Motor sequence learning in children with recovered and persistent developmental stuttering: preliminary findings

Tendera A.[a], Wells R.[b], Belyk M.[k], Varyvoda D.[a], Boliek C.A.[b,c], & Beal, D.S[a-g].

[a] Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada

[b] Department of Communication Sciences and Disorders, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada

[c] Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Canada

[d] Department of Speech-Language Pathology, Faculty of Medicine, University of Toronto, Toronto, Canada

[e] Rehabilitation Sciences Institute, Faculty of Medicine, University of Toronto, Toronto, Canada

[f] Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada

[g] Institute for Stuttering Treatment and Research, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada

[k] Department of Speech, Hearing, and Phonetic Sciences, Faculty of Psychology and Language, University College London, UK.

Corresponding Author:

Deryk S. Beal, Ph.D., Reg. CASLPO, CCC-SLP

Clinician-Scientist, Bloorview Research Institute
Assistant Professor, Department of Speech-Language Pathology, University of Toronto
Holland Bloorview Kids Rehabilitation Hospital | 150 Kilgour Road, Toronto, ON M4G 1R8
Phone: 416-425-6220 x3582
Email: dbeal@hollandbloorview.ca
Web: http://research.hollandbloorview.ca/scientist/dbeal

Abstract

Purpose: Previous studies have associated developmental stuttering with difficulty learning new motor skills. We investigated non-speech motor sequence learning in children with persistent developmental stuttering (CWS), children who have recovered from developmental stuttering (CRS) and typically developing controls (CON).

Methods: Over the course of two days, participants completed the Multi-Finger Sequencing Task, consisting of repeated trials of a10-element sequence, interspersed with trials of random sequences of the same length. We evaluated motor sequence learning using accuracy and response synchrony, a timing measure for evaluation of sequencing timing. We examined error types as well as recognition and recall of the repeated sequences.

Results: CWS demonstrated lower performance accuracy than CON and CRS on the first day of the finger tapping experiment but improved to the performance level of CON and CRS on the second day. Response synchrony showed no overall difference among CWS, CRS and CON. Learning scores of repeated sequences did not differ from learning scores of random sequences in CWS, CRS and CON. CON and CRS demonstrated an adaptive strategy to response errors, whereas CWS maintained a high percentage of corrected errors for both days.

Conclusions: Our study examined non-speech sequence learning across CWS, CRS and CON. Our preliminary findings support the idea that developmental stuttering is not associated with sequence learning per se but rather with general fine motor performance difficulties.

Keywords:

Stuttering; sequencing; finger tapping; skill learning; non-speech motor learning; children

1. **Introduction**

Stuttering is a neurodevelopmental communication disorder characterized by frequent and protracted speech dysfluencies that impair the production of speech, first presenting in children aged 2–6 years (Cavenagh, Costelloe, Davis, & Howell, 2015; Yairi & Ambrose, 1999). While approximately 75% of preschoolers who stutter recover in childhood, the disorder persists into adulthood for the remaining 25% (Yairi & Ambrose, 1999). The nature of and cure for stuttering are still unknown. To date, stuttering research has focused on how speech and language are produced given that it is a disorder of speech execution with symptomatic motor breakdowns characterized by involuntary blocks, prolongations, and repetitions in speech (for a review, see Ludlow & Loucks, 2003; Etchell et al., 2018). It is still unclear what factors affect stuttering persistence and its recovery (Smith & Weber, 2017).

1.1 Stuttering as a motor problem. Numerous studies have demonstrated that adults who stutter (AWS) and children who stutter (CWS) produce slower, more variable, and error-prone speech (for review see Bloodstein & Bernstein Ratner, 2008; Kleinow & Smith, 2000; Max, 2004; Smith et al., 2012). Importantly, these atypical markers of stuttering are also evident in childhood (Smith et al., 2012; Walsch, Mettel, & Smith, 2015). CWS often show delays in speech motor coordination, characterized by slower articulatory rate, and increased variability of speech movements compared to children who recovered and controls (CON; Usler, Smith, & Weber, 2017; Walsch, Mettel, & Smith, 2015; MacPherson & Smith, 2013).

The characteristics of limited motor control in stuttering are likely not exclusive to speech, but rather, shared across motor systems (e.g., Busan et al., 2009; Neef et al., 2011; Max et al., 2004). Stuttering symptoms have been proposed to stem from a more general disorder, which compromises the entire motor system, including speech motor pathways (e.g., Jones et al., 2002; Busan et al., 2003; Busan et al., 2009), leading to slower, less stable and longer in duration

speech and non-speech movements (e.g., manual, hand, non-speech orofacial movements) in people who stutter. AWS have demonstrated poorer timing during finger tapping tasks and nonspeech oral motor tasks compared to CON (Brown et al., 1990; Forster & Webster, 2001; Ward, 1997). AWS also showed greater accuracy and speed variability during finger, lip, and jaw movements (Caruso et al., 1988; Forster & Webster, 2001; Max et al., 2001; Max et al., 2003). The examination of motor abilities in children also showed CWS perform worse on finger tapping (Falk et al., 2014) and hand-clapping tasks (Olander et al., 2010). CWS and adolescents who stutter also showed poor synchronization with both metronome and musical stimuli, suggesting impaired timing during motor control tasks in stuttering (Falk, Müller, Della Bella, 2015; 2014). On the contrary, Zelaznik et al. (1997) and Zelaznik, Smith, and Franz (1994) showed no timing deficit in AWS; however, revealed poorer bimanual finger coordination than in AWNS.

Only a few studies present contradictory findings to the predominant belief that stuttering is a domain-general motor deficit. For example, in the timing synchronization-continuation experiment Max and Yudman (2003) showed that AWS and CON had similar timing abilities across repetitive non-speech orofacial movements, assessing oral opening and closing without the voiced speech component, and finger movements, moving the index finger to contact the thumb. Variability also did not differ among the conditions, suggesting that timing ability across multiple movement systems is normal in people who stutter. A large-scale cross-sectional study by Hilger, Zelaznik, & Smith (2016) measured bimanual motor timing performance in 115 children (70 CWS). Hilger, Zelaznik, & Smith (2016) hypothesized that the early stages of stuttering are associated with typical general motor system development. Preschool CWS showed no difference in rate and variability of clapping movements compared to age-matched fluent controls, suggesting no association of stuttering with performance of clapping tasks in the

early stages of development. Finally, it was argued that simple motor timing mechanisms engaged in the clapping task is not predictive of persistence or recovery from stuttering.

There have been several attempts to develop a framework which is able to justify motor deficits as the cause of stuttering. The Speech Motor Skill (SMS) theory proposed by van Lieshout, Hulstijn, & Peters (2004) posits that people who stutter are less skilled at performing fine motor behaviors. Specifically, van Lieshout et al. (2004) suggests that deficits in the preparation and execution of complex motor actions is influenced by other cognitive demands, such as cognitive-linguistic, socio-emotional or environmental domains, as a causal mechanism of stuttering. They emphasized that behavioral evidence is mostly apparent in speech because of its increased demand on the timing, accuracy, and sequencing of movements. However, other motor systems might be deviant in stuttering but to reveal these differences more complex experimental designs which are sensitive to such behavioral differences. Finally, the SMS theory proposes that the motor control deficits found in stuttering emerge in early childhood and arise from an innate inability to learn a motor task, which in turn limits the presupposed benefit of motor practice and learning. Thus, the speech disorder is associated with motor control development which may directly influence stuttering onset, maintenance and recovery, given that these subtypes are determined during a stage of development when sequence motor learning goes through profound refinements (Savion-Lemieux et al., 2009).

1.2 Motor learning in stuttering. Another factor critical in defining speech and non-speech motor skills is motor learning, the ability to acquire a new motor skill as a result of practice or experience. The theories surrounding motor learning deficits for speech and non-speech tasks have been tested empirically in various experimental arrangements, but findings have been inconclusive (Tumanova et al., 2015; Smits-Bandstra & Gracco, 2014; Smits-Bandstra & De Nil, 2006). AWS demonstrated slower reaction and lower accuracy performance in

pseudoword production (Ludlow, Sire, & Zirika, 1997). Additionally, CWS required more practice to accurately perform a repetition task (Cooper & Allen, 1977). Berman Hakim and Bernstein Ratner (2004) also reported that CWS showed lower accuracy when they were asked to repeat multiple nonsense syllable sequences.

Adaptation occurs in response to motor learning. Max and Baldwin (2010) investigated motor learning adaptation by testing practice effects from situation effects based on reading of repeated and novel sentences and learning effects based on initial adaptation and retention tests after 2 hours and on the second day after the initial test. They reported that 8 out of the 10 participants showed an adaptation effect of the repeated sentences for the first approach. Fluency improvement was observed 2 hours after the initial adaptation session but a day later the fluency effect was washed out for novel sentences but was still present for repeated sentences. suggesting relatively strong "retention of adaptation" of familiar material. However, they focused on the adaptation and learning effects solely on the group of people who stutter and did not have a control group in their study. On the other hand, Tumanova et al. (2015) employed a visuomotor tracking task, with both the jaw and hand, to test non-speech motor practice effects in AWS and CON. AWS showed no difference in timing and amplitude of hand and jaw movement over practice, suggesting typical practice effects between groups.

Importantly, Namasivayam & van Lieshout (2008) showed reduced retention of newly learned motor skills and suggested retention of motor skills as a potential source of deficit in stuttering. In their study, Namasivayam & van Lieshout (2008) examined short-term practice within the same day and long-term retention after one week in the speech motor context. They reported similar practice effects in terms of amplitude and duration of movements; however, AWS variability of trials did not decrease over time that reflected differences in strength and stability of coordination between groups. Similarly, Smits-Bandstra and Gracco (2013) tested

retention of implicit sequence learning skills in AWS and CON. One week after the initial test, AWS showed slower reaction times on early trials of the retention test, suggesting subtle delay in retention of implicit sequencing skills.

Slower learning of more complex motor sequences has also been discussed as a key deficit factor in stuttering. To clarify this idea, a number of researchers employed motor sequences in both speech and nonspeech domains with a highly cited serial reaction time task (Nissen & Bullemer, 1987), in which participants are asked to complete repeated and random sequences. Repeated motor sequence learning was previously tested only in AWS and revealed some interesting but heterogeneous results. Bauerly and De Nil (2011) reported that AWS initiated sequences more slowly and had longer sequence durations both early in practice and after a consolidation period relative to CON; however, the overall learning of repeated sequences, evaluated for accuracy, response anticipation or sequence duration, was intact. Results of the explicit syllabic sequencing task reported by Smits Bandstra et al. (2006) were also mixed, indicating that AWS were slower but equally accurate as controls, supporting the argument for a subtle explicit learning difficulty at the earlier learning stages in stuttering. However, during the retention task, AWS performed similarly to CON in terms of both timing and accuracy. In their non-speech motor version of the sequencing task, Smits-Bandstra et al. (2006) found that AWS were not able to improve their sequence response time compared to CON during the initial testing phase and showed an overall reduced skill retention after a 40-minute delay. Similarly, AWS showed slower response times and reduced accuracy in initiating nonsense syllable sequences while doing a colour recognition as a dual task paradigm (Smits Bandstra & De Nil, 2009). A comparable trend was observed in a finger sequence learning task by Smits Bandstra, De Nil and Rochon (2006), in which AWS were slower and less accurate even after practice under single and dual task conditions, suggesting poorer skill learning

strategy in stuttering. Finally, Smits-Bandstra and De Nil (2013) examined finger tapping during 10-element sequences in AWS and CON to examine whether stuttering was associated with differences in the sequence organization and specific sequence elements, which the authors referred to as "chunking." Their findings showed longer inter-sequencing durations in AWS that might arise from decreased motor processing efficiency or the need for more practice to automatize sequencing. These observations indicate less proficient learning of complex sequential motor tasks in AWS. The question of whether implicit learning is impaired in AWS has undergone limited investigation. Smits Bandstra and Gracco (2013) employed a nonsense syllable serial reaction time task to understand solely implicit learning in AWS. They reported that AWS were slower to acquire the implicit skill compared to controls, suggesting an implicit learning difficulty. In their follow-up study, Smits Bandstra and Gracco (2013b) examined the retention effect of implicit sequence learning. AWS showed a general learning effect but the responses were slower than by CON, suggesting some subtle difficulties in implicit sequencing skills.

In summary, findings from studies in speech and non-speech motor control in stuttering showed some considerable support for the hypothesis that people who stutter are less effective at learning motor skills and performing complex motor tasks. The deficit in motor control includes slower onset speed, high variability, subtly lower accuracy, and weaker learning and retention effects. As such, it has been speculated that persistent stuttering might be a subtle domain-general motor deficit specifically associated with more complex motor sequence learning for both speech and non-speech movements. A limitation in the current literature suggests that although there is some indication of a speech-motor and non-speech motor skill learning deficit in stuttering, previous work has utilized methodologies predominantly focused on a repeated sequence in the absence of a control random sequence (e.g., Smits Bandstra & De Nil, 2009;

Smits Bandstra et al., 2006; Smits Bandstra, De Nil & Rochon, 2006). As a result, it is difficult to conclude what type of sequence learning is more susceptible to breakdowns or whether the differences in motor learning, identified in those works, are specific to sequence learning per se or rather stem from a more generalized motor performance deficit.

Another gap in the previous literature is that we still do not know the detailed characteristics of the sequence performance errors. Motor sequence tasks are more complex and require more cognitive and motor resources which makes it similar to speech. The speech symptomology of stuttering varies from prolongations to repetitions; however, the patterns of motor sequencing errors are still not known. A detailed analysis of sequence errors might reveal important error trends and further describe behavior of people who stutter. Finally, and most importantly, there is still lack of evidence concerning the potential impairment of sequence learning in CWS. This gap in the literature might be explained by the fact that sequencing paradigms are more complex than simple reaction time paradigms which makes it difficult to administer them on young children. Examination of sequence learning with CWS can advance our understanding of developmental stuttering, including its onset and potential recovery, given persistence and recovery groups are determined during a stage of development when sequence motor learning goes through profound refinements (Savion-Lemieux et al., 2009). Furthermore, there is still no evaluation of sequence learning in children with recovered stuttering (CSR). It is therefore unknown if the differences in motor sequence learning are present closer to the onset of childhood stuttering, if they play a role in recovery and if the disorder is linked to an early-stage motor sequence learning deficit. A further important step in testing these research questions is to examine whether sequence learning is present in childhood and how it is related to stuttering persistence versus recovery.

1.3 The purpose of the present study. We aimed to collect preliminary data on motor sequence learning and performance in CWS, children who had recovered from stuttering (CRS), and CON. We adapted a multi-finger sequencing task (MFST), a variant of the traditional serial reaction-time task of Nissen and Bullemer (1987) and a serial reaction-time task for children (Thomas & Nelson, 2001) modified by Savion-Lemieux, Bailey, and Penhune (2009) to adequately test young children. The key measurements of our experiment were. (a) response accuracy of motor sequence learning, and (b) response synchronization, which involves a component of motor sequencing requiring fine-grained sensorimotor integration and timing. While accuracy is a well-accepted measure in the field and is understood as the percentage of correctly completed responses from all completed trials, response synchronization is a relatively novel measure that provides a more task-appropriate evaluation of timing. Response synchronization reflects a response time relative to a stimulus onset and thus, provides a more precise timing measure for evaluation of sequencing timing. Response synchrony differs from previous reaction time measures, because anticipatory responses are included in its calculation. Because anticipation is expected to increase with learning of the task, it follows that responses to the learned sequence become more synchronous (Savion-Lemieux & Penhune, 2005). Testing of response synchronization and accuracy of motor sequence learning can potentially inform us about stuttering onset and maintenance or recovery. As the majority of children who will recover from stuttering will have done so prior to age ten years old, we hypothesize that accuracy and error type will differ across CWS, CRS and CON.

2. Methods

2.1 Participants

We recruited 42 participants (11 CWS, 14 CRS, 17 CON) through the Institute for Stuttering Treatment and Research (ISTAR) at the University of Alberta and via local advertising. The final

sample, after exclusions, consisted of 31 right-handed children including 8 CWS (M = 9 years, 7 months; range = 7 years, 10 months – 11 years, 7 months; SD = 18.9 months), 12 CRS (M = 10 years, 0 months; range = 8 years, 4 months – 11 years, 5 months; SD = 12.4 months) and 11 CON (M = 10 years, 7 months; range = 8 years, 1 month – 12 years, 9 months; SD = 18.1 months). Eleven children were excluded: three due to behavioral issues during the task (2 CON, 1 CRS); three due to comorbid disorders other than a diagnosis of persistent developmental stuttering for CWS (1 each in the CON, CWS, and CRS groups); one for being left-handed (1 CON); and 4 for falling below average limits on either the PPVT-4 or Digit Span subtest WISC-4 (2 CON, 2 CWS). All children in the final sample had no reported neurological or developmental disorders other than a diagnosis of persistent developmental stuttering in the CWS group and a documented history of recovery from developmental stuttering among CRS. the CRS had been in remission for at least 2 years. The diagnostic criteria for persistent developmental stuttering were that the children had a preschool aged onset of the disorder and a continued level of stuttering, defined as >3% syllables stuttered, that remained of concern to the family and speech-language pathologist. The diagnostic criteria for recovered stuttering were that the children had a preschool onset of the disorder with a past level of stuttering, defined as >3% syllable stuttered, that was of significant concern to the family but that had now been in remission, defined as <1 % syllables stuttered, and of no concern for the family for the past two years. Percentage syllable stuttered was determined based on reading the Caterpillar passage and short conversation of minimum of 300 syllables. All children had received speech therapy for their stuttering but none were in active therapy at the time of the study. As such, in the case of CRS, it was impossible to determine to what extent their recovery and remission was natural versus therapy-assisted.

All participants had hearing within normal limits, as confirmed by a pure tone hearing screen and/or parent report of a previously passed hearing screen. In addition, we matched the groups on receptive language and cognitive measures, assessed by scores on the Peabody Picture Vocabulary Test Fourth Edition (PPVT-4; Dunn & Dunn, 1981) and Digit Span subtests of the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-4; Wechsler, 2003). All participants included in the present study scored within the average to above average range on both the PPVT-4 and the Digit Span subtest of the WISC-4. Given that previous research suggests differences in the motor skill performance of musicians versus non-musicians, and novice versus experienced child pianists, (Palmer & Meyer, 2000; Watanabe, Savion-Lemieux, & Penhune, 2006) we asked parents about the music training of their children. All participants in the current study had less than three years of musical training. The University of Alberta Health Research Ethics Board approved the study. A parent provided written informed consent for each child participant, and each child provided informed assent.

2.2 Multi-finger sequence task (MFST) and stimuli

A well-established experimental protocol, the MFST (Savion-Lemieux, Bailey, & Penhune, 2009), was employed to evaluate sequence learning in CWS, CRS and CON participants via measurement of the accuracy and synchrony of their responses during the task. Accuracy was measured as percentage of correct responses; while synchrony was measures as time proximity after the target is presented. In our implementation of the MFST, participants reproduced 10-element sequences of key presses on an Apple MacBook Air keyboard using the four fingers of the right hand. Similar to Savion-Lemieux, Bailey and Penhune (2009), the visual display was made up of four columns (5 cm^2), which remained on screen for the entirety of each trial. The visual cue for each key press was a cartoon hamster (4.5 cm^2) presented in one of the frames. The cue was presented for 500 ms and was followed by a 400 ms inter-stimulus interval.

Colored stickers marked the four keys corresponding to a particular frame, with participants told to "catch" the hamster by pressing the appropriate key. We instructed participants to wait until the animal appeared in the frame before catching it to minimize anticipation of responses and optimize response synchronization, as per Savion-Lemieux, Bailey and Penhune (2009). Each participant completed a minimum of two familiarization blocks before the MFST test session. The familiarization blocks consisted of 15 repetitions of simple and predictable sequences (key sequences 1-2-3-4: index, middle, ring, pinkie; and 4-3-2-1: pinkie, ring, middle, index). Regardless of group membership, some younger children required additional familiarization trials to ensure they were able to coordinate attention to the stimulus while moving each of their four fingers independently.

Each block consisted of six trials of the 10-element sequences. Four of the trials were a repetition (REP) of the same sequence and two were quasi-random (RAN) sequences. REP and RAN sequences were of equal difficulty as per Savion-Lemieux, Bailey and Penhune (2009). Over the course of the experiment, the REP sequence was performed 24 times, one of the RAN sequences was performed three times, two of the RAN sequences two times and the remaining RAN sequences were performed once each. This distribution of RAN sequences allowed us to test if repetition of a sequence could significantly improve performance. The blocks were quasi-randomly designed to space out presentation of REP and RAN sequences. The blocks never started or ended with a RAN sequence and no two RAN sequences were presented consecutively. There was a 1300 ms delay between trials.

The two familiarization trials and three testing blocks took a combined total time of 5 minutes and 30 seconds, with the testing blocks lasting approximately one minute each. At the end of the task participants were rewarded with a dancing animal animation and verbal praise.

2.3 Behavioural measures

A 900 ms window was used to identify key presses corresponding to each visual stimulus (Fig. 1). The response window extended from 100 ms prior to stimulus onset to 300 ms after stimulus offset to allow for anticipatory and delayed responses. The stimulus itself was presented for a total of 500 ms. Only the first key pressed within the response window was scored but all responses were recorded.

We used accuracy and response synchronization scores to assess learning and calculated both within each block independently for REP and RAN sequences. We defined accuracy as the average percentage of correct key presses. We recorded response synchrony only for correct key presses during the 900 ms response window. We then calculated the average response synchrony for each sequence based on the correct responses during that sequence and from this, calculated the average within each block for RAN and REP sequences.   Additionally, five categories of the errors were defined: delayed responses, delayed self-corrected errors, self-corrected errors, uncorrected errors, and failures to respond. A delayed response is one given beyond the appropriate response time window described above (see Figure 1). A delayed self-correction consisted of an initial incorrect response followed by a correct response completed beyond the appropriate response time window. A self-corrected error was an initial incorrect response followed by a correct response within the acceptable response time for that stimulus. An uncorrected error consisted of an incorrect response with no attempt to correct the error or an unsuccessful attempt. Lastly, failures to respond in the given time block were simply coded as a "no response" from the participant.

2.4 Procedure

13

All participants completed the study over two consecutive days. On the first day, participants completed the consent and assent processes and screening to determine eligibility. Children then completed two familiarization blocks (15 trials) to ensure that they understood the task and were comfortable with the set-up and positioning of the equipment. Once the children were familiar with the experimental set-up they completed the three testing blocks (18 trials) of practice. On day two, all participants completed two more familiarization blocks (15 trials) and then three final testing blocks (18 trials) of the MFST, after which they completed recall and recognition testing. In the recall test, children were asked to reproduce the REP sequence on the computer keyboard, with no assistance from the visual stimulus. Children were scored based on percentage of correct key presses. In the recognition task, children were presented with three sequences (one REP, two RAN) and were asked to identify the REP sequence. The percentage of children who correctly identified the REP sequence was calculated for all groups.

2.4 *Statistical Analysis*

Accuracy & Response Synchrony: Data were analyzed in R v3.6.0 (R Core Team, 2019) running on OS v10.11.6. Data handling and visualization were supplemented with the tidyverse (Wickham, 2017), mixed models were fit with lme4 (Bates, Maechler, Bolker, & Walker, 2015), bootstrapped confidence intervals were computed using merTools (Knowles & Frederick, 2019), and bootstrapped inferential statistics were computed using pbkrtest (Halekoh & Højsgaard, 2014).

Response synchrony and accuracy data were fitted by a Linear Mixed Model (LMM) and Generalized Linear Mixed Models (GLMMs), respectively. The statistical approach has been employed in speech, language, and hearing research (e.g., Quene & Van den Bergh, 2004) and was selected because of its suitability for repeated-measures data. LMM allows for accounting for random effects to analyze repeated measures and GLMM allows for accounting for binary

responses. Models were fit with fixed effects of condition (repeated or pseudorandom), group (CON, CWS, or CRS), test day (1 or 2), and block number (1, 2, or 3). All two-way interactions were included in the model. Initial model fitting with the Akaike Information Criterion (AIC) suggested that models could be meaningfully simplified by omitting three-way and four-way interactions. However, the three-way interactions between i) group, condition, and day, as well as ii) group, condition, and block were none-the-less retained from theoretical interest. Both models were fit with the maximal random effects structure that did not lead to singular model fits (Barr, Levy, Scheepers, & Tily, 2013).

The accuracy GLMM included random intercepts of participants with random slopes of both day and condition. The response synchrony LMM had the same random effects structure save the random slope of condition, which had to be omitted. Since only trials with correct responses were included in the response synchrony LMM, some participants may have had too few surviving data points in one or more conditions to reliably estimate random slopes.

Error Type & Motor Sequence Learning: Data were analyzed in SPSS version 22.0 (SPSS, 2001). The error analysis was completed via a chi-square test of the proportions of error types in the task. The proportions of error types across groups and across days was compared to examine whether or not there were group differences in errors made and a correlation between error type and practice time. Similarly, the effects of REP versus RAN learning were analyzed with a chi-square analysis to see if the percentage of children able to identify the repeated sequence was the same across groups. Additionally, a one-way ANOVA was employed to compare the percentage of sequence elements correctly recalled by the children.

Additional Analyses: Data was explored for Pearson correlations between response synchrony and accuracy as potential indicators of different outcomes during best performance. The correlations were calculated between age and all performance metrics to identify if the

expected age-related trajectory of improvement from Savion-Lemieux (2009) was present in CWS, CRS and CON participants.

A note on inference

There is growing concern that the application of "bright line" *p*-value thresholds, or their widespread and persistent misapplication by the scientific community, provides a poor foundation for scientific inference (Wasserstein & Lazar, 2016). We therefore adopt the recommendations of the American Statistical Association in approaching the inherent uncertainty of data in the honest spirit of open science (Wasserstein, Schirm, & Lazar, 2019). We have chosen to report *p*-values since they may continue to provide useful information, but base our conclusions on effect estimates informed by confidence intervals. We have also adopted the language of Dushoff, Kain, & Bolker (2019) to highlight effects for which the data provide clear evidence.

Results

3.1 Accuracy

The GLMM predicting mean accuracy revealed clear evidence that performance improved on day 2 (LRT = 30.05, *p* = 0.001), but whether performance improved over blocks was less apparent (LRT = 5.57, *p* = 0.068). As shown in Figure 2, there was an interaction between Day and Block which may reflect that while participants improved over the first day of training they plateaued after the second day of training (LRT = 8.45, *p* = 0.019).

Performance was higher during the REP condition than the RAN condition (LRT 8.44, *p* = 0.002), but an interaction suggests that this difference was more apparent on Day 1 (LRT =11.0, *p* = 0.003). There was no clear difference between groups (LRT = 2.42, *p* = 0.392), however there is some evidence for an interaction (LRT = 5.53, *p* = 0.085) reflecting the lower

16

performance of CWS on Day 1. There was no evidence for higher level interactions. See

Appendix A and B for effect estimate tables.

[insert Figure 2 around here]

3.2. Response Synchrony

The LMM predicting Response Synchrony revealed clear evidence that speed increased

on day 2 (LRT = 8.871, p = 0.005), but there was little evidence for improvement over

experimental blocks (LRT = 3.994, p = 0.13). There was however an interaction between Day

and condition, demonstrating greater improvements for the REP over the RAN condition (5.68, p

= 0.016, see Figure 3).

There was no clear difference between groups (LRT = 3.24, p = 0.242), unlike response

accuracy there was no evidence of interaction between group and day (LRT = 52.84, p = 0.277).

There was an interaction between group and block (LRT = 10.58, p = 0.034). However, the

direction of this effect was not consistent with reduced motor learning among CWS. There was

no evidence for higher level interactions. See Supplementary Materials for effect estimate tables.

[insert figure 3 around here]

3.3 Error Type

The chi-square test revealed significant differences in the proportions of error types

between groups. On the first day, there were significant differences in the proportions of errors

made by the three groups for both the REP (X^2(4, N = 1237) = 24.28, p = .0021) and RAN

sequences (X^2(4, N = 679) = 42.69, p = < .001). On the second day, there were significant

differences among the three groups for REP ($X^2(4, N = 700) = 18.53, p = .018$) but not RAN sequences ($X^2(4, N = 467) = 11.52, p = .17$).

To assess how practice affects the proportions of error type, we compared the proportions of errors within each group across both days (see Figure 4). For CON and CRS, there were significant differences in the types of errors made for both REP (CON: $X^2(4, N = 496) = 27.41, p < 0.001$ and CRS $X^2(4, N = 764) = 15.72, p = .0034$) and RAN (CON: $X^2(4, N = 319) = 13.97, p = .0074$ and CRS $X^2(4, N = 454) = 20.73, p < 0.001$) sequences. However, for CWS, there were no differences in the types of errors made between day one and day two for either REP or RAN sequences (REP: $X^2(4, N = 677) = 4.467, p = .35$, RAN: $X^2(4, N = 373) = 7.51, p = .11$). For both CON and CRS, there was a >10% increase in the proportion of self-corrected errors driven largely by a decrease in uncorrected errors, and therefore a decrease in total errors, between Day 1 and Day 2.

[insert figure 4 around here]

3.4 Associations between accuracy and response synchrony

CWS and CRS had no significant correlation between the accuracy and response synchrony in the REP sequence ($r(187) = 0.0063, p = .9316$ and $r(285) = 0.049, p = .4107$, respectively). However, CON had a moderate negative correlation ($r(238) = -0.25, p = 7.0\times10^{-5}$) between accuracy and response synchrony.

For the RAN sequences, CON and CWS displayed moderate correlations in opposite directions. As with the REP sequence, CON had a negative correlation ($r(117) = -0.19, p = 0.039$) between accuracy and response synchrony during RAN trials. However, CWS displayed a positive correlation ($r(90) = 0.28, p = 0.0065$) between accuracy and response synchrony. In other words—consistent with the REP sequences—CON were more synchronous for more accurately

completed RAN sequences but CWS less synchronous. CRS had no significant correlation (r(140)=0.10, p = 0.2169).

3.5 Impact of age

As shown in Table 1, within the CON group, we found a significant positive correlation between age and both REP (r (9) = 0.63, p = 0.038) and RAN (r (9) = 0.74, p = 0.0097) sequence accuracy, suggesting typically developing children perform more accurately with age. However, there were no significant correlations for CRS for either the REP sequence r(10) = 0.21, p = 0.51) or RAN sequences (r(10) = 0.33, p =0.30). The same was true for CWS, with the correlation for REP sequences being r(6)=0.60 (p = 0.12) and for RAN sequences being r(9) = 0.63 (p = 0.095). There were no significant correlations between age and either REP or RAN response synchrony for any groups.

3.6 Recognition and Recall

There were clear differences in the number of participants across groups who could correctly recognize the REP sequence (χ^2 = 8.99, p = 0.011), in which only CON were able to benefit from recognition in REP sequences but not CRS and CWS. There were no clear differences between the percentages of correctly recalled elements among groups (F (2,28) = 0.52, p = 0.60). There were no clear differences between the accuracy or response synchrony outcomes whether children recognized the sequence or not.

4. Discussion

We examined non-speech motor sequence learning in CWS, CRS and CON during the MFST performance. The study results revealed four main findings. First, CWS demonstrated lower performance accuracy than CON and CRS on the first day of the finger tapping experiment but improved to the performance level of CON and CRS on the second day. Second, the analysis of response synchrony showed no overall difference between CWS, CRS and CON.

CWS displayed less synchrony than CRS and CON during some experimental blocks, but the direction of this effect was not consistent with reduced motor learning. Third, every group exhibited practice effects on accuracy measures during the second day of testing and achieved a similar accuracy percentage; however, synchrony practice effects were present only on the first day and plateaued on the second day. Fourth, CON and CRS demonstrated an adaptive strategy to response errors, shifting from a high percentage of non-corrected errors to corrected errors across days, whereas CWS maintained a high percentage of corrected errors for both days. Qualitatively, only CON demonstrated explicit recognition of the REP sequence at the end of practice whereas CWS and CRS showed no such recognition.

4.1 Stuttering as a general non-speech motor disorder

Accuracy of performance. There is a lack of consensus on whether developmental stuttering is specific to general motor control or, if it involves motor subsystems which regulate more complex sequencing (Bauerly & De Nil, 2015; Olander, Smith, & Zelaznik, 2010; Smits-Bandstra & De Nil , 2007), causing of a motor breakdown that can result in disfluency. The disparate findings can be explained by the comparison of task complexities that can differentially impinge on motor control. Simpler or more repetitive tasks might not place unique demands on motor control and, thus, not explain the behavioural group differences. Our study adds to a handful of research attempts which allow adequate assessment of the temporal and spatial aspects of movements (e.g., Caruso et al., 1988; Forster & Webster, 2001; Max et al., 2001; Max et al., 2003). Here, we employed a task that allowed us to evaluate the temporal and motor aspects of non-speech sequence movements. In this way, we addressed a gap in the literature that examined single movements of a specific non-speech effector that made it difficult to compare to the complexity of speech-related tasks. Overall, our findings revealed lower accuracy and slightly less synchronous responses in REP and RAN sequences in CWS compared to CON and

CRS. These results serve as a supporting argument for stuttering being a generalized motor performance deficit rather than exclusive motor sequencing issue.

On the first day of testing, CWS performed less accurately than CRS and CON. Our findings corresponded with previous studies which showed that CWS perform worse on finger tapping (Falk et al., 2014) and hand-clapping tasks (Olander et al., 2010). This trend was also present in previous studies on persistent stuttering in adulthood. AWS demonstrated decreased performance in finger tapping tasks or nonspeech oral motor tasks compared to adult controls (Brown et al., 1990; Forster & Webster, 2001; Ward, 1997). The mechanism underlying the more inaccurate performance in stuttering may be caused by ineffective selection of competing motor commands, resulting in atypical spatial coding (Smith, 1989; Walsch, Mettel, & Smith, 2015). In the present study, CWS clearly showed a lower accuracy percentage than CRS and CON indicating that recovery may possibly be driven by motor maturation processes.

Interestingly, although the performance of CRS was not statistically clearly different from CON, their performance was slightly less accurate than CON in both REP and RAN sequences (Figure 2). This indicates a subtle difference in accuracy even after the recovery. A finding that is harder to reconcile is that this group difference was supported by the correlation analysis between age and performance accuracy in REP and RAN sequences. Only within the CON group, older children had better accuracy while completing the task in both REP and RAN conditions; whereas both CWS and CRS groups did not show the improvement with maturation. CWS have been shown to have abnormal trajectories of development in brain regions critical to motor control relative to CON (Beal et al., 2015; Chow & Chang, 2017). Over the course of recovery, CRS may have developed compensatory strategies for this deficit, allowing them to perform more accurately at the start; however, compared to fluent speakers, a subtle motor deficit is still present and is not fully alleviated with age. Future work is warranted to investigate

if the number of years since recovery is linked with fine motor performance and the associated neural correlates.

Timing synchrony. Previous studies have suggested that a motor timing deficit in stuttering contributes to the development and persistence of this disorder (e.g., Boutsen, Brutten, & Watts, 2000; Kent, 1983). Motor timing deficits for both verbal and non-verbal tasks are well documented in people who stutter of all ages (Bauerly & De Nil, 2015; Borden, 1983; Falk, Müller, & Dalla Bella, 2015; Olander, Smith, & Zelaznik, 2010), and neuroimaging findings have indicated neural timing abnormalities in the grey and white matter regions (Beal et al., 2010, 2011; Etchell, Johnson, & Sowman, 2014; Chang et al., 2019). In the context of response synchrony, all groups performed similarly. Only in some experimental blocks, CWS tended to respond less synchronously than CON and CRS. This finding might suggest that CWS are able to achieve a typical synchronization pattern; however, they had difficulty demonstrate it consistently compared to CRS and CON. Falk, Müller, & Dalla Bella (2014) also reported that older children and adolescents who stutter showed lower synchronization consistency and higher rate of early responses as compared to CON; moreover, poor synchronization in stuttering was associated with the highest degree of stuttering severity. Similar to our findings, their study showed that age as a factor was not correlated with performance.

A recent large-scale study by Hilger et al. (2016) also revealed no differences in timing between CWS and CON on the basis of timing coordination in the bimanual clapping task. Although their task and timing analysis differ from those employed here, findings of both studies align, indicating that a sole timing deficit might be not the overarching issue in CWS. Max and Yudman (2003) also showing no differences in the ability of AWS to generate temporal speech and non-speech related movement patterns with a simple isochronous rhythm.

Motor learning in stuttering. With regard to the question of whether stuttering is associated with a motor learning deficit, previous studies provided disparate results. Some argue that stuttering is associated with motor skill learning limitations (e.g., Smits-Bandstra et al., 2006; Namasivayam & van Lieshout, 2011; Smits-Bandstra & Gracco, 2014); while other researchers reported no clear group differences in motor learning between people who stutter and controls (e.g., Namasivayam & van Lieshout, 2008; Bauerly & De Nil, 2011; Tumanova et al., 2015). Our findings are consistent with the latter. The initial accuracy performance in CWS was the lowest among groups on the first day. On the second day CWS accuracy of performance reached the score of CRS and CON in both REP and RAN sequences, demonstrating no practice deficit in the motor learning ability of CWS and CRS.

Previous studies have also evaluated learning of temporal characteristics in various experimental tasks, measured as reaction or response time after practice. The findings of these studies differ, showing either compromised temporal learning effect in people who stutter (Smits-Bandstra, 2013; Smits-Bandstra, 2007; Ludlow, Sire, & Zikira, 1997; Cooper & Allen, 1977) or similar to that of fluent speakers (Tumanova et al., 2015). Our study evaluated learning effects with reference to the synchronization speed. The temporal learning effect demonstrated a similar pattern across groups. Only in some experimental blocks, our data suggested that CWS were slightly slower and less synchronous in their responses than CON and CRS on the first and second day of testing. However, the trend of this effect did not align with reduced motor learning among CWS. Overall, all groups improved on the first and second days of testing in both REP and RAN sequences. The consistency across sequence conditions suggests that the improvement is not specific to sequence learning, per se, but instead reflects normal general motor learning in CWS and CRS.

CWS and CRS may not have the neural efficiency to respond either as quickly or accurately as the CON and may need more resources to process the stimulus to ensure a correct response. This can be potentially explained by a deficit in automatization development, suggested by Smits-Bandstra and De Nil (2006). People who stutter might over-rely on sensory feedback which taxes their attentional resources. As a result, movement sequences might be chunked to optimize the task completion and increase accuracy, concurrently with a loss in time precision. Qualitatively, it was interesting to observe the differences in the optimal response times between the groups in both the RAN and REP sequences. For CWS and CRS, the most accurate responses occurred in the window 500-600 ms, while for CON the most accurate responses were within approximately 400 ms. This is consistent with the idea that delayed neural development leads to less efficient neural firing in the motor system for children who have stuttered at some point in their lives (Smith & Weber, 2017).

A unique contribution of this study to the motor learning literature was the analysis of error types that allowed for learning strategies to be compared on the initial and second days of testing. Across all groups and both days, the two most common types of errors were self-corrected and uncorrected errors. For CON and CRS, there was a shift from primarily uncorrected errors on the first day to primarily self-corrected errors on the second day, which corresponds with a decrease in non-responses. This shift was larger in CON in the REP sequence, in which the proportion of self-corrected errors increased from 31% to 53%. This shift was approximately twice as large as that of the CRS group, which exhibited increases in self-corrected errors from 28% to 37% on the REP sequence. CWS did not show significant differences in error types across both days for either the RAN or REP sequence. Interestingly, on Day 1, CWS made more self-corrected errors than the other groups (36% for RAN, 38% for REP). This aligns with theories implicating self-regulation of motor control, such as hyper self-

monitoring of execution, in developmental stuttering (Vasic & Wijnen, 2005). Whereas CWS over-monitor, CRS and CON may become more comfortable with the task as they develop the timing skills to predict whether they will be able to correct their responses in the appropriate time frame. CRS and CON may be able to change their strategy as they adapt to the situation with practice, while CWS might be unable to implement the same strategy.

      Repeated and randomized sequencing. For both the practiced repeated sequence and the unpracticed random sequences, CON performed with similar accuracy to the participants in Savion-Lemieux's (2009) experiment, indicating that the group performed the task at an age-appropriate level. We hypothesized that RAN sequences would be repeated with less accuracy and synchrony across the groups on the first and second day of testing. Additionally, we expected RAN sequencing to be specifically challenging for CWS and CRS. Indeed, we observed a greater improvement for REP sequences in both accuracy and response synchrony across all groups. All three groups learned RAN sequences less effectively compared to REP sequences. Under both REP and RAN learning conditions, CWS performed less accurately than CRS and CON on the first day. These findings differed from the conclusions put forth by Smits-Bandstra & Gracco (2013) which suggest that random sequence learning is impaired in stuttering. In their study, Smits-Bandstra & Gracco (2013) reported slower reaction times for syllables learned in the random condition than syllables in the repeated conditionacross the initial two-hour training in AWS. Importantly, AWS were able to benefit from random sequence learning after a week and to perform the task similarly to fluent speakers on the retention test. The implications of these findings are harder to synthesize with our own findings because our study did not evaluate retention effects. As a next step, a retention effect should be tested in CWS and CRS to evaluate how novel skill learning is over an extended period.

Conclusions, limitation and future studies

Studies of non-speech motor learning in CWS and CRS are critical to inform models of stuttering which are almost exclusively based upon data from only CWS. Although the current study has not been designed to test any particular theory, its novel findings contributed to better understanding whether developmental stuttering is associated with a deficit in motor learning. Our preliminary analysis of multiple measures collectively suggests that CWS did not deviate from CON and CRS in non-speech motor sequence learning but rather are consistent with theories that propose a subtle general motor performance deficit that underlies the onset and persistence of developmental stuttering. An ability to overcome this performance deficit appears to contribute to recovery. Our results might contribute to the development and evaluation of stuttering treatment programs that target improved motor control, as we demonstrated that CWS approach initial CON-like accuracy after just two-days of motor practice.

To understand motor learning requires multiple visits resulting in recruitment and attrition challenges. Our sample size of 31 children provides a starting point for understanding the baseline general motor skills and motor learning in developmental stuttering. Importantly, recovered stutters are also difficult to recruit given that families often lose touch with clinics when concerns subside. Our unique context allowed us to an opportunity to overcome these barriers. As a next step, future work with complex motor tasks and increased sample size is needed to replicate our findings and finalize whether general non-speech motor performance can be used to infer about the causal effect of this speech disorder and predict stuttering onset and maintenance or recovery. Continued efforts in this area might inform the potential predictive value of motor accuracy for recovery or the development of a metric for assessing efficacy of clinical interventions. Another consideration is whether motor control skills interact with therapeutic efficacy. It would be beneficial to test whether individual variability interferes with therapy outcomes or long-term maintenance.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Bauerly, K. R., & Nil, L. F. De. (2011). Journal of Fluency Disorders Speech sequence skill learning in adults who stutter. *Journal of Fluency Disorders*, *36*(4), 349–360. https://doi.org/10.1016/j.jfludis.2011.05.002

Beal, D. S., Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., & De Nil, L. F. (2010). Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. *NeuroImage*, *52*(4), 1645–1653. https://doi.org/10.1016/j.neuroimage.2010.04.277

Beal, D. S., Lerch, J. P., Cameron, B., Henderson, R., Gracco, V. L., & DeNil, L. F. (2015). The trajectory of gray matter development in Broca's area is abnormal in people who stutter. *Frontiers in Human Neuroscience, 9*, 1–10.

Beal, D. S., Quraan, M. A., Cheyne, D. O., Taylo, M. J., Gracco, V. L., & Nil, L. F. De. (2011). Speech-induced suppression of evoked auditory fields in children who stutter. *NeuroImage*, *54*(4), 2994–3003. https://doi.org/10.1038/jid.2014.371

Boutsen, F. R., Brutten, G. J., & Watts, C. R. (2000). Timing and intensity variability in the metronomic speech ofstuttering and nonstuttering speakers. *Journal of Speech Language and Hearing Research, 43*, 513–520.

Brown, C. J., Zimmerman, J. P., Hegmann, J. P., and Linville, R. N. (1990). Variations in self-paced behaviors in stutterers and nonstutterers. *J. Speech Hear. Res. 33*, 307–316. doi: 10.1044/jshr.3302.317

Busan, P., Battaglini, P. P., & Sommer, M. (2017). Clinical Neurophysiology Transcranial magnetic stimulation in developmental stuttering : Relations with previous neurophysiological research and future perspectives. *Clinical Neurophysiology*, *128*(6), 952–964. https://doi.org/10.1016/j.clinph.2017.03.039

Caruso, A. J., Abbs, J. H., & Gracco, V. L. (1988). Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain, 111*(2), 439–455. http://dx. doi.org/10.1093/brain/111.2.439.

Cavenagh, P., Costelloe, S., Davis, S., & Howell, P. (2015). Characteristics of young children close to the onset of stuttering. *Communication Disorders Quarterly*, *36*(3), 162–171. https://doi.org/10.1177/1525740114549955

Chang, S., Garnett, E. O., Etchell, A. (2019). Functional and neuroanatomical bases of developmental stuttering: current insights. *Physiology & Behavior*, *25*(6), 566–582. https://doi.org/10.1016/j.physbeh.2017.03.040

Chow, H. M., & Chang, S. E. (2017). White matter developmental trajectories associated with persistence and recovery of childhood stuttering. *Human Brain Mapping*, *38*(7), 3345–3359. https://doi.org/10.1002/hbm.23590

Cooper, M.H. & Allen, G.D. (1977). Timing control accuracy in stutterers and nonstutterers. *Journal of Speech and Hearing Research, 20*, 55–71.

Dushoff, J., Kain, M. P., & Bolker, B. M. (2019). I can see clearly now: Reinterpreting statistical significance. *Methods in Ecology and Evolution*, *2019*(November 2018), 3–6. https://doi.org/10.1111/2041-210X.13159

Etchell, A. C., Johnson, B. W., & Sowman, P. F. (2014). Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: a hypothesis and theory. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2014.00467

Falk, S., Müller, T., & Bella, S. D. (2014). Sensorimotor Synchronization in Stuttering Children and Adolescents. *Procedia - Social and Behavioral Sciences*, *126*(0), 206–207. https://doi.org/10.1016/j.sbspro.2014.02.375

Forster, D.C. & Webster, W.G. (2001) Speech-motor control and interhemispheric relations in recovered and persistent stuttering. *Dev. Neuropsychol., 19*, 125–145.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Rogers Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30. https://doi.org/10.18637/jss.v059.i09

Hilger, A.I., Zelaznik, H., & Smith, A. (2016). Evidence That Bimanual Motor Timing Performance Is Not a Significant Factor in Developmental Stuttering Allison. *American Journal of Speech-Language Pathology*, *59*(August), 674–685. https://doi.org/10.1044/2016

Knowles, J. E., & Frederick, C. (2019). merTools: Tools for Analyzing Mixed Effects Regression Models. Retrieved from http://cran.r-project.org/package=merTools

Ludlow, C. L., & Loucks, T. (2003). Stuttering: A dynamic motor control disorder. *Journal of Fluency Disorders*, *28*(4), 273–295. https://doi.org/10.1016/j.jfludis.2003.07.001

MacPherson, M., & Smith, A. (2013). Influences of sentence length and syntactic complexity on the speech motor control of children who stutter. *Journal of Speech, Language, and Hearing Research, 56*, 89–102.

Max L. (2004). Stuttering and internal models for sensorimotor control: a theoretical perspective to generate testable hypotheses. In: Maassen B, Kent R, Peters HFM, van Lieshout P, Hulstijn W, (Eds.), Speech motor control in normal and disordered speech (p. 357–87). Oxford, UK: Oxford University Press.

Max, L., Caruso, A., Gracco, V. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech, Language, & Hearing Research, 46*, 215–32.

Max, L., & Yudman, E. M. (2003). Accuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. *Journal of Speech, Language, and Hearing Research*, *46*(1), 146–163. https://doi.org/10.1044/1092-4388(2003/012)

Namasivayam, A. K., & van Lieshout, P. (2008). Investigating speech motor practice and learning in people who stutter. *Journal of Fluency Disorders, 33*, 32–51.

Neef, N. E., Paulus, W., Neef, A., Gudenberg, A. W. Von, & Sommer, M. (2011). Clinical Neurophysiology Reduced intracortical inhibition and facilitation in the primary motor tongue representation of adults who stutter. *Clinical Neurophysiology*, *122*(9), 1802–1811. https://doi.org/10.1016/j.clinph.2011.02.003

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from

    performance measures. *Cognitive Psychology, 19*, 1–32.

Olander, L., Smith, A., & Zelaznik, H. (2010). Development of Stuttering. *Journal of Speech

    Language and Hearing Research*, *53*(4), 876–886. https://doi.org/10.1044/1092-

    4388(2009/09-0007).

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria:

    R Foundation for Statistical Computing. Retrieved from http://cran.r-project.org/

Savion-Lemieux, T., Bailey, J. A., & Penhune, V. B. (2009). Developmental contributions to

    motor sequence learning. *Experimental Brain Research*, *195*, 293–306.

    https://doi.org/10.1007/s00221-009-1786-5

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities

    of preschool children who stutter: evidence from behavioral and kinematic indices of

    nonword repetition performance. *Journal of Fluency Disorders, 37*, 344–58.

Smith, A., & Kleinow, J. (2000). Kinematic correlates of speaking rate changes in stuttering and

    normally fluent adults. *Journal of Speech, Language and Hearing Research, 43*, 521–

    536.

Smits-Bandstra, S., De Nil, L. F., & Rochon, E. (2006). The transition to increased automaticity

    during finger sequence learning in adult males who stutter. *Journal of Fluency Disorders,

    31*, 22–42.

Smits-Bandstra S, Gracco V. Speech sequence learning in persons who stutter and persons with

    Parkinson's disease. Journal of Motor Behavior. 2013; 45(5):381–393.

Smits-Bandstra, S., & Gracco, V. (2013). Retention of Implicit Sequence Learning in Persons

    who Stutter and Persons with Parkinson's Disease Sarah. *Journal of Motor Behavior*,

    45(5), 381–393. https://doi.org/10.1038/jid.2014.371

SPSS for Windows. (2001). (Version 11.0) [Computer software]. Chicago: Author

Thomas, K.M & Nelson, C.A. (2001) Serial reaction time learning in preschool- and school-age

    children. *Journal of Experimental Child Psychology, 79*, 364–387.

Tumanova, V., Zebrowski, P.M., Goodman, S.S. & Arenas, R.M. (2015). Motor Practice Effects

    and Sensorimotor Integration in Adults who Stutter: Evidence from Visuomotor Tracking

    Performance. *Journal of Fluency Disorders, 45*, 52–72.

Taylor, J.A., Ivry, R.B. (2011). Flexible cognitive strategies during motor learning.

    PLoS Comput Biol, 7:e1001096, 1-13. doi: 10.1371/journal.pcbi.1001096

Usler, E., Smith, A., & Weber, and C. (2017). A Lag in Speech Motor Coordination During

    Sentence Production Is Associated With Stuttering Persistence in Young Children. *Journal*

    of Speech Language and Hearing Research, *60*(January), 51–62.

    https://doi.org/10.1044/2016

van Lieshout, PHHM.,Hulstijn, W., Peters, HFM. Searching for the weak link in the speech

    production chain of people who stutter: A motor skill approach.. In: Maassen, B.; Kent,

    RD.; Peters, HFM.; van Lieshout, PHHM.; Hulstijn, W., editors. Speech motor control in

    normal and disordered speech. Oxford University Press; Oxford: 2004. p. 313-357.

Walsh, B., Mettel, K. M., & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. *Journal of Neurodevelopmental Disorders*, 1–12. https://doi.org/10.1186/s11689-015-9123-8

Ward, D. (1997). Intrinsic and extrinsic timing in stutterers' speech: Data and implications. *Journal of Language and Speech, 40*, 289–310.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on *p* -Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p < 0.05." *American Statistician*, *73*, 1–19. https://doi.org/10.1080/00031305.2019.1583913

Webster, W.G. (1986). Response sequence organization and reproduction by stutterers. *Neurpsychologia, 24*, 813–821.

Weinstein, J., Caruso, A. J., Severing, K., & VerHoeve, J. (1989). Abnormalities of coulometer control in stutterers. Investigative Ophthalmology and Visual Science, 30, 480 (Abstract).

Wickham, H. (2017). tidyverse: Easily Install and Load the "Tidyverse."

Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering I: Persistency and recovery rates. *Journal of Speech, Language, and Hearing Research*, *42*, 1097–1112. https://doi.org/10.1086/250095

Zelaznik, H. N., Smith, A., Franz, E. A., & Ho, M. (1997). Differences in bimanual coordination associated with stuttering. *Acta Psychologica, 96*(3), 229-243.

Table 1. Correlations between age and various performance metrics for REP and RAN sequence across participant groups. * denotes statically significant results ($p < 0.05$).

	CWS	CRS	CON
Accuracy (REP)	$r(9) = 0.60, p = 0.12$	$r(10) = 0.21, p = 0.51$	$r(6) = 0.63, p = 0.038*$
Accuracy (RAN)	$r(9) = 0.63, p = 0.095$	$r(10) = 0.33, p = 0.30$	$r(6) = 0.74, p = 0.0097*$
Response Synchrony (REP)	$r(9) = -0.33, p = 0.43$	$r(10) = -0.52 p = 0.084$	$r(6) = -0.55 p = 0.083$
Response Synchrony (RAN)	$r(9) = -0.089, p = 0.83$	$r(10) = -0.37, p = 0.24$	$r(6) = -0.51, p = 0.11$

Figure 1(a,b). Schematic representation of the task (a) and timing of a single sequence element (b): each element started with a 100 ms blank screen, then a visual cue (hamster) appeared, a participant responded by selecting the appropriate response button; the visual cue disappeared, ending the trial. The response time was continuously recorded for 900 ms starting at the beginning of the element and ending 300 ms after the stimulus presentation. If the response was provided after 900 ms, it was considered a late response.
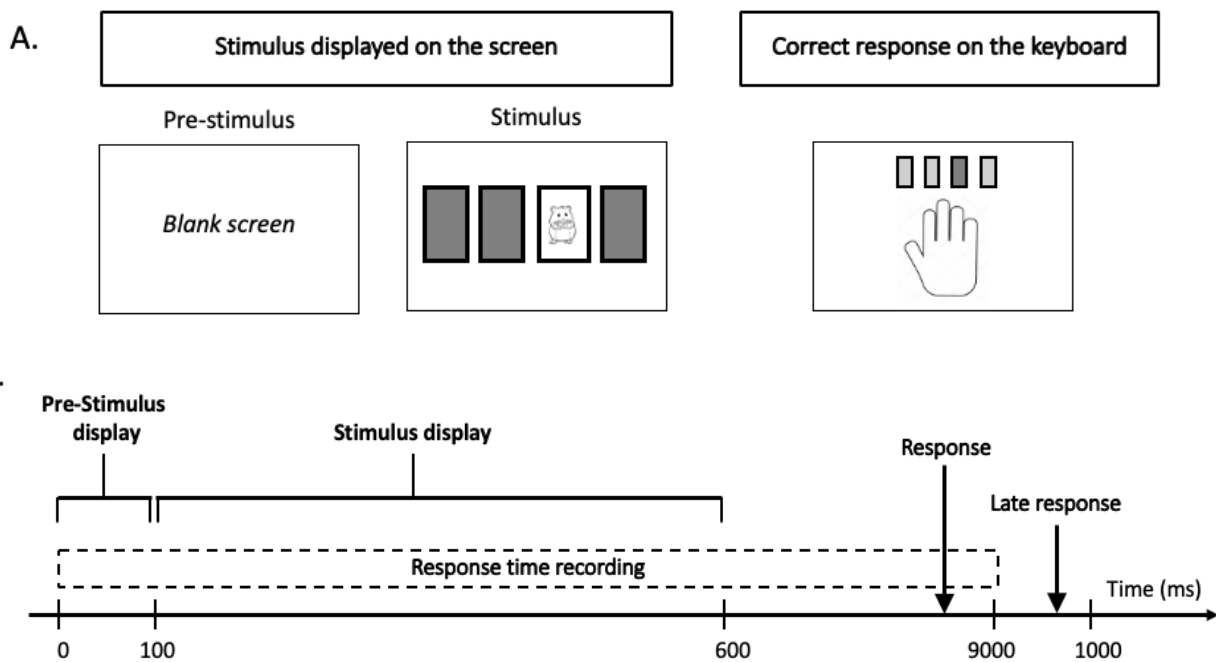
Figure 2(a-d). Estimated response accuracy with 95% confidence intervals for all groups across 3 blocks of practice on Day 1 and Day for RAN (a,b) and REP (c,d) sequences.
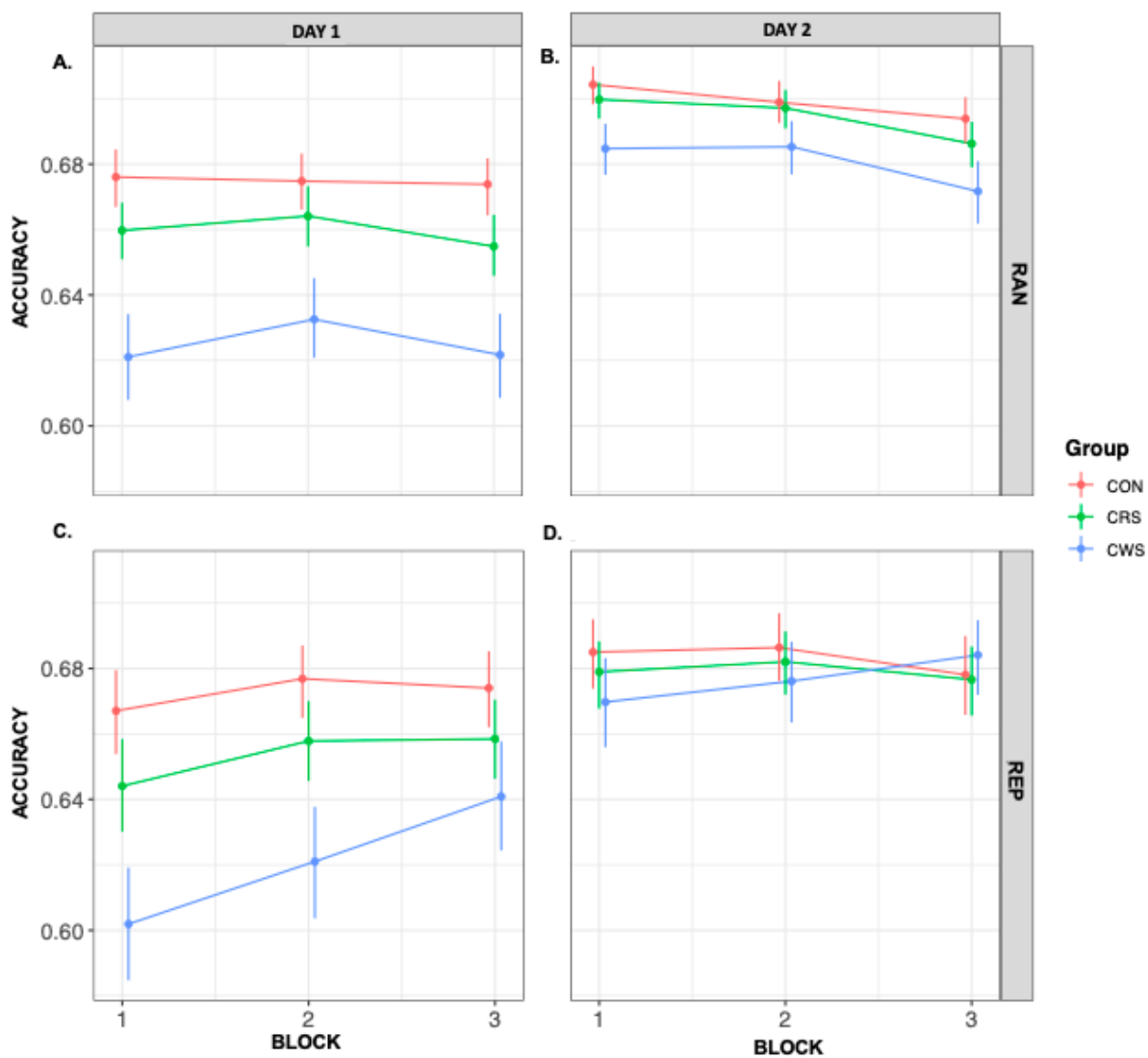
Figure 3(a-d). Estimated response synchronization time with 95% confidence intervals for all groups across 3 blocks of practice on Day 1 and Day for RAN (a,b) and REP (c,d) sequences.
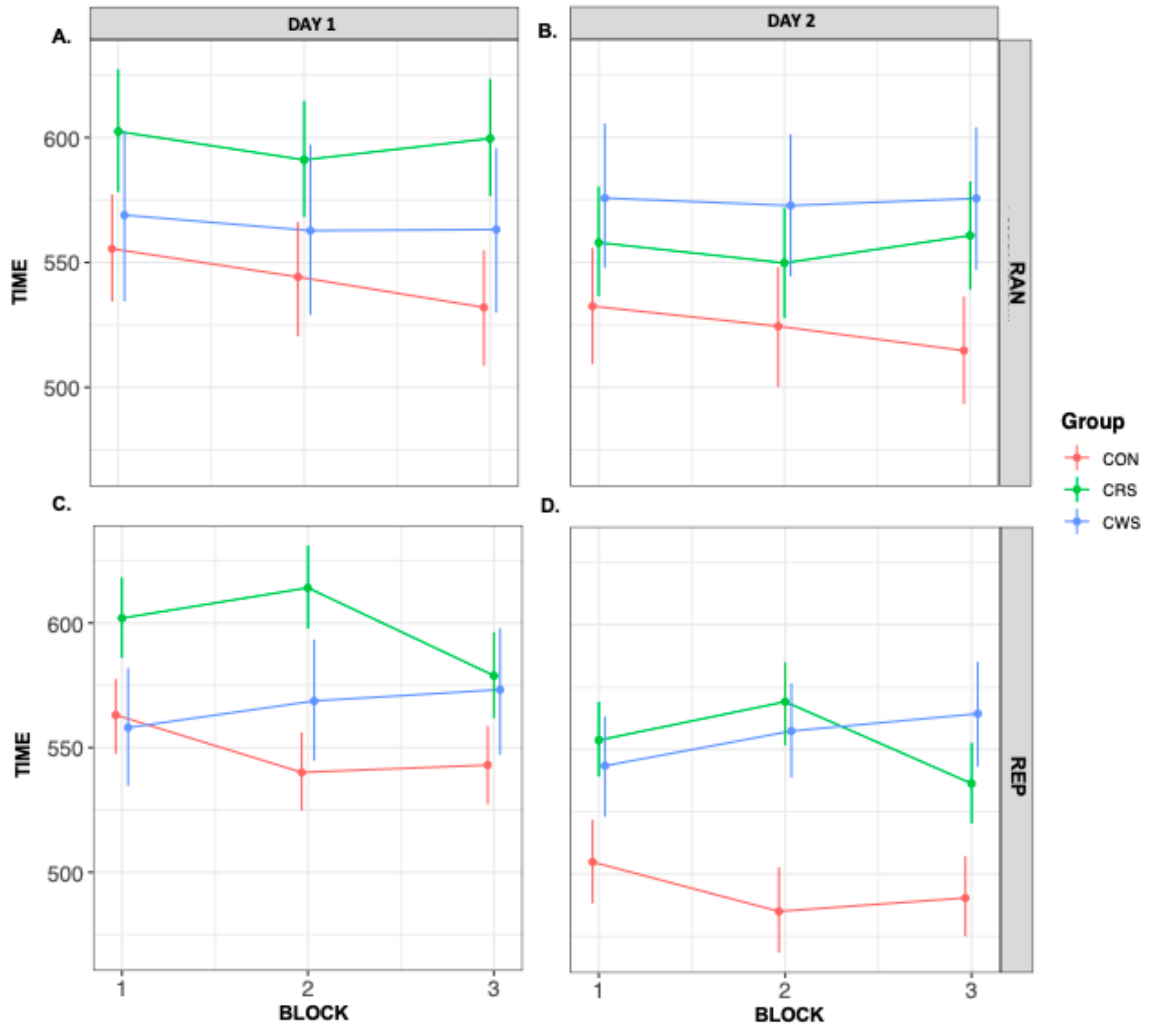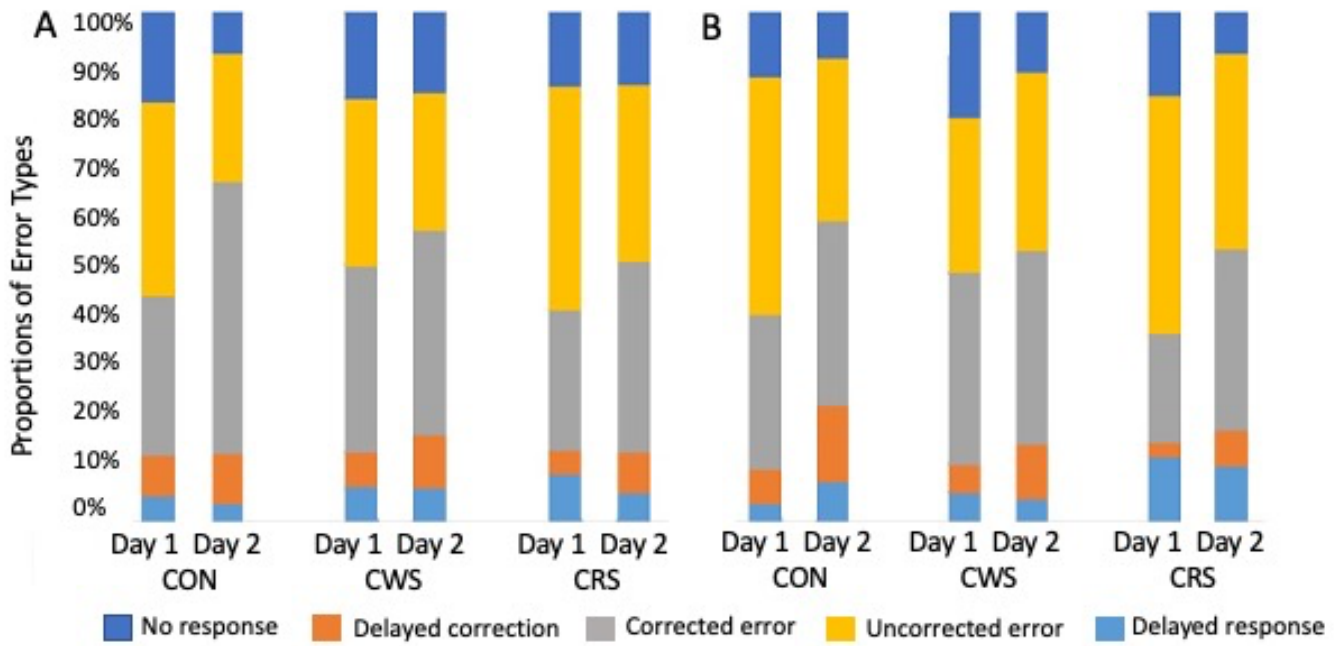
Figure 4. Proportions of ty pes of errors made on Day 1 (left) and Day 2 (right) by CON (top), CWS (middle) and CRS (bottom) in the A) REP and B) RAN sequences.

Accuracy Estimates for children who stutter (CWS; intercept), children with recovered stuttering

(CRS), controls (CON), age, experimental blocks (block1; block2), testing day (Day1; Day 2).

Predictors	Estimate	CI
(Intercept)	-2.08	[-3.92, -0.21]
RAN	-0.1	[-0.16, -0.04]
CON	0.3	[-0.04, 0.65]
CRS	0.06	[-0.27, 0.38]
Day 1	-0.37	[-0.46, -0.28]
block 1	-0.03	[-0.1, 0.04]
block 2	0.06	[-0.006, 0.13]
Age	0.3	[0.11, 0.48]
RAN:CON	-0.025	[-0.11, 0.06]
RAN:CRS	-0.04	[-0.12, 0.04]
RAN:Day 1	0.06	[0.015, 0.11]
RAN:block 1	-0.1	[-0.16, -0.03]
RAN:block 2	-0.01	[-0.08, 0.06]
CON:Day 1	0.15	[0.02, 0.29]
CRS:Day 1	0.018	[-0.11, 0.14]
CON:block 1	0.06	[-0.03, 0.16]
CRS:block 1	0.025	[-0.06, 0.11]
CON:block 2	-0.02	[0.114, 0.08]
CRS:block 2	0.02	[-0.07, 0.11]
Day 1:block 1	-0.08	[-0.15, -0.02]
Day 1:block 2	0.005	[-0.06, 0.07]
RAN:CON:Day 1	0.04	[-0.03, 0.11]
RAN:CRS:Day 1	0.02	[-0.05, 0.08]
RAN:CON:block 1	0.02	[-0.07, 0.18]
RAN:CRS:block 1	0.01	[-0.08, 0.1]
RAN:CON:block 2	0.055	[-0.04, 0.15]
RAN:CRS:block 2	0.003	[-0.09, 0.09]

Note. CON = control group; CRS = children with recovered stuttering; RAN= randomized condition

Response Synchrony for children who stutter (CWS; intercept), children with recovered

stuttering (CRS), controls (CON), age, experimental blocks (block1; block2), testing day (Day1;

Day 2).

Predictors	Estimate	CI
(Intercept)	825	[630.9, 1019]
RAN	3.86	[-0.58, 8.3]
CON	-28.6	[-64.27, 7.44]
CRS	20.1	[-13.34, 53.54]
Day 1	12.78	[2.51, 23.03]
block 1	3.35	[-2.89,
block 2	0.054	[-6.1, 6.21]
Age	-26.49	[-45.63, -7.42]
RAN:CON	2.78	[-3.07, 8.63]
RAN:CRS	-3.15	[-9.04, 2.74]
RAN:Day 1	-5	[-9.44, -0.57]
RAN:block 1	2.1	[-4.14, 8.28]
RAN:block 2	-2.95	[-9.09, 3.18]
CON:Day 1	5.12	[-8.97, 19.18]
CRS:Day 1	9	[-5.53, 21.75]
CON:block 1	8.25	[0.02, 16.52]
CRS:block 1	-0.82	[-9.1, 7.45]
CON:block 2	-4.33	[-12.52, 3.86]
CRS:block 2	4.55	[-3.63, 12.73]
Day1:block 1	1.5	[-4.07, 7.07]
Day1:block 2	-0.24	[-5.77, 5.3]
RAN:CON:Day 1	-4	[-9.85, 1.85]
RAN:CRS:Day 1	3.98	[-1.91, 9.86]
RAN:CON:block 1	-3.14	[-11.36, 5.09]
RAN:CRS:block 1	-1.26	[-9.53, 7.01]
RAN:CON:block 2	7.51	[-0.64, 15.67]
RAN:CRS:block 2	-8.17	[-16.35, 0.01]

Note. CON = control group; CRS = children with recovered stuttering; RAN= randomized
condition