An exploration of the role of executive functions in preschoolers' phonological development

Catherine Torrington Eaton & Nan Bernstein Ratner

Taylor & Francis
Taylor & Francis Group

An exploration of the role of executive functions in preschoolers' phonological development

Catherine Torrington Eaton [a] and Nan Bernstein Ratner[b]

[a]Department of Communication Disorders, Rockhurst University, Kansas City, MO, USA; [b]Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, USA

ABSTRACT
There is limited yet compelling evidence that domain-general processes may contribute to speech sound change. This study explored whether executive functions contribute to the achievement of adult-like speech production. Children who are 4 to 5 years old, 42 with high-average speech production skills, 11 with low-average and nine with speech sound disorder (SSD), participated in a battery of executive function and speech production tasks. Performance accuracy was compared across groups and also correlated with speech sound accuracy from a single-word naming task. Children with SSD demonstrated poorer performance than other groups on forward digit span, whereas children with low-average speech skills underperformed their peers on the Flexible Item Selection Task (FIST). These preliminary results suggest that children with speech errors may have less mature working memory than peers who have mastered phonological targets earlier in development.

Introduction

Most children acquire and then refine the sounds of their language without instruction. The majority of typically developing (TD) children achieve adult-like speech production prior to age 8 (see Bernthal, Bankson, & Flipsen, 2013 for summary). Yet for some children, speech sound accuracy must be explicitly targeted in speech therapy. The question addressed in this study is why most children spontaneously refine their speech production, while children with speech sound disorder (SSD) require assistance to achieve adult-like output.

Although there is an abundance of domain-specific theories on how children 'fine-tune' their speech production (e.g. speech perception, underlying phonological representations), there are no existing theories that consider if and how domain-general cognitive processes might be involved (Munson, Edwards, & Beckman, 2012; Rvachew, Rafaat, & Martin, 1999). Working memory, inhibitory control and cognitive flexibility, often considered the core executive functions (EFs) in adults (Diamond, 2013; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), might underpin the replacement of early-developing, inaccurate production patterns with adult-like articulation. Working memory (WM), defined as the ability to temporarily store information and mentally manipulate it (Davidson, Amso, Cruess Anderson, & Diamond, 2006), could theoretically underlie the ability to hold incoming evidence needed to then efficiently update

---

CONTACT Catherine Torrington Eaton ✉ cathyeaton12@gmail.com 📧 Department of Communication Disorders, Rockhurst University, 407b Van Ackeren Hall, 1100 Rockhurst Road, Kansas City, MO 64151, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/iclp.

phonological rules. Inhibitory control, often referred to as the ability to stop an inappropriate response (Simpson & Riggs, 2007), might be involved in suppressing old or aberrant speech patterns. Cognitive flexibility—defined as the ability to provide an appropriate response and then, when the task changes, to quickly shift to a different, but also appropriate response (Deak, Ray, & Pick, 2004)—might enable a child to flexibly adopt a new production pattern that more accurately matches the input.

In the adult literature, the three core EF constructs are often recruited in the same tasks and are seen as mutually supportive; however, research suggests that they can also be identified as distinct processes (Badre, 2008; Davidson et al., 2006; Miyake et al., 2000; Østby, Tamnes, Fjell, & Walhovd, 2011). Though historically fewer in number, behavioural and neurobiological studies of EFs in young children suggest that rapid changes in the EF system occur between three and six years of age (Amso & Casey, 2006; Davidson et al., 2006). For this reason, both the nature and relationship of EF components in preschool children are largely open to debate (Chevalier et al., 2012). Statistical methodologies such as factor and confirmatory factor analyses have been used to compare models of best fit, whereby performance on tasks factor together according to latent variables or the underlying cognitive processes. To date, studies have supported several models of EFs in preschoolers, including a unitary construct of executive control (Shing, Lindenberger, Diamond, Li, & Davidson, 2010; Wiebe, Andrews Espy, & Charak, 2008), a system similar to that of adults in which all three core EF components can be isolated (Diamond, 2013; Garon, Bryson, & Smith, 2008), and a two-construct system with only inhibitory control and WM (Miller, Giesbrecht, Muller, McInerney, & Kerns, 2012). In order to explore whether domain-general cognitive mechanisms do contribute to the achievement of adult-like speech production, it is important to have an understanding of the underlying organisation of EFs during this stage of development.

A number of studies have examined verbal short-term memory in children with SSD compared to TD children (Lewis et al., 2006; Shriberg et al., 2009; Tkach et al., 2011). Several researchers have found reduced short-term capacity in the SSD population using performance on a forward digit span task. Considerably more evidence has been found using non-word repetition tasks. Children with SSD consistently show lower non-word repetition accuracy than children who are TD, even when articulation is controlled (Lewis et al., 2011; Munson, Edwards, & Beckman, 2005; Shriberg et al., 2009). These results may suggest a deficit in WM that could affect children's ability to store and then manipulate phonological input.

Inhibitory control in preschoolers with SSD has been an underexplored area. Murphy et al. (2014) recently published a study demonstrating higher rates of false alarms on an auditory attention task by children with SSD as compared to controls, which the researchers interpreted as suggestive of a deficit in either inhibitory control or selective attention. In addition, treatment evidence has shown that correcting sounds in error by training unfamiliar words or non-words facilitates target learning (Cummings & Barlow, 2011; Gierut, Morrisette, & Ziemer, 2010). These results could suggest that the inability to inhibit early production patterns associated with known words may, in fact, interfere with speech sound change in children with SSD. In other words, it is possible that the use of unfamiliar words lessens the demand on impaired or inefficient inhibitory control processes.

Results by Dodd and colleagues suggest that some children with SSD perform more poorly on cognitive flexibility tasks than TD children (Crosbie, Holm, & Dodd, 2009; Dodd & McIntosh, 2008). In a task known as the Flexible Item Selection Test (FIST), children are asked to identify two cards on a particular dimension (i.e. colour, size, shape

or number) out of a set of three (Jacques & Zelazo, 2001). After choosing the first pair, participants are asked to select two cards from the same set that are related on a different dimension. The groups showed significant differences on this task; children with SSD were less successful on both the first and second selections of the FIST than the TD group. Dodd and her colleagues' interpretation of these findings is that children with a particular subtype of SSD have deficits in the ability to abstract rules—such as the rules governing native-language phonology—rather than an impairment of cognitive flexibility *per se*. Nonetheless, these results suggest the value of follow-up study to examine whether cognitive flexibility might be implicated in the process of typical speech sound development and the slower progress of some children in mastering adult-like speech production.

Based on evidence to date, in this study we asked whether the domain-general processes of WM, inhibitory control and/or cognitive flexibility contribute to the acquisition of adult-like speech sound production by exploring whether preschool children with varying speech production abilities—high-average, low-average and disordered—perform differently on EF tasks. We also examined whether specific EF tasks were strongly associated with speech sound accuracy. In order to better interpret our findings, we explored the underlying organisation of EFs.

Methods

Participants

This study was approved by the Institutional Review Board at the University of Maryland. Participants were recruited through a variety of sources in two metropolitan areas in the USA, Washington, D.C., and Kansas City, Missouri, and included university clinics, private and school-based SLP referral, community-based listservs, word of mouth and personal referral. Interested families were pre-screened to rule out children with hearing loss, fluency or motor issues, or bilingualism. Because high proficiency in more than one language has been shown to improve EFs (Martin-Rhee & Bialystok, 2008), children who had greater than 20% exposure to a second language were excluded. Families that met the pre-screening criteria provided written informed consent prior to child participation.

For inclusion in the study, all children were required to score within the typical range on two standardised assessments of receptive language and one test of non-verbal intelligence, and pass hearing and oral-motor screenings. Although not considered as inclusion or exclusion criteria, a 10-minute play session was incorporated in the study to further describe participants' expressive language skills. To explore differences in EFs that could be related to differential progress in mastering adult-like speech production, children were assigned to one of three groups based on performance on the Goldman-Fristoe Test of Articulation (GFTA; Goldman & Fristoe, 2000). We operationally defined the high-average group as children who scored higher than the 66th percentile, the low-average group as children who scored between the 33rd and 15th percentiles and the SSD group as children who scored below the 15th percentile. It was felt that creating distinct, non-overlapping levels of speech sound mastery would best address the general relationship between progress in children's speech sound accuracy and EFs. Overall study eligibility criteria were designed to create groups of children differing only in speech production abilities.

Procedures and measures

Experimental sessions took place approximately two weeks after eligibility testing. All testing sessions were audio-recorded using a Shure SM51 microphone placed approximately 8–12 inches from the participant and connected to a Marantz PMD600 digital recorder. One task was also video-recorded using a Sony 120× digital Handycam. Two orders were created for each experimental task and counterbalanced across participants. The order of administration of tasks used a set sequence; EF tasks were pseudo-randomised followed by administration of the speech task. This sequence was chosen so that the most cognitively demanding tasks were administered before fatigue became a potential factor. After the session, participants chose a small toy, and families received standardised test scores and interpretation administered by a certified speech-language pathologist as well as modest monetary compensation.

Digit span tasks

The digit span forward task assesses verbal short-term memory capacity, but the backward span task is proposed to tap into both capacity and mental manipulation processes (Diamond, 2013). Digit span forward and backward tasks were always administered consecutively so that participants could apply knowledge of the forward task to better understand what was being asked of them in the backward condition. Stimuli were presented as audio files in order to ensure consistency of rate, prosody and volume. Two practise trials were administered to teach each task. For the forward span task, participants were instructed to repeat numbers exactly as they were heard. Test trials started with two-digit strings. If the child correctly repeated the first string, the researcher increased the digit string by one (i.e. three-digit string followed by a four-digit string). If the first string was incorrect, the child was presented a second string with the same number of digits (i.e. three-digit string followed by a second three-digit string). This procedure continued until the child was unable to repeat either the first or second string. Similar procedures were used in the backward span, only participants were instructed to repeat numbers in reverse order. For the digit span tasks, the highest number of digits that were accurately repeated functioned as the raw score.

Animal span task

The third WM task was modelled after Willoughby et al.'s (2012) WM span task. The task was administered on a laptop using Power Point and began by having participants name 12 animals and four colours to ensure 100% familiarity with test items. For practise and experimental trials, one colour tile and one animal were pictured in a house line drawing. Participants were prompted to name first the colour and then the animal pictured in the house. The slide was then advanced to a picture of an empty house where participants were prompted to recall the animal name (not colour) that used to be in the house (Figure 1). The purpose of producing colour names during the encoding phase, but not in the recall stage, was to further tax WM (Fatzer & Roebers, 2012). The number of houses in each trial gradually increased from one to five. Every successfully recalled animal in a trial—regardless of the order in which it was named—counted towards the total raw score.
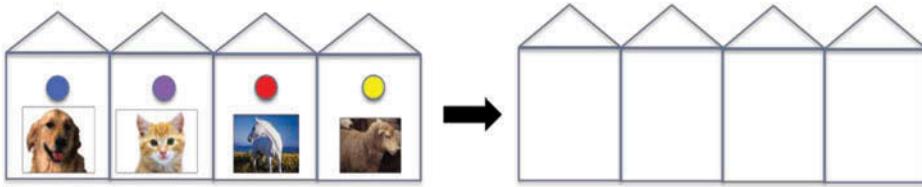
Figure 1. Animal span task (based on Willoughby et al., 2012).

Modified day-night stroop (DN-Stroop) task

This task, adapted from Pasalich et al. (2010), was intended to assess inhibitory control, specifically verbal response inhibition. Children were asked to verbalise semantically opposite labels, referred to as the 'silly name', for the images they saw (e.g., [picture of dog] 'this is Cat'). The experiment included 20 trials of four characters (dog, cat, boy, girl) that were displayed for 2000 msec followed by a 300 msec fixation cross. The task was administered on a MacBook Pro using Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993); trials were pseudo-randomised such that each of the four characters occurred once in every four trials, and no character was presented twice in a row. The total number of correctly named responses functioned as the raw score. Online responses were later re-scored by the researcher using video-recordings to ensure that delayed responses (i.e. responses initiated just before the next trial onset) were accurately coded.

Hearts and Flowers tasks

In these tasks children were taught to press a key on either the same or opposite side of the keyboard as the symbol they saw on the screen (Davidson et al., 2006). There were three blocks or tasks that were administered in a set order based on the following design: In the first block (the Hearts task), the prepotent motor response was reinforced when the child was asked to press a key on the same side as the stimulus (a heart). This task was a training precursor to the Flowers task. The second block (the Flowers task) required the child to inhibit the prepotent, habituated congruent response and instead press a key on the opposite side to the stimulus (a flower). This block assessed non-verbal inhibitory control. The third block (the Hearts and Flowers task) mixed the two types of responses, thus testing the child's flexibility in switching motor responses to match the stimulus (heart or flower) being presented. This block tested non-verbal cognitive flexibility.

Stimuli consisted of one blue heart and one blue flower from Microsoft shapes (3/4-inches in diameter), displayed on a white background. The [a] and ['] keys on the testing laptop were each covered by different stickers, which marked the response keys. Instructions prior to all test blocks emphasised both speed and accuracy. Each of the first two blocks (or tasks) consisted of 3 practise trials and 20 test trials. For these tasks, the stimuli occurred equally on the left or right side of a fixation cross in pseudo-random order such that no more than two consecutive trials on one side were presented. Participants had 1500 msec to respond with a 500 msec inter-stimulus interval (ISI) marked by a plain fixation cross. As per Davidson et al. (2006) there were no practise trials for the Hearts and Flowers task (block 3), but the experimenter verbally reviewed the

motor response for each symbol and then administered 30 trials, only 20 of which were switch trials. Switch trials, which required the participant to change rule sets from congruent to incongruent or incongruent to congruent responses, were the only trials calculated for purposes of this study (Deak et al., 2004). The response time increased to 2000 msec in the third block.

Accuracy and reaction times (RTs) of the Flowers task (block 2) and of the Hearts and Flowers task (block 3) were calculated from Psyscope output. Button presses under 200 msec were removed from both accuracy and RT measures (Davidson et al., 2006), which resulted in the exclusion of 3.7% and 1.7% of the data in blocks two and three, respectively. Inaccurate responses were not included in RT calculations.

Flexible item selection task (FIST)

The second cognitive flexibility task was the FIST, which was specifically designed for children aged three to five (Jacques & Zelazo, 2001). Stimuli and procedures were the same as described in Jacques and Zelazo's study. Children were explicitly taught how to point to cards that could 'go together' according to four possible attributes, and then practised the task with explicit feedback. For test trials, after participants made a first selection, regardless of accuracy, they were prompted to point to two *other* cards out of the set of three that went together (Figure 2). Though both participants' first and second selections were recorded, the measure relevant to cognitive flexibility was the total number of accurate second selections. That is, if the first selection was correct, the second selection was scored as accurate or inaccurate. If the first selection was incorrect, the second selection was automatically counted as inaccurate.

Spontaneous language production

A minimum of 40 child utterances from each participant's play session were transcribed and analysed using the software program CLAN (MacWhinney, 2000). Single-word utterances were not included in the total. Language samples were transcribed from the beginning of each session. The command DSS (Developmental Sentence Score) was run to obtain a measure of syntactic complexity, where higher values indicate greater competence with linguistic structures. The command VOC-D was used to measure vocabulary diversity; this algorithm is similar to type-token ratio, but is proposed to control for variations in sample size (see MacWhinney, 2000, for a review).

Figure 2. Flexible item selection task (based on Jacques & Zelazo, 2001).

Speech sound production task

Participants were administered the Picture Naming Task (PNT), a 125-single-word naming task that features each consonant at least twice in initial-, medial- and final-word positions and yields a total of 480 consonants if all word forms are produced as intended (Preston & Edwards, 2010). If the child was unable to spontaneously name the item, the target was elicited via delayed imitation. Morphological variations (e.g. affixes) were accepted.

Responses were broadly transcribed using Phon (Rose et al., 2006). Although diacritic marks and allophonic variations were often transcribed or noted, they did not factor into the primary measure of interest, Percent Consonants Correct-Revised (PCC-R; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). Distortions that were judged within the phonemic boundary for the target phoneme were counted as accurate. Accuracy judgments were determined using both perceptual and acoustic information in PRAAT (Boersma & Weenink, 2009). The PCC function in Phon was used to calculate the dependent measure of speech sound accuracy.

Fifteen per cent of the PNT (a total of 4898 consonants) was re-transcribed; segments that agreed in place, manner and voicing were counted as agreements (Tyler, Williams, & Lewis, 2006). Overall inter-rater reliability for the PNT was 0.825 using Cohen's kappa (Pearson's correlation $r = 0.827$), which was slightly lower than what is expected for broad transcription (Shriberg & Lof, 1991). Closer inspection revealed that one of the low-average and two of the SSD transcripts yielded significantly lower agreement between coders. Those transcripts were reviewed using consensus procedures described by Shriberg et al. (1984). Analyses were run a second time on all but the three transcripts, yielding a Cohen's kappa of 0.856 (Pearson's correlation $r = 0.856$), which is considered an acceptable level of agreement.

Results

Participant characteristics

A total of 62 children ages 4;0 to 5;11 (mean age = 60 months) met the final criteria and completed experimental testing. The three experimental groups were well-matched on several dimensions (see Table 1). There were no significant differences in age (SSD mean in months: 62.7 SD 5.6; low-average: 58.6 SD 6.8; high-average: 60.1 SD 6.2; $F(2, 59) = 0.926$, $p = 0.561$). Similarly, there was no between-group difference in gender distribution (number of females per group: SSD = 3; low-average = 5; high-average = 18; χ^2 (2, n = 62) = 0.931, $p = 0.628$).

In terms of standardised test performance, there was a large difference in percentile ranks on the articulation assessment ($F(2, 59) = 541.48$, $p < 0.001$, $\eta^2 = 0.948$), but no statistically significant difference on receptive language and cognitive measures. Although the mean of the SSD group on the receptive vocabulary measure (PPVT-4) was lower than the two other groups, this difference did not reach statistical significance ($F(2, 59) = 2.29$, $p = 0.110$, $\eta^2 = 0.072$).

The two expressive language measures, DSS and VOC-D, were compared between groups using an analysis of variance. Data from three participants (two SSD, one TD) were excluded because the child did not produce at least 40 greater than single-word utterances and/or speech production was too unintelligible to accurately transcribe. As can be seen in Table 1, the groups were equally matched in syntactic usage ($F(2, 56) = 0.210$, $p = 0.811$). On average, SSD participants demonstrated more restricted lexical inventories in

Table 1. Test score data.

Measure	SSD (n = 9)			Low-average (n = 11)			High-average (n = 42)		
	M	SD	Range	M	SD	Range	M	SD	Range
Articulation	9	3.6	3–13	23	6.7	15–33	82	8.0	66–96
Receptive lang	63	13.4	50–91	78	15.2	37–95	70	15.2	37–99
Receptive vocab	74	13.7	58–95	83	14.0	50–95	86	15.6	45–99
Non-verbal IQ	69	13.3	50–91	67	22.1	34–90	70	20.9	34–99
DSS	8	1.4	6–11	8.3	1.7	6–12	8	1.3	6–11
VOC-D	55	13.1	37–79	64	12.6	53–94	71	17.0	44–119
PNT PCC-R	64.8	0.1	49–75	73.0	0.1	51–84	94.6	2.7	88–99

Percentile ranks of standardised assessments: Articulation (Goldman & Fristoe, 2000); Receptive language (Wiig, Secord & Semel, 2004); Receptive vocabulary (Dunn & Dunn, 2007); Non-verbal intelligence (Kaufman & Kaufman, 2012). Non-standardised assessments: DSS, Developmental Sentence Score; VOC-D, Vocabulary Diversity; PNT PCC-R, Picture Naming Task Percent Consonants Correct- Revised.

play samples ($F(2, 56) = 5.585$, $p = 0.006$); pairwise comparisons using Bonferroni adjusted alpha levels showed the statistical difference was between the high-average and SSD groups' means (SSD vs. high-average: $p < 0.007$; low-average vs. high-average: $p < 0.392$; SSD vs. low-average: $p < 0.330$). These findings are consistent with studies demonstrating lower expressive vocabulary scores and lower lexical diversity in spontaneous speech in children with SSD (Edwards, Fox, & Rogers, 2002; Shriberg et al., 2009).

Group means from the PNT, the primary measure of speech sound accuracy, are also listed in Table 1. Results of a one-way analysis of variance showed a significant difference between groups in PCC-R scores on the PNT ($F(2, 59) = 128.83$, $p < 0.001$, $\eta^2 = 0.814$). Pairwise comparisons confirmed differences between all of the groups (SSD vs. high-average: $p < 0.001$; low-average vs. high-average: $p < 0.001$; SSD vs. low-average: $p < 0.010$). The overlap in PNT scores between the low-average and SSD group can best be explained by the fact that the PNT was not standardised; for example, the child in the low-average group with the lowest PNT accuracy fell into the low-typical range on the GFTA-2 for a 4.1-year-old child. Importantly, there was a high correlation between PNT scores and GFTA-2 standard scores ($r = .933$, $p < .001$) for all participants.

Differences in WM

Separate ANCOVAs were run on each of the two digit span tasks and the animal span task using age in months as a covariate to statistically control for differences based solely on maturation (Chevalier, Huber, Weibe, & Andrews Espy, 2013). Research has shown that young 4-year-olds perform differently on EF tasks than older 5-year-olds (Carlson, 2005). As seen in Table 2, overall children in the SSD and low-average groups underperformed their high-average peers on all tasks of WM. There was only a small difference, however, between group means on the backward digit span task, which proved difficult for most participants ($F(2, 59) = 0.792$, $p = 0.458$, $\eta^2 = 0.027$). Thirty-three per cent of children with SSD, 9% in the low-average group and 14% of the high-average group were unable to manipulate even two digits. Group differences on the animal span task were also not statistically significant ($F(2, 59) = 1.125$, $p = 0.332$, $\eta^2 = 0.037$).

In contrast, the forward digit span task showed a significant difference between group means ($F(2, 59) = 3.878$, $p = 0.026$, $\eta^2 = 0.118$). The effect size indicates that almost 12% of the variability is accounted for by performance differences on this task. Pairwise analyses indicated statistically significant differences between only the SSD and

Table 2. EF task mean raw scores.

Task	SSD		Low-average		High-average	
	M	SD	M	SD	M	SD
Forward digit span	3.6	0.7	3.9	0.7	4.2	.8
Max score: n/a						
Backward digit span	1.7	1.3	2.1	0.8	2.1	1.1
Max score: n/a						
Animal span	28.7	3.9	26.5	4.7	29.0	4.0
Max score: 37						
Flowers (accuracy)	12.2	4.4	9.8	4.9	12.1	4.7
Max score: 20						
Flowers (RT)	985.0	152.6	959.8	151.1	968.5	132.8
Max score: n/a						
Modified DN-Stroop	11.7	5.7	11.9	3.2	12.7	4.3
Max score: 20						
Hearts and Flowers (accuracy)	11.3	3.6	10.5	5.3	11.8	3.7
Max score: 20						
Hearts and Flowers (RT)	1154.5	362	1167.2	272	1196.1	270
Max score: n/a						
FIST	8.4	1.7	4.3	3.3	7.5	3.2
Max score: 12						

Accuracy data are reported in raw scores. RTs are in milliseconds.

high-average group means (SSD vs. high-average, $p = 0.027$; SSD vs. low-average, $p = 0.530$; low-average vs. high-average, $p = 0.808$). This difference is in agreement with studies showing lower digit span performance in children with moderate-severe SSD (Lewis et al., 2011; Tkach et al., 2011). In addition, although the mean performance of children with low-average speech was lower than children with high-average speech, digit span performance was not statistically different among the TD children. In total, results imply that deficits in WM specific to storage capacity may be implicated in the process of speech development.

Differences in inhibitory control

Three ANCOVAs were run using age as a covariate. The Flowers task – both accuracy and RT – and the modified DN-Stroop were examined to test whether performance on inhibitory control differed between groups (see Table 2 for all EF data). There were no statistically significant differences between group means on the Flowers task for accuracy ($F(2, 59) = 0.762$, $p = 0.471$, $\eta^2 = 0.026$) or RTs ($F(2, 59) = 0.132$, $p = 0.876$, $\eta^2 = 0.005$). Similarly, a comparison of group means on the DN-Stroop did not show a statistically significant difference ($F(2, 59) = 0.450$, $p = 0.640$, $\eta^2 = 0.015$). These results suggest that inhibitory control, as measured by these tasks, may not underlie the process of refining speech sounds in preschool-aged children.

Differences in cognitive flexibility

As with previous tasks, three analyses were run using age in months as a covariate on the measures of cognitive flexibility. Results from an ANCOVA demonstrated no statistically significant differences between group means on accuracy ($F(2, 59) = 0.450$, $p = 0.640$, $\eta^2 = 0.015$) or RTs ($F(2, 59) = 0.210$, $p = 0.811$, $\eta^2 = 0.007$) for the Hearts and Flowers task. It should be noted that accuracy for all participants in this task was close to chance

performance (chance = 10; SSD mean = 11.3, low-average mean = 10.5, high-average mean = 11.8), so any group differences in RTs likely had little to do with efficiency of processing.

A rank analysis of covariance was used to compare group means on the FIST because the data failed to meet assumptions of homogeneity (Quade, 1967). The non-parametric omnibus analysis demonstrated a statistically significant difference between groups on FIST scores ($F(2, 59) = 4.355$, $p = 0.017$, $\eta^2 = 0.133$). Post hoc analyses indicated that the low-average group mean was significantly different from the high-average group ($p = 0.018$) and approached statistical significance when compared to the SSD group ($p = 0.072$). Eight out of 11 children in the low-average group responded with less than 50% accuracy. There was no statistically significant difference between SSD and high-average group means ($p = 1.0$), although notably the SSD group mean was the highest among the three groups.

In sum, these findings indicated that children with low-average speech production skills significantly underperformed their SSD peers and peers with above-average speech productions skills on one task of cognitive flexibility. Findings are in partial agreement with those of Crosbie et al. (2009), whose data suggest that subgroups of children with SSD demonstrate less cognitive flexibility than children with more typical speech development. Further interpretation of these results, particularly as related to EF organisation, will be presented in the discussion.

Relationships between specific EF tasks and speech sound accuracy

To test the relationships between each of the three core EFs and speech sound development, Pearson's partial correlations, using age as a control variable, were run between tasks and the PNT PCC-R. This analysis used z-scores for the EF tasks rather than raw scores.

Results (see Table 3) indicated that only the forward digit span task was related to speech production accuracy as measured by the PNT ($r = 0.379$, $p = 0.003$). The correlation was positive; children who repeated more digits had greater speech sound accuracy. These findings are in agreement with the results of this study and others, suggesting that children with SSD have deficits in short-term memory capacity.

There were other relationships of interest among EF tasks. As seen in Table 3, forward digit span was moderately correlated with performance on both backward digit span ($r = 0.307$, $p = 0.016$) and the FIST ($r = 0.289$, $p = 0.024$). The Flowers task was moderately correlated with both the Hearts and Flowers task ($r = 0.376$, $p = 0.003$) and animal span ($r = 0.400$, $p = 0.001$).

Organization of EFs

A factor analysis was conducted to examine whether the experimental tasks measured specific core EF constructs as designed or whether they might group according to other latent variables. As described earlier, an understanding of the underlying organisation of core EFs in children is important for interpreting this study's findings. Results from the component matrix using varimax rotation with Kaiser normalisation showed a two-factor solution, with the first factor explaining 29.3% and the second explaining 23.9% of the variance. This analysis agreed with results from the correlation matrix: Digit span tasks

Table 3. Partial correlation matrix between EF tasks and PNT with age as a control variable.

	1	2	3	4	5	6	7	8
1 PNT	–	0.379**	0.226	0.072	0.085	0.088	0.133	0.054
2 Forw	0.379*	–	0.307	−0.119	0.098	−0.119	0.289	−0.042
3 Back	0.226	0.307	–	0.160	0.052	−0.204	0.086	−0.078
4 Anim	0.072	−0.119	0.160	–	0.073	0.400*	0.050	0.257
5 DNS	0.085	0.098	0.052	0.073	–	−0.040	0.063	0.195
6 Flow	0.088	−0.119	−0.204	0.400*	−0.040	–	0.125	0.376*
7 FIST	0.133	0.289	0.086	0.050	0.063	0.125	–	0.065
8 H&F	0.054	−0.042	−0.078	0.257	0.195	0.376*	0.065	–

*Indicates p-value is <.01; PNT = Picture Naming Task; Forw = Forward digit span; Back = Backward digit span; Anim = Animal span; DNS = Modified Day-Night Stroop task; Flow = Flowers task; FIST = Flexible Item Selection Task; H&F = Heart and Flowers task.

and FIST loaded onto one construct, whereas the Hearts and Flowers tasks, animal span task and DN-Stroop loaded onto a second construct.

Discussion

This study asked whether domain-general processes underlie the achievement of adult-like speech production by examining EF profiles of children with high-average, low-average and disordered (SSD) speech production profiles. Results demonstrated between-group performance differences in only two tasks: Decreased capacity on forward digit span in children with SSD and poorer performance on the FIST in children with low-average speech skills. Of these two tasks, only forward digit span was significantly associated with speech sound accuracy. Interestingly, forward and backward digit spans and the FIST were correlated with each other and also factored together, suggesting a similar underlying cognitive component. The modified DN-Stroop, animal span task and Hearts and Flowers tasks were not related to speech sound accuracy, but they factored together under a different latent variable.

Organisation of EFs in preschoolers

In order to more effectively interpret the relationship between speech sound change and core EFs, it is important to elucidate the nature of EFs in young children. Research in the preschool population has supported a number of different models of core EF organisation during development, from a single or unitary EF construct to three distinct constructs similar to adult models (Garon et al., 2008; Wiebe et al., 2008; Shing et al., 2010; Miller et al., 2012; Diamond, 2013). Data analysis from participants in the current study supported a dual-construct system of organisation consisting of WM and inhibitory control (Miller et al., 2012). As was evident from factor analysis, children likely accomplished tasks through reliance on alternate cognitive skills rather than those that were targeted (Dauvier, Chevalier & Blaye, 2012; Ramscar, Dye, Gustafson, & Klein, 2013); specifically, cognitive flexibility tasks seemed to rely on the core EFs of WM and inhibitory control to perform.

The first group of tasks identified in the factor analysis consisted of forward and backward digit spans and the FIST. These results suggest that, for this age group, WM may underlie these three tasks. Forward digit span is believed to measure storage capacity,

whereas the backward digit span task likely measures both capacity and manipulation of information (Diamond, 2013). Although the FIST was intended to provide a measure of cognitive flexibility, preschoolers appeared to rely on WM to make their second selections. Specifically, successful performance on the FIST requires temporarily storing one's first selection while considering alternate features that would constitute an accurate second response.

The second group of tasks that factored together under a different latent variable included the Hearts and Flowers tasks, animal span and the modified DN-Stroop. It seems likely that inhibitory control was the EF construct underlying these tasks. Both the Flowers and Hearts and Flowers tasks require cognitive control to inhibit the congruent or previous response. Animal span, although it was selected as a WM task, likely required inhibition due to interference from previous trials. The modified DN-Stroop, which was specifically designed to tap into inhibitory control, was perhaps surprisingly the least weighted task in this construct and did not correlate with any other task. This result is consistent with that of Pasalich et al. (2010), who found that the DN-Stroop was not associated with performance on the more traditional preschool Stroop task.

The present study's results support a dual-construct model and lead to a number of questions about cognitive flexibility and task-switching in children. It is possible that the tasks used in this study did not provide an adequate measure of cognitive flexibility. Card-sorting tasks have been used in the adult literature to tap into this domain (Stuss et al., 2000) and perhaps the inclusion of a child-friendly version would have yielded different results in the factor analysis.

On the other hand, the ability to adeptly switch between tasks may be so reliant on WM and inhibitory control abilities in preschoolers that cognitive flexibility should not be considered a separate construct until later in development. Even in Diamond's model of core EFs (2013), which includes all three components, the interaction between WM and inhibitory control serves as the foundation for cognitive flexibility. Although the tasks used in this study were originally designed to test hypotheses for three core EFs, we used a dual construct model to interpret results.

WM and speech sound change

Overall, results of this study suggest not only that WM underlies speech sound change, but also that differentially affected components—storage capacity versus manipulation—may correspond to distinct subtypes of development. We discuss this further below.

Age-appropriate short-term capacity appears to be logically and empirically essential in the development of one's native language sound system. Results indicated that children with SSD were more limited in digit span recall than children with speech delay and TD children. These findings are consistent with evidence from the SSD literature more broadly (Lewis et al., 2006; Shriberg et al., 2009; Tkach et al., 2011). Results also showed that forward digit span was significantly correlated with speech sound accuracy, which again suggests a fundamental role of short-term memory. Similar to what has been proposed for SLI and word learning (Gathercole, 2006), short-term memory is theoretically involved in the process of forming stable phonological representations or native speech sound categories over time (Munson et al., 2012). If input representations decay more quickly in short-term memory, templates for accurate production may be inaccurate

or less available. Furthermore, the child's ability to compare his own productions to the ambient targets would also be impaired. Without appropriate stored models, production accuracy would be difficult to achieve.

A second component of WM, mental manipulation, may also play a role in speech sound change. As indicated by the correlational and factor analyses, the FIST was a WM rather than a cognitive flexibility task for this age group. There were limited demands on short-term memory capacity because the cards were in front of participants for the duration of each trial. Rather, it appears that mental manipulation was necessary in formulating a second set of features that was distinct from the first. This function could conceivably underlie speech sound change, whereby mental manipulation is used to isolate differences between the child's own productions and mature target representations, and then update or overwrite phonological rules. A weakness in this cognitive skill could conceivably affect speech sound change.

The fact that backward digit span was only marginally related to speech sound accuracy might have been due to extraneous factors such as task difficulty and age. This task proved to be so difficult that up to a quarter of participants were unable to manipulate even two items during instructed practise. Thus, floor effects likely affected the results for this task. Although backward digit span is a sensitive measure of WM in older children and adults, findings from this study suggest that it may be an inappropriate task for evaluating this age group (Carlson, 2005).

Inhibitory control and speech sound development

According to results from the factor analysis, four tasks measured the construct of inhibitory control. Children with SSD performed similarly as their peers on these tasks. In addition, none of the inhibitory control tasks was related to speech sound accuracy. These findings provide evidence against the hypothesis that poor inhibitory control underlies changes in speech sound accuracy.

On the other hand, a closer look at the tasks identified in this construct offers a different interpretation. The Hearts and Flowers tasks and the animal span task were heavily reliant on visuo-spatial processing. One strategy that children used when completing the latter task was to cue animal recall by location. For instance, if participants had difficulty recalling a particular animal, they frequently pointed to the specific house where it was located. It is possible that performance on these tasks was a greater reflection of children's visuo-spatial skills than inhibitory control. Future research should consider alternate tasks such as conventional Stroop tasks or more verbal rather than visual tasks with this population before the hypothesised relationship between inhibitory control and maturation of speech sound accuracy is rejected.

Limitations and future directions

Several limitations of this study were related to the composition of the experimental groups. It is possible that larger groups might have affected the results by adding statistical power. Clinical participants were not recruited according to profiles such as error typicality or speech sound consistency, two variables that have been shown to influence performance on the FIST specifically. Likewise, there was no attempt to

control for differences in attention; co-morbid diagnoses such as attention deficit disorder could contribute to differences in speech sound or EF profiles. Future work that controls these factors may lead to better specification of distinct underlying deficits or subgroups of children with significantly impaired profiles of speech sound development, which could in turn result in more targeted interventions to bring these children to a more appropriate level of speech sound accuracy (Crosbie et al., 2009; Tyler, Lewis, & Welch, 2003).

The tasks used in this study were based on the most current literature on EFs in young children. There is considerable work to be done towards finding or refining tasks that measure exactly what they intend to measure in this challenging population. It is also possible that performance differences on these EF tasks may relate more to linguistic than general cognitive skills. Future work with young children should continue to explore different tasks to ensure that these preliminary results are replicable.

Acknowledgments

Declaration of interest

ORCID

Catherine Torrington Eaton http://orcid.org/0000-0002-1874-0854

References

Amso, D., & Casey, B. (2006). Beyond what develops when: Neuroimaging may inform how cognition changes with development. *Current Directions in Psychological Science*, *15*, 24–29.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*, 193–200.

Bernthal, N., Bankson J., & Flipsen, P. (2013). *Articulation and phonological disorders in children* (7th ed.). Boston: Pearson Publishing.

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer*, version 5.1.05. Retrieved from http://www.praat.org.

Carlson, S. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*, 595–616.

Chevalier, N., Huber, K., Wiebe, S., & Andrews Espy, K. (2013). Qualitative change in executive control during childhood and adulthood. *Cognition*, *128*, 1–12.

Chevalier, N., Sheffield, T., Nelson, J., Clark, C., Wiebe, S., & Andrews Espy, K. (2012). Underpinnings of the costs of flexibility in preschool children: The roles of inhibition and working memory. *Developmental Neuropsychology*, *37*, 99–118.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.

Crosbie, S., Holm, A., & Dodd, B. (2009). Cognitive flexibility in children with and without speech disorder. *Child Language Teaching and Therapy*, *25*, 250–270.

Cummings, A., & Barlow, J. (2011). A comparison of word lexicality in the treatment of speech sound disorders. *Clinical Linguistics and Phonetics*, *25*, 265–286.

Dauvier, B., Chevalier, N., & Blaye, A. (2012). Using finite mixture of GLMs to explore variability in children's flexibility in a task-switching paradigm. *Cognitive Development*, *27*, 440–454.

Davidson, M., Amso, D., Cruess Anderson, L., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078.

Deak, G., Ray, S., & Pick, A. (2004). Effects of age, reminders and task difficulty on young children's rule-switching flexibility. *Cognitive Development*, *19*, 385–400.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Dodd, B., & McIntosh, B. (2008). The input processing cognitive linguistic and oro-motor skills of children with speech difficulty. *International Journal of Language and Communication Disorders*, *10*, 169–178.

Dunn, L., & Dunn, D. (2007). *Peabody picture vocabulary test*, 4th edition. Circle Pines, MN: American Guidance Service Publishing/Pearson Assessments.

Edwards, J., Fox, R., & Rogers, C. (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and articulatory accuracy. *Journal of Speech, Language, and Hearing Research*, *45*, 231–242.

Fatzer, S., & Roebers, C. (2012). Language and executive functions: The effect of articulatory suppression on executive functioning in children. *Journal of Cognition and Development*, *13*, 454–472.

Garon, N., Bryson, S., & Smith, I. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60.

Gathercole S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543.

Gierut, J., Morrisette, M., & Ziemer, S. (2010). Nonwords and generalization in children with phonological disorders. *American Journal of Speech-Language Pathology*, *19*, 167–177.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe test of articulation*. Minnesota: American Guidance Services, Inc.

Jacques, S., & Zelazo, P. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, *20*, 573–591.

Kaufman, A., & Kaufman, N. (2012). *Kaufman Brief Intelligence Test*, 2nd edition. San Antonio, TX: Pearson Education.

Lewis, B., Avrich, A., Freebairn, L., Taylor, H., Iyengar, S., & Stein, C. (2011). Subtyping children with speech sound disorders by endophenotypes. *Topics in Language Disorders*, *31*, 112–127.

Lewis, B., Shriberg, L, Freebairn, L, Hansen, A., Stein, C., Taylor, H., & Iyengar, S. (2006). The genetic bases of speech sound disorders: Evidence from spoken and written language. *Journal of Speech, Language, and Hearing Research*, *49*, 1294–1312.

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

Martin-Rhee & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and cognition*, *11*, 81–93.

Miller, M., Giesbrecht, G., Muller, U., McInerney, R., & Kerns, K., (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, *13*, 395–423.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., & Howerter, A. (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.

Munson, B., Edwards, J., & Beckman, M. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language and Hearing Research*, 48, 61–78.

Munson, B., Edwards, J., & Beckman, M. (2012). Phonological representations in language acquisition: Climbing the ladder of abstraction. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The oxford handbook of laboratory phonology* (pp. 288–309). Oxford, NY: Oxford University Press.

Murphy, C., Pagan-Neves, L., Wertzner, H., & Schochat, E. (2014). Auditory and visual sustained attention in children with speech sound disorder. *PLOS One*, 9, 1–9.

Østby, Y., Tamnes, C., Fjell, A., & Walhovd, K. (2011). Morphometry and connectivity of the fronto-parietal verbal working memory network in development. *Neuropsychologia*, 49, 3854–3862.

Pasalich, D., Livesey, D., & Livesey, E. (2010). Performance on Stroop-like assessments of inhibitory control by 4- and 5-year-old children. *Infant and Child Development*, 19, 252–263.

Preston, J., & Edwards, M. (2010). Phonological awareness and types of sound errors in preschoolers with speech sound disorders. *Journal of Speech, Language and Hearing Research*, 53, 44–60.

Quade, D. (1967). Rank analysis of covariance. *Journal of the American Statistical Association*, 62, 1187–1200.

Ramscar, M., Dye, M., Gustafson, J., & Klein, J. (2013). Dual routes to cognitive-flexibility: Learning and response-conflict resolution in the dimensional change card sort task. *Child Development*, 84, 1308–1323.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2006). Introducing Phon: A software solution for the study of phonological acquisition. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 489–500). Somerville, MA: Cascadilla Press.

Rvachew, S., Rafaat, S. & Martin, M. (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology*, 8, 33–43.

Shing, Y., Lindenberger, U., Diamond, A., Li, S., & Davidson, M. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology*, 35, 679–697.

Shriberg, L., Austin, D., Lewis, B., McSweeny, J., & Wilson, D. (1997). The Percentage of Consonant Correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language and Hearing Research*, 40, 708–722.

Shriberg, L., Kwiatkowski, J., & Hoffman, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research*, 27, 456–465.

Shriberg, L. & Lof, G. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*, 5, 225–279.

Shriberg, L, Lohmeier, H., Campbell, T., Dollaghan, C., Green, J., & Moore, C. (2009). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Task (SRT). *Journal of Speech, Language and Hearing Research*, 52, 1189–1212.

Simpson, A. & Riggs, K. (2007). Under what conditions do young children have difficulty inhibiting manual actions? *Developmental Psychology*, 43, 417–428.

Stuss, D., Levine, B., Alexander, M., Hong, J., Palumbo, C., Hamer, L.,... Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, 38, 388–402.

Tkach, J., Chen, X., Freebairn, L, Schmithorst, V., Holland, S., & Lewis, B. (2011). Neural correlates of phonological processing in speech sound disorder: A functional magnetic resonance imaging study. *Brain and Language*, 119, 42–49.

Tyler, A., Lewis, K. & Welch, C. (2003). Predictors of phonological change following intervention. *American Journal of Speech-Language Pathology*, 12, 289–298.

Tyler, A., Williams, M. & Lewis, K. (2006). Error consistency and the evaluation of treatment outcomes. *Clinical Linguistics & Phonetics*, *20*, 411–422.

Wiebe, S., Andrews Espy, K. & Charak (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*, 575–587.

Wiig, E., Secord, W. & Semel, E. (2004). *Clinical evaluation of language fundamentals in preschool* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Willoughby, M., Blair, C., Wirth, R., Greenburg, M. & The Family Life Project Investigators (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment, 24*, 226–239.