Running head: INDIVIDUAL DIFFERENCES IN DISFLUENCY

Individual differences in the production of disfluency: a latent variable analysis of memory

ability and verbal intelligence

Paul E. Engelhardt[1], Mhairi E.G. McMullon[2], and Martin Corley[3]

[1]University of East Anglia

[2]University of Northumbria

[3]University of Edinburgh

**Word count: 9085**

Address for correspondence:
Paul E. Engelhardt
School of Psychology
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ
Email: p.engelhardt@uea.ac.uk

Abstract

Recent work has begun to focus on the role that individual differences in executive function and intelligence have on the production of fluent speech. However, isolating the underlying causes of different types of disfluency has been difficult given the speed and complexity of language production. In this study, we focused on the role of memory abilities and verbal intelligence, and we chose a task that relied heavily on memory for successful performance. Given the task demands, we hypothesised that a substantial proportion of disfluencies would be due to memory retrieval problems. We contrasted memory abilities with individual differences in verbal intelligence as previous work highlighted verbal intelligence as an important factor in disfluency production. Seventy-eight participants memorized and repeated 40 syntactically complex sentences, which were recorded and coded for disfluencies. Model comparisons were carried out using hierarchical structural equation modeling. Results showed that repetitions were significantly related to verbal intelligence. Unfilled pauses and repairs, in contrast, were marginally ($p < .09$) related to memory abilities. The relationship in all cases was negative. Conclusions explore the link between different types of disfluency and particular problems arising in the course of production, and how individual differences inform theoretical debates in language production.


**Keywords:** speech fluency, verbal intelligence, executive function, disfluency, individual differences

Individual differences in the production of disfluency: a latent variable analysis of memory

ability and verbal intelligence

There is surprisingly little research looking at the relationship between individual

differences and the fluency of language outputs. However, understanding differences

associated with individual variation is crucially important as theories of language production

seek to account for and model individual differences. In the current study, we examined the

role of executive function and intelligence on disfluency production. The goals of this study

were to examine (1) the relationship between individual differences in cognitive control and

fluent language outputs, and (2) to take a closer look at one particular cause of covert

disfluencies (i.e. pauses and repetitions). Previous research has drawn a distinction between

overt and covert disfluency: Overt disfluencies often provide information about the

underlying cause of disruption. In contrast, covert disfluencies, such as an unfilled (or silent)

pause, do not reveal anything about the nature of the problem that caused the disfluency to

occur (cf. Fraundorf & Watson, 2013). Hypothesized causes of covert disfluency include

lapses of attention, planning/thinking demands, lexical retrieval difficulty, faulty monitoring,

etc. (Barr, 2001; Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Clark, 1994; Clark &

Fox Tree, 2002; Clark & Wasow, 1998; Garrett, 1982; Levelt, 1983; Maclay & Osgood,

1959; O'Connell & Kowal, 2005). However, it is likely that all of these "causes" affect

production in some situations. Thus, there is no one-to-one mapping between problems in

production and the manner in which speech is disrupted (Arnold, Tanenhaus, Altmann, &

Fagnano, 2004; Foxtree & Clark, 1997; Postma, Kolk, & Povel, 1990; Shriberg, 1996).

However, from a theoretical standpoint it is important to understand what goes wrong, when

it goes wrong, and how the system recovers from a problem (Levelt, 1983). Answers to these

questions would help differentiate competing models of production (Deese, 1984; Kempen,

1978; Levelt, 1999; Vosse & Kempen, 2000) and elucidate the monitoring mechanism associated with language production (Blackmer & Mitton, 1991; Hartsuiker & Kolk, 2001; Levelt, 1989; Nooteboom, 1980; Postma, 1997, 2000).

The remainder of the Introduction covers a diverse set of topics. We begin by reviewing models of language production, and then "causes" of different types of disfluency with a view towards explaining language-based causes of disfluency and different stages of language production. We next turn to reviewing the literature on individual-based causes of disfluency (i.e. individual differences). Because the majority of work examining the influence of executive function and intelligence on disfluent speech outputs was conducted on a clinical population, we review disfluency production in Attention-Deficit/Hyperactivity Disorder (ADHD). The Introduction finishes with a broad overview of executive function and intelligence, and the rationale and methodology used in the current study.

*Models of Language Production*

Many theories of language production assume a stage-based approach, and arguably, the most prominent model was proposed by Levelt (1989). This model assumes three main stages (i.e. conceptualization, formulation, and articulation) with each consisting of sub-stages (Bock & Levelt, 1994; Ferreira & Engelhardt, 2006; Fromkin, 1973; Garrett, 1975; Pickering et al., 2002). Conceptualization is the non-linguistic thought or message that the speaker wants to convey. Formulation involves grammatical encoding, which is the process of selecting lemmas to convey the message, and in the event of multi-word utterances, assigning grammatical roles, creating a linear sequence, and placement of function words. Formulation also includes phonological encoding, which is the process of retrieving the sounds and organizing them into a phonetic plan. Finally, the articulation stage is where the phonetic plans are sent to the articulators. Most models of production also assume the

existence of a monitoring mechanism, which is responsible for detecting problems within pre-articulated speech plans (for reviews see Blackmer & Mitton, 1991; Postma, 2000).

Speech errors have long been used to inform theoretical models of production (e.g. Dell, 1986; Garrett, 1975). Speech errors and overt disfluencies often provide some clue about the underlying cause of the error/disruption. For example, antonym-type repairs, such as *turn left…right at the junction*, point toward incorrect lemma selection (Dell, 1986; Dell & O'Seaghdha, 1992; Hartsuiker & Notebaert, 2010). In contrast, covert disfluencies, such as unfilled pauses, do not reveal anything about the nature of the problem that caused the disfluency to occur. (The label "covert" is used to refer to the fact that the *cause* of the speech disruption remains hidden, or perhaps, more specifically there is no behavioural evidence about what went wrong.) The idea that covert disfluency may be a signal to particular difficulties in production is an issue that has recently attracted some empirical attention (e.g. Hartsuiker & Notebaert, 2010). However, it is important to keep in mind that there are independent causes of disfluency (e.g. lapses of attention), which do not have anything to do with the stages or sub-stages of language production.

*Causes of Disfluency*

With regard to conceptualization, it has been shown that filled and unfilled pauses occur frequently in cases where the speaker experiences planning difficulty and needs more time (Clark & Wasow, 1998).[1] This was confirmed in an empirical study that examined constituent ordering in a production task that examined given vs. new constituents (Arnold et al., 2000). When speakers produced fillers (i.e. experiencing production difficulty), they were more likely to first say an easy to produce goal-given noun phrase, followed by a more difficult theme-new noun phrase. In cases in which there were no sentence-initial fillers, speakers were less likely to produce goal-early structures. This pattern is consistent with the

---

[1] Similar results have been shown with prolongations, e.g., producing *the* as *theee* (Fox Tree & Clark, 1997).

*Minimal-Load* principle (Levelt, 1981; 1982), in that speakers experiencing production difficulty were more likely to produce given constituents first, whilst gaining additional time to plan and produce a new and more difficult constituent second. A second type of study that has made inferences about fillers as a cue to conceptualization is referred to as "the feeling of knowing".[2] In these studies, listeners were asked to judge (or rate) the knowledge state of speakers. Brennan and Williams (1995) used a question-answering paradigm and argued that *uh* and *um* were not premeditated (speaker) choices but an emergent aspect of production (cf. Clark & Fox Tree, 2002). This is because listeners did not distinguish between *uh*s and *um*s, and both led to lower (knowing) ratings than unfilled pauses (Brennan & Williams, 1995; Smith & Clark, 1993; Swerts & Krahmer, 2005).

Grammatical encoding occurs after conceptualization and is typically assumed to consist of two sub-stages, called functional- and positional-level processing (Bock & Levelt, 1994), and has received more modelling work and empirical investigation, especially in terms of picture naming (Forster & Chambers, 1973; Levelt, 2001; Levelt & Roelofs, 1999; Levelt et al., 1991; Schriefers et al., 1990). The first stage of grammatical encoding involves lemma selection. For lemma selection, there is assumed to be activation of a cohort of semantically-related lemmas (Kempen & Huijbers, 1983). In phonological encoding, the sound segments for the selected lemma are retrieved (Garrett, 1975). Evidence for the grammatical encoding stage of production is supported by speech errors, which often reveal semantic similarity with an intended word (e.g. saying "stummy" instead of "stomach" or "tummy", and relatedly, evidence for the phonological encoding stage of production, is supported by phonologically-based speech errors (i.e. words with similar phonology but semantically unrelated).

---

[2] In this literature, the "feeling of knowing" does not mean actual knowing, but instead listeners perception of the knowledge state of the speaker, and is sometimes referred to as "feeling of another's knowing" (Brennan & Williams, 1995).

Perhaps the most robust evidence of lexical access difficulty leading to increased probability of disfluency comes from a study by Hartsuikeer and Notebaert (2010). Those authors utilized a network description task, in which they manipulated the name agreement of objects in the network. Name agreement refers to the number of possible names a given object can be described with, and name agreement produces a marked effect on naming latencies (Griffin, 2001; Paivio, Clark, Digdon, & Bons, 1989; Severens, Van Lommel, Raticnckx, & Hartsuiker, 2005; Vitkovitch & Tyrell, 1995). In the Hartsuiker and Notebaert (2010) study, participants had to name objects with either high- or low-name agreement. Results showed that pauses and self-corrections occurred more frequently for low-name agreement pictures compared to high-name agreement pictures (see also Schnadt & Corley, 2006). When the most frequent name required an infrequent (neuter gender) determiner there were more self-corrections and repetitions compared to names requiring more common gender determiner. Determiner selection takes place after the content noun, which it must agree with. The distinct pattern of disfluencies (pauses vs. repetitions) was interpreted as evidence of different responses to different processes within production. In this case, there was a more-or-less consistent mapping between type of difficulty and type of disfluency.

In more recent study, Fraundorf and Watson (2013) conducted a study utilizing a story re-telling task. Participants were told an excerpt from Alice in Wonderland, and they were then given a series of 14 plots points which they needed to incorporate into the story that they had to "re-tell". By analyzing the places where disfluencies occurred, Fraundorf and Watson made inferences about the underlying cause(s) of distinct types of disfluency (i.e. fillers, silent pauses, and repetitions). Interestingly, they argued that different types of disfluency reflect different strategies for dealing with production difficulty. Fraundorf and Watson concluded that filled pauses were produced in response to conceptual difficulties as evidenced by their frequent occurrence at new plot points and the beginnings of utterances.

Unfilled pauses occurred at similar places, but were less reliably produced at particular locations. However, they were also produced in response to other factors affecting grammatical and phonological processes (e.g. lexical frequency and first mention). There were systematic patterns with repetitions, and Fraundorf and Watson argued that these are dependent on the availability of the material to be repeated. In contrast, when errors occurred or when corrections required such things as being more informative, then the corrections occurred relatively late in the production process (i.e. phonological encoding and/or problems detected by the monitoring mechanism) (Hartsuiker & Kolk, 2001; Levelt, 1989; Postma, 2000). Now that we have reviewed *language-based* causes of disfluency, we now turn to *person-based* causes of disfluency, and the majority of existing literature in this area has focuses on a clinical population (i.e. Attention-Deficit/Hyperactivity Disorder).

*Disfluency in ADHD*

One motivation for the current study came from a series of papers that investigated sentence production in ADHD (Engelhardt, Corley, Nigg, & Ferreira, 2010; Engelhardt, Ferreira, & Nigg, 2009, 2011; Engelhardt, Veld, Nigg, & Ferreira, 2012). In particular, these papers focused on the role of *inhibitory control* in sentence production, as many of the main theories of ADHD focus on deficiencies in behavioural-response inhibition (e.g. Barkley, 1997; Barkley & Murphy, 2006; Martel, Nikolas, & Nigg, 2007; Nigg, 2001; Nigg, Carr, Martel, & Henderson, 2007; Pennington & Ozonoff, 1996; Schachar, Tannock, Marriott, & Logan, 1995; Tannock & Schachar, 1996). The most robust findings with respect to inhibitory control were shown for repair disfluencies (i.e. when the speaker stops and then starts over with a new word or phrase). In these studies, participants saw two pictures and a verb and they had to produce a sentence. Individuals diagnosed with the combined subtype of ADHD (i.e. those with symptoms of both inattention and hyperactivity-impulsivity) and those with partially-remitted ADHD produced more repairs compared to typically-developing

controls (Engelhardt et al., 2010; Engelhardt et al., 2012). Approximately two-thirds of the repair disfluencies were cases in which participants made a structural revision, that is, they switched from active to passive voice (e.g. *the girl ......the bicycle was ridden by the girl*), and approximately one-third showed clear evidence of a production error (e.g. *the boy ... girl had ridden the bicycle*), which would be consistent with lexical selection difficulty (Shao, Roelofs, & Meyer, 2013). This result was later extended to individual differences in typically-developing individuals. Engelhardt et al. (2013) showed that performance on the Stroop task and stop-signal reaction time (both primarily inhibition tasks) can account for nearly one-third of the variance in repair disfluency production, and this finding held even when individual differences in intelligence and set shifting were controlled for.

*Executive Function and Intelligence*

Executive functions are often described as low-level *cognitive control* mechanisms that govern thoughts and actions in the service of achieving goals and monitoring performance (Burgess, 1997; Denckla, 1996; Friedman & Miyake, 2004; Logan, 1985; Miyake et al., 2000; Rabbit, 1997; Stuss & Benson, 1986). The most commonly postulated executive functions are updating/monitoring of working memory, set shifting, and inhibition (Miyake et al., 2000; 2001). It is widely assumed that these executive functions play a role in most, if not all, cognitive processing, including language production (Roelofs, 2003). A large literature has emerged concerning how different executive functions are related to different types of intelligence. For example, Friedman et al. (2006) reported that working memory was highly related to both fluid and crystalized intelligence, and that inhibition and set shifting share less variance with intelligence (see also Baddeley & Logie, 1999). If there are any criticisms of the executive function/intelligence work it is that it is difficult to operationally define and empirically dissociate the hypothesized executive functions from one another and from more general individual difference variables, such as intelligence or processing speed

(Ardila, Pineda, & Rosselli, 2000; Duncan, Johnson, Swales, & Freer, 1997; Friedman et al., 2006; Jester et al., 2009; Teuber, 1972). The reason executive functions and intelligence are difficult to dissociate is because of shared variance, that is, these abilities tend to correlate within individuals (Friedman et al., 2007; Kline, 1991; Miyake et al., 2000). The higher the shared (as opposed to unique) variance the more difficult it is to dissociate different constructs, and problematically, the ratio of shared-to-unique variance differs between abilities and sometimes between samples (Miyake & Friedman, 2012).

Investigations of the role of executive functioning in more complex cognitive tasks often use individual differences paradigms. The basic idea is that if an ability, such as language production, relies on a (low-level) executive function, then individuals varying in that executive function will also vary, for example, in language production performance. If there is no relationship, then correlations between tasks should be at or near zero. Research into the role of executive functioning on the fluency of language outputs interestingly began with the study of clinical populations (e.g. ADHD, Autism Spectrum Disorders (ASD)) and individuals with *deficits* in particular abilities (e.g. older adults).[3] More recently however, research has shifted to typically-developing individuals (e.g. Engelhardt, Nigg, & Ferreira, 2013; Shao, Roelofs, & Meyer, 2012).

Intelligence is widely assumed to reflect functioning across broader and wider neural networks compared to executive function. All participants taking part in the Engelhardt et al. ADHD studies were also given an assessment of fullscale intelligence as part of the experimental protocol. Specifically, each participant completed five subscales from the Wechsler Adult Intelligence Scale (Wechsler, 1997a, 1997b), and so, several previously

---

[3] Older adults have been shown to experience word finding difficulties, which suggests problems with lexical retrieval (Cooper, 1990; Kemper, 1992; Le Dorze & Bedard, 1998; Mortensen, Meyer, & Humphreys, 2006; Sandson, Obler, & Albert, 1987; Schmitter-Edgecombe, Vesneski, & Jones, 2000).

published papers reported the correlations between intelligence and various sorts of disfluency.

*Current Study*

In the current study, we were primarily interested in covert disfluencies (i.e. *repetitions* and *unfilled pauses*). Repetitions are when a speaker stops and then repeats something they just said with no functional benefit, and unfilled pauses are silent pauses in speech. Repetitions were shown to be significantly correlated with fullscale intelligence (Engelhardt et al., 2010), are produced more frequently by participants with the inattentive subtype of ADHD (Engelhardt et al., 2011), and are significantly correlated with Stroop performance (Engelhardt et al., 2013). Unfilled pauses were shown to be significantly correlated with two of the Wechsler subscales: vocabulary and matrix reasoning, and were produced more frequently in participants with the inattentive subtype of ADHD. These correlational findings are intriguing, but at this point, it is not known how much of the variance associated with these correlations is shared and how much is unique. As mentioned previously, working memory tends to correlate highly with intelligence, whereas inhibition and set shifting do not (Friedman et al., 2006).

Covert disfluencies present a unique challenge to empirical investigation because, as mentioned above, there is no behavioural evidence concerning the nature of the underlying problem (Levelt, 1983). In the present study, we were interested in the relationship between memory abilities and the rate of covert disfluency production. Our primary hypothesis was that in cases where memory retrieval was too slow to keep up with production, unfilled pauses and repetitions would be a likely outcome (similar hypotheses have been proposed for stuttering, see Howell & Au-Yueng, 1999). We employed a memorize-and-repeat sentence production task, which as noted in earlier work (e.g. Ferreira, 1991), reduces the need for conceptual-level processing, lexical selection, and syntactic planning. Therefore, this task

eliminates much of the difficult work of both conceptualization and grammatical encoding (Bock, 1996; Bock & Levelt, 1994; Ferreira & Engelhardt, 2006; Levelt, 1989). Instead, the memorize-and-repeat production task is primarily about memory encoding, memory retrieval, and speech production.[4] Based on previous research, we might expect covert disfluencies, in particular repetitions, to be related to verbal intelligence (e.g. Engelhardt et al., 2010). However, our primary hypothesis (given memory-associated task demands) was that a substantial number of covert disfluencies would be due to (slow) memory retrieval. One possibility is that the correlation between verbal intelligence and repetitions is because of shared variance between (verbal) intelligence and working memory ability (Miyake & Friedman, 2012). In this study, we assessed both verbal intelligence and working memory, and we also obtained two measures of executive functioning (i.e. Stroop task and Wisconsin Card Sorting task).

*Methodology.* We used six sub-tests from the Wechsler Adult Intelligence Scale (1997a) to create a two-factor latent variable model (see Figure 1). Latent variable modelling has several advantages given the goals of the study and the nature of the dataset. The first is that latent variables represent shared variance from multiple tasks used to tap the same underlying construct. Therefore, latent variables are less susceptible to idiosyncratic task properties (i.e. task impurity issues). The second advantage is that because measurement error is separated from the latent variable, the latent variable provides a purer measure of the constructs of interest. The third advantage is that latent variables are allowed to correlate, and thus, the shared variance between constructs is incorporated and quantifiable within the model. We employed a nested model fitting procedure to ascertain whether particular types of disfluency are associated with variability in verbal intelligence or working memory abilities.

---

[4] The only study we are aware of to examine the role of working memory load on production difficulty showed an increased number of subject-verb agreement errors in low-span participants, which was interpreted as evidence of memory resources being critical at a key stage of **syntactic planning** (Hartsuiker & Barkhuysen, 2006).

To do so, we added another observed variable (disfluency) to the model and drew pathways from each latent variable onto disfluency. We then ran hierarchical analyses in which we tested whether omitting one of these two pathways resulted in a significant decrease in model fit. If it does, then it suggests that that pathway is important and indicates substantial shared variance. In terms of sample size, it is recommended practice is to have a minimum of ten participants for each observed variable, and a minimum of three observed variables for each latent variable (Stevens, 2002).
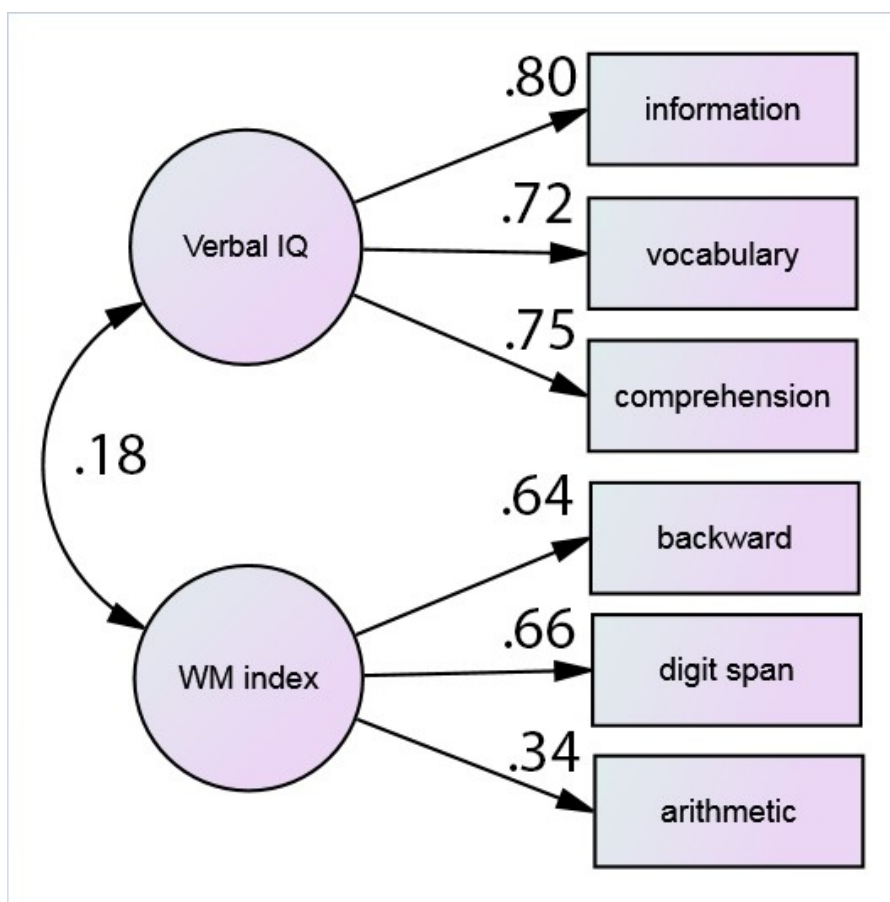


*Figure 1.* Two factor structural equation model.

*Summary.* In this study, participants memorized and repeated syntactically complex sentences. Each sentence contained a main and subordinate clause, and the order of clauses was reversed in half of the items (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Christianson, & Hollingworth, 2001: for examples, see Table 1). The memorize-

and-repeat production task does not require the speaker to generate semantic and syntactic content, but it does require the speaker to generate phonological and motor programmes. Importantly, this task increases memory demand compared to previous work, and our primary hypothesis was that covert disfluencies (i.e., unfilled pauses and repetitions) are due to (slow) memory retrieval, and if so, they should be related to individual differences in working memory. If not, then they should be related to individual differences in verbal intelligence, as reported by Engelhardt et al. Thus, our main interest was the relationship between covert disfluency and memory ability. However, we also coded several other types of disfluency as well. We noted filled pauses (i.e. *uh's, um's,* and *er's*) and repairs.

Table 1

*Example sentences.*

|  | Modifier |
| --- | --- |
| *Subordinate-Main* |  |
| 1. While the student read the notes blew off the desk. | No |
| 2. As Angela cleaned the dog that was spotted stood in the yard. | Yes |
| *Main-Subordinate* |  |
| 3. The politician gave a speech while the activist protested the war. | No |
| 4. The child that was annoyed sat on the bed while the mother calmed down. | Yes |

Method

*Participants*

Participants were 78 adults with a mean age of 20.0 years ($SD$ = 2.36, range 18.2 – 34.90). Sixty-nine were female, and nine were male. Approximately half were compensated £20 for taking part in the study, and half received participation credits from the undergraduate psychology pool at Northumbria University.[5]

---

[5] Seven participants were excluded from the analysis because too many of their utterances were either incomplete or were not recorded properly. These participants are not included in the total sample ($N$ = 78).

*Standardized Measures*

*Intelligence and Working Memory.* Participants completed seven subtests from the Wechsler Adult Intelligence Scale 3$^{rd}$ edition (Wechsler, 1997a). The verbal intelligence subtests were comprehension, information, similarities, and vocabulary, and the working memory subtests were arithmetic, backward digit span, and digit span. Vocabulary requires participants to provide the definitions of words and measures the degree to which one has learned and is able to express meanings verbally. Similarities requires participants to describe how two words are similar, with the more difficult items typically describing the opposite ends of a "unifying continuum". Thus, it measures abstract verbal reasoning. Comprehension requires participants to answer questions based on understanding of general world knowledge and social situations. The test is designed to assess verbal reasoning, verbal comprehension, and use past experience to demonstrate knowledge and judgements. Information requires the participant to address a broad range of general knowledge topics. It assesses the ability to retrieve general information. Arithmetic requires the participant to mentally solve arithmetic (story type) problems. It assesses a variety of abilities associated with numerical reasoning, short-term memory, and mental manipulation. It also requires concentration, attention, and avoidance of distraction. Digit span requires participants to encode and then recall a sequence of numbers in the same order as given by the examiner. The backward digit span requires participants to recall items given in the reverse order. These tasks assess working memory, manipulation, attention, and encoding.

*Wisconsin Card Sorting Test.* A computerized version of the Wisconsin Card Sorting Test was administered. The task requires participants to match a card to one of four other cards based on one of four attributes (shape, color, quantity, or design). There were four kinds of shapes (i.e. stars, crosses, triangle, and circles) and there were four colors (i.e. red, yellow, blue, and green). Participants were given feedback after every decision. After 5

correct decisions, the correct match attribute changed. This was repeated through nine cycles.

The dependent measure was number of perseveration errors, that is, the number of incorrect

decisions based on the previous match attribute (Anderson, Damasio, Jones, & Tranel, 1991;

Heaton, Chelune, Talley, Kay, & Curtiss, 1993).

*Stroop Task.* The Stroop task assesses the ability to monitor response conflict and

suppress a competing response in order to successfully execute the task requirements (i.e.

inhibition). Participants completed a computerized version of the Stroop task (Golden, 1978;

Stroop, 1935), which contained 160 trials in total. There were six conditions (i.e. 20 color-

congruent trials, 20 color-incongruent trials, 40 patch trials, 40 written trials, 20 word

congruent trials, and 20 word incongruent trials). Responses were recorded automatically

through the use of a voice trigger, and to compute an interference score, we subtracted the

mean reaction time (msec) of the color-congruent trials from the mean of the color-

incongruent trials (Lansbergen, Kenemans, & van Engeland, 2007).

*Sentence Production*

*Materials.* The 40 experimental items were based on sentences from Christianson et

al. (2001). There were 20 subordinate-main sentences which were all ambiguous and

contained optionally-transitive subordinate clause verbs. Half of these had a modifier

between the main clause subject and main clause verb (see Supplementary Materials, Section

A, for a list of critical items). The mean length of these sentences was 10.4 words (46.35

characters) and they ranged in length from 8 to 13 words. There were 20 main-subordinate

sentences, and the subordinate clause for these were all transitive. Of the 20, five contained

reflexive subordinate clause verbs, and half had a modifier between the main clause subject

and main clause verb. The mean length of these sentences was 11.15 words (49.4 characters)

and they ranged from 9 to 14 words. There were not length differences between the

subordinate-main and main-subordinate sentences (both *p's* > .16). None contained commas

separating the subordinate and main clauses, and there were 421 words in total in the experimental items.

      *Procedure.* The task was based on the procedure from Ferreira (1991). Participants were instructed that they would see a sentence that they had to memorize and repeat back, and that it was important that they spoke the sentence exactly as it was written and in a natural manner. Participants pressed the space bar and a fixation cross appeared for 1 second. The fixation cross was followed by the sentence, and it was presented in the centre of the computer screen. After participants had memorized the sentence, they pressed the space bar, and a question appeared on the screen (i.e. "What happened?"). Participants spoke the sentence out loud, and when they were finished speaking they pressed the space bar to start the next trial. There were three practice trials and 40 experimental items. The order of trials was randomly determined for each participant. If participants forgot the sentence on a particular trial, they could press the "R" key to go back and re-view the sentence. Partial recordings were not saved. Participants spoke into a condenser microphone in a sound dampened testing cubicle and the experiment was programmed with E-prime experimental software. The sentences were automatically recorded and saved as .wav files.

*Utterance Coding*

      *Memory errors.* Any errors in the utterance affecting content words were counted as memory errors. These included omissions of content words, incorrect inclusions, and incorrect substitutions (e.g. *archivist* vs. *activist, large* vs. *big, floor* vs. *ground,* etc.). Minor differences (e.g. *eating* vs. *eatin*, *book* vs. *books*) and differences involving function words (e.g. *the* vs. *a, have* vs. *has*) were not counted as memory errors.

*Disfluency.* Three main types of disfluency were examined: unfilled pauses, repetitions, and repairs.[6] Repetitions refer to unintended repeats of a word or string of words with no functional benefit. Repairs occur when a speaker suspends articulation, and then starts over with a new word or phrase. We assessed the lengths of all unfilled pauses that were 250 msec or greater. We viewed the threshold for an unfilled pause as a somewhat subjective decision because often researchers will utilize a higher threshold (e.g. $1 - 3$ seconds), so as to exclude prosodic pausing (Kormos & Denes, 2004; Lake, Humphreys, & Cardy, 2011). However, a recent study by De Jong and Bosker (2013) that investigated perceptions of fluency in L2 learners and accounted for speech rate, argued that 250 msec is the best threshold for unfilled pauses, and this is consistent with the original work of Goldman-Eisler (1968) (see also, Harley, 2013; Harley & MacAndrew, 2001; Redford, 2013). With a 250 msec threshold, approximately 16% of sentences contained at least one unfilled pause. The dataset was coded twice, once by the second author and once by a trained research assistant. The first author compared the two data files and resolved discrepancies. In cases in which the length of an unfilled pause differed by more than 50 ms, it was reassessed by the first author.[7] For the remainder, we averaged the two durations. The corpus contained 3120 sentences (approximately 33,000 words in total), and the proportion of sentences with a particular type of disfluency is shown in Figure 2.

---

[6] We also coded filled pauses, interjections, and blends. However, these affected less than ½ percent of the total utterances produced, and therefore, were excluded from all analyses.

[7] The correlations for the different types of disfluency between the undergraduate coders was extremely high (i.e. presence of disfluency = .99, filled pause = 1.00, unfilled pause = .88, repetitions = 1.00, repairs = .94, recall errors = .87). Thus, we did not calculate inter-rater reliabilities. There no cases in which the third rater (i.e. the first author) disagreed with the other two coders, and so, consensus was achieved in all cases.
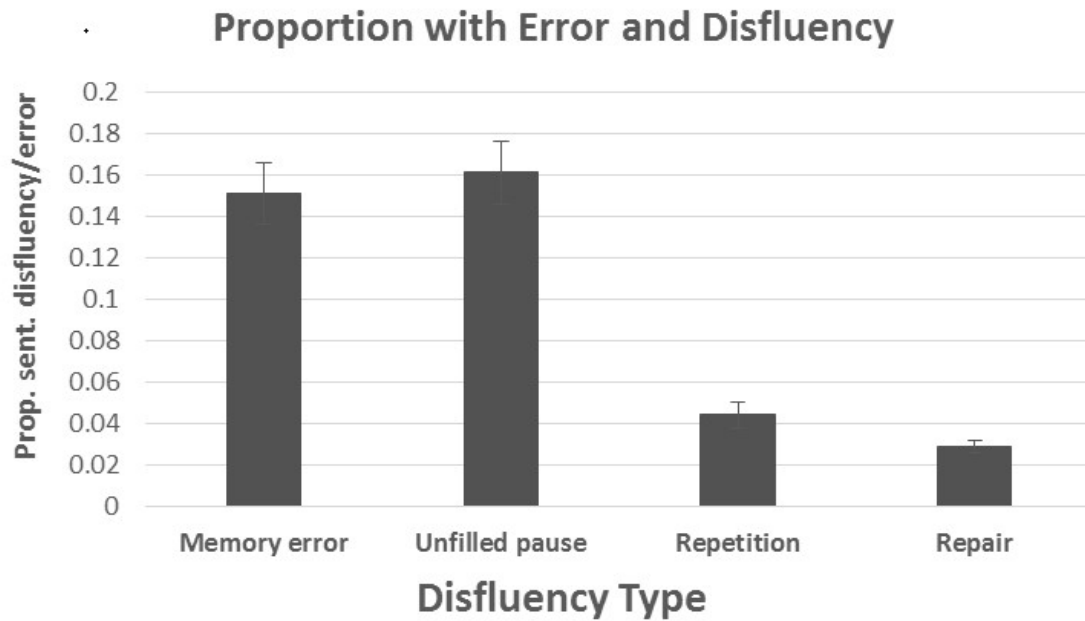
## Proportion with Error and Disfluency



*Figure 2.* Proportion of sentences containing a disfluency broken down by type and proportion of sentences with memory error. Error bars show the standard error of the mean.

*Procedure*

Participants were recruited via fliers posted on university grounds and by advertisement on the psychology undergraduate participation pool. Upon entering the lab, participants provided written informed consent and basic demographic information. They then completed each of the tests in the battery (verbal intelligence, working memory, Wisconsin Card Sort task, Stroop task, and the sentence production task). Tasks were completed in different rooms and in different testing cubicles, and participants were given obligatory breaks between tasks to avoid fatigue. Each participant completed the tasks in the same order. (Participants completed several additional tasks, including the autism spectrum quotient (AQ) and a second sentence production task.) The entire testing session lasted approximately 2 hours.

*Reliability*

The standardized measures used in the current study are all well-established tests with widely accepted reliability. The Wechsler intelligence tests (and the subscales) typically have

reported reliabilities in the .85 - .95 range (Friedman et al., 2006; Friedman et al., 2007; Wechsler, 1997a,b). The Stroop task and the Stop task have reported reliabilities in the .80 - .90 range (Friedman et al., 2007; Friedman & Miyake, 2004), and the Trails task and the Wisconsin Card Sort task typically have lower/borderline acceptable reliability ~.70 (for extended discussions of reliability in standardized executive function tasks see Denckla, 1996; Friedman & Miyake, 2004; Rabbitt, 1997). For the non-standardized measure (i.e. the sentence production task), we computed split-half reliabilities, and we used Spearman-Brown prophecy formula corrected coefficients (Brown, 1910; Spearman, 1910). The mean reliability for unfilled pause was $\alpha = .83$, repetitions $\alpha = .61$, repairs $\alpha = .37$, and memory errors $\alpha = .82$. The only type of disfluency that fell outside of traditionally accepted levels was repairs (Nunnally, 1978).

*Analysis procedures*

Structural equation models were created and run using AMOS. In the analyses, we report several fit indices for our models, which is recommended practice (Anderson & Gerbing, 1988; Gonzalez & Griffin, 2001; Hu & Bentler, 1995, 1998, 1999; Kane et al., 2004; Kline, 1998; Miyake et al., 2001). The chi-square statistic reflects significant differences between the observed covariance matrix and the reproduced covariance matrix. With chi-squared tests, a non-significant value is desirable (Satorra & Bentler, 2001). We also report the confirmatory fit index (CFI), which reflects improvement of the model fit relative to a baseline model in which all covariances are zero. This test also reflects the proportion of the observed covariance matrix explained by the model, and so, it reflects how well the model fits the data. The acceptability level of the CFI is .90 (Stevens, 2002). Finally, we report the root mean square error of approximation (RMSEA). Here values less than .05 indicate good fit (Kline, 1998). In cases where we wanted to compare (nested) models, we

utilized a chi-square difference test in order to determine whether one model fit the data significantly better than another.

<div align="center">Results</div>

*Data Screening and Preparation*

Data points greater than 3.0 standard deviations from the mean for each variable in the data set were defined as outliers. Outliers were replaced with the mean of that variable (McCartney, Burchinal, & Bub, 2006). This avoids listwise deletion and the corresponding reduction in power (Shafer & Graham, 2002). There were three outliers in the dataset, which were assessed via standardized values (vocabulary had one outlier -3.03, backward digit span had one outlier -3.13, and digit span had one outlier -4.89). [8] Finally, multivariate tests are sensitive to deviations from normality, and therefore, we applied transformations (i.e. square root, logarithm, or inverse) to the skewed variables in the dataset (Kline, 1998).

The descriptive statistics and bivariate correlations between variables are presented in Table 2 and Table 3, respectively. The first step in the analysis was to ensure that a two factor model fit the verbal intelligence and working memory measures better than a one factor model. Results showed that a two-factor model was generally a good fit $\chi^2(8) = 9.812$, $p = .28$, CFI = .980, RMSEA = .054. A one-factor model, in contrast, was not a good fit $\chi^2(9) = 23.674$, $p = .005$, CFI = .834, RMSEA = .146. A chi-squared difference test showed that model fit was significantly better with the two-factor model than with the one-factor model $\Delta\chi^2(1) = 13.86$, $p < .001$. (Factor loadings are shown in Figure 1.) Thus, there was statistical evidence supporting the dissociation of the two main hypothesized constructs (see also Ackerman, Beier, & Boyle, 2005; Martin, 2001).

---

[8] We acknowledge that this method of dealing with outliers has been criticized because it tends to underestimates the variance and makes it more likely to reject the null hypothesis (for discussion see Schafer & Graham, 2002). However, given the lower number of outliers and the nature of the analyses (i.e. SEM), we do not think this is an issue for our study, nor does it substantially affect our results and conclusions.

Table 2

*Descriptive statistics for the intelligence sub-tests, executive function measures, and the sentence production task*

| Measure | *N* | Mean | SD | Min | Max | Skew | Kurtosis |
|---|---|---|---|---|---|---|---|
| **Verbal Comprehension Index** | | | | | | | |
| Comprehension | 78 | 18.01 | 5.29 | 5.0 | 29.0 | -.114 | -.581 |
| Information | 78 | 15.71 | 3.65 | 7.0 | 24.0 | .012 | -.280 |
| Similarities | 78 | 22.28 | 3.91 | 13.0 | 30.0 | -.127 | -.405 |
| Vocabulary | 78 | 40.64 | 7.45 | 23.0 | 58.0 | -.108 | -.028 |
| **Working Memory Index** | | | | | | | |
| Arithmetic | 78 | 14.26 | 2.67 | 6.0 | 19.0 | -.491 | 0.080 |
| Backward Digit | 78 | 6.92 | 2.00 | 3.0 | 12.0 | 0.219 | -.722 |
| Digit Span | 78 | 10.40 | 1.74 | 6.0 | 15.0 | 0.215 | -.193 |
| **Executive Function Tasks** | | | | | | | |
| Stroop | 78 | 59.1 | 50.97 | -126.1 | 192.6 | -.123 | 1.913 |
| Perseveration Errors [b] | 78 | .06 | 0.03 | 0.01 | 0.13 | .067 | -.588 |
| **Proportion of Disfluencies** | | | | | | | |
| Unfilled Pauses [a] | 78 | .36 | .17 | 0.0 | 0.77 | -.06 | -.318 |
| Repetitions [a] | 78 | .16 | .14 | 0.0 | 0.47 | .236 | -.875 |
| Repairs [a] | 78 | .13 | .12 | 0.0 | 0.39 | .151 | -1.314 |
| Memory Errors [a] | 78 | .34 | .18 | 0.0 | 0.79 | .196 | 0.033 |

*Note.* [a]square root transformation, [b]inverse transformation, and [c]logarithm transformation.

Table 3

*Bivariate correlations between verbal intelligence, executive function, and disfluencies*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Age | - | -.05 | -.03 | -.12 | .01 | .00 | -.04 | -.04 | .06 | -.13 | .00 | .31** | -.14 | .14 | .12 |
| 2. Gender | | - | .24* | .20 | .08 | -.05 | -.07 | .22$^{\#}$ | .10 | -.17 | .13 | .12 | -.11 | -.07 | -.15 |
| 3. Comprehension | | | - | .59** | .40** | .55** | .27* | .05 | -.02 | .01 | -.03 | -.09 | -.25* | .09 | -.10 |
| 4. Information | | | | - | .28* | .57** | .29** | .15 | .10 | -.02 | -.07 | -.23* | -.18 | -.09 | -.09 |
| 5. Similarities | | | | | - | .45** | .17 | .14 | .02 | .14 | .12 | -.13 | -.29* | -.20 | -.09 |
| 6. Vocabulary | | | | | | - | .15 | .02 | -.02 | .05 | .15 | -.18 | -.26* | .07 | -.03 |
| 7. Arithmetic | | | | | | | - | .18 | .22* | .22$^{\#}$ | -.10 | -.12 | .16 | -.12 | -.19 |
| 8. Backward Digit | | | | | | | | - | .44** | .09 | .05 | -.13 | -.12 | -.19 | -.32** |
| 9. Digit Span | | | | | | | | | - | -.13 | -.05 | -.22$^{\#}$ | -.10 | -.10 | -.37** |
| 10. Stroop | | | | | | | | | | - | .07 | -.09 | .17 | .04 | -.13 |
| 11. Perseveration Errors | | | | | | | | | | | - | .02 | -.08 | -.20 | -.19 |
| 12. Unfilled Pauses | | | | | | | | | | | | - | .17 | .32** | .63** |
| 13. Repetitions | | | | | | | | | | | | | - | .12 | .15 |
| 14. Repairs | | | | | | | | | | | | | | - | .30** |
| 15. Memory Errors | | | | | | | | | | | | | | | - |

*Note.* *p* < .05, **p* < .01. Gender coded 0 = male and 1 = female

*Memory Errors*

The second step in the data analysis focused on memory errors or failures to correctly repeat the sentence. The bivariate correlations (in Table 3) revealed that memory errors were significantly correlated with the two working memory span tasks. We tested the proportion of sentences with a memory error in a two-factor (nested) path model, in which there were factor loadings from both verbal intelligence and working memory loading onto memory errors (see Figure 3). Results from the hierarchical tests showed that model fit was significantly worse when the regression weight from working memory to memory errors was set to zero (see Table 4). Models fits were good when memory errors loaded on working memory and were poor when memory errors were loaded onto verbal intelligence. This suggests that the memorize-and-repeat paradigm was operating as intended. We observed that successful task performance was related to working memory, and in particular, the span tasks.
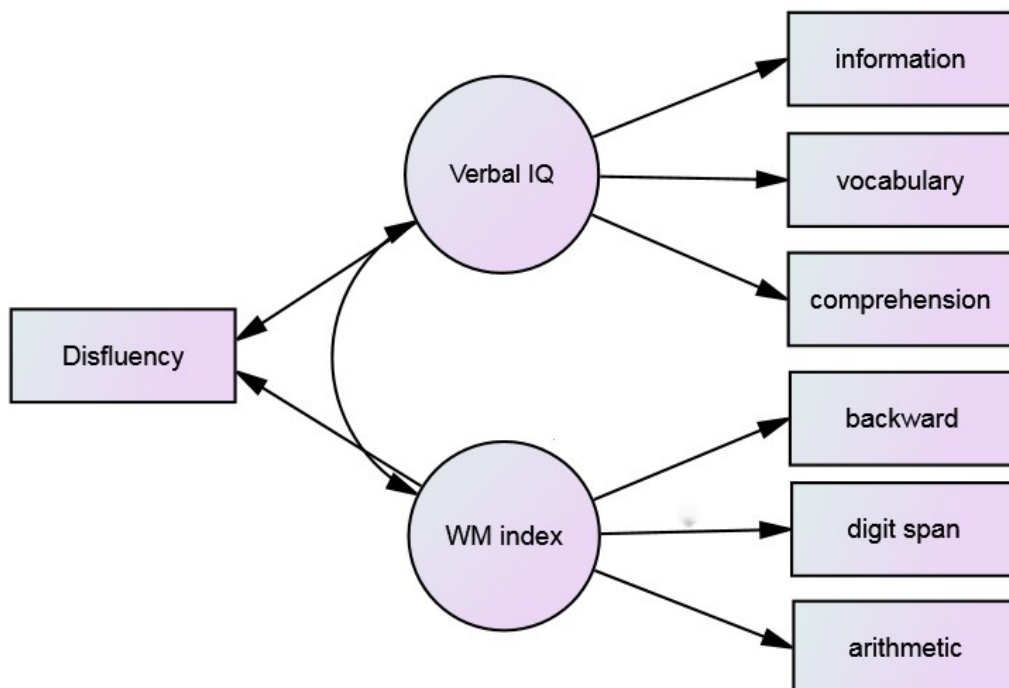


*Figure 3.* Nested path model.

*Disfluency*

We tested unfilled pauses and repetitions in hierarchical path models. We also tested repairs even though our main hypotheses focused on covert disfluency. The results are shown in Table 4.[9] For repetitions, we found that model fit was significantly worse when the regression weight from verbal intelligence to repetitions was set to zero. Model fits were generally good when repetitions loaded onto verbal intelligence (cf. RMSEA), and were poor when the verbal intelligence pathway was set to zero. This indicates that the tendency to produce repetitions is related to more general verbal abilities and not to individual differences in working memory. For unfilled pauses and repairs, there was a marginal decrease in model fit when the regression weight from working memory was set to zero (see Table 4), and in general, model fits were quite good in both sets of analyses. The marginality of these two findings is obviously unsatisfying from an empirical standpoint, especially given the size of the dataset. However, there are a couple of points that should be kept in mind. The first is that unfilled pauses were significantly correlated with memory errors ($r = .63$)[10], which was the highest correlation in the dataset. Repairs were also significantly correlated with memory errors, and repairs and unfilled pauses were significantly correlated with each other. Also, recall that memory errors clearly loaded onto the span tasks. Thus, the results indicate that there is shared variance between the working memory (span) tests, memory errors, and two types of disfluency (i.e. unfilled pauses and repairs). Repetitions, in contrast, were clearly related to verbal intelligence.

---

[9] We also looked at the proportion of sentences with different types of disfluency based on structure type (i.e. subordinate-main and main-subordinate). The results are presented in Section B of the Supplementary Materials, but there was little difference based on clause order. Thus, we collapsed results across the full set of items.

[10] This correlation was also positive ($r = .21$, $p < .001$) on sentence-level data. Thus, trials in which participants made a memory error were more likely to have an unfilled pause, suggesting that pauses accompany memory retrieval issues rather than a strategy for ultimately producing more correct recalls.

As a final analysis of the proportion of sentences with a disfluency, we conducted a posthoc model re-specification in which we removed arithmetic from the model and substituted memory errors. The factor loading for arithmetic was the lowest of the six Wechsler tests, and it did not correlate with the two span tasks, whereas memory errors were significantly correlated with both. Results of this follow up analysis were similar to the results presented in Table 4. There was a significant decrease in model fit $\Delta\chi^2(1) = 5.01$, $p <$ .05 when the regression weight from verbal intelligence to repetitions was set to zero. However, the results for unfilled pauses and repairs were both significant: Model fit significantly decreased when the regression weight from working memory was set to zero (unfilled pauses $\Delta\chi^2(1) = 27.50$, $p < .00001$ and repairs $\Delta\chi^2(1) = 4.46$, $p < .05$). We acknowledge the problems associated with model re-fitting post data collection. However, in this case, a minor increase in model fit led to a much clearer picture, especially in terms of statistical significance, of the relationship between individual differences in working memory and both unfilled pauses and repairs. In particular, it made a substantial impact on unfilled pauses. The factor loading from memory errors to the latent variable representing working memory was -.52, whereas the factor loading from arithmetic to working memory was .34. (The other factor loadings were virtually unchanged when arithmetic was replaced with memory errors.) So across both sets of hierarchical analyses, we see that repetitions are related to verbal intelligence and not to working memory, and unfilled pauses and repairs are more related to working memory than to verbal intelligence.

*Unfilled Pause Length*

Recall that we used a relatively low threshold for unfilled pauses, and results, at the worst, showed a marginal relationship between unfilled pauses and memory ability. However, the analyses thus far have only considered the presence/absence of an unfilled pause. As part of the data coding procedure, we also assessed the length of all unfilled pauses in the dataset.

If there were multiple pauses greater than 250 msec in length, then the length of each was noted. No sentence contained more than four unfilled pauses, and only two had four unfilled pauses. A histogram showing the distribution of pause length is shown in Figure 4. As can be seen, the distribution was positively skewed with the vast majority of unfilled pauses being shorter than 1 second, and the overall mean was 887 msec. To investigate the relationship between the *length of unfilled pauses* and the individual difference variables, we ran correlations between the individual difference variables and the mean length of first pauses, the mean length of second pauses, and the mean length of third pauses (see Table 5). Results from this follow up analysis revealed some interesting differences compared to the previous analyses. The previous analyses showed that the presence/absence of unfilled pauses was marginally-to-significantly related to working memory. However, the mean length of those unfilled pauses seem to be most closely related to the two executive function measures (Stroop task and perseveration errors). The Stroop task is most commonly used to assess inhibition and perseveration errors used to assess set shifting. However, there were also two marginally significant correlations between two of the verbal intelligence subtests and the length of first pauses. The relationship between pause length and the two executive function measures is interesting because the Stroop task and perseveration errors were not correlated to any of the other variables in the dataset. We return to this issue in the General Discussion.

Table 4

*Hierarchical structural equation analyses examining memory errors and disfluencies against verbal intelligence and memory ability*

| | Model fit | CFI | RMSEA | Factor Loadings | | $\Delta\chi^2$ difference |
|---|---|---|---|---|---|---|
| **Memory Errors** | | | | | | |
| *Saturated Model* | $\chi^2(12) = 10.34, p = .59$ | 1.00 | .000 | IQ = -.008 | WM = -.529 | |
| Verbal IQ set to 0 | $\chi^2(13) = 10.34, p = .67$ | 1.00 | .000 | IQ = .000 | WM = -.532 | $\Delta\chi^2(1) = 0.00$ |
| WM index set to 0 | $\chi^2(13) = 24.38, p = .03$ | .884 | .107 | IQ = -.116 | WM = .000 | $\Delta\chi^2(1) = 14.04$*** |
| **Repetitions** | | | | | | |
| *Saturated Model* | $\chi^2(12) = 18.74, p = .10$ | .931 | .085 | IQ = -.289 | WM = -.072 | |
| Verbal IQ set to 0 | $\chi^2(13) = 23.75, p = .03$ | .890 | .104 | IQ = .000 | WM = -.149 | $\Delta\chi^2(1) = 5.01$* |
| WM index set to 0 | $\chi^2(13) = 18.98, p = .12$ | .939 | .077 | IQ = -.302 | WM = .000 | $\Delta\chi^2(1) = 0.24$ |
| **Unfilled Pauses** | | | | | | |
| *Saturated Model* | $\chi^2(12) = 11.73, p = .47$ | 1.00 | .000 | IQ = -.187 | WM = -.247 | |
| Verbal IQ set to 0 | $\chi^2(13) = 13.82, p = .39$ | .991 | .029 | IQ = .000 | WM = -.298 | $\Delta\chi^2(1) = 2.09$ |
| WM index set to 0 | $\chi^2(13) = 14.60, p = .33$ | .982 | .040 | IQ = -.236 | WM = .000 | $\Delta\chi^2(1) = 2.87$# |
| **Repairs** | | | | | | |
| *Saturated Model* | $\chi^2(12) = 13.84, p = .31$ | .979 | .045 | IQ = .066 | WM = -.251 | |
| Verbal IQ set to 0 | $\chi^2(13) = 14.09, p = .37$ | .988 | .033 | IQ = .000 | WM = -.233 | $\Delta\chi^2(1) = 0.25$ |
| WM index set to 0 | $\chi^2(19) = 16.60, p = .22$ | .960 | .060 | IQ = .011 | WM = .000 | $\Delta\chi^2(1) = 2.76$# |

*Note.* #$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

Discussion

The focus of the current paper was the role of individual differences in the production of covert disfluencies, which occur when there is no overt error or observable correction but yet a clear disruption of speech output. As mentioned in the Introduction, covert disfluencies are empirically difficult to study because both the problem and correction of the problem remain hidden. To get around this difficulty, we selected a task that had a relatively narrow cognitive-demand profile, and specifically, we used a memorize-and-repeat sentence-production task (Ferreira, 1991). In previous work, we reported that repetitions (Engelhardt et al., 2010) and unfilled pauses (Engelhardt et al., 2013) were related to individual differences in (verbal) intelligence. Much of the previous literature suggested that covert disfluencies reflect any one of a number of problems in the course of speaking, but what we were interested in was the relationship between memory ability and covert disfluency. That is, do memory retrieval difficulties (or memory retrieval slowdowns) lead to higher rates of disfluency in sentence production? If they do, then it suggests that individual differences in memory can affect the fluency of speech outputs. It does not mean however, that it will in every situation or when memory demands are lower. The obvious caveat, which applies to all laboratory-based production studies, is that everyday speaking situations are much different as various cognitive demands (e.g. attention, memory, turn taking, comprehension, etc.) fluctuate in real time to impact ongoing speech planning and output. However, our main interest in this study was the relationship between working memory and covert disfluency. We wanted to isolate that relationship, and in particular, in the context of previous work (e.g. Engelhardt et al., 2013), contrast individual differences in memory ability with individual differences in verbal intelligence. Thus, we felt that a controlled speaking situation was necessary given the specificity of the hypotheses, and thus, all participants in our study produced the same words and the same syntactic structures, which ensured that production

demands were equal across participants. In general, we observed that the relationships between disfluency production and the individual difference variables were negative (i.e., lower ability individuals tended to produce more disfluencies). In the remainder of the discussion, we cover the proportion of sentences with a particular type of disfluency, the effect of executive function on the length of hesitations, the strengths and limitations, and end with conclusions.

Table 5

*Bivariate correlations between individual differences measures and unfilled pause length*

| Variable | Length 1[a] | Length 2[b] | Length 3[c] |
|---|---|---|---|
| 1. Comprehension | -.09 | -.09 | -.19 |
| 2. Information | -.03 | -.04 | -.07 |
| 3. Similarities | -.21[#] | -.21 | -.14 |
| 4. Vocabulary | -.23[#] | -.17 | .21 |
| 5. Arithmetic | -.07 | -.20 | .42 |
| 6. Backward Digit | -.16 | -.27 | -.03 |
| 7. Digit Span | -.17 | -.12 | -.18 |
| 8. Stroop | -.22[#] | -.37* | -.48[#] |
| 9. Perseveration | -.29* | -.32[#] | -.33 |

*Note.* [#]$p < .08$, *$p < .05$, **$p < .01$. [a]Seventy-three participants produced at least one unfilled pause. [b]Thirty-eight participants produced at least two unfilled pauses on one trial. [c]Fourteen participants produced at least three unfilled pauses on one trial.

*Types of Disfluency*

To summarize the main findings, we found that repetitions were significantly related to verbal intelligence, and not to working memory. Unfilled pauses and repairs, in contrast, were (at worst) marginally related to working memory. A follow up analysis involving a (post-hoc) model tweak resulted in significant effects for both unfilled pauses and repairs on working memory. Setting the post-hoc significance issues aside for the moment, repairs showed a nearly identical regression weight as unfilled pauses on working memory, and both

repairs and repetitions showed no relationship with the "other" latent variable. That is, repetitions were not related to working memory, and repairs were not related to verbal intelligence (both regression weights < .072). Unfilled pauses were different, as there was some systematic variance between unfilled pauses and verbal intelligence. In the Introduction, we made an argument about the importance of being able to account for shared versus unique variance. For researchers working on the inter-relations of executive functioning and intelligence (e.g. Miyake & Friedman, 2012) this may seem obvious. However, as individual difference studies of language production (and comprehension) have begun to appear, it is not always clear that language researchers fully appreciate the issues associated with shared variance. For example, Vuong and Martin (2013) had participants complete a sentence comprehension task and a Stroop task, and based on a correlation between the two, argued that executive control is important in language comprehension. However, from one correlation it is impossible to determine whether the variance is shared or unique. The Stroop variance may well be shared variance with working memory, and thus, to make claims about executive function-*x* vs. executive function-*y,* it is advisable to simultaneously model both. Our data has an example of how bi-variate correlations can be somewhat misleading. From Table 3, it is clear that repetitions are more related to intelligence than to working memory, and repairs are more related to working memory. However, unfilled pauses showed slightly higher correlations with intelligence, but the hierarchical tests revealed slightly more systematic (shared) variance with working memory. The main reason this oddity can happen is because our model contained both working memory and verbal intelligence, and it also estimated the correlation between the two. Thus, by modelling working memory and verbal intelligence, we were able to gain what we think is a clearer picture compared to studies that consider only one ability or those that examine individual difference variables in separate analyses (i.e. through simple correlations).
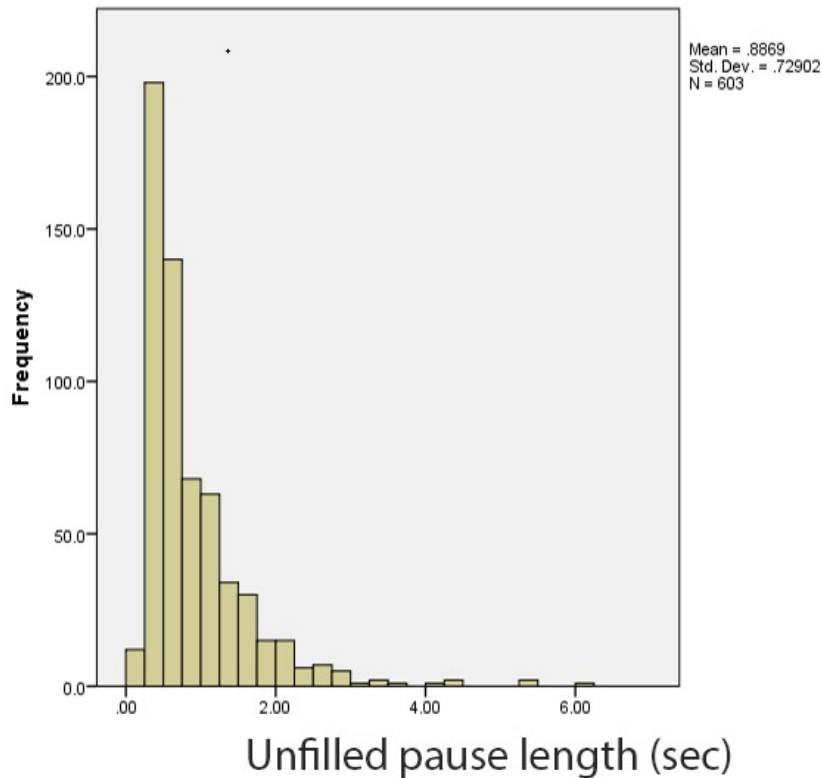
*Figure 4.* Histogram showing the distribution of the length of all unfilled pauses.

There are several points worth considering, before we move onto the length of unfilled pauses. Recall that consistent with our previous research, repetitions were related to verbal intelligence, and the regression path for the full model was -.29, which according to Cohen (1988) represents a medium effect size. An obvious issue that arises from this finding is what exactly is verbal intelligence and why might it be important to avoid repeating oneself? Answers to these questions usually begin by acknowledging that verbal intelligence is primarily based on crystallized knowledge, or knowledge acquired through prior experience (e.g. Bates & Stough, 1997; Carroll, 1993; Deary, 2001; Spearman, 1927). Explanations (e.g. Luke, Henderson, & Ferreira, 2015) tend to then highlight the quality of experience leading to quality of lexical representations, including a greater number of lemmas, a greater specificity of individual lemmas, faster access, etc. However, performance on individual verbal intelligence subtests might be slightly distinct.

The second point concerns a study by Lake et al. (2011), who assessed individuals with high-functioning forms of ASD in order to investigate speaker-oriented vs. listener-oriented disfluency (see also, Belardi & Williams, 2009; Shriberg et al., 2001; Suh et al., 2014; Tager-Flusberg, Paul, & Lord, 2005; Thurber & Tager-Flusberg, 1993). Lake et al. showed a dissociation in which individuals with HFA produced fewer filled pauses and repairs, and more unfilled pauses and repetitions when compared to typically-developing controls (cf. Engelhardt et al., 2017). They argued that the former are listener-oriented and the latter are speaker-oriented. Before delving into the dissociation between the different types of disfluency, it is important to note that Lake et al. examined production in dialogue and employed different methods for assessing disfluencies, most notably their threshold for unfilled pauses was much longer. If we consider first repetitions and repairs, which were assessed similarly in the two studies, Lake et al. argued that they are due to different underlying causes, repetitions to speaker-internal factors and repairs to listener-oriented factors. Our results showed that distinct individual difference variables were related to repetitions and repairs (verbal intelligence and working memory, respectively); however, the direction of the effects were the same (i.e. lower functioning individuals produced more disfluency). If speakers produce repairs for the benefit of the listener, then one might expect the direction of the effect to be reversed: Higher-ability individuals should be more attuned and accommodating to listeners' needs compared to lower-ability individuals. Again, it is important to bear in mind that there may be task differences that have a role in these discrepancies, but at the same time, our task is the presumably easier of the two. The Lake et al. task was much more spontaneous and involved interacting with an experimenter, whereas our task was essentially scripted monologue.[11] There are two other bits of evidence that support a negative relationship between cognitive abilities and repair disfluencies. First,

---

[11] We are aware of only one study (Finlayson & Corley, 2012) that compared monologue and dialogue in the same participants. However, results showed little difference in rates and types of disfluency between conditions.

Engelhardt et al. (2013) showed a negative relationship between repairs and inhibitory control. Second, in the current study, most of the participants also completed the autism quotient (AQ) questionnaire, which is widely used to assess autism spectrum symptoms in typically-developing individuals. However, results show no correlation between repairs and total AQ scores ($r = -.001$, $p = .99$, $N = 75$), and all six of the subscales were likewise non-significant (all $r's < .15$, $p's > .20$). Thus, the Lake et al. study is the only one we are aware of to suggest that repairs are listener-oriented.

Finally, we think it is important to highlight what these results mean moving forward both in terms of practical and theoretical significance. What we have shown in this study is that there are some reasonably strong associations between individual difference variables and the production of different types of disfluency. At this point, we cannot say for certain whether the two are causally related or merely correlational. However, given the clinical work looking at ADHD and ASD, and the increased rates of disfluency in both disorders, we strongly suspect that there are causal links between individual differences in executive functioning and intelligence and the ability to produce fluent language outputs. Similar, causally-based conclusions are routinely made in the executive function literature (e.g. Miyake & Friedman, 2012), which we reviewed in the Introduction. Future work is needed to conclusively establish causal relationships, as well as greater specificity regarding which stages of language production are affected/related to individual differences.

*Unfilled Pause Length*

As part of the analysis protocol, we also assessed the length of all silent pauses longer than 250 msec. Results revealed that the length of unfilled pauses correlated with the Stroop task and perseveration errors, whereas the proportion of sentences with a disfluency did not. We found the correlations with length quite surprising. We chose a task that relied on memory with the idea that pausing and repeating would be associated with individual

differences in memory ability, and if so, then we also assumed that the length of pauses should also be related to memory (or perhaps, speed of memory retrieval). Given that the executive function measures correlated with pause length, and given the fact that the executive function measures did not correlate with proportion of sentences with a disfluency (Stroop task, $r = -.09$ and perseveration errors, $r = .02$), we speculate that (1) pausing behaviour is moderately associated with memory (as one would expect with a high-demand memory task) and (2) that how the production system gets "back on track" following a disruption is more related to individual differences in cognitive control. Consistent with our speculations, a recent study by Hussey, Ward, Christianson, and Kramer (2105) found increased performance on a sentence comprehension task following stimulation of the left lateral prefrontal cortex, an area strongly suspected of being involved in executive control in the language domain. Unfortunately, in the current study, we did not have multiple measures of inhibition and set shifting, and so, we are not able to model these executive functions in the same way we did with working memory and verbal intelligence. Thus, we are not in a position to comment on how much of the variance is shared and how much is unique.

*Strengths and Limitations*

We think the main strengths of this study are (1) the relatively large sample and (2) that we modelled two individual differences variables and tested them within a single model. The size of the sample ($N = 78$) and the size of test battery are unique for language production research, which tends to require more time consuming analyses. There are however several limitations. The first one is that the task we used required rote memorization of a sentence, which is very much unlike everyday speaking situations. However, we thought that a controlled speaking task was necessary given our goals, and thus, we employed a task in which all participants produced the same words and the same syntactic structures, which ensures equal task demands across the entire sample. A second limitation is that we are not in

a position to model inhibition or set shifting. This is important given the correlations with the length of unfilled pauses. However, when designing individual differences studies, there are usually limitations on the number of measures that can be administered, and thus, some sacrifices must be made. However, the correlations between Stroop performance and perseveration errors provide an obvious future direction for looking at the role of cognitive control in the length of speech pauses. Another obvious avenue for future research is to consider the role of attention, which has a tendency to be the odd man out in investigations of executive function (e.g. Jou & Harris, 1992; Miyake et al., 2000).

*Conclusions*

In the current study, we found that repetitions were significantly related to individual differences in verbal intelligence, which is consistent with Engelhardt et al. (2010). Unfilled pauses and repairs, in contrast, were marginally more related to the working memory measures compared to verbal intelligence. That is, model fits were marginally decreased when the pathway from disfluency to working memory was eliminated. To our knowledge, no study has previously investigated the length of pauses in speech and individual differences. The extent to which language production is influenced by individual differences has only recently become the topic of empirical studies. In our view, individual difference studies have the potential to shed new light and to expand theoretical debates in language production. As Underwood (1975) eloquently argued more than 40 years ago, individual differences studies are one of the key drivers of theory construction (Levelt, 1989) and modelling efforts (Dell, Chang, & Griffin, 1999; Dell, Burger, & Svec, 1997; Samuelson, Jenkins, & Spencer, 2015). However, there are a range of important research questions relating to individual differences. For example, do individual differences in executive function (e.g. inhibition) affect particular stages or processes of sentence production (e.g. lemma selection) (Berg & Schade, 1992; Wheeldon & Levelt, 1995)? With regards to

monitoring, does the production system rely on a domain-general error monitoring mechanism (Holroyd & Coles, 2002) or is the language monitoring device domain-specific (Ganushchak & Schiller, 2008; Oomen & Postma, 2001, 2002)? We think these results shed new light on the factors influencing the production of covert disfluency in sentence production, and our hope is that this study sparks future research that considers how individual differences in executive function and intelligence impact on the fluency of speech outputs.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin, 131,* 30-60.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423.

Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin card sorting performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology, 13,* 909–922.

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 15,* 31–36.

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76(1)*, 28- 55.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new - Disfluency and reference resolution. *Psychological Science, 15*, 578-582.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multicomponent model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge Univ. Press.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 1,* 65-94.

Barkley, R. A., & Murphy, K. R. (2006). *Attention deficit hyperactivity disorder: A clinical workbook-3rd Edition*. New York: Guilford Press.

Barr, D. J. (2001).Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.) *Oralité et gestualité: Interactions et comportements multimodaux dans la communication*, pp. 597-600.

Paris: L'Harmattan.

Bates, T. C., & Stough, C. (1997). Processing speed, attention, and intelligence: Effects of spatial attention on decisions time in high and low IQ subjects. *Personality & Individual Differences, 23(5),* 861-868.

Berg, T., & Schade, U. (1992). The role of inhibition in a spreading activation model of language production, Part 1: The psycholinguistic perspective. *Journal of Psycholinguistic Research, 22,* 405–434.

Belardi, K. M., & Williams, D. L. (2009). *Disruptions in spontaneous language in HFA: Indicators of linguistic processing challenges.* Presented at the International Meeting for Autism Research, Chicago.

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and timing of repairs in spontaneous speech. *Cognition, 39,* 173-194.

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3,* 395-421.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in spontaneous speech: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123-147.

Brennan, S. E., & Williams, M. (1995). The feeling of Another′ s Knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34(3)*, 383-398.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3,* 296–322.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, UK: Psychology Press.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42,* 368-407.

Clark, H. H. (1994). Managing problems in speaking. *Speech Communication, 15,* 243-250.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73-111.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37,* 201-242.

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

Cooper, P. V. (1990). Discourse production and normal aging: Performance on oral picture description tasks. Journal of Gerontology. *Psychological Sciences, 45,* 210-214.

Deary, I. J. (2001). *Intelligence: A very short introduction*. Oxford: Oxford University Press.

Deese, J. (1984). Pauses, prosody, and the demands of production in language. In H. W. Dechert & M. Raupach (Ed.), *Temporal Variables in Speech* (pp. 69-84). The Hague: Mouton.

De Jong, N.H., & Bosker H.R. (2013). Choosing a threshold for silent pauses to measure second language fluency. *Proceedings of the $6^{th}$ Workshop on Disfluency in Spontaneous Speech (DISS)* (pp. 17-20).

Dell G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review, 93,* 283–321.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104,* 123-147.

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science, 23,* 517–542.

Dell, G. S. & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287-314.

Denckla, M. B. (1996). A theory and model of executive function: A neuropsychological perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 263–278). Baltimore, MD: Brookes.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology, 14,* 713–741.

Engelhardt, P. E., Alfridijanta, O., McMullon, M., Corley, M. (2017). Speaker-versus listener-oriented disfluency: A re-examination of arguments and assumptions from Autism Spectrum Disorder. *Journal of Autism Developmental Disorders,* 1-14.

Engelhardt, P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory & Cognition, 38,* 617-628.

Engelhardt, P. E., Ferreira, F., & Nigg, J. T. (2009). Priming sentence production in adolescents and adults with attention-deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology, 37,* 995-1006.

Engelhardt, P. E., Ferreira, F., & Nigg, J. T. (2011). Language production strategies and disfluencies in multi-clause network descriptions: A study of adult attention-deficit hyper-activity disorder. *Neuropsychology, 25(4),* 442-453.

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica, 144,* 424-432.

Engelhardt, P. E., Veld, S., Nigg, J. T., & Ferreira, F. (2012). Are language production problems apparent in adults who no longer meet diagnostic criteria for Attention-Deficit/Hyperactivity Disorder? *Cognitive Neuropsychology, 29,* 275-299.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30,* 210-233.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: implications for models of reanalysis. *Journal of Psycholinguistic Research, 30,* 3-20.

Ferreira, F., & Engelhardt, P. E. (2006). Syntax and Production. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 61-91). Oxford, UK: Elsevier Inc.

Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin and Review, 19,* 921-928.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12(6)*, 627-635.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*, 151.167.

Fraundorf, S. H., & Watson, D. G. (2013). Alice's adventures in *um*-derland: psycholinguistic sources of variation in disfluency production. *Language, Cognition, and Neuroscience, 29,* 1083-1096.

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescents. *Psychological Science, 18,* 893-900.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control processes: A latent variable analysis. *Journal of Experimental Psychology: General, 133,* 101–135.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17,* 172-179.

Fromkin, V. (1973). *Slips of the tongue.* San Francisco: WH Freeman.

Ganushchak, L. Y., & Schiller, N. O. (2008). Brain error-monitoring activity is affected by semantic relatedness: An event-related brain potentials study. *Journal of Cognitive Neuroscience, 20,* 927-940.

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation, 9,* 133-177.

Garrett, M. F. (1982). Production of speech; Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19-76. London: Academic Press.

Golden, C. J. (1978). Stroop Color and Word Test: A Manual for Clinical and Experimental Issues. Wood Dale, IL: Stoelting.

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

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every ''one'' matters. *Psychological Methods, 6,* 258–269.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82(1)*, B1-B14.

Harley, T.A. (2013). *The psychology of language: From data to theory.* Psychology Press, New York.

Harley, T.A., & MacAndrew, S.B.G. (2001). Constraints upon word substitutions speech errors. *Journal of Psycholinguistic Research, 30,* 395-418.

Hartsuiker, R.J., & Barkhuysen, P.H. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes, 21,* 181-204.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology, 42,* 113-157.

Hartsuiker R. J., & Notebaert L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology, 57*, 169-171.

Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G. & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and expanded.* Odessa, FL; Psychological Assessment Resources.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine and the error-related-negativity. *Psychological Review, 109,* 679-709.

Howell, P., & Au-Yueng, J. (1999). The EXPLAIN theory of fluency control applied to the diagnosis of stuttering. *Amsterdam Studies in the the Theory and History of Linguisitic Science Series, 4,* 75-94.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to under-parameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L., & Bentler, P. M. (1999). Cuttoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1-55.

Hussey, E. K., Ward, N., Christianson, K., & Kramer, A. F. (2015). Language and memory improvements following tDCS of executive-control brain regions. *PLoS ONE,* Nov. 3, 2015, DOI:10.1371/journal.pone.0141417.

Jester, J. J., Nigg, J. T., Puttler, L. I., Long, J., Fitzgerald, H. E., & Zucker, R. A. (2009). Intergenerational transmission of neuropsychological executive functioning. *Brain and Cognition, 70,* 145-153.

Jou, J., & Harris, R.J. (1992). The effect of divided attention on speech perception. *Bulletin of the Psychonomic Society, 30,* 301-304.

Kane, M. J., Hambrick, D. Z., Wilhelm, O., Payne, T., Tuholski, S., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189-217.

Kempen, G. (1978). Sentence construction by a psychologically plausible formulator. In Campbell, R., & Smith, P. (Eds.). Recent advances in the psychology of language: Formal and experimental approaches. New York: Plenum Press.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14(2)*, 185-209.

Kemper, S. (1992). Adults' sentence fragments Who, What, When, Where, and Why. *Communication Research, 19,* 444-458.

Kline, P. (1991). *Intelligence: the psychometric view.* London: Routledge.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32,* 145-164.

Lake, J.K., Humphreys, K.R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin and Review, 18,* 135-140.

Lansberge, M.M., Kenemans, J.L., & van Engeland, H. (2007). Stroop interference and attention-deficit/hyperactivity disorder: a review and meta-analysis. *Neuropsychology, 21,* 251-262.

Le Dorze, G., & Bedard, C. (1998). Effects of age and education on the lexico-semantic content of connected speech in adults. Journal of Communication Disorders, 31, 53-71.

Levelt, W. J. M. (1981). Linearization in describing spatial networks. In S. Peters & E. Saarinen (Eds.), *Processes, beliefs, and questions (pp.199-220).* London, England: D. Reidel Publishing Company.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint for the speaker. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83-122). Oxford: Oxford University Press.

Levelt, W. J. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences, 98(23)*, 13464-13471.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22(1),* 1-38.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica, 60,* 193–210.

Luke, S.G., Henderson, J.M., & Ferreira, F. (2015). Children's eye-movements during reading reflects the quality of lexical representations: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomenon in spontaneous English speech. *Word, 15,* 19-44.

Martin, N.G. (2001). Genetic covariance among measures of information processing speed, working memory, and IQ. *Behavior Genetics, 31,* 581–592.

Martel, M., Nikolas, M., & Nigg, J. T. (2007). Executive function in adolescents with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 1437-1444.

McCartney, K., Burchinal, M. R., & Bub, K. L. (2006). Best practices in quantitative methods for developmentalists. *Monographs of the Society for Research in Child Development, 71,* 1 – 145.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21,* 8-14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex ''frontal lobe'' tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related changes in object naming: A review. *Language and Cognitive Processes, 21,* 238-290.

Nigg, J. T. (2001). Is ADHD a disinhibtion disorder? *Psychological Bulletin, 127,* 571-598.

Nigg, J. T., Carr, L. A., Martel, M. M., & Henderson, J. M. (2007). Concepts of inhibition and developmental psychopathology. In C. MacCleod & D. Gorfein (Eds.), *Inhibition in Cognition*, (pp. 259-277). American Psychological Association Press.

Nooteboom, S. G. (1980). Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand.* New York: Academic Press.

Nunnally, C.J. (1978). *Psychometric Methods.* New York: Harper and Row.

O'Connell, D. C., & Kowal, S. (2005). *Uh* and *Um* Revisited: Are They Interjections for Signaling Delay? *Journal of Psycholinguistic Research, 31,* 555-576.

Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research, 30,* 163-184.

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes, 17,* 163-184.

Paivio, A., Clark, J. M., Digdon, N., & Bons, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition, 17(2)*, 163-174.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37,* 51-87.

Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one  stage. *Journal of Memory and Language, 46(3)*, 586-605.

Postma, A. (1997). On the mechanisms of speech monitoring. In W. Hulstijn, H. Peters, &P. Van Lieshout (Eds.), *Speech production: motor control, brain research, and fluency disorders.* Amsterdam: Elsevier.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77,* 97-131.

Postma, A., Kolk, H., & Povel, D. J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech, 33,* 19-29.

Rabbitt, P. (Ed.). (1997). *Methodology of frontal and executive function.* Hove, UK: Psychology Press.

Redford, M.A. (2013). A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics, 34,* 569-589.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modelling attentional control in the Stroop task. *Psychological Review*, *110*, 88-125.

Samuelson, L.K., Jenkins, G.W., & Spencer, J.P. (2015). Grounding cognitive-level processes in behavior: The view from dynamic systems theory. *Topics in Cognitive Science, 7,* 191-205.

Sandson, J., Obler, L. K., & Albert, M. L. (1987). Language changes in healthy aging and dementia. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (pp. 264-292). New York: Oxford University Press.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514.

Schachar, R., Tannock, R., Marriott, M., & Logan, G. D. (1995). Deficient inhibitory control and attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology,*

*23,* 411-437.

Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W. R. (2000). Aging and word-finding: A comparison of spontaneous and constrained naming tests. *Archives of Clinical Neuropsychology, 15,* 479–493.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors  on disfluency production. *In Proceedings of the 28th Cognitive Science Society,* 1-6.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29(1),* 86-102.

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms  for 590 pictures in Dutch. *Acta psychologica, 119(2),* 159-187.

Shafer, J. L, & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147-177.

Shao, Z., Meyer, A.S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory and Cognition, 41,* 1200-1211.

Shao, Z., Roelofs, A., & Meyer, A.S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology, 65,* 1927-1944.

Shriberg, E. (1996). Disfluencies in Switchboard. *Proceedings, International Conference on Spoken Language Processing (ICSLP),* Vol. Addendum, 11-14. Philadelphia, PA, 3 – 6 October.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning

autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research, 44,* 1097–1115.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32(1),* 25-38.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3,* 271- 295.

Spearman, C. (1927). *The abilities of man.* London: Macmillan.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences.* Hillsdale, N.J.: Earlbaum.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Suh, J., Eigsti, I., Naigles, L., Barton, M., Kelly, E., & Fein, D. (2014). Narrative performance of optimal outcome children and adolescents with a history of an autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders, 44,* 1681-1694.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language, 53(1),* 81-94.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and Communication in Autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders*, Vol. 1: Diagnosis, development, neurobiology, and behavior (3rd ed., pp. 335–364). Hoboken, NJ: Wiley.

Tannock, R., & Schachar, R. (1996). Executive dysfunction as an underlying mechanism of behavior and language problems in attention deficit hyperactivity disorder. In J. Beitchman, N. Cohen, M. M. Konstantearas, and R. Tannock (Eds.) *Language,*

*learning, and behavior disorders* (pp. 128-155). Cambridge: University Press.

Teuber, H. L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis, 32,* 615–656.

Thurber, C., & Tager-Flusberg, H. (1993). Pauses in the narratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand. *Journal of Autism and Developmental Disorders, 23,* 309–322.

Underwood, B.J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30,* 128-134.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *Quarterly Journal of Experimental Psychology, 48(4)*, 822-848.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and lexicalist grammar. *Cognition, 75,* 105-143.

Vuong, L. C., & Martin, R. C. (2013). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition, and Neuroscience, 29,* 312-325.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale (3rd ed.)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Intelligence Scale for Children (4th ed.)*. San Antonio: The Psychological Corporation.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34,* 311-334.

Acknowledgements

Supplementary Materials

Section A

List of experimental items, which were based on Christianson et al. (2001). Struct = structure type, SM = subordinate-main, MS = main-subordinate, Gpath = garden path sentence yes/no, Mod = modifier yes/no, Verb = verb type, opt = optionally transitive, ref = reflexive, L(word) = length in terms of number of words, L(characters) = length in terms of number of characters.

| Number | Struct | Sentence | Gpath | Mod | Verb | L(Word) | L(Characters) |
|---|---|---|---|---|---|---|---|
| 1 | SM | While the secretary typed the memo neared completion. | y | n | opt | 8 | 45 |
| 2 | SM | While the student read the notes blew off the desk. | y | n | opt | 10 | 41 |
| 3 | SM | As the artist painted the model sat in the chair. | y | n | opt | 10 | 40 |
| 4 | SM | While the director filmed the actor recited the lines. | y | n | opt | 9 | 45 |
| 5 | SM | While the puppy sniffed the kitten sat on the sofa. | y | n | opt | 10 | 41 |
| 6 | SM | As the professor lectured the students took notes. | y | n | opt | 8 | 42 |
| 7 | SM | While Jack ordered the fish cooked in a pot. | y | n | opt | 9 | 35 |
| 8 | SM | As the mare fed the colt stamped its hoof. | y | n | opt | 9 | 33 |
| 9 | SM | As the maid dusted the picture tipped over. | y | n | opt | 8 | 35 |
| 10 | SM | While Rick drove the car veered into a ditch. | y | n | opt | 9 | 36 |
| 11 | SM | While the skipper sailed the boat that was leaky veered off course. | y | y | opt | 12 | 55 |
| 12 | SM | As Angela cleaned the dog that was spotted stood in the yard. | y | y | opt | 12 | 49 |
| 13 | SM | As Mark vacuumed the drapes that were pleated hung in the window. | y | y | opt | 12 | 53 |
| 14 | SM | As the cowboy rode the horse that was big sweated profusely. | y | y | opt | 11 | 49 |
| 15 | SM | While the clown juggled the balls that were bright fell on the ground. | y | y | opt | 13 | 57 |
| 16 | SM | As the champion raced the challenger that was strong stumbled and fell. | y | y | opt | 12 | 59 |
| 17 | SM | While the man hunted the deer that was graceful ran into the woods. | y | y | opt | 13 | 54 |
| 18 | SM | While the chef stirred the soup that was spicy boiled over. | y | y | opt | 11 | 48 |
| 19 | SM | While the scientists explored the cave that was dark grew cold. | y | y | opt | 11 | 52 |
| 20 | SM | As the athlete wrestled the opponent that was sweaty shouted insults. | y | y | opt | 11 | 58 |
| **Means** | | | | | | **10.4** | **46.4** |
| 21 | MS | The rocket sat on the launch pad while the reporter photographed the astronaut. | n | n | opt | 13 | 66 |

| 22 | MS | The hot dog began to burn while Tom grilled the hamburger. | n | n | opt | 11 | 47 |
| 23 | MS | The child stood on the pavement while the caricaturist drew a lady's portrait. | n | n | opt | 13 | 65 |
| 24 | MS | The adventurer lead the group while the navigator followed his instincts. | n | n | opt | 11 | 62 |
| 25 | MS | The parent shouted as the child rudely imitated his grandfather. | n | n | opt | 10 | 54 |
| 26 | MS | The politician gave a speech while the activist protested the war. | n | n | opt | 11 | 55 |
| 27 | MS | The girl sliced some ham while the boy drank some milk. | n | n | opt | 11 | 44 |
| 28 | MS | The phone rang as the girl called for her dog. | n | n | opt | 10 | 36 |
| 29 | MS | The teacher helped the boy while the girl ran away. | n | n | opt | 10 | 41 |
| 30 | MS | The instructor danced while the students observed the film. | n | n | opt | 9 | 50 |
| 31 | MS | The dog leapt as the boy hit the door. | n | n | opt | 9 | 29 |
| 32 | MS | The car broke down while the punks stole the bike. | n | n | opt | 10 | 40 |
| 33 | MS | The child that was annoyed sat on the bed while the mother calmed down. | n | y | ref | 14 | 57 |
| 34 | MS | The child that was pudgy giggled while Jim bathed. | n | y | ref | 9 | 41 |
| 35 | MS | The gorilla that was large sat in the grass while the chimp groomed. | n | y | ref | 13 | 55 |
| 36 | MS | The baby that was cute spit up on the bed while Anna dressed. | n | y | ref | 13 | 48 |
| 37 | MS | The dog that was furry barked while the boy washed. | n | y | ref | 10 | 41 |
| 38 | MS | The horse that was brown stood in the stall while the jockey settled down. | n | y | opt | 14 | 60 |
| 39 | MS | The dog found a bone in the garden while the boy played. | n | y | opt | 12 | 44 |
| 40 | MS | The frog kicked vigorously while the snake swallowed the mouse. | n | y | opt | 10 | 53 |
| **Means** | | | | | | **11.15** | **49.4** |

Section B

We investigated whether there were any differences in the frequency of disfluencies based on clause order. As can be seen in Table A below, there were very little differences based on the clause order.

Table A

Comparison of disfluency and memory results based on structure type.

| Structure type | Unfilled | Repetition | Repair | Memory |
|---|---|---|---|---|
| Subordinate-main | .159(.02) | .045(.01) | .034(.01) | .141(.02) |
| Main-subordinate | .163(.02) | .042(.01) | .024(.004) | .161(.02) |
| By subjects (*df* = 77) | *t*=-.37, *p*=.71 | *t*=.40, *p*=.69 | *t*=1.73, *p*=.09 | *t*=-1.44, *p*=.15 |
| By items (*df* = 38) | *t*=-.21, *p*=.84 | *t*=.40, *p*=.69 | *t*=1.23, *p*=.23 | *t*=-1.17, *p*=.26 |