>AWS</th>
<th>Dual Spatial</th>
<th>AWNS</th>
<th>AWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>1.00 (0.02)</td>
<td>0.99 (0.02)</td>
<td>S2</td>
<td>0.93 (0.16)</td>
<td>0.91 (0.15)</td>
</tr>
<tr>
<td>S3</td>
<td>0.99 (0.03)</td>
<td>1.00 (0.00)</td>
<td>S3</td>
<td>0.87 (0.22)</td>
<td>0.89 (0.14)</td>
</tr>
<tr>
<td>S4</td>
<td>1.00 (0.02)</td>
<td>0.99 (0.05)</td>
<td>S4</td>
<td>0.87 (0.22)</td>
<td>0.89 (0.13)</td>
</tr>
<tr>
<td>L2</td>
<td>1.00 (0.00)</td>
<td>0.98 (0.04)</td>
<td>L2</td>
<td>0.83 (0.26)</td>
<td>0.90 (0.15)</td>
</tr>
<tr>
<td>L3</td>
<td>1.00 (0.02)</td>
<td>0.98 (0.04)</td>
<td>L3</td>
<td>0.84 (0.21)</td>
<td>0.87 (0.18)</td>
</tr>
<tr>
<td>L4</td>
<td>1.00 (0.00)</td>
<td>0.99 (0.02)</td>
<td>L4</td>
<td>0.83 (0.21)</td>
<td>0.87 (0.20)</td>
</tr>
<tr>
<td>Baseline Verbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>1.00 (0.02)</td>
<td>0.99 (0.03)</td>
<td>S2</td>
<td>0.88 (0.16)</td>
<td>0.85 (0.20)</td>
</tr>
<tr>
<td>S3</td>
<td>0.99 (0.03)</td>
<td>0.98 (0.04)</td>
<td>S3</td>
<td>0.85 (0.16)</td>
<td>0.79 (0.22)</td>
</tr>
<tr>
<td>S4</td>
<td>0.98 (0.04)</td>
<td>0.99 (0.05)</td>
<td>S4</td>
<td>0.76 (0.28)</td>
<td>0.75 (0.26)</td>
</tr>
<tr>
<td>L2</td>
<td>1.00 (0.02)</td>
<td>0.99 (0.02)</td>
<td>L2</td>
<td>0.84 (0.21)</td>
<td>0.85 (0.13)</td>
</tr>
<tr>
<td>L3</td>
<td>1.00 (0.00)</td>
<td>0.99 (0.03)</td>
<td>L3</td>
<td>0.77 (0.24)</td>
<td>0.74 (0.23)</td>
</tr>
<tr>
<td>L4</td>
<td>0.99 (0.04)</td>
<td>0.99 (0.03)</td>
<td>L4</td>
<td>0.76 (0.27)</td>
<td>0.76 (0.21)</td>
</tr>
</tbody>
</table>

Note: AWNS=Adults who do not stutter; AWS=Adults who stutter; S=short ISI (1000 ms), L=long ISI (3000 ms); 2, 3, 4 = WM loads.
Table 7. Model fit of sequential multilevel logistic regression model of secondary task performance.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>-2LL</th>
<th>X²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1: Individual Intercept Only</td>
<td>2</td>
<td>4963</td>
<td>4977</td>
<td>2479</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2: Added Task effect</td>
<td>3</td>
<td>4039</td>
<td>4060</td>
<td>2016</td>
<td>926.12</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>M3: Added Speaker</td>
<td>4</td>
<td>4040</td>
<td>4069</td>
<td>2016</td>
<td>0.18</td>
<td>.67</td>
</tr>
<tr>
<td>M4: Added Task * Speaker</td>
<td>5</td>
<td>4039</td>
<td>4075</td>
<td>2014</td>
<td>3.05</td>
<td>.08</td>
</tr>
<tr>
<td>M5: Added Domain</td>
<td>6</td>
<td>3982</td>
<td>4025</td>
<td>1985</td>
<td>59.72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>M6: Added Task * Domain * Speaker</td>
<td>9</td>
<td>3982</td>
<td>4046</td>
<td>1982</td>
<td>5.96</td>
<td>.11</td>
</tr>
<tr>
<td>M7: Added ISI</td>
<td>10</td>
<td>3975</td>
<td>4047</td>
<td>1977</td>
<td>8.38</td>
<td>.004</td>
</tr>
<tr>
<td>M8: Task * Speaker * Domain * ISI</td>
<td>17</td>
<td>3977</td>
<td>4099</td>
<td>1971</td>
<td>11.91</td>
<td>.10</td>
</tr>
<tr>
<td>M9: Added Load1 &amp; Load2</td>
<td>19</td>
<td>3950</td>
<td>4086</td>
<td>1956</td>
<td>31.24</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>M10: Added full interactions (Task * Domain * ISI * Load1 &amp; Load2)</td>
<td>49</td>
<td>3980</td>
<td>4330</td>
<td>1941</td>
<td>30.61</td>
<td>.43</td>
</tr>
<tr>
<td>M11: Task + Domain + ISI + Load1 + Load2 (removed Speaker and all interactions)</td>
<td>7</td>
<td>3948</td>
<td>3998</td>
<td>1967</td>
<td>21.39</td>
<td>.045</td>
</tr>
</tbody>
</table>

Note: Logistic models were fit using Maximum Likelihood Estimation and included a random intercept for each subject.

Note: Model was fit using Maximum Likelihood Estimation and included a random effect of an individual-level intercept (Var = 1.22). Model was based on 9360 observations across 39 subjects.

Table 8. Results of final model of secondary task performance (M11).

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unstandardized Coefficient</th>
<th>SE</th>
<th>Incidence Ratio</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.72</td>
<td>.04</td>
<td>1.13</td>
<td>2.92</td>
<td>.004</td>
</tr>
<tr>
<td>Task</td>
<td>-1.69</td>
<td>.08</td>
<td>.18</td>
<td>20.46</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Domain</td>
<td>-0.33</td>
<td>.04</td>
<td>.72</td>
<td>7.71</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>ISI</td>
<td>.12</td>
<td>.04</td>
<td>1.13</td>
<td>2.92</td>
<td>.004</td>
</tr>
<tr>
<td>Load1</td>
<td>-.28</td>
<td>.05</td>
<td>.75</td>
<td>5.36</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Load2</td>
<td>.08</td>
<td>.04</td>
<td>1.08</td>
<td>1.75</td>
<td>.08</td>
</tr>
</tbody>
</table>

Note: Model was fit using Maximum Likelihood Estimation and included the random effect of an individual-level intercept (Var = 1.22). Model was based on 9360 observations across 39 subjects.
Figure 3. Secondary task performance by task type and group based on: (a) WM domain; (b) WM load; (c) ISI variations. Note: AWS=Adults who stutter; AWNS=Adults who do not stutter; ISI=inter-stimulus interval. All figures based on raw data.
3.4 Speech rate

Effects of experimental conditions and stuttering status on speaking rate were examined by estimating multilevel regression models using the total number of syllables spoken as the dependent variable. Descriptive data (presented as raw syllable counts and adjusted counts based on 10 second intervals) are presented in Table 9. Models were offset by the total duration of each trial (in seconds) to control for differences in the length of speaking trials in different experimental conditions.

Model comparisons (Table 10) suggested that a model including Speaker (AWS vs. AWNS), Task (baseline vs. dual), and their interaction best described the data. Adding remaining experimental conditions within the dual task (Domain, ISI, Loads) did not improve fit. The intercept for the final model (Model 4, see Table 11) indicated that on average, AWNS produced approximately 140 syllables over a period of 32.14 seconds, which was the average duration of speaking trials. Coefficients suggested that speaker type had an overall effect on total syllables
spoken, with AWS producing approximately 28 fewer syllables than AWNS, controlling for total
duration ($\beta = -28.22, SE = 5.72, t = 4.93, p = .00002$). The overall effect of Task (dual vs.
baseline) on number of syllables spoken was also significant, resulting in a reduction of
approximately 50 syllables in dual compared to baseline task conditions ($\beta = -49.61, SE =
1.03, t = 48.09, p < .00001$). The Speaker x Task interaction further indicated that AWNS
showed a greater reduction in syllable counts under dual task conditions compared to AWS, with
AWS demonstrating a smaller dual task effect by approximately 10 syllables ($\beta = 10.38, SE =
1.48, t = 7.02, p < .00001$). See Figure 4 for a visual representation of these patterns. Overall,
speech rate results were consistent with Hypothesis 2, which predicted a general reduction in
speaking rate under dual task conditions but less pronounced slowing of rate in AWS, relative to
AWNS.

Table 9. Syllable counts (raw and adjusted) by group and condition. Mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>Total syllables</th>
<th>Syllables/10 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AWNS</td>
<td>AWS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest Baseline</td>
<td>192.77 (37.02)</td>
<td>152.00 (37.47)</td>
</tr>
<tr>
<td>Vigilance Task</td>
<td>195.73 (41.28)</td>
<td>159.47 (42.85)</td>
</tr>
<tr>
<td>Dual Spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>95.93 (26.42)</td>
<td>76.02 (26.48)</td>
</tr>
<tr>
<td>S3</td>
<td>95.17 (22.55)</td>
<td>76.53 (22.19)</td>
</tr>
<tr>
<td>S4</td>
<td>92.90 (39.60)</td>
<td>77.02 (34.56)</td>
</tr>
<tr>
<td>L2</td>
<td>93.25 (21.44)</td>
<td>75.31 (22.00)</td>
</tr>
<tr>
<td>L3</td>
<td>91.48 (27.28)</td>
<td>74.60 (26.53)</td>
</tr>
<tr>
<td>L4</td>
<td>92.76 (20.77)</td>
<td>74.73 (21.73)</td>
</tr>
<tr>
<td>Dual Verbal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>94.74 (28.59)</td>
<td>77.39 (27.34)</td>
</tr>
<tr>
<td>S3</td>
<td>96.97 (23.05)</td>
<td>75.64 (22.83)</td>
</tr>
<tr>
<td>S4</td>
<td>91.90 (39.71)</td>
<td>75.21 (36.09)</td>
</tr>
<tr>
<td>L2</td>
<td>96.37 (20.04)</td>
<td>77.12 (24.56)</td>
</tr>
<tr>
<td>L3</td>
<td>93.32 (26.85)</td>
<td>78.53 (30.69)</td>
</tr>
<tr>
<td>L4</td>
<td>91.92 (21.16)</td>
<td>74.59 (24.17)</td>
</tr>
<tr>
<td>Posttest Baseline</td>
<td>194.87 (40.41)</td>
<td>158.44 (44.43)</td>
</tr>
</tbody>
</table>

Note: AWNS=Adults who do not stutter; AWS=Adults who stutter; S=short ISI (1000 ms),
L=long ISI (3000 ms); 2, 3, 4 = WM load.
Table 10. Model fit of sequential multilevel regression models of speech rate.

<table>
<thead>
<tr>
<th>Model Description</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>-2LL</th>
<th>$X^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1: Baseline Model</td>
<td>3</td>
<td>47278</td>
<td>47298</td>
<td>-23636</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2: Added Speaker</td>
<td>4</td>
<td>47270</td>
<td>47296</td>
<td>-23631</td>
<td>10.51</td>
<td>.002</td>
</tr>
<tr>
<td>M3: Added Task</td>
<td>5</td>
<td>44575</td>
<td>44607</td>
<td>-22283</td>
<td>2696.64</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M4: Added Task * Speaker</td>
<td>6</td>
<td>44528</td>
<td>44567</td>
<td>-22258</td>
<td>49.11</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M5: Added Domain * ISI * Loads</td>
<td>28</td>
<td>44549</td>
<td>44731</td>
<td>-22247</td>
<td>22.83</td>
<td>.42</td>
</tr>
</tbody>
</table>

Note: Models were fit using Maximum Likelihood Estimation and included a random intercept for each subject.

Table 11. Results of final model of speech rate (M4).

<table>
<thead>
<tr>
<th>Unstandardized Coefficient</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>140.11</td>
<td>3.99</td>
<td></td>
</tr>
<tr>
<td>Speaker</td>
<td>-28.22</td>
<td>5.72</td>
<td>4.93</td>
</tr>
<tr>
<td>Task</td>
<td>-49.61</td>
<td>1.03</td>
<td>48.09</td>
</tr>
<tr>
<td>Speaker * Task</td>
<td>10.38</td>
<td>1.48</td>
<td>7.02</td>
</tr>
</tbody>
</table>

Note: Model was fit using Maximum Likelihood Estimation and included the random effect of an individual-level intercept (Var = 297.7). Residual variance was equal to 486.6, and model was based on 4914 observations across 39 subjects. Outcome variable represented number of syllables spoken offset by the duration of each speaking trial, which varied by condition.

Figure 4. Speech rate (syllables/second) by task type and group. Note: AWS=Adults who stutter; AWNS=Adults who do not stutter. Figure based on raw data.
3.5 Fluency

3.5.1 Between-subject effects

Multilevel generalized linear models with a negative binomial link function (Hardin & Hilbe, 2007) were used to test hypotheses related to fluency effects. The outcome variable for estimated models was the number of disfluencies produced during each speaking trial. Descriptive data for typical and atypical disfluencies are provided in Tables 12 and 13, respectively. To control for unequal durations of speaking trials across different experimental conditions, as well as for varying speaking rates, models were offset by the total number of syllables produced during each speaking trial. The dependent variable therefore represented the ratio of disfluencies to fluent syllables. Similar results were obtained when models were offset by the total duration (in seconds) of each speaking trial. All speaking trials during dual-task conditions were included, regardless of accuracy on secondary task. Identical analyses in which erroneous responses were excluded (17% of total trials for AWNS, 16% total trials for AWS) produced the same results. A series of successive models (see Table 14) was estimated and compared using AIC, BIC, and log-likelihood statistics.

Results indicated that a model including Speaker (AWNS vs. AWS), Disfluency (typical vs. atypical), Task (baseline vs. dual) and two two-way interactions (Speaker x Disfluency and Disfluency x Task) provided the best fit (Model 5). Addition of Trial number (to test order effects) marginally improved fit (Model 6); however, AIC/BIC was similar or higher, Pseudo-R² was small (.00009) and coefficients were barely significant (only the Disfluency x Trial x Task interaction had a p-value <.05 and indicated a very slight increase in atypical disfluencies under dual task conditions as trial number increased). More complex models that included three-way interactions (Model 7) and remaining experimental conditions (Domain, ISI, Loads; see Model
8) were not significantly better than Model 5, and displayed poorer fit based on similar or larger AIC/BIC values. These nonsignificant effects for manipulations of WM domain, ISI, and WM load were inconsistent with Hypothesis 3c, which predicted greater dual task effect for verbal vs. nonverbal secondary tasks, as well as Hypothesis 3d, which anticipated increased dual task effect with shorter ISIs and increases in WM load.

Table 12. Atypical disfluencies (raw counts and as a proportion of syllables) by group and condition. Mean (SD).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Counts</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AWNS</td>
<td>AWS</td>
</tr>
<tr>
<td>Pretest Baseline</td>
<td>2.03 (2.34)</td>
<td>9.51 (6.80)</td>
</tr>
<tr>
<td>Vigilance Task</td>
<td>1.52 (2.10)</td>
<td>6.37 (5.33)</td>
</tr>
<tr>
<td>Dual Spatial S2</td>
<td>0.81 (1.26)</td>
<td>3.09 (2.61)</td>
</tr>
<tr>
<td>S3</td>
<td>0.68 (1.14)</td>
<td>2.89 (2.47)</td>
</tr>
<tr>
<td>S4</td>
<td>0.73 (1.13)</td>
<td>2.96 (2.73)</td>
</tr>
<tr>
<td>L2</td>
<td>0.83 (1.25)</td>
<td>2.79 (2.54)</td>
</tr>
<tr>
<td>L3</td>
<td>0.94 (1.30)</td>
<td>2.76 (2.73)</td>
</tr>
<tr>
<td>L4</td>
<td>0.73 (1.00)</td>
<td>2.92 (2.39)</td>
</tr>
<tr>
<td>Dual Verbal S2</td>
<td>0.68 (1.06)</td>
<td>3.13 (2.51)</td>
</tr>
<tr>
<td>S3</td>
<td>0.69 (1.03)</td>
<td>3.10 (2.40)</td>
</tr>
<tr>
<td>S4</td>
<td>0.71 (0.96)</td>
<td>2.70 (2.50)</td>
</tr>
<tr>
<td>L2</td>
<td>0.76 (1.05)</td>
<td>3.08 (2.51)</td>
</tr>
<tr>
<td>L3</td>
<td>0.76 (1.28)</td>
<td>2.63 (2.31)</td>
</tr>
<tr>
<td>L4</td>
<td>0.82 (1.14)</td>
<td>2.61 (2.28)</td>
</tr>
<tr>
<td>Posttest Baseline</td>
<td>2.02 (2.27)</td>
<td>7.05 (5.26)</td>
</tr>
</tbody>
</table>

Note: AWNS=Adults who do not stutter; AWS=Adults who stutter; S=short ISI (1000 ms), L=long ISI (3000 ms); 2, 3, 4 = WM load.
Table 13. Typical disfluencies (raw counts and as a proportion of syllables) by group and condition. Mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>Counts</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AWNS</td>
<td>AWS</td>
</tr>
<tr>
<td>Pretest Baseline</td>
<td>6.72 (4.55)</td>
<td>6.98 (4.61)</td>
</tr>
<tr>
<td>Vigilance Task</td>
<td>7.17 (4.50)</td>
<td>6.46 (3.61)</td>
</tr>
<tr>
<td>Dual Spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>3.67 (2.39)</td>
<td>3.30 (2.41)</td>
</tr>
<tr>
<td>S3</td>
<td>3.66 (2.27)</td>
<td>3.51 (2.30)</td>
</tr>
<tr>
<td>S4</td>
<td>4.01 (2.89)</td>
<td>3.26 (2.52)</td>
</tr>
<tr>
<td>L2</td>
<td>3.48 (2.13)</td>
<td>3.30 (1.99)</td>
</tr>
<tr>
<td>L3</td>
<td>3.51 (2.17)</td>
<td>3.25 (2.37)</td>
</tr>
<tr>
<td>L4</td>
<td>3.60 (2.23)</td>
<td>3.43 (2.19)</td>
</tr>
<tr>
<td>Dual Verbal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>3.59 (2.60)</td>
<td>3.48 (2.46)</td>
</tr>
<tr>
<td>S3</td>
<td>3.60 (2.47)</td>
<td>3.66 (2.59)</td>
</tr>
<tr>
<td>S4</td>
<td>3.62 (2.60)</td>
<td>3.42 (2.62)</td>
</tr>
<tr>
<td>L2</td>
<td>3.81 (2.62)</td>
<td>3.65 (2.54)</td>
</tr>
<tr>
<td>L3</td>
<td>3.84 (2.76)</td>
<td>3.37 (2.20)</td>
</tr>
<tr>
<td>L4</td>
<td>3.57 (2.39)</td>
<td>3.58 (2.15)</td>
</tr>
<tr>
<td>Posttest Baseline</td>
<td>7.98 (4.07)</td>
<td>6.79 (3.73)</td>
</tr>
</tbody>
</table>

Note: Note: AWNS=Adults who do not stutter; AWS=Adults who stutter; S=short ISI (1000 ms), L=long ISI (3000 ms); 2, 3, 4 = WM load.

Examination of the coefficients for the final model (Model 5, see Table 15) indicated that controlling for number of syllables (speech rate), atypical disfluencies were significantly less frequent than typical disfluencies, overall. The rate of atypical disfluencies was 28% of the rate of typical disfluencies ($\beta = -1.26, SE = .06, t = 20.96, p < .00001$); thus, there were 28 atypical disfluencies produced for every 100 typical disfluencies. The interaction between Disfluency and Speaker was also significant, with AWS producing approximately four times more atypical disfluencies compared to AWNS ($\beta = 1.43, SE = .03, t = 41.83, p < .00001$); thus, 420 atypical disfluencies were produced by AWS for every 100 atypical disfluencies produced by AWNS. Finally, the interaction between Disfluency and Task indicated that a
significantly lower frequency of atypical disfluencies occurred during dual compared to baseline tasks. This was true regardless of speaker type (AWS vs. AWNS) and regardless of experimental condition (Domain, ISI, Load) within the dual task. The rate of atypical disfluencies under dual task conditions was 70% of the rate of atypical disfluencies under non-dual task \((\beta = -0.35, SE = .06, t = 6.04, p < .00001)\); thus, for every 100 atypical disfluencies produced during baseline conditions, 70 were produced during dual task conditions. This Disfluency x Task interaction supported the primary hypothesis of the study, Hypothesis 3a, which predicted a reduction in stutter-like disfluencies under dual task conditions, although results indicate that this effect was not specific to AWS (Figure 5). Typical forms of disfluency (coded as the reference level for Disfluency predictor) did not show any changes as a result of experimental manipulations (Figure 6). Thus, Hypothesis 3b, which predicted that typical disfluencies would increase under dual compared to baseline conditions for both groups of speakers, was not supported.

Table 13. Model fit of sequential multilevel negative binomial regression models of disfluencies.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>-2LL</th>
<th>(X^2)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1: Individual Intercept Only</td>
<td>3</td>
<td>40916</td>
<td>40937</td>
<td>-20455</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>M2: Added Speaker</td>
<td>4</td>
<td>40904</td>
<td>40933</td>
<td>-20448</td>
<td>13.88</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>M3: Added Disfluency</td>
<td>5</td>
<td>39181</td>
<td>39217</td>
<td>-19586</td>
<td>1724.34</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M4: Added Task</td>
<td>6</td>
<td>39168</td>
<td>39211</td>
<td>-19578</td>
<td>15.15</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>M5: Added 2-way interactions</td>
<td>8</td>
<td>37435</td>
<td>37493</td>
<td>-18710</td>
<td>1737.10</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M6: Added Trial and its interactions</td>
<td>14</td>
<td>37429</td>
<td>37530</td>
<td>-18701</td>
<td>37401</td>
<td>.007</td>
</tr>
<tr>
<td>M7: Added 3-way interactions (removed Trial)</td>
<td>10</td>
<td>37438</td>
<td>37510</td>
<td>-18709</td>
<td>1.48</td>
<td>.48</td>
</tr>
<tr>
<td>M8: Added Domain * ISI * Loads</td>
<td>30</td>
<td>37466</td>
<td>37681</td>
<td>-18703</td>
<td>12.08a</td>
<td>.91</td>
</tr>
</tbody>
</table>

Note: Models were fit using Maximum Likelihood Estimation (Laplace Approximation), utilized a negative binomial logarithmic link function \((\theta = 6.07)\), and included the random effect of an individual-level intercept \((\text{Var} = .27)\).

\(a\) compares M5 to M7.
Table 14. Results of final model of disfluencies (M5).

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unstandardized Coefficient</th>
<th>SE</th>
<th>Incidence Ratio</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disfluency</td>
<td>-1.26</td>
<td>.06</td>
<td>.28</td>
<td>20.96</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Speaker</td>
<td>0.21</td>
<td>.16</td>
<td>1.23</td>
<td>1.30</td>
<td>.19</td>
</tr>
<tr>
<td>Task</td>
<td>0.03</td>
<td>.04</td>
<td>1.04</td>
<td>.90</td>
<td>.37</td>
</tr>
<tr>
<td>Disfluency * Speaker</td>
<td>1.43</td>
<td>.03</td>
<td>4.20</td>
<td>41.83</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Disfluency * Task</td>
<td>-0.35</td>
<td>.06</td>
<td>.70</td>
<td>6.04</td>
<td>&lt;.00001</td>
</tr>
</tbody>
</table>

Note: Model was fit using Maximum Likelihood Estimation (Laplace Approximation) and utilized a negative binomial logarithmic link function ($\theta = 6.07$). It included the random effect of an individual-level intercept (Var = .27). Residual variance was equal to 1.11, and the model was based on 9826 observations across 39 subjects.

Figure 5. Proportion of disfluent syllables (atypical disfluencies) by task type and group. Note: AWS=Adults who stutter; AWNS=Adults who do not stutter. Figure based on raw data.
3.5.2 Subject level variables

Although between-subject differences in disfluency rates are accounted for by the random intercept included in the models above, additional analyses were conducted to assess the degree to which these differences could be explained by specific subject-level factors. Subject-level factors included measures of expressive vocabulary, verbal WM (OSpan), and measures of stuttering severity and impact, based on the potential influence of these specific factors on experimental effects. In addition, cross-level interactions effects were tested to see if subject-level differences differentially influenced performance on experimental tasks.
Cognitive Variables. An estimated series of models is described in Table 16. Examination of fit statistics indicated that a model including Vocabulary, OSpan, Disfluency, Task, and Speaker (Model 12) displayed best fit. However, coefficients suggested that interactions with the dual task effect were not significant and a more parsimonious model (Model 13) that removed all interactions with Task other than Disfluency x Task and Speaker x Task (which were included in the baseline M5 model) did not explain significantly less variance and displayed better fit.

Table 15. Model fit of sequential multilevel negative binomial regression model of disfluencies including subject-level variables.

<table>
<thead>
<tr>
<th>Model Description</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>-2LL</th>
<th>(X^2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5: Baseline Model</td>
<td>8</td>
<td>36371</td>
<td>36449</td>
<td>-18188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M9: Added Vocabulary and OSpan</td>
<td>10</td>
<td>36393</td>
<td>36465</td>
<td>-18187</td>
<td>2.07</td>
<td>.36</td>
</tr>
<tr>
<td>M10: Added interactions by Disfluency</td>
<td>12</td>
<td>36151</td>
<td>36237</td>
<td>-18063</td>
<td>246.55</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M11: Added interactions by Task</td>
<td>16</td>
<td>36156</td>
<td>36271</td>
<td>-18062</td>
<td>2.27</td>
<td>.69</td>
</tr>
<tr>
<td>M12: Added interactions by Speaker</td>
<td>26</td>
<td>35946</td>
<td>36133</td>
<td>-17947</td>
<td>229.92</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M13: Removed interactions by Task</td>
<td>16</td>
<td>35931</td>
<td>36046</td>
<td>-17950</td>
<td>4.87</td>
<td>.90</td>
</tr>
</tbody>
</table>

Note: Models were fit using Maximum Likelihood Estimation (Laplace Approximation) and utilized a negative binomial logarithmic link.

\(^a\) Baseline model was identical to Model 5, Table 15 and included Disfluency, Speaker, Task and 2-way interactions but was re-estimated using 38 subjects (9576 observations) who completed subject-level measures (one subject was excluded from this analysis due to missing data for cognitive covariates).

Variance components for the final model (M13, Table 17) indicated that Vocabulary, OSpan and interactions explained 10% of the between subject variance that was unaccounted for in M5 (variance in M5 = .2805, variance in Model 13 = .2512, Pseudo-R\(^2\) = .10). Coefficients suggested that for every one-unit increase in expressive vocabulary (M=103.42, SD=11.05), total typical disfluencies decreased 3% among AWNS but remained the same among AWS, and atypical disfluencies decreased 1% in both speaker types. Similar results were obtained for
OSpan (M=44.29, SD=16.6), such that for every one-unit increase in OSpan, typical disfluencies decreased 1% among AWNS but increased 1% among AWS, and atypical disfluencies decreased 1% and 2% in AWNS and AWS, respectively.

Table 16. Results of final model of disfluencies including subject-level variables (M13).

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unstandardized Coefficient</th>
<th>SE</th>
<th>Incidence Ratio</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.30</td>
<td>0.11</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disfluency</td>
<td>-1.46</td>
<td>0.04</td>
<td>0.23</td>
<td>39.83</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Speaker</td>
<td>0.15</td>
<td>0.16</td>
<td>1.16</td>
<td>0.91</td>
<td>0.36</td>
</tr>
<tr>
<td>Task</td>
<td>0.01</td>
<td>0.02</td>
<td>1.01</td>
<td>0.76</td>
<td>0.45</td>
</tr>
<tr>
<td>Disfluency * Speaker</td>
<td>1.44</td>
<td>0.04</td>
<td>4.20</td>
<td>41.19</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Disfluency * Task</td>
<td>-0.17</td>
<td>0.03</td>
<td>0.84</td>
<td>6.00</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.97</td>
<td>2.35</td>
<td>0.02</td>
</tr>
<tr>
<td>OSpan</td>
<td>-0.006</td>
<td>0.007</td>
<td>0.99</td>
<td>0.83</td>
<td>0.40</td>
</tr>
<tr>
<td>Disfluency * Vocabulary</td>
<td>0.02</td>
<td>0.003</td>
<td>1.02</td>
<td>7.98</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Speaker * Vocabulary</td>
<td>0.03</td>
<td>0.02</td>
<td>1.03</td>
<td>1.89</td>
<td>0.06</td>
</tr>
<tr>
<td>Disfluency * OSpan</td>
<td>0.004</td>
<td>0.002</td>
<td>1.00</td>
<td>2.39</td>
<td>0.02</td>
</tr>
<tr>
<td>Speaker * OSpan</td>
<td>0.02</td>
<td>0.01</td>
<td>1.02</td>
<td>1.73</td>
<td>0.08</td>
</tr>
<tr>
<td>Disfluency * Speaker * Vocabulary</td>
<td>-0.04</td>
<td>0.003</td>
<td>0.97</td>
<td>10.18</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>Disfluency * Speaker * OSpan</td>
<td>-0.03</td>
<td>0.002</td>
<td>0.97</td>
<td>12.51</td>
<td>&lt;.00001</td>
</tr>
</tbody>
</table>

Note: Model was fit using Maximum Likelihood Estimation (Laplace Approximation) and utilized a negative binomial logarithmic link function ($\theta = 6.71$). It included the random effect of an individual-level intercept (Var = .25). Residual variance was equal to 1.10, and the model was based on 9576 observations across 38 subjects (one subject was excluded from this analysis due to missing data for cognitive covariates). Vocabulary and OSpan were grand-mean centered.

Stuttering Severity. An estimated series of models is described in Table 18. Examination of fit statistics indicated that Model 15, which included OASES (total impact score), SSI (raw score),
Disfluency, Task, and no cross-level interactions displayed best fit. Although a model including Task and interactions between dual task conditions (Domain, ISI, Loads) increased model fit significantly (Model 17), it displayed poorer AIC and BIC, coefficients were not significant or marginally significant, and Pseudo-$R^2$ for the additional within-subject variance explained by these interactions was low (M15 variance = 1.077, M17 variance = 1.0685, $R^2 = .008$).

Table 17. Model fit of sequential multilevel negative binomial regression models of disfluencies including subject-level severity for AWS only.

<table>
<thead>
<tr>
<th>Model</th>
<th>$df$</th>
<th>$AIC$</th>
<th>$BIC$</th>
<th>-2LL</th>
<th>$X^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5: Baseline Model$^a$</td>
<td>6</td>
<td>19792</td>
<td>19830</td>
<td>-9890</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M14: Added OASES and SSI</td>
<td>8</td>
<td>19794</td>
<td>19845</td>
<td>-9890</td>
<td>1056.03</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M15: Added interactions by Disfluency</td>
<td>10</td>
<td>19103</td>
<td>19167</td>
<td>-9541</td>
<td>695.12</td>
<td>&lt;.00001</td>
</tr>
<tr>
<td>M16: Added interactions by Task</td>
<td>14</td>
<td>19104</td>
<td>19194</td>
<td>-9537</td>
<td>7.02</td>
<td>.14</td>
</tr>
<tr>
<td>M17: Added Domain * ISI * Loads</td>
<td>80</td>
<td>19108</td>
<td>19622</td>
<td>-9474</td>
<td>134.97</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note: Models were fit using Maximum Likelihood Estimation (Laplace Approximation) and utilized a negative binomial logarithmic link.

$^a$ Baseline model was parallel to Model 5, Table 15 and included Disfluency, Task, and their interaction. However, it was re-estimated using only the 18 stuttering participants (4536 observations) who completed these measures, and therefore did not include Speaker factor.

Variance components for the final model (Model 15, Table 19) indicated that OASES, SSI and interactions explained an additional 15% of the between subject variance that was unaccounted for in M5 (variance in M5 = .2483, variance in Model 15 = .2122, Pseudo-$R^2 = .15$). Coefficients suggested that controlling for SSI, OASES had only a marginal relationship with disfluencies. In contrast, SSI had significant predictive value, such that for every one-unit increase in SSI, typical disfluencies decreased by 3% and atypical disfluencies increased by 7%.
Table 18. Results of final model of disfluencies including subject-level stuttering severity for AWS only (M15).

<table>
<thead>
<tr>
<th>Unstandardized Coefficient</th>
<th>SE</th>
<th>Incidence Ratio</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.77</td>
<td>.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disfluency</td>
<td>.29</td>
<td>.11</td>
<td>1.34</td>
<td>2.56</td>
</tr>
<tr>
<td>Task</td>
<td>.01</td>
<td>.03</td>
<td>1.01</td>
<td>.36</td>
</tr>
<tr>
<td>OASES</td>
<td>.23</td>
<td>.21</td>
<td>1.25</td>
<td>1.06</td>
</tr>
<tr>
<td>SSI</td>
<td>-.03</td>
<td>.01</td>
<td>0.97</td>
<td>2.36</td>
</tr>
<tr>
<td>Disfluency * Task</td>
<td>-.19</td>
<td>.04</td>
<td>.83</td>
<td>4.65</td>
</tr>
<tr>
<td>Disfluency * OASES</td>
<td>-.10</td>
<td>.04</td>
<td>0.91</td>
<td>2.24</td>
</tr>
<tr>
<td>Disfluency * SSI</td>
<td>.07</td>
<td>.003</td>
<td>1.08</td>
<td>26.37</td>
</tr>
</tbody>
</table>

Note: Model was fit using Maximum Likelihood Estimation (Laplace Approximation) and utilized a negative binomial logarithmic link function ($\theta = 5.88$). It included the random effect of an individual-level intercept (Var = .21). Residual variance was equal to 1.08, and the model was based on 4536 observations across 18 stuttering participants. OASES and SSI were grand-mean centered.

### 3.6 Speech errors

Descriptive statistics for speech errors (raw counts and as a proportion of syllables) are presented in Table 20. A series of multilevel negative binomial regression models (Table 21) was estimated to examine the effects of experimental conditions and stuttering status on speech errors. As in preceding models for disfluency counts, speech error models were offset by the total number of syllables in each speaking trial to control for unequal durations of speaking trials across different experimental conditions and varying speaking rates. The dependent variable therefore represented the ratio of speech errors to fluent syllables.
Results revealed that no model was a significant improvement over a baseline model that consisted of a random intercept only, indicating that speech errors were not influenced by either experimental conditions or stuttering status (Figure 7). Results of the baseline intercept-only model suggested that individual differences explained 17% of the variance in speech errors (Pseudo-$R^2 = .17$). Overall, findings for error counts are consistent with Hypothesis 4, which predicted no change in overt speech error rates under dual compared to baseline conditions.

Table 19. Errors (raw counts and proportion of syllables) by group and condition. Mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>Error counts</th>
<th></th>
<th>Errors/syllables</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AWNS</td>
<td>AWS</td>
<td>AWNS</td>
<td>AWS</td>
</tr>
<tr>
<td>Pretest Baseline</td>
<td>1.47 (1.71)</td>
<td>1.05 (1.64)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>Vigilance Task</td>
<td>1.33 (1.27)</td>
<td>0.89 (1.30)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>Dual Spatial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.74 (1.07)</td>
<td>0.43 (0.77)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>S3</td>
<td>0.61 (0.89)</td>
<td>0.59 (0.82)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>S4</td>
<td>0.66 (0.80)</td>
<td>0.43 (0.70)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>L2</td>
<td>0.65 (0.92)</td>
<td>0.57 (0.84)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>L3</td>
<td>0.60 (0.88)</td>
<td>0.45 (0.78)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>L4</td>
<td>0.63 (0.88)</td>
<td>0.42 (0.67)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>Dual Verbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.67 (0.85)</td>
<td>0.48 (0.75)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>S3</td>
<td>0.67 (1.01)</td>
<td>0.45 (0.79)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>S4</td>
<td>0.76 (1.09)</td>
<td>0.53 (0.83)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>L2</td>
<td>0.62 (0.93)</td>
<td>0.43 (0.74)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>L3</td>
<td>0.64 (0.92)</td>
<td>0.48 (0.75)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>L4</td>
<td>0.65 (0.97)</td>
<td>0.53 (0.78)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>Posttest Baseline</td>
<td>1.53 (1.27)</td>
<td>0.96 (1.24)</td>
<td>0.01 (0.01)</td>
<td>0.01 (0.01)</td>
</tr>
</tbody>
</table>

Note: AWNS=Adults who do not stutter; AWS=Adults who stutter; S=short ISI (1000 ms), L=long ISI (3000 ms); 2, 3, 4 = WM load.
Table 20. Model fit of sequential multilevel negative binomial regression models of speech errors.

<table>
<thead>
<tr>
<th>Model Description</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>-2LL</th>
<th>$\chi^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1: Baseline Model</td>
<td>3</td>
<td>10089</td>
<td>10109</td>
<td>-5042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2: Added Speaker</td>
<td>4</td>
<td>10090</td>
<td>10116</td>
<td>-5041</td>
<td>1.34</td>
<td>.25</td>
</tr>
<tr>
<td>M3: Added Task</td>
<td>5</td>
<td>10092</td>
<td>10124</td>
<td>-5041</td>
<td>.15</td>
<td>.70</td>
</tr>
<tr>
<td>M4: Added Speaker * Task</td>
<td>6</td>
<td>10093</td>
<td>10132</td>
<td>-5041</td>
<td>.60</td>
<td>.44</td>
</tr>
<tr>
<td>M5: Added Domain * ISI * Loads</td>
<td>28</td>
<td>10114</td>
<td>10297</td>
<td>-5029</td>
<td>22.78</td>
<td>.41</td>
</tr>
</tbody>
</table>

Note: Models were fit using Maximum Likelihood Estimation, were offset by number of syllables spoken, and included a random intercept for each subject.

Figure 7. Proportion of syllables containing errors by task type and group. Note: AWS=Adults who stutter; AWNS=Adults who do not stutter. Figure based on raw data.
Chapter IV. Discussion

The key finding emerging from this study was that dual task conditions had a facilitative effect on the fluency of speech production. This effect was specific to atypical forms of disfluency but was not specific to speakers with stuttering disorders, and was not influenced by specific manipulations within the secondary task (WM domain, WM load, and ISI). The observed fluency benefit was accompanied by a reduction in speech rate, which was greater in fluent compared to stuttering speakers, but there was no corresponding change in the frequency of overt speech errors. Secondary task performance was compromised in the dual compared to baseline condition and was affected by manipulations of WM domain, WM load, and ISI; however, these effects did not differ between the two speaker groups. Results for individual sets of analyses are discussed below, followed by a general discussion of the broader theoretical implications of these findings.

4.1 Perceived dual task effects

Overall, subjective measures of cognitive effort, speech anxiety, and stuttering severity for the four main speaking task conditions (pretest baseline, dual verbal, dual spatial, and posttest baseline) confirmed that key aspects of the task design functioned as expected. Cognitive effort ratings showed that participants perceived dual tasks within both WM domains (verbal and spatial) as more cognitively demanding than the baseline speaking tasks. This finding was anticipated, given the considerable mental effort required for dual tasking, particularly when tasks involve ongoing processing over an extended period of time, as in the present study. Ratings for the two dual tasks (dual verbal and dual spatial) were comparable; however, participants perceived posttest speaking as less demanding than the identical pretest task. Similar
results were obtained by Metten and colleagues (Metten et al., 2011, Experiment 1) in a dual task study with stuttering adults and were interpreted as a practice effect. This explanation can apply to the present study as well, although based on this logic, a corresponding change in speech fluency should have occurred on the final task, and this was not observed. A simpler but probable explanation is that the posttest reduction in perceived mental effort reflected participants’ increasing familiarity and comfort with experimental procedures, in particular, the experience of producing 60-second recorded monologues, which may have initially imposed a certain amount of discomfort for some speakers.

Speech anxiety ratings were similarly lower on the final task (posttest speech) compared to all three previous ones (pretest speech baseline, dual verbal, dual spatial). As suggested above, this result likely indicated increasing familiarity with experimental procedures and associated reduction of general anxiety or unease.

Subjective ratings of fluency were comparable across tasks, indicating that although objective reductions in disfluency did occur during the dual task conditions, these effects were not perceptible to stuttering participants. The lack of a perceived effect in stuttering was somewhat surprising, particularly in light of recent data showing strong correlations between subjective ratings of stuttering severity by participants and objective counts of stuttering frequency performed by clinicians (Karimi et al., 2013). One reason for the discrepancy in obtained results could be that participants’ ability to reliably estimate stuttering severity was compromised due to ongoing suppression of attentional resources by the secondary task. Poor awareness of fluency changes may reflect participants’ diminished ability to monitor their speech or focus on speech-related skills during dual task performance, which was precisely the intent of these conditions. Another consideration is that the objective fluency measure essentially
consisted of frequency counts and did not take into account other aspects of stuttering severity, such as the presence of secondary symptoms, degree of facial tension, duration of disfluencies, and experience of covert stuttering. It is possible that these features are more noticeable to individuals who stutter or are more easily perceived, especially when full attentional resources are not available.

4.2 Secondary task performance patterns

Analyses of secondary task performance revealed a main effect of task type and of each manipulation within the WM task (WM domain, WM load, and ISI), indicating that the WM task generally worked as intended. Overall performance on the WM task was less accurate for: (1) dual compared to baseline conditions, (2) verbal compared to spatial WM, (3) WM load of 4 compared to load of 2, and (4) long ISI (3000 ms) compared to short ISI (1000 ms).

Poorer secondary task performance under dual compared to baseline conditions is an expected result of the dual task paradigm (Bourke, 1996; Pashler, 1994), particularly when concurrent tasks draw upon similar cognitive resources (Leclercq, 2002) or cannot be performed automatically (Pashler, 1999; Poldrack et al., 2005). Previous dual task studies combining speech production with a wide variety of secondary tasks (e.g., color recognition, rhyme/category judgments, random finger tapping, mental calculation) likewise showed performance decrement on these tasks (e.g., Bosshardt, 1999; Declerck & Kormos, 2012; De Nil & Bosshardt, 2000; Smits-Bandstra & De Nil, 2009), suggesting that producing speech is never completely automatic and requires some degree of ongoing access to central resources in all speakers.

Within the secondary task, WM load showed the clearest effect on performance, with the largest load (4 items) resulting in significantly poorer accuracy than the smallest load (2 items).
This is consistent with findings of the original study from which the task was derived (Salthouse et al., 1991) and general data on set size and updating within the WM literature (Engle, 2010; Ikowska & Engle, 2010; Oberauer, 2009; Nash, Unsworth & Engle, 2007).

The effect of domain was expected to manifest as greater interference for the verbal compared to the spatial WM task under dual task conditions. Instead, results showed a main effect of domain, with verbal WM tasks being more difficult than spatial tasks overall, but no Domain x Task interaction. Salthouse and colleagues (1991) showed a strong correlation between the original versions of these verbal and spatial tasks; however, it is possible that subtle modifications in the structure of the task, as adapted for the present study, influenced probabilities of correct responses somewhat differently for each domain. Within the verbal task, consistent updating of the numeric total was required in order to arrive at the correct response. By contrast, because the target stimulus was presented in a 2x2 grid during the spatial task, it was possible to narrow down the correct outcome to 2/4 options based on any one directional arrow within a sequence of arrows, and to a single correct response based on the subsequent arrow. It is not clear, however, whether this type of strategy necessarily contributed to performance differences based on domain, particularly since there was no evidence of a Domain x Load interaction.

ISI affected performance but not in the anticipated direction. Brief ISI was expected to impose more continuous demands on central resources and result in increased competition between concurrent tasks (Pashler, 1994). Results showed the opposite pattern, such that longer ISI resulted in poorer recall. The most logical interpretation of this finding is that the ISI condition essentially served as an additional, unintended manipulation of WM, with longer ISI rendering memory representations more susceptible to decay or interference, as described by...
many different theoretical models of WM (Baddeley, 2003; Cowan, 2005, 2008; Engle, Tuholski, Laughlin, & Conway, 1999; Oberauer, 2009).

Dual task manipulations were expected to interact with task type (dual vs. baseline) and more significantly affect performance under dual conditions, but they did not. It is possible that interactions were not detected due to the relative ease of the secondary task, which was designed to engage WM to an extent that was sufficient to induce measurable effects on speech production but not enough to significantly reduce WM performance. As a result, some participants continued to achieve very high levels of accuracy on the WM task under dual task conditions, while others clearly had more difficulty managing simultaneous task demands. This variability may have obscured interactions between task type and WM conditions. The emphasis on maintaining high levels of performance during dual tasking was based on the consideration that an excessively difficult distractor task might have lead to arbitrary prioritization of one task or the other rather than division of attentional resources between tasks, and would have made results difficult to interpret. Ultimately, analyses revealed identical fluency effects when secondary responses were accurate or inaccurate, suggesting that these effects were not necessarily contingent upon a correct response to each WM trial; however, this outcome could not be assumed a priori.

Results showed no overall differences between groups and no interactions with the group factor. Thus, Hypothesis 1, which predicted greater dual task decrement for AWS compared to controls, was not supported. This prediction was based on evidence that production of speech is less automatized for AWS than for fluent speakers (Saltuklaroglu et al., 2009; Smits-Bandstra & De Nil, 2009) as well as several dual task studies showing increased disruption of secondary task performance in AWS compared to controls during dual task conditions (Jones et al., 2012; De
Nil & Bosshardt, 2000; Sussman, 1982). Results of the present study may have differed due to the relative ease of the secondary task, as previously noted. It is possible that task demands did not strain WM resources enough to expose group differences, which are often quite subtle (e.g., note discussion related to Jones et al., 2012 study, below). Nevertheless, the current finding that speaker groups performed comparably on the secondary task suggested that AWS and controls were prioritizing tasks similarly and central resources were being suppressed in a similar manner. Ultimately, this outcome made interpretation of dual task effects on fluency more straightforward.

4.3 Dual task effects on speaking rate

Speech rate results were consistent with Hypothesis 2 and indicated that: (1) AWS produced speech at a slower rate than controls; (2) all speakers showed a reduction in speech rate during dual compared to baseline tasks; and (3) compared to controls, AWS showed a smaller reduction in speech rate under dual task conditions.

Data related to rate changes in dual task conditions provided a context for understanding fluency effects and addressed an important limitation of many previous studies on this topic, which omit these details (Bosshardt, 1999; Kamhi & McOsker, 1982; De Nil & Bosshardt, 2000; Oomen & Postma, 2001b; Postma et al., 1991; Vasić & Wijnen, 2005). The cause for rate reduction under dual task conditions is not entirely clear but may be related to dual task effects on linguistic productivity (e.g., see Bosshardt, 2006; Jou & Harris, 1992) and seems to occur specifically when secondary tasks clearly engage central resources. Consistent with this interpretation, an early study combining speech production with external forms of distraction (visual and auditory noise) showed no resulting changes in speech rate (Mallard & Webb, 1980);
however, the production of random finger tapping sequences while speaking (designed to disrupt central executive functioning) significantly reduced syllable counts (Oomen & Postma, 2002).

The smaller reduction in speech rate for stuttering speakers was an interesting finding that has not been reported in prior studies, and likely reflects combined effects of a habitually slower speech rate (due to excessive disfluency) and enhanced fluency under dual task conditions. Based on available data, it is not possible to determine whether reductions in rate potentially underlie changes in fluency or whether both changes are independently caused by dual task conditions but are not causally linked to each other. Either way, findings suggest that if there is an association between rate and fluency of speech production, this relationship is specific to atypical disfluencies, as typical disfluencies were unaffected by dual task conditions and by the corresponding reductions in rate.

A final point related to speaking rate results (and also pertinent to fluency and error results below) is the apparent change in variability between baseline and dual task conditions for both speaker groups on graphed raw data (see Figures 4, 5, 6, and 7, which represent speaking rate, atypical disfluencies, typical disfluencies, and errors, respectively). Differences in error bars can be attributed to variability as well as to sample size; however, descriptive data for each speech-related outcome measure, when adjusted for differences in trial length (see right-sided portions of Tables 9, 12, 13, and 14), suggest comparable variability across individual experimental conditions. Thus, differences in graphed error bars most likely reflect the considerable difference in sample size between the two conditions represented (the baseline condition represented the mean for 6 speaking trials whereas dual conditions represented the collapsed mean for 60 trials and encompassed various experimental manipulations within the secondary task).
4.4 Dual task effects on fluency

One of the central questions motivating the present study was whether suppressing WM resources during the production of speech would result in a measurable enhancement of speech fluency based on overt disfluency counts. Analysis of fluency data indicated a strong dual task effect, although performance patterns for the two speaking groups showed some unexpected results. Three key findings emerged: (1) dual task effects occurred specifically on atypical forms of disfluency and to a comparable extent in both speaker groups; (2) typical forms of disfluency were not influenced by dual tasking in either group; and (3) dual task effects on fluency were not influenced by manipulations of WM domain, WM load, or ISI in the secondary task.

The dual task effect for atypical disfluencies was generally consistent with Hypothesis 3a, which predicted a reduction in atypical disfluencies (specifically for AWS); but not with Hypothesis 3b, which predicted increased typical disfluencies across speakers. Overall, neither of the speech monitoring accounts for stuttering behavior was fully supported. The similarity of fluency effects between speaker groups was partially in line with the CRH, which claims that stuttering and fluent speakers differ in the quantity but not the quality of disfluencies (Postma et al., 1991; Postma & Kolk, 1993). Based on this account, however, dual task effects should also have been observed for typical disfluencies and should have been accompanied by an increase in speech errors, both of which did not occur. Results were also inconsistent with the VCH, which predicts dual task effects for stuttering speakers only, based on the assumption that hypervigilant monitoring is a unique characteristic of this speaker group. Although the CAH and associated hypotheses related to explicit motor control were not designed to explain any specific phenomena associated with stuttering, findings generally concurred with these models. In line with this view, effects were comparable for the two speaker groups and were evident specifically
in atypical disfluencies, which most clearly resemble the *dechunking* behaviors described in this literature.

Although the categorization of disfluencies is sometimes overlooked, the differential effect of dual tasking on the two forms of disfluency (atypical vs. typical) in the present study has important implications for both research and clinical work. Many studies that report unchanged disfluency (e.g., Oomen & Postma, 2002) or increased disfluency (e.g., Bosshardt, 1999) under dual task conditions do not clearly distinguish between typical and atypical forms of disfluency. The present findings suggest that combining disfluencies indiscriminately may obscure an effect that is specific to atypical disfluencies. This possibility could also explain the general lack of consistency in this literature. More broadly, the issue of whether categorizing disfluencies provides useful diagnostic information and how specific disfluency types should be categorized, has been the subject of considerable debate (Cordes & Ingham, 1995; Cordes, 2000a, 2000b; Einarsdottir & Ingham, 2005; Wingate, 2001). Although the present study does not address specific questions at the heart of this controversy (e.g., Can disfluency types distinguish typical nonfluency from incipient stuttering?), the findings do imply that: (1) stuttering and fluent speakers are clearly differentiated by the frequency of atypical disfluencies but not typical disfluencies; and (2) atypical and typical forms of disfluency may involve distinct underlying processes. Specifically, results suggest that attentional control or performance monitoring may be relevant to atypical forms of disfluency but are most likely unrelated to typical disfluencies.

The question of whether observed fluency changes were due to a practice or order effect was considered; however, this interpretation was rejected based on several considerations. First, a preliminary comparison of typical and atypical disfluencies (as a proportion of total syllables)
in the pretest- versus the posttest-baseline speaking tasks showed no significant differences in either disfluency type within either group. Results of multilevel modeling further indicated that a model including trial number (for all speaking trials in the experiment, across conditions) as a predictor did not significantly improve fit. Last, predicted effects for this model (that included trial number) showed the reverse pattern (disfluencies increased with increasing trial number), which is more consistent with possible fatigue than with a practice effect.

Measures of speech fluency showed no effect of WM domain, WM load, or ISI. It was expected that dual task effects would be specific to WM tasks within the verbal domain (Hypothesis 3c), based on the consideration that the speaking and verbal WM tasks drew upon similar resources and were more likely to result in interference (Cocchini et al., 2002; Leclercq, 2002) than concurrent performance of the speaking and spatial WM task. Results did not support this prediction. The lack of a differential effect based on WM domain indicated that similar fluency changes occurred when WM was taxed, regardless of the nature of the stimuli being processed. An alternative possibility is that the spatial task may have unintentionally taxed verbal WM. Observation of participants during the baseline condition of the WM tasks revealed that speakers frequently applied verbal rehearsal strategies to the verbal WM task as well as to spatial WM tasks. For the verbal WM task, rehearsal naturally took the form of adding and reciting numbers; however, for the spatial task, participants sometimes rehearsed the sequence of movements within the grid (e.g., “up down up”) or verbally updated the target location (“cell 3”). It is not clear whether the majority of participants utilized such strategies and whether they were able to apply these strategies effectively during dual task conditions but if this did occur, both secondary tasks may have essentially taxed verbal WM resources and similar fluency effects would be a logical outcome.
The prediction that dual task effects (on fluency) would be stronger for WM tasks involving higher WM loads and shorter ISI (Hypothesis 3c) was also not supported. Instead, findings indicated similar fluency changes whenever WM was engaged, regardless of the extent to which it was engaged or how frequently it was accessed. It is possible that more extreme changes in WM load or ISI were necessary in order to observe an effect of these manipulations on speech fluency. In a recent dual task study involving concurrent rhyme judgments and a WM task, for example, vulnerability to interference (in the stuttering group) was apparent only in the most demanding WM condition, which involved recalling items in a set size of five (Jones, 2012).

Interestingly, dual task effects did not differ based on stuttering severity, as demonstrated by the lack of an interaction between subject-level factors, which included SSI and OASES scores, and task type (baseline vs. dual). This result was consistent with the finding that dual task effects did not differ by speaker type and indicated that a proportionally similar reduction in stuttering was observed across participants, even within the stuttering group. The use of different stuttering measures accounted for different perspectives on whether stuttering severity is best represented via: (1) objective measures of overt speech behaviors, as measured by the SSI; or (2) subjective ratings of the experience of stuttering, as measured by the OASES. Ultimately, neither index of stuttering severity determined the extent to which speakers benefited from dual task conditions. One interpretation of this finding is that although the secondary task suppressed a general resource that contributes to speech disfluencies, it may not have precisely targeted the cognitive mechanism that most clearly differentiates stuttering and fluent speakers and that differentiates stuttering speakers from each other. This possibility is discussed in further detail below.
A related point was that the SSI predicted the frequency of typical and atypical disfluencies but the OASES did not. This is not surprising, considering that the SSI score is largely based on stuttering frequency; however, it does confirm that the variable representing disfluency in the study was consistent with a widely used and established measure of stuttering severity. The finding that OASES scores were not related to disfluency highlights the principle that measures of overt stuttering behaviors and subjective experience of stuttering can provide two diagnostic perspectives that do not necessarily align with each other.

4.5 Dual task effects on overt errors

Error results supported Hypothesis 4, which predicted no change in the frequency of errors under dual compared to baseline conditions. This finding did not differ based on speaker group, and indicated that performing the WM task did not interfere with speakers’ ability to simultaneously monitor and correct covert errors arising during speech planning. Based on the CRH, reduction in disfluencies under dual task conditions should have been accompanied by an increase in the number of overt errors produced (Kolk, 1991; Postma & Kolk, 1993); however, this did not occur. Overall, previous findings related to this question are inconsistent, with some evidence showing increased error rates under dual task conditions (Oomen & Postma, 2002), but other data showing reduced error rates under similar conditions (Maraist & Hutton, 1957) or no change at all (Oomen & Postma, 2001b). This lack of agreement is at least partly due to vague specifications of the types of speech output constituting an error in each study but is also likely a result of the wide variation in speaking and secondary tasks, as previously discussed. Another relevant consideration is that the number of errors in the present data was exceedingly small (mean number of total errors per trial was approximately 1-2 errors across conditions in both
groups). Vasic and Wijnen (2005) reported similarly small values for their error counts and on the basis of this outcome, discarded error measures from their analyses altogether.

These error results, combined with previous research on this topic highlight the considerable challenges involved in studying speech errors, particularly covert errors, which by definition, cannot be reliably detected or measured, and are poorly understood. Overall, the relationship of covert errors to stuttering is supported by minimal evidence, and although there are data suggesting that subtle phonological encoding deficits are associated with stuttering (Anderson & Byrd, 2008; Howell et al., 2000; Jones et al., 2012; Sasisekaran et al., 2013; Sasisekaran & Weisberg, 2014), the implications of these differences are not clear, especially since there is no indication that these phonological or linguistic weaknesses are functionally significant (see Nippold, 2002 for review).

4.6 Does a specific cognitive mechanism contribute to speech disfluency?

Although dual task conditions in the present study had a clear effect on fluency, the underlying cognitive basis for the observed reduction in disfluency is less apparent. Results showed that suppressing WM reduced atypical disfluency, suggesting that WM resources contribute to the occurrence of stutter-like behavior. What specific function associated with WM might explain this link and how could this link be used to explain the excessive disfluency occurring in individuals who stutter?

One possibility considered a priori was that engaging WM might reduce stuttering by suppressing inner-loop speech monitoring functions associated with WM. If disfluencies arise as a byproduct of detecting and repairing covert errors, fewer disfluencies should occur when secondary tasks interfere with speech monitoring resources. This prediction was not supported.
by error data, which did not show the expected increase in overt errors under dual task conditions and suggested that monitoring of prearticular speech was not affected.

Alternatively, WM may function as part of a broader performance monitoring system that contributes to disfluency in a way that does not involve covert errors. In general, the phenomenon of covert errors and the cognitive basis for their repair are not well understood, as noted above. In his original model, Levelt implied that error signals during speech planning were detected and processed by WM (Levelt, 1989). More recent event-related potentials (ERP) data, however, suggest that a substantial part of speech monitoring is the outcome of medial frontal processes (e.g., ACC) that reflect a general monitoring mechanism and that are not inherent to the language system (Riès et al., 2011). An ERP study examining neural evidence for hypervigilant monitoring in people who stutter similarly found increased ERN and Pe amplitudes for stuttering speakers compared to controls during performance of linguistic as well as nonlinguistic tasks (Arnstein et al., 2011). These findings suggest that stuttering is associated with excessive monitoring activity that is not necessarily specific to speech. Thus, it is possible that the role of WM in disfluency involves actions that are initiated by the speaker in response to signals generated by the performance monitoring system, but that these actions do not represent attempts to repair covert errors.

Instead, recruitment of WM resources may reflect the speaker’s transition to an explicit, rule-based processing system in an effort to consciously control movement sequences involved in speech production. The assumption that explicit performance control occurs through WM is supported by dual task studies demonstrating performance benefits for skilled participants when WM resources are engaged and tasks are executed implicitly, particularly under conditions involving pressure (e.g., Beilock et al., 2002; Masters, 1992; Poolton et al., 2006). A recent study
utilizing transcranial direct current stimulation further showed that cathodal (inhibitory) stimulation to the left dorsolateral prefrontal cortex suppressed verbal WM resources and enhanced motor learning on a golf-putting task. In the present study, engagement of WM resulted in a fluency benefit for both speaker groups, suggesting that applying these resources to speech production tends to interfere with fluency, regardless of whether a stuttering disorder is present.

4.7 Resource allocation in individuals who stutter

The finding that fluency effects were comparable across speaker types and levels of stuttering severity can be interpreted in two different ways. One possibility is that the secondary task suppressed resources that generally contribute to speech disfluencies but did not affect the key mechanism that sets stuttering speakers apart from fluent ones, as suggested earlier. An alternative conclusion is that speaker groups showed parallel effects because atypical disfluencies in all speakers arise from a similar underlying process in which maladaptive attentional tendencies interfere with automaticity. This explanation is consistent with Bloodstein’s Anticipatory Struggle Hypothesis (Bloodstein, 1972, 1975, 1984), which attributed stuttering to early speech-related anxiety, excessive “speech consciousness,” and resulting fragmentation of sequenced articulatory movements that eventually becomes habitual.

In line with this concept, a recent study using electromyography to compare perioral muscle activity in stuttering and fluent children showed no neurophysiological differences between the groups (Walsh & Smith, 2013). As interpreted by the authors, this finding suggests that early stuttering is not characterized by inherently atypical motor patterns, but rather that disfluencies arise from “interruptions” to otherwise normal command signals. The Dual-
Diatheses Stressor framework of stuttering (Walden et al., 2012) further proposes that specific attentional and emotional tendencies (along with language weaknesses) constitute sources of vulnerability that lead to childhood stuttering when specific stressors are activated. This view is supported by several studies indicating that children who stutter are more emotionally reactive and have more difficulty shifting their attention and regulating emotional reactions than their fluent peers (Anderson, Pellowski, Conture, & Kelly, 2003; Eggers, De Nil, & Van den Bergh, 2010; Felsenfeld, van Beijsterveldt, & Boomsma, 2010; Karrass et al., 2006; however, see Alm, 2014 for a different conclusion based on these results). Assuming early differences in temperament do exist, stuttering children may be more liable to react strongly to breakdowns in speech performance, and have difficulty shifting their attention away from these breakdowns and modulating emotional responses. Parallel patterns were recently demonstrated in stuttering adults, who showed evidence of an attentional bias to threat words on the emotional Stroop task (Hennessey, Dourado, & Beilby, 2013; Lieshout, Ben-David, Lipski, & Namasivayam, 2014). Interestingly, Lieshout and his colleagues (2014) found similarities in the responses of stuttering and fluent adults to emotional stimuli, specifically, in the rate of lower lip movements, which increased in both speaker groups. Thus, all participants showed subtle changes in motor activity during responses to threat words. Within the same study, however, only AWS showed a reduction in upper lip variability and between-lips phase differences in response to threat words. As interpreted by the authors, this finding indicates that stuttering speakers adapted their motor responses in a unique way, most likely as an attempt to control the destabilizing effects of anxiety. These findings, combined with those reported for stuttering children, suggest that stuttering is associated with an overactive evaluation mechanism for detecting threat and directing attentional resources toward these stimuli. Importantly, increased attention to perceived
threat appears to result in efforts to stabilize the motor system that are ultimately counterproductive.

4.8 Optimal levels of cognitive control: When less can be more and more can be less

The present findings and related literature on attention and performance (Beilock & Carr, 2005; Beilock & Gray, 2007, 2012; DeCaro & Beilock, 2010; Gray, 2004; Masters, 1992; Wulf & Lewthwaite, 2010) demonstrate that maximizing cognitive control does not always benefit task performance. This principle is the central premise of the Matched Filter Hypothesis (Chrysikou, Novick, Trueswell, & Thompson-Schill, 2011; Chrysikou, Weber, & Thompson-Schill, 2013; Thompson-Schill, Ramscar, & Chrysikou, 2009) which proposes that optimal levels and patterns of resource allocation vary based on specific task demands, goals, and contexts. Based on this framework, efficient filtering of sensory information via top-down control (associated primarily with prefrontal cortex [PFC] activity) supports performance across a variety of tasks that are rule-driven, involve conflict, or require abstraction of concepts. The same form of control, however, hinders performance on tasks (e.g., early language learning, implicit motor learning) that are habitual and automatic and best served by subcortical (e.g., basal ganglia) neural systems. Thus, optimal performance relies on dynamic adjustments to the filtering mechanism based on task requirements.

Relatedly, Shine and Shine (2014) describe the ability to delegate control of routine tasks from the cortex to lower neural circuits (such as the basal ganglia) as a critical evolutionary development that originally arose to manage information processes associated with bipedalism but was co-opted for a variety of other tasks. According to the authors, this transfer of behavior from a volitional to automatic mode may have been the driving force for the rapid expansion of
human cognition, as this change freed the cortex for more complex tasks, increased overall efficiency, and allowed activities to be performed well even when cortical networks are overwhelmed by extreme stress.

Consistent with this view, neuroimaging data indicate that early stages of task learning are characterized by engagement of frontal regions and specific subcortical areas (associative striatum) that receive input from PFC; but that these regions show lower activation once a task has been automatized (Ashby, Turner, & Horvitz, 2010; Poldrack et al., 2005). Poldrack and colleagues (2005) further reported that regions within the basal ganglia showed decreased activity on trials that formed a sequence compared to pseudorandom trials, reflecting the more effective chunking of information that accompanies automatization. Two key findings from neuroimaging studies within the stuttering literature suggest that stuttering is associated with differences in this neural circuitry. Giraud and colleagues (Giraud et al., 2008) found a positive correlation between basal ganglia activity and stuttering severity, implying that representations were less efficiently organized and resembled early stages of motor learning in these speakers. ERP and fMRI studies have further demonstrated increased ACC activity in stuttering compared to fluent speakers during conflict tasks, indicating that individuals who stutter demonstrate excessive monitoring activity, even when their behavioral performance does not show impairment relative to controls (Arnstein et al., 2011; Liu et al., 2014). The interpretation of higher activation patterns in these regions is not consistent from study to study – Arnstein and colleagues interpret larger ERN as hyperactive error-monitoring, whereas Liu and colleagues suggest that these findings reflect compensation for neural insufficiency in other frontal regions. Nevertheless, overall findings suggest that stuttering may be associated with inappropriate matching between task demands and the extent or types of cognitive resources utilized to meet
these demands. In line with this view, results of the present study indicated that full availability of WM resources was associated with greater propensity toward forms of disfluency most closely associated with stuttering, whereas limiting available resources resulted in fluency enhancement. This outcome suggests that fluent speech production, like skilled motor performance across a variety of tasks, benefits from a state of minimized executive control and maximized automatization. The finding that typical forms of disfluency did not increase under dual task conditions for either group further implies that suppressing WM resources did not significantly compromise speakers’ ability to formulate and organize linguistic output. Other aspects of performance, such as linguistic or conceptual complexity, may have been affected but is a question for future research.

4.9 Is stuttering associated with underlying deficits in WM?

Although a causal link between stuttering and WM deficits has been proposed (see Bajaj, 2007 for review); support for this notion is based on studies using nonword repetition tasks, which specifically target phonological WM (Anderson, Wagovich, & Hall, 2006; Byrd, Vallely, Anderson, & Sussman, 2012; Hakim & Ratner, 2004) and contradictory findings have been reported (Sasisekaran, 2013; Smith et al., 2010). The use of complex span tasks has not been reported in previous studies of WM in stuttering speakers. Results of the present study indicated comparable performance patterns for the two speaker groups across all manipulations of the secondary task (WM domain, WM load, ISI), and on two widely-used span tasks (OSpan and SSpan; developed by Unsworth et al., 2005), which measured verbal and spatial spans, respectively. The combined findings indicated that stuttering is not associated with weaknesses in either verbal or nonverbal WM capacity. Results of a recent dual task study (Jones et al.,
involving concurrent rhyme judgments and WM recall suggested that WM in AWS may be more vulnerable to interference; however, no baseline recall data were reported, making it difficult to determine whether WM recall was poorer for AWS overall or was specific to dual task conditions. Further, the level of WM at which performance deteriorated (load of 5) exceeded the maximum load (4) used in the present study, which may explain the apparent discrepancy.

An important point related to WM findings is that although no group differences were found in any aspect of performance (absolute span as well as accuracy at individual set sizes) on either span measure, analyses of subject-level variables contributing to fluency effects revealed that OSpan performance was related to disfluencies. Specifically, results indicated that increases in OSpan scores predicted lower overall rates of atypical disfluency in both speaker groups. This finding is not consistent with the conclusion that stuttering and WM show no association at all and indicates the need for future research examining the nature of this relationship.

4.10 Limitations

Several limitations in the present design require consideration. First, although the secondary task clearly taxed WM resources, more closely matched stimuli and task structure for the verbal and spatial versions of the task may have resulted in a clearer domain effect. Secondary tasks could also have been set at individual capacity (e.g., see Cocchini et al., 2002) to ensure that the task was fully engaging the targeted mechanism in each participant. Relatedly, very high levels of performance on the secondary task, even under dual task conditions, suggests that both WM tasks were performed with some degree of automaticity and may have diminished dual task effects. Future studies that incorporate less automatic secondary tasks and on which performance patterns do not approach ceiling, could potentially demonstrate more pronounced
effects on speech-related variables. Importantly, a control condition in the secondary task with no WM load would have helped clarify whether effects observed necessarily involved WM per se or whether they reflected more general resources, such as sustained attention. Although the vigilance task essentially resembled a sustained attention task, it was not designed as a control condition and was too different in general structure (e.g., it involved more frequent responses and a variety of ISIs) to be used for comparison purposes.

Another issue that was not considered in the present study was the possibility that dual task conditions could have subtly affected aspects of linguistic productivity in addition to influencing secondary task accuracy. For example, studies by Bosshardt (2002, 2006) showed that generating sentences while performing a secondary task resulted in fewer propositions. Bosshardt also found evidence for a tradeoff between the rate of disfluency and linguistic productivity in stuttering speakers as cognitive processing load increased. Under certain dual task conditions, stuttering speakers maintained a low level of disfluency but produced significantly fewer propositional units compared to controls; in other cases, stuttering speakers were more disfluent under dual conditions but produced propositions at a rate similar to controls. Dual task conditions in the present study could have similarly influenced specific aspects of linguistic productivity (e.g., semantic diversity, syntactic complexity, word frequency); however, these were not the focus of current analyses and were not considered.

Additionally, results indicated a reduction in both atypical disfluencies and overall speaking rate under dual task conditions but did not clarify whether the two changes may have been causally linked. Speech rate calculations were based on the number of fluent syllables produced over the total duration of each speaking trial (as recommended by Guitar, 2006; see Data Processing section); thus, this measure inherently reflected some of the same effects
demonstrated by disfluency data. More detailed analysis of articulation rate based on syllable counts occurring specifically during fluent stretches of speech could have provided a more precise measure of articulatory speed and clarified whether speech movements were in fact produced at a slower rate. Either way, however, overall speaking rate and articulatory rate tend to be closely correlated (Kelly & Conture, 1992), and the finding of slowed articulation rate under dual task conditions, if confirmed, would still be insufficient to establish a causal relationship between rate and fluency outcomes.

Within the speaking task, speech preparation time was not controlled. This aspect of the design was necessary in order to allow participants to select topics for each speaking trial but made it difficult to determine whether language formulation and conceptual planning may have differed between groups or contributed to results in any way. A comparison of mean topic selection time for the two speaker groups within baseline and dual conditions suggested that this was not the case, $t(37) = -0.15, p = .88$ and $t(37) = 0.09, p = .93$ for the baseline and dual task conditions, respectively. Nevertheless, individual differences in speech preparation time could have contributed to performance in ways that were not detected in aggregated data. Additionally, although the spontaneous speaking task was intentionally selected in order to induce stuttering rates that resembled naturally occurring rates of stuttering, important aspects of spoken output (emotionality of content, vocabulary, syntactic complexity, level of abstraction) could not be controlled and may have introduced noise to resulting data. Finally, the considerable length of the experiment resulted in the distinct possibility of fatigue influencing outcomes. Predicted effects of the model that included trial number, as described above (increasing trial numbers were associated with higher rates of disfluency), suggested that this may have occurred. Similar designs with fewer manipulations might have eliminated this additional source of noise.
4.11 Future research

The present findings indicate several directions for future research. A straightforward interpretation of the data implied that observed effects were contingent upon WM involvement; however, the similarity of results across all manipulations of WM (and lack of a control condition requiring no WM) suggests that general attentional resources could also have driven these outcomes. The specific role of processes related to performance monitoring and self-evaluation also remains unclear. Underlying neurocognitive mechanisms contributing to these functions, and potentially to disfluency, require further elucidation. Dual task paradigms that specifically target functions associated with the ACC or PFC (other than WM) have not been attempted and could potentially provide an interesting perspective to these questions. Neurophysiological techniques that enable cortical modulation of specific neural networks (e.g., transcranial direct current stimulation) can be particularly useful in this regard.

Potential application of this methodology to a younger population is an important direction for future studies on this topic. If maladaptive forms of conscious control contribute to atypical disfluencies, examining whether such patterns are evident near the onset of stuttering would clarify the role of these processes in the development of fluency disorders. Late maturation of executive functions may make dual tasking difficult to implement in very young children; however, simple dual task designs appropriate for school-age children (5-8 years) are reported in the literature (e.g., Gautier & Droit-Volet, 2002) and might be feasible for younger children, particularly if WM demands are not critical for inducing fluency effects. Future studies examining the role of emotional and attentional biases during early stuttering are also necessary to fully understand how and when such biases arise, whether they are related to temperament, and what their implications might be for stuttering persistence or recovery.
4.12 Clinical implications

Although preliminary, the present findings suggest that clinical intervention for stuttering may be enhanced by strategies that induce more implicit forms of speech production. Neuroimaging studies of post-treatment speech fluency have shown a reduction in ACC activity in treated adults (De Nil & Kroll, 2001; Kroll, De Nil, & Houle, 1997; Neumann et al., 2005), suggesting positive changes in anticipatory behaviors and self-monitoring tendencies immediately following intensive treatment. It is not clear, however, whether these changes are maintained. Subjective reports from long-term post-treatment follow up indicate that many individuals who stutter continue to consciously attend to their speech after participating in intensive treatment programs (Boberg and Kully, 1994). Therapeutic applications of dual tasking are emerging in general rehabilitation research (Brauer & Morris, 2010; Plummer, Villalobos, Vayda, Moser, & Johnson, 2014) but have only been examined in one study of stuttering adults (Metten et al., 2011, Experiment 3). The outcome of this particular study was not very encouraging; however, results only consisted of subjective reports and were based on an insufficient sample (N=3). Further research is needed to determine whether and how dual task methods can supplement traditional approaches to stuttering treatment.

These results can also potentially inform early intervention approaches for stuttering. If incipient stuttering stems, in part, from excessive attention and conscious control of speech production, it is possible that this cyclic process could be interrupted at its start. The role of cognitive and attentional mechanisms in the emergence of atypical disfluencies, and potential effects of diverting attention from speech production, have obvious and immediate implications for the treatment of young children. Clarifying these associations could provide both parents and clinicians with a novel and practically useful perspective to early stuttering disorders.
4.13 Summary and conclusion

Over a century ago, William James asserted:

“Our lower centres know the order of movements... but our higher thought-centres know hardly anything about the matter... In an habitual action, mere sensation is a sufficient guide, and the upper regions of the brain and mind are set comparatively free” (James, 1890, pp.115-116).

The present study explored applications of this perspective to stuttering disorders, based on theoretical frameworks of stuttering suggesting that minimizing cognitive resources while speaking could enhance fluency. The central finding confirmed these hypotheses and showed that continuously engaging WM while speech was being produced resulted in a significant reduction of atypical disfluency in all speakers. Overall, the data did not support either of the speech monitoring accounts of stuttering (CRH, VCH). Suppressing WM resources while producing speech resulted in fluency changes but no accompanying changes in the frequency of speech errors. This finding contradicts the CRH, which implies a tradeoff between disfluencies and overt speech errors (Kolk, 1991; A. Postma & Kolk, 1993). The similarity in outcomes for the different speaker groups was also inconsistent with the VCH, which claims that excessive disfluency arises from hypervigilant speech monitoring but that this phenomenon is unique to speakers who stutter (Vasić & Wijnen, 2005). Much like the axiomatic view that stuttering is essentially what a speaker does to avoid stuttering, the findings demonstrate that speech fluency across speaker types is enhanced by less effort, but compromised by more. Further research is needed to fully understand the cognitive mechanisms involved in this effect and how these findings can potentially inform novel approaches to stuttering intervention.
Appendices

Appendix A: Topic prompts for speaking task

1. How do you use Facebook?
2. Is it ethical to eat meat?
3. Which is more important: talent or hard work?
4. What will you remember from 2012?
5. Have you ever interacted with the police? Describe.
6. What's your favorite holiday? Why?
7. New Year's resolutions
8. Describe a favorite place.
9. Describe a fantasy vacation.
10. Is home-schooling a good idea?
11. Describe things you created when you were a child.
12. What makes a great conversation?
13. Why do we share photos? (Do you?)
14. Where were you during Hurricane Sandy?
15. Your first date: What do you remember?
16. First day at college: What do you remember?
17. Why does (almost) everyone love babies?
18. Entrance exams for kindergartners?
19. Do smart phones make us smarter?
20. What makes a perfect date?
21. Are computers intelligent?
22. Are children intelligent?
23. Describe a TV show (or shows) you loved as a child.
24. What's your favorite time of year? What do you do to enjoy it?
25. What are your thoughts about this experiment? (You can be honest.)
26. How do male and female roles differ in your family?
27. What have you done to earn money?
28. How well do standardized tests measure ability?
29. What's the coolest thing you've ever seen in a museum?
30. What would you do with a million dollars?
31. Describe a movie or TV show you saw recently.
32. Give a book recommendation. Why would you recommend it?
33. Describe a favorite restaurant.
34. DON'T see this movie: __________
35. How capable are you in the kitchen? What can you make well?
36. What are your thoughts about the Super Bowl?
37. Websites I regularly visit
38. What's in a name?
39. Thoughts or memories related to pets
40. Should we unplug regularly? Imagine life without TV, internet, phone calls, or texts.
41. Why are soap operas so popular?
42. The worst vacation
43. Coping with brothers and sisters
44. Foods I love, foods I hate
45. Facebook: Effective social networking tool or waste of time?
46. A teacher that made a difference
47. Describe your parents.
48. Should kids have cell phones?
49. Describe what a typical day is like for you.
50. Why I (love/hate) watching the news
51. Describe the city or town where you grew up.
52. How do you usually spend your weekends (or how do you wish you'd spend them?)
53. Describe one of your close friends.
54. Describe your profession or job.
55. If you were to change your occupation, what might you choose to do? Why?
56. Which animals do you like to see at the zoo? Why?
57. Describe a class or course you enjoyed.
58. What do you like to do for fun?
59. What is something you wish you had done differently?
60. A favorite fairy tale (retell it)
61. Ways to entertain a young child
62. What do you do to relax?
63. Describe something you learned how to do when you were a child.
64. Describe a game you played or activity you enjoyed as a child.
65. Describe a special celebration in your life.
66. Describe a room in your home.
67. Describe your dream job.
68. Describe a book you liked when you were a child.
69. Are you green? Describe ways you protect the environment.
70. Time wasters: Describe things you do to procrastinate.
71. What are some of your favorite rides or activities at amusement parks?
72. Describe something you'd like to learn how to do.
73. Describe your daily commute.
74. If you could meet any historical figure, who would you like to meet and why?
75. Describe a personal pet peeve.
76. Describe the first car you ever drove.
77. Describe a perfect Sunday.
78. Describe your first girlfriend/boyfriend.
79. Describe your first job.
References


