JSLHR

Research Article

Cognitive Flexibility for Semantic and Perceptual Information in Developmental Stuttering

Julie D. Anderson,[a] Stacy A. Wagovich,[b] and Levi Ofoe[a]

Purpose: The purpose of this study was to examine cognitive flexibility for semantic and perceptual information in preschool children who stutter (CWS) and who do not stutter (CWNS).
Method: Participants were 44 CWS and 44 CWNS between the ages of 3;0 and 5;11 (years;months). Cognitive flexibility was measured using semantic and perceptual categorization tasks. In each task, children were required to match a target object with two different semantic or perceptual associates. Main dependent variables were reaction time and accuracy.
Results: The accuracy with which CWS and CWNS shifted between one semantic and perceptual representation to another was similar, but the CWS did so significantly more slowly. Both groups of children had more difficulty switching between perceptual representations than semantic ones.
Conclusion: CWS are less efficient (slower), though not less accurate, than CWNS in their ability to switch between different representations in both the verbal and nonverbal domains.

In recent years, there has been increasing interest in examining the role of executive function (i.e., working memory, inhibition, and cognitive flexibility [CF]) in developmental stuttering. This interest has been fueled, at least in part, by the fact that some studies have reported that the spoken language processing abilities of children who stutter (CWS) may not be as robust as those of children who do not stutter (CWNS; e.g., Anderson & Conture, 2004; Byrd et al., 2007; Mohan & Weber, 2015; Sasisekaran et al., 2013; Weber-Fox et al., 2013) and that language processes are inextricably linked to executive function (Müller et al., 2009; Pisoni et al., 2010). While the study of executive function in CWS is still in its infancy, evidence has emerged to suggest that CWS may differ from their typically developing peers in their ability to temporarily store (short-term memory) and manipulate (working memory) information and suppress a dominant response or irrelevant information (inhibition; e.g., Anderson & Wagovich, 2017; Anderson et al., 2019; Eggers et al., 2013; Hakim & Ratner, 2004). Whether differences also exist between CWS and CWNS in CF is less than clear, given that few investigators have, thus far, examined this component of executive function in CWS.

CF refers to the ability to flexibly alternate or switch between multiple mental states, representations, perspectives, or rules (Chevalier et al., 2011; Deák, 2003; Garon et al., 2008; Jacques & Zelazo, 2005; Zelazo et al., 2008). In young children, CF is thought to emerge from the development of working memory and inhibition, because being able to switch from one mental set (i.e., a stimulus–response association) to another requires that multiple mental sets be held in memory and previous mental sets be suppressed (Carroll et al., 2016; Cragg & Chevalier, 2012; Diamond, 2013; Garon et al., 2008). Thus, since CWS have been shown in some studies to be less skilled than CWNS in components of working memory and inhibition, it seems reasonable to expect that they might also have weaknesses in CF. These findings serve to motivate this study, the focus of which is to examine CF in preschool CWS and their typically fluent peers. By way of background, we begin with a brief review of the literature on memory and inhibition, including a discussion of the findings, to date, from studies of CWS. We then consider CF as a divergent construct, along with the few studies that have been conducted on the topic in CWS.

[a]Department of Speech and Hearing Sciences, Indiana University, Bloomington
[b]Department of Speech, Language and Hearing Sciences, University of Missouri, Columbia

Correspondence to Julie D. Anderson: judander@indiana.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Courtney T. Byrd

Received August 6, 2019
Revision received February 13, 2020
Accepted August 16, 2020
https://doi.org/10.1044/2020_JSLHR-19-00119

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Working Memory and Inhibition

Working memory involves the temporary storage (short-term memory) and manipulation of auditory and visual information (Baddeley, 2000, 2003). Several theoretical models have been proposed to explain the nature, structure, and function of working memory, but among the most influential is Baddeley's (2000, 2003) multiple-component model. According to Baddeley, there are four components of working memory: the central executive, which is the superordinate attentional control system that integrates working memory with other cognitive processes; the phonological loop and visuospatial sketchpad, which store and manipulate verbal and visuospatial information, respectively; and the episodic buffer, which provides temporary storage for multidimensional information, connecting it to long-term memory.

Of the four working memory components, the phonological loop (i.e., verbal short-term memory), which consists of the phonological store and an articulatory rehearsal mechanism, has been of most interest to researchers in developmental stuttering, presumably because of its strong association with language acquisition and learning (Archibald, 2017; Ellis & Sinclair, 1996; Gathercole et al., 1997; Martin & Gupta, 2004). Verbal short-term memory has been examined in CWS using nonword repetition tasks. Findings from these studies have been contradictory, with some reporting that CWS are less successful than CWNS at repeating nonwords (e.g., Anderson et al., 2006; Hakim & Ratner, 2004; Pelczarski & Yaruss, 2016) and others reporting no such differences (e.g., Sasisekaran & Byrd, 2013; Smith et al., 2012). One problem with using nonword repetition as a measure of working memory, however, is that it also places demands on other skills besides verbal short-term memory, such as auditory perception, phonological processing, and motor planning and execution (Estes et al., 2007). Thus, for those studies for which differences have been found, it is not entirely clear what the nature of the problem is for CWS, although verbal short-term memory is a likely culprit as nonword repetition places a heavy demand on this component of executive function.

Verbal short-term memory in CWS has also been examined using forward span tasks, and while these tasks generally place fewer demands on other processes compared to nonword repetition, the findings have also been conflicting (e.g., Anderson et al., 2019; Pelczarski & Yaruss, 2016). In the most recent of these studies, Anderson et al. (2019) examined the verbal short-term memory skills of preschool CWS and CWNS using two forward span tasks, which focused on the influence of phonological and semantic similarity. The authors found differences between groups in memory span, particularly when the phonological makeup of words was manipulated in the word lists to be repeated; more subtle differences in patterns of performance were observed when semantic qualities of words were manipulated. Findings were interpreted to suggest that CWS have weaknesses in verbal short-term memory and that these weaknesses are apparent in their use of phonological and, to a lesser extent, semantic processing, as an aid to memory.

Another component of executive function that has received attention in recent years in the developmental stuttering literature is inhibition. Inhibition can be defined as the ability to suppress a dominant (prepotent) thought, behavior, or emotion in favor of a less dominant but more appropriate thought, behavior, or emotion (Diamond, 2013; Garon et al., 2008). According to Friedman and Miyake (2004), there are three main types of inhibition: simple and complex response inhibition, resistance to distractor interference, and resistance to proactive interference. Simple prepotent response inhibition involves the suppression of a natural, prepotent response, whereas complex response inhibition also requires an alternative, less dominant response to be executed. Resistance to distractor interference involves suppressing irrelevant information, whereas resistance to proactive interference task involves suppressing previously relevant information.

Several investigators have examined inhibition, primarily response inhibition, in CWS, and like verbal short-term memory, the findings have been mixed. In particular, Anderson and Wagovich (2017) and Eggers et al. (2013) found evidence to suggest that inhibition may be an area of weakness for CWS, whereas Piispala et al. (2016) and Eggers et al. (2018) did not. However, in subsequent studies, Piispala et al. (2017, 2018) reanalyzed the data from their original 2016 event-related potential study and arrived at a different conclusion. In the 2016 study, school-age CWS and CWNS completed the Go/NoGo subtest of the Amsterdam Neuropsychological Tasks (ANT; De Sonneville, 2009), and N2 and P3 activity was measured over nine electrodes. In this task, children press a button when they see a symbol of a man running (the Go stimulus), but not when they see a symbol of a man standing (the NoGo stimulus). It is this latter stimulus—the NoGo stimulus—that requires inhibition, in that children must refrain from pressing the button in response to this stimulus. In the 2017 reanalysis, Piispala et al. reexamined the original event-related potential data by calculating a measure of global brain activation over 36 electrodes. Contrary to their initial report, they found that the CWS exhibited a weak or absent NoGo P3 component, which they interpreted as a deficit in inhibitory control. In the 2018 reanalysis, they also reported that CWS had reduced occipital alpha activity in the NoGo condition, suggesting difficulties inhibiting visual attention.

In summary, although there are clearly discrepancies in the literature, evidence would seem to suggest that, when compared to CWNS, CWS may have difficulty with components of working memory and inhibition, skills that are foundational for CF. This supposition is further supported by the findings of a recent meta-analysis in which CWS were not only found to score lower than CWNS on measures of nonword repetition and forward span but also rated by parents as having weaker inhibition and attentional focus/persistence skills than the CWNS (Ofoe et al., 2018).

CF: Construct and Measurement

CF is a poorly understood construct that has been referred to in the literature using a variety of terms, such as set shifting or switching, task shifting/switching, attention shifting/switching, attentional flexibility, mental flexibility, mental set shifting/switching, flexible categorization, and representational flexibility (Ionescu, 2012). Some investigators have attempted to bring more clarity to the construct by categorizing it in different ways. For example, Garon et al. (2008) categorize CF based on the type of shift required—that is, whether it is in response to some aspect of the stimuli (attention shifting or set shifting/switching) or requires a change in motor response (task or response shifting/switching)—whereas Jacques and Zelazo (2005) classify CF as deductive or inductive based on how much information is given to children. When children are given all the information they need to perform a switching task, then it is deductive, but when they are not told how to perform the task, it is inductive.

Regardless of how one goes about classifying CF, most CF tasks require children to form an association between a stimulus and response (i.e., mental set), maintain that association in working memory, and then inhibit the currently formed association in favor of a new, conflicting mental set (Bennett & Müller, 2010; Garon et al., 2008). Several tasks have been developed to assess CF in children between 3 and 5 years of age, including two that are relevant to this study: the Flexible Item Selection Task (FIST; Jacques & Zelazo, 2001) and the Double Categorization Task (DCT; Blaye & Jacques, 2009). The FIST measures children's ability to flexibly switch between different perceptual categories, whereas the DCT measures children's ability to switch between different semantic representations. In the FIST, children are shown 12 sets of three cards, each set containing a simple picture (e.g., an orange fish, an orange phone, and a purple fish). One of the cards is the target picture (e.g., the orange fish), whereas the other two cards match the target picture in a single perceptual dimension (e.g., color or shape). Children are asked to point to two cards that are the same in one dimension (e.g., the *orange* fish and *orange* phone) and then point to another two cards that are the same in a second dimension (e.g., the orange *fish* and purple *fish*).

Jacques and Zelazo (2001) reported that children were able to identify a single dimension in which two cards were similar, which represents the abstraction component of the task, with a high degree of accuracy (M = 87%). However, they had much more difficulty selecting a second perceptual dimension in which the two cards matched, which represents the CF component of the task (M = 41%). Bennett and Müller (2010), however, noted that it is possible that the children had difficulty selecting a second match not because they could not flexibly shift between dimensions, but because they had difficulty abstracting a second, less salient dimension, after having identified the more salient one. After all, while the second match requires CF, some degree of abstraction is also needed. In a follow-up study, Jacques et al. (2009) reported that children continued to have difficulty selecting a second match even when the degree of abstraction was controlled by having the experimenter select the first match. This suggests that the difficulty children had in making the second match was due to their inability to switch between dimensions, not their ability to select a less salient dimension.

In the DCT, children are presented with one target picture (e.g., dog) centered on top of a computer screen and a trio of associates centered below it that are thematic (e.g., bone), taxonomic (e.g., bear), and unrelated (e.g., telephone) to the target. Children are asked to identify two matches for the target picture. According to Blaye and Jacques (2009), children's first match measures their ability to select a salient semantic match, whereas their second match measures semantic flexibility because they must switch to a different semantic representation while the most salient match is still present. On average, children were able to correctly select a first match with 88% accuracy, with the 5-year-old children outperforming the 3- and 4-year-old children. This demonstrated, according to the investigators, that preschool children have adequate categorical knowledge to be able to select at least one salient match. Like the FIST, children had more difficulty selecting a second semantic match, as they correctly selected a first and second match with approximately 66% accuracy on average.

In this study, we modified the DCT to examine CF for semantic information in CWS and CWNS. We then juxtaposed semantic flexibility with perceptual flexibility using a task modeled after both the FIST and the DCT. To put this study into context, however, we first review the relatively scant literature on CF in CWS.

CF: Developmental Stuttering

To date, only two performance-based studies have been conducted to examine the CF skills of CWS. In the first study, Eggers and Jansson-Verkasalo (2017) examined attentional flexibility/switching in 16 CWS and 16 CWNS between the ages of 6;4 and 9;10 (years;months) using the Shifting Set Task of the ANT program. This subtest consists three parts. In Part I, children hear a low-frequency single or double tone and then respond by pressing a button once or twice, respectively. This part of the task establishes a baseline prepotent response—that is, it primes children to respond in a certain way. In Part II, children are again presented with a single or double tone, but this time, at a high frequency instead of a low frequency. When they hear the high-frequency tone, children must do the opposite of what they hear: When they hear a single tone, they press the button twice, and when they hear a double tone, they press the button once. This part of the task requires children to inhibit the prepotent response in favor of a conflicting, subdominant response. In Part III, children are presented with both low- and high-frequency single and double tones. When the single or double tone is presented at a low frequency, children must respond as they did in

Part I, but when the tones are presented at a high frequency, they respond as they did in Part II. Thus, this part of the task measures CF, as children must switch from one mode of responding to another, depending on the frequency of the tone. Using this task, Eggers and Jansson-Verkasalo found that, when compared to Part I, both groups of children responded more slowly in Parts II and III, but there were no between-groups differences in reaction time (RT). However, while the CWNS exhibited similar accuracy rates in Parts I and III, the CWS produced significantly more errors in Part III (compared to Part I and the CWNS). Thus, by responding more slowly in Part III, the CWNS were able to maintain their level of accuracy, whereas the CWS did not—they produced 10 times more errors as the CWNS. This finding suggests, according to the authors, that the effectiveness with which CWS can flexibly switch from one mode of responding to another may be less than optimal.

In the second study, Eichorn et al. (2018) examined CF in 16 preschool CWS and 30 preschool CWNS using a modified version of the Dimensional Change Card Sort task (Frye et al., 1995). This task consisted of two phases: the preswitch and postswitch phases. In the preswitch phase, a picture appeared in the center of the screen (e.g., a yellow hexagon), and children were asked to select one of three target pictures (e.g., a purple hexagon, a blue parallelogram, and a yellow curved rectangle) that matched it in color (e.g., the yellow curved rectangle) or shape (e.g., the purple hexagon). The postswitch phase was the same as the preswitch phase, except that children selected the target picture using the opposite rule. That is, if a child selected a target picture based on color in the preswitch phase, they would have to identify the target based on shape in the postswitch phase and vice versa. With this modified version of the Dimensional Change Card Sort task, the authors found that the CWNS were significantly slower and less accurate in the postswitch phase compared to the preswitch phase, as expected, whereas the CWS exhibited comparable accuracy rates in both phases and were even slower in the postswitch phase than the CWNS. The authors also reported that the two groups of children exhibited different patterns of speed–accuracy trade-offs, with the CWS responding more slowly in the postswitch phase perhaps as a means of safeguarding accuracy. These findings were interpreted to suggest that CWS have difficulty with CF (as indicated by the increase in postswitch slowing) and may be more concerned about the prospect of making errors (as indicated by their comparable levels of accuracy in the two phases and their speed–accuracy trade-off pattern).

The CF skills of CWS have also been examined using a questionnaire-based measure. Questionnaire-based measures differ from those of performance-based measures in that they measure aspects of executive function more broadly in complex, goal-directed everyday situations, whereas performance-based measures assess processing efficiency (i.e., how much effort is required) of cognitive abilities under controlled conditions (Samyn et al., 2015; Toplak et al., 2013). Thus, questionnaire-based measures are not considered to be interchangeable with those of performance-based measures; rather, they assess different aspects of CF independently (Samyn et al., 2015; Toplak et al., 2013). With this in mind, Ntourou et al. (2018) administered the Behavior Rating Inventory of Executive Function–Preschool Version (Gioia et al., 2003) to the parents of 150 preschool CWS (n = 75) and CWNS (n = 75). They found that the parents of CWS rated their children higher (more poorly) than the parents of CWNS on the Behavior Rating Inventory of Executive Function–Preschool Version Shift scale. Similar findings were also reported by Eggers et al. (2010). These findings suggest that CWS may have difficulty with CF in real-life, goal-directed activities (e.g., transitioning from one situation, activity, or topic to another), in which their behavior is not being regulated by an external examiner.

Taken together, findings from the two performance-based studies suggest that the CF skills of preschool and school-age CWS may be less effective (as indexed by reduced accuracy/error rates) and/or efficient (as indexed by reduced RT or speed) than their fluent peers even in ideal (i.e., structured) performance situations, whereas findings from Ntourou et al. (2018) suggest that these difficulties may extend to more complex, unstructured everyday situations that are goal directed.

Rationale for and Purpose of the Study

CF and other aspects of executive function and attention are critical for the proper functioning of domain-specific processes, including those associated with speech, language, motor, sensory, and emotional functions (Pisoni et al., 2010, 2008). Importantly, these interactions are bidirectional (Kronenberger & Pisoni, 2018; Müller et al., 2009). This means that weaknesses in executive function could have a negative effect on the development and/or functioning of domain-specific processes and vice versa. Thus, as we have suggested elsewhere (Anderson & Ofoe, 2019), the study of executive function in developmental stuttering is of theoretical importance in that it could explain not only the multifactorial nature of developmental stuttering but also the considerable variability among individuals who stutter.

In this study, we examined the ability of preschool CWS and CWNS to flexibly switch between different semantic and perceptual representations using two tasks: the Double Semantic Categorization Task (DSCT) and the Double Perceptual Categorization Task (DPCT). The DSCT and the DPCT are computer-based tasks, which were modeled after the DCT and/or the FIST. By including a measure of semantic flexibility, we can examine the extent to which both groups of children can use their semantic knowledge to aid them in shifting between aspects of an object's meaning.

This study is motivated by two considerations. First, as previously indicated, several studies have suggested that CWS may have weaknesses in inhibition (e.g., Anderson & Wagovich, 2017; Eggers et al., 2013; Piispala et al., 2017,

2018) and working memory (e.g., Anderson et al., 2019; Hakim & Ratner, 2004; Pelczarski & Yaruss, 2016). Since inhibition and working memory provide a scaffold for the development of CF (Carroll et al., 2016), difficulties in one or both processes could, therefore, result in weaknesses in CF. Second, as also noted above, while only two performance-based studies have examined CF in CWS (Eggers & Jansson-Verkasalo, 2017; Eichorn et al., 2018), both found evidence to suggest that CWS may have difficulty with CF. This study extends this previous work in one important way: Whereas the previous studies examined CF using nonverbal stimuli, this study does so using both nonverbal and verbal stimuli. Thus, findings from this study will provide insight into whether CWS, as a group, also have weakness with CF in the verbal domain. If so, this finding would suggest that reduced CF in CWS is a domain-general difference. We hypothesized that the CWS would perform less accurately and more slowly than the CWNS on both the semantic and perceptual flexibility tasks.

Method

Participants

Eighty-eight CWS (n = 44) and CWNS (n = 44) between the ages of 3;0 and 5;11 participated in the study. Of these 88 children, 40 were between the ages of 3;0 and 3;11, 24 were between the ages of 4;0 and 4;11, and 24 were between the ages of 5;0 and 5;11. All children were native speakers of American English, with no history of neurological, hearing, intellectual, articulation/phonology, language, or any other developmental problem (e.g., autism spectrum disorder) per parent report and examiner observation/testing. The parents of the participants were made aware of the study through advertisements, flyers, and referrals from other parents, speech-language pathologists, and/or preschool/day care centers in Indiana and Missouri.

Group Classification Criteria

The two groups of children were classified as CWS or CWNS based on the frequency of stuttered disfluencies (part-word repetitions, single-syllable word repetitions, sound prolongations, and/or blocks; Yairi & Seery, 2015) produced during a parent–child conversational interaction (described further below).

Each child in the CWS group exhibited three or more stuttered disfluencies, on average, across three 100-word conversational speech samples (M = 6.12, SD = 2.70, range: 3.33–14.00) and received an estimated score of 12 or higher on the Stuttering Severity Instrument–Fourth Edition (Riley, 2009), with 29 CWS having “mild” stuttering and 15 having “moderate” stuttering. The mean parent-reported time since stuttering onset (Yairi & Ambrose, 1992) for the CWS was 14.40 months (SD = 9.60). Two of the 44 CWS were receiving therapy for stuttering during the time in which they participated in this study (one for 9 months and the other for 18 months).

Each child in the CWNS group exhibited less than three stuttered disfluencies, on average, across three 100-word conversational speech samples (M = 0.58, SD = 0.61, range: 0.00–2.00). As expected, a Mann–Whitney test revealed that the CWS (Mdn = 5.15, M rank = 66.50) exhibited significantly more stuttered disfluencies than the CWNS (Mdn = 0.33, M rank = 22.50), U = 0.00, z = −8.10, p < .001.

Group Matching Criteria

CWS and CWNS were matched at each data collection location (see below) by age (± 4 months) and gender (31 boys and 13 girls in each group). The mean age was 50.57 months (SD = 11.10) for children in the CWS group and 51.34 months (SD = 11.03) for children in the CWNS group, a nonsignificant difference, t(86) = −0.33, p = .99. The two groups of children were also equated by family socioeconomic status (SES), which was measured using Hollingshead’s Four-Factor Index of Social Position (Hollingshead, 1975). Using this measure, each child’s family social position score is based on parental marital status, education level, and occupation, with scores ranging from 8 (Class V, lower) to 66 (Class I, upper). The mean family social position score was 51.39 (Class II, upper-middle; SD = 12.34) for the CWS (Mdn = 52.50, M rank = 48.92) and 47.56 (Class II, upper-middle; SD = 12.19) for the CWNS (Mdn = 47.56, M rank = 40.08), a nonsignificant difference, U = 773.50, z = −1.63, p = .10.

Procedure

Testing was completed at two data collection sites: Indiana University and University of Missouri. At each data collection site, children spent 2–3 hr over the course of two separate visits engaged in the following procedures: (a) parent–child conversational interaction, (b) standardized speech-language testing and hearing screening, (c) simple auditory detection task (SADT), and (d) semantic and perceptual CF tasks. Children also completed several other tasks unrelated to the present investigation. The presentation of all tasks was randomized across participants, and the study protocol was approved by an institutional review board at each institution. Written informed consent was obtained from all parents or guardians prior to their child’s participation.

Parent–Child Conversational Interaction

For group classification purposes (see above), children and their parents were seated at a table with a set of toys appropriate for the child’s age. The parents were instructed to converse with their children naturally and avoid questions that could be answered with a one-word response (e.g., yes/no questions). Each conversational interaction lasted for approximately 20 min. A minimum sample of 300 words was collected from each child and analyzed for stuttered disfluencies and, for CWS only, stuttering severity.

Speech-Language Testing and Hearing Screening

To ensure that children were within normal limits in their speech and language skills, four standardized, norm-referenced speech and language tests were administered: (a) Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007), a receptive vocabulary measure; (b) Expressive Vocabulary Test–Second Edition (Williams, 2007), an expressive vocabulary measure; (c) Test of Early Language Development–Third Edition (Hresko et al., 1999), a receptive/expressive language measure; and (d) "Sounds-in-Words" subtest of the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000), a speech sound articulation measure. Children who received a standard score of 85 or higher on each speech and language test participated in the study. An omnibus multivariate analysis of variance revealed no significant difference between the CWS and CWNS groups on all four speech and language tests, $F(4, 83) = 0.62$, $p = .65$, $\eta_p^2 = .03$, with p values from univariate analyses of variance tests for each speech and language test ranging from .23 to .55.

To ensure that children had hearing within normal limits, bilateral pure-tone testing was conducted at 20 dB SPL for 1000, 2000, and 4000 Hz (American Speech-Language-Hearing Association, 1997). All but one child, a child who stutters, passed the hearing screening. The single child who stutters who did not pass the screening did so because she refused to participate. Despite the lack of objective screening data, this child remained in the study because the parents had no concerns about the child's hearing and the child had no difficulty following instructions throughout testing.

SADT

To control for potential differences in basic auditory and motor processing skills, children completed an SADT, whereby they pressed a button whenever they heard a 2000-Hz tone. The SADT was developed using E-Prime 2.0 software by Psychology Software Tools, Inc. (PST). A PST Serial Response Box, which features five button keys, was directly connected to the computer via the serial port. The first button key on the response box was framed in red, with the four remaining button keys covered.

Like the two experimental tasks described below, children were seated in front of the computer with their hands on a mark in front of the PST Serial Response Box. The children were then told that they were going to play a "listening game" in which the objective was to press a button as fast as they could whenever they heard a "beep." Before starting the SADT, children completed two practice phases: vocal simulation and computer practice. During the vocal simulation phase, the examiner vocally simulated a beep on three separate occasions, and after each occasion, children had to respond by pressing the button framed in red. During the computer practice phase, the vocally simulated beep was replaced with a computer-generated 2000-Hz tone, and children had to press the button whenever they heard the tone on three consecutive trials.

Following practice, children completed the test items, which were identical to the computer practice phase, without any feedback on their performance from the examiner. The experiment consisted of 13 trials in which each 2000-Hz tone was presented for 1,000 ms. To reduce predictability, the amount of time in between each trial varied randomly from 1,500 to 3,000 ms. RT was measured in milliseconds from the onset of the 2000-Hz tone to the onset of the children's button presses and recorded on the computer using E-Prime. Children's responses were scored as correct (child correctly pressed the button in response to the tone) or incorrect (child failed to respond to the tone or responded before it was presented).

Semantic and Perceptual CF Tasks

CF was measured using two tasks: DSCT and DPCT. While both tasks had the same format (target picture on top, with three associates below it), they differed in that the initial association and switch is between semantic representations in the DSCT and perceptual representations in the DPCT. Thus, performance on the DSCT is dependent on and facilitated by semantic knowledge, whereas the DPCT is neither dependent upon nor facilitated by semantic knowledge.

DSCT. The DSCT is a computerized adaptation of the DCT, which Blaye and Jacques (2009) used with French children between the ages of 3;2 and 5;11. In this task, children must match a target object with two different semantic associates.

Stimuli. The original stimuli used by Blaye and Jacques (2009) contained several late-acquired stimulus items and some vocabulary that would not be typical for young American English–speaking children (e.g., cot, pressure cooker, parasol). Thus, the stimuli used in this study were selected from among several different sources (Baldwin, 1992; Blaye & Jacques, 2009; Markman & Hutchinson, 1984; Waxman et al., 1997) to ensure that they were acquired early in life, based on the age of acquisition values of Kuperman et al. (2012), and regionally appropriate. Fifty-two stimulus items were selected from these sources and grouped into 13 experimental sets, each containing a target object, thematic associate (i.e., an item that occurs together with the target object as part of a scene, theme, or event), taxonomic associate (i.e., an item that is a superordinate of the target object), and a semantically unrelated associate of the target object. See Appendix A for the full list of experimental stimuli.

Age-appropriate grayscale drawings of 45 (86.5%) of the 52 stimulus items were obtained from the computerized image corpus of Rossion and Pourtois (2004), which are based on the widely used Snodgrass and Vanderwart (1980) drawings. The remaining seven (13.5%) grayscale drawings were acquired from clip art, as they were not in the Rossion and Pourtois image corpus. Normative values for young children on name agreement, picture familiarity, word frequency, and visual complexity were obtained for each stimulus item from the database of Cycowicz et al. (1997),[1] which is also based on the Snodgrass and Vanderwart drawings. An omnibus multivariate analysis of variance revealed no significant difference in age of acquisition, name agreement, familiarity, frequency, and complexity across all four types of stimulus items (target, thematic, taxonomic, and unrelated), $F(5, 40) = 1.27$, $p = .23$, $\eta_p^2 = .13$, with separate univariate analyses of variance p values ranging from .33 to .50 for each psycholinguistic variable.

The target object was framed in a black box and always appeared centered at the top of the computer screen for each set of pictured stimulus items. The three associates (thematic, taxonomic, and unrelated) were centered below the target object. The placement of the three associates (left, middle, right) was counterbalanced across the 13 experimental sets of stimulus items, with the leftmost stimulus item framed in a purple box, the middle in a green box, and the right in an orange box (see Figure 1).

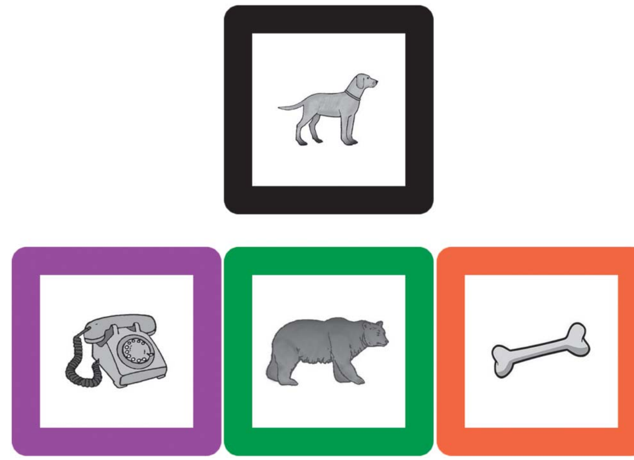
An additional 20 pictured stimulus items were selected from the image corpus of Rossion and Pourtois (2004) and grouped into five target/associate sets. These stimulus items, which were arranged on the computer screen in the same way as the experimental stimulus sets, were used in the two practice phases described below.

Procedure. The DSCT was created using the E-Prime 2.0 software program, and responses were recorded using the PST Serial Response Box. The latency of children's responses (i.e., RT), which was measured in milliseconds from the onset of the three associates to the onset of the child's first and second responses, was recorded directly onto the computer with the E-Prime software. Like the stimulus items depicted on the computer screen, the first button key on the response box was framed in purple, the third button key was framed in green, and the fifth button key was framed in orange, with the two button keys in between covered.

Children were seated in front of the computer and instructed to place their hands on a mark in front of the response box. They were then given these instructions: "We are going to play a matching game. You will see one picture on top of the screen and three pictures below it (examiner points to an example on the screen). Your job is to press the button of the two pictures that go with the top picture as fast as you can. We are going to practice first so you can see how the game is played." Following the instructions, children completed two practice phases: familiarization (demonstration) and training.

Consistent with the procedures of Blaye and Jacques (2009), children were presented with two sample stimulus sets, each consisting of a target object and three (thematic, taxonomic, and unrelated) associates, during the *familiarization phase*. For both stimulus sets, they were first asked to identify one of the three pictures that goes with the target object ("Look at this one [examiner points to the target object]. Can you point to the one, among these three, that goes best with this one? [examiner points to the three associates]"). They were then asked to identify another picture that goes with it ("Ok, now there are only two pictures left. Can you show me the one that goes best with this one [examiner points to the target object] among these two [examiner points to the two remaining associates]?"). If the child selected the unrelated associate, they received explicit (verbal and visual) corrective feedback. This phase was repeated until the child correctly responded to both stimulus sets.

Figure 1. Sample of a set of stimulus items in the Double Semantic Categorization Task.



During the *training phase,* the children were presented with three sample target/associate stimulus sets one at a time and instructed to press the buttons (purple, green, orange) on the response box of the two associates (thematic and taxonomic) that matched the target object. For example, given the target object "dog" shown in Figure 1, correct responses were to press the green button for the taxonomic associate "bear" and the orange button for the thematic associate "bone." If the child responded correctly to the thematic and taxonomic pictured associates, each associate would turn into a smiley face, but if the child responded to the incorrect (unrelated) associate, a large "X" appeared over the unrelated associate. After this initial stimulus set, children completed two more training sets in which they had to identify the thematic and taxonomic associates with visual feedback (smiley face or "X"). The training phase was repeated until the child correctly responded to all three stimulus sets.

Following the training phase, children were told that they were "now…going to play the game for real" and then were presented, in a fixed random order, the 13 experimental target/associate combinations one at a time. For each experimental stimulus set, children had to identify two matches for the target object by pressing the button associated with each semantic associate. No verbal or visual feedback was given to the children about the accuracy of

[1]Four stimulus items were not included in the Cycowicz et al. (1997) study and, thus, did not have values for name agreement, familiarity, frequency, and complexity.

their responses. However, because the task was challenging for some children, particularly the younger ones, children were prompted by the examiner to select a second associate (e.g., the examiner points to the target object and says, "Which other one goes with this one?") or, when they occasionally pointed to the screen rather than the button, to "press the button," if needed. The target object appeared on the computer screen first, followed 1,000 ms later by the three associates. The target/associate stimulus set appeared on the computer screen until the child responded with a second button press, and if the child failed to respond, the examiner advanced the program to the next item. The intertrial interval (the time between each set of experimental stimuli) was 2,500 ms. The experiment lasted approximately 5–10 min.

Scoring. Children's responses to each stimulus set were scored as correct (child correctly selected both matches) or incorrect (child selected the unrelated associate or the same associate for both matches).[2]

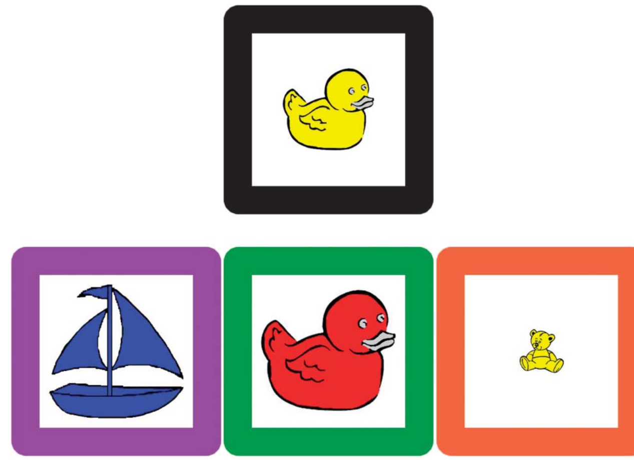
DPCT. The DPCT is similar to the DSCT, except that perceptual stimuli are used instead of semantic stimuli. As previously indicated, the task was modeled after the DSCT and the FIST, which was developed by Jacques and Zelazo (2001) for use with children between 2 and 5 years of age. In this task, children must match a target object with two possible perceptual associates.

Stimuli. Experimental stimuli included 12 target objects paired with 12 sets of stimulus items consisting of one unrelated associate and two possible perceptual associates that matched the target by color, shape, or size.[3] The color associate was expressed by the colors blue, red, and yellow; the shape associate was expressed by a teddy bear, boat, and duck; and the size associate was expressed by small (1 in.), medium (2 in.), and large (3 in.) objects. Each perceptual associate appeared in eight of the 12 sets of experimental stimulus items. See Appendix B for the full list of experimental stimuli.

Like the DSCT, the target object was framed in a black box centered on top of the computer screen. The three potential associates (one unrelated, two associates) were centered below the target object, with the leftmost stimulus item framed in a purple box, the middle stimulus item framed in a green box, and the rightmost stimulus item framed in an orange box (see Figure 2). The placement of the three associates (left, middle, right) was counterbalanced across the 12 sets of stimulus items.

Five different sets of stimulus items, each consisting of a target object and three potential associates (20 stimulus items in total), were also created for the practice phases. Each set of practice items was arranged on the computer

Figure 2. Sample of a set of stimulus items in the Double Perceptual Categorization Task.



screen in the same way as the experimental stimuli described above.

Procedure. The PST Serial Response Box was used to record children's responses to the DPCT, with the first button key framed in purple, the third button key framed in green, and the fifth button key framed in orange. The second and fourth button keys on the response box were covered.

Children were seated in front of the computer with their hands on a mark in front of the response box. The examiner began by telling the children, "We are going to play a matching game. Let's first look at some pictures." Children were then shown a picture containing three teddy bears, three ducks, and three boats in a 9 × 9 grid, as part of the prefamiliarization phase. The first row of the grid contained a small, medium, and large red teddy bear; the second row contained a small, medium, and large blue duck; and the third row contained a small, medium, and large yellow boat. The examiner pointed out the three different colors, shapes, and sizes to the child.

Following the prefamiliarization phase, the examiner said, "Let's try something different. Now you will see one picture on top of the screen and three pictures below it (examiner points to an example on the screen). Your job is to press the button of the two pictures that go with the top picture as fast as you can. We are going to practice first so you can see how the game is played." Following these instructions, children completed two more practice phases: familiarization and training. The instructions and procedures for the familiarization and training phases were conducted in the same manner as described for the DSCT, except that perceptual stimuli were used instead of semantic stimuli, as described in the preceding stimulus section.

Following practice, children were presented with 12 experimental target/associate combinations in a fixed random order and selected two matches for the target object by

[2]Although we noted the accuracy of children's first and second responses, children had to select both relevant matches to get credit for the response, because in this study, we were only interested in the CF component of the task, and to demonstrate CF in the second response, the first response had to be correct.

[3]The DPCT contained 12 experimental sets of stimulus items, while the DSCT contained 13 due to counterbalancing procedures.

pressing the button associated with each perceptual associate without feedback about their performance (like the DSCT, children were given prompts to select an associate or press the button, if needed). For each stimulus set, the interstimulus interval (time between the presentation of the target object and the three associates) was 1,000 ms, and the intertrial interval was 2,500 ms. The stimulus set appeared on the computer screen until the child pressed the second button key. If the child failed to respond with a second button press, the examiner advanced the program to the next item. Children completed the experiment in approximately 5–10 min.

Scoring. The scoring for the DPCT was identical to the DSCT described above.

Data Analysis

All statistical analyses were conducted using IBM SPSS Statistics for Windows, Version 25 (IBM Corp.) and SAS software, Version 9.4 of the SAS System for Windows (SAS Institute, Inc.). Prior to analysis, all dependent variables were assessed for normality and homogeneity of variance to determine if they met the assumptions of parametric statistics.

The main dependent variables for the auditory and motor control task, the SADT, were response accuracy and RT for correct responses. Response accuracy was not normally distributed. Thus, between-groups differences (CWS, CWNS) in response accuracy on the SADT were examined using the distribution-free Mann–Whitney *U* test. Between-groups differences in RT were analyzed using a univariate analysis of covariance (ANCOVA), with chronological age added as a covariate, as the RT data were normally distributed, and variances were equal across groups. Chronological age was added as a covariate because it was highly correlated with RT for both groups of children. Prior to analysis, RT values that were more than 2 *SD*s above and below the mean for each group of children were considered outliers and, thus, removed from the data corpus (Ratcliff, 1993).

The main dependent variables for the DSCT and DPCT were response accuracy (correct/incorrect for each set of stimulus items), an index of flexibility, and RT for correct second responses, an index of efficiency. Since response accuracy at the individual trial level is a dichotomous outcome variable (and, hence, bimodal in its distribution), generalized estimating equations (GEEs) with binary logistic models were used to examine differences in accuracy on the DSCT and DPCT, with the participants' identification number as the subject variable.[4] Independent variables included participant (CWS, CWNS) and age (3-, 4-, and 5-year-olds) group. Spearman's rank correlation coefficients indicated that receptive vocabulary, represented by PPVT-4 standard scores, was moderately correlated with overall response accuracy (i.e., total number correct) on the DSCT. Thus, this variable was added as a covariate to all GEE analyses to control for its influence on response accuracy in the DSCT. Simple contrasts and pairwise comparisons, with Bonferroni correction, were used to further explore differences between means (the reported *p* values reflect the Bonferroni adjustment).

RT was measured from the onset of the three associates to the onset of the child's first match and then second match in both tasks. However, only RT for the second correct match was statistically analyzed, as these values index efficiency of CF (see introduction for additional information). For each set of stimulus items, the recorded RT values for the second response were cumulative (i.e., they included the RT values for the first response). Thus, to measure the amount of time it took the child to select a second match, the RT values for the second match were subtracted from the RT values for the first match. Like the SADT, RT values that were 2 *SD*s above or below the mean for each group of children were deemed outliers and, thus, removed from the final data corpus, as per the guidelines established by Ratcliff (1993). The removal of outliers is a standard practice in RT studies, as extreme values can increase mean and error variances, decrease power, and/or increase the likelihood of Type I or II errors (Cousineau & Chartier, 2010; Osborne & Overbay, 2004). Consistent with previous RT studies (e.g., Anderson & Wagovich, 2017; Catts et al., 2002; Karalunas et al., 2013; Nardini et al., 2016),[5] each child was also required to have usable RT values for approximately half of the experimental trials to be included in the analyses, as mean RTs cannot be meaningfully calculated with fewer than six trials. For example, if a child responded correctly to six out of 12 trials on the DPCT and the RT values for two of these correct responses had been outliers, then this child was excluded from the RT analysis as less than half of the data (four out of 12 or 33%) were usable. The mean number of usable trials in the DSCT (n = 13) was 10.61 (SD = 1.91) for the CWS and 10.35 (SD = 2.04) for the CWNS. Likewise, the mean number of usable trials in the DPCT (n = 12) was 7.71 (SD = 1.55) for the CWS and 8.19 (SD = 1.97) for the CWNS. To maintain equal sample sizes and pair matching across groups, if a child was excluded from the analysis due to an insufficient number of usable RT values, then their matched pair was also removed from the final data corpus.

Mean RT data for both the DSCT and DPCT met all parametric assumptions and were, thus, analyzed using separate ANCOVAs with participant group (CWS, CWNS) and age group (younger, older) as between-subjects variables. Unlike accuracy where age group classification was based on 3-, 4-, and 5-year-olds, age group for the RT

[4]While response accuracy could have been analyzed across trials (e.g., overall response accuracy or proportion correct) using nonparametric statistics, the GEE analysis was chosen because the model can accommodate covariates.

[5]These studies required participants to have 25%–80% usable RT data to be included in the RT analyses, with most having a criterion of 50% (three of the four had a criterion of 50%; one had a criterion of 80%; one of the three studies with a 50% criterion also had a 25% criterion for some tasks).

analyses was based on a median split (*Mdn* = 56.50 months for both tasks), and children were classified into younger (≤ 56 months) and older (≥ 57 months) age groups. Age group was defined in this way for the RT analyses because there were an insufficient number of children in the 3-, 4-, and 5-year-old age groups.[6] SADT RT, which was strongly correlated with RT for the CWNS in both tasks but only weakly correlated with RT for the CWS, was added as a covariate to control for basic auditory and motor processing skills.

Finally, to compare children's response accuracy across experimental tasks, responses to the DSCT and DPCT were converted to proportions. Thus, if a child accurately selected both matches in, for example, eight out of 13 of the DSCT trials and six out of the 12 DPCT trials, then their proportion correct would be 62% and 50%, respectively. Proportion correct was not normally distributed, and thus, the differences in proportion correct between the two tasks were analyzed using Wilcoxon signed-ranks tests for both the CWS and CWNS groups. The absolute difference in proportion correct (i.e., proportion correct in the DSCT − proportion correct in the DPCT = difference) between the two groups of children was examined using an ANCOVA with chronological age serving as a covariate (chronological age was moderately correlated with the difference score for the CWS), as the distribution and variance of the difference scores were normally distributed and equal across groups. A repeated-measures ANCOVA was used to examine within-group differences in RT, with chronological age and SADT RT scores serving as covariates. Given the parametric nature of the data, the between-groups difference in RT difference scores (i.e., RT in the DSCT − RT in the DPCT = difference) was examined using an analysis of variance (neither chronological age nor SADT RT scores were added as covariates to the model, as they were not associated with the RT difference scores). To be included in these analyses, children had to have enough usable RT values for both experimental tasks.

As a measure of the strength of the association, Pearson's correlation coefficient *r* or the phi coefficient ϕ was used as the effect size measure for all nonparametric tests, with an *r* or a ϕ of .50 representing a "large" effect, .30 representing a "medium" effect, and .10 representing a "small" effect (Cohen, 1988, 1992). The effect size indicator partial eta squared (η_p^2) is reported for all parametric comparisons, with .14 representing a "large" effect, .06 representing a "medium" effect, and .01 representing a "small" effect (Cohen, 1988).

Results

The purpose of this study was to examine CF for semantic and perceptual information in preschool CWS versus CWNS. This was accomplished by examining the accuracy and speed of children's responses to two computerized tasks: the DSCT and the DPCT.

SADT

Although not a main focus of the study, children's performance on the SADT was analyzed to determine if there were any differences between the CWS and CWNS groups in their ability to press a button in response to auditory stimuli. Findings from the Mann–Whitney *U* test revealed no significant difference between the CWS (*Mdn* = 13.00, *M* rank = 43.52) and CWNS (*Mdn* = 13.00, *M* rank = 45.48) in response accuracy on the SADT, $U = 1011.00$, $z = 0.60$, $p = .55$, $r = .06$. An ANCOVA, with chronological age as the covariate, further revealed no significant difference between the CWS (adjusted *M* = 1,338.71, *SE* = 47.61) and CWNS (*M* = 1,318.43, *SE* = 47.61) in RT on the SADT, $F(1, 85) = 0.09$, $p = .76$, $\eta_p^2 = .001$. These findings indicate that the two groups of children were comparable in their basic auditory and motor processing abilities.

DSCT

Accuracy

Response accuracy was analyzed using a GEE analysis for binary outcomes (correct or incorrect), with participant group (CWS, CWNS) and age group (3-, 4-, and 5-year-olds) as between-subjects factors (see Figure 3). PPVT-4 scores were added as a covariate to the analysis. Results revealed no significant difference in the probability of producing an accurate response between the CWS (adjusted *M* = .78, *SE* = .03) and CWNS (adjusted *M* = .78, *SE* = .02), $\chi^2(1, N = 88) = .13$, $p = .72$. The main effect of age group was statistically significant, $\chi^2(1, N = 88) = 54.05$, $p < .001$, as was the covariate, PPVT-4, $\chi^2(1, N = 88) = 6.25$, $p = .01$. However, the interaction between participant group and age group was not significant, $\chi^2(1, N = 88) = 1.75$, $p = .42$.

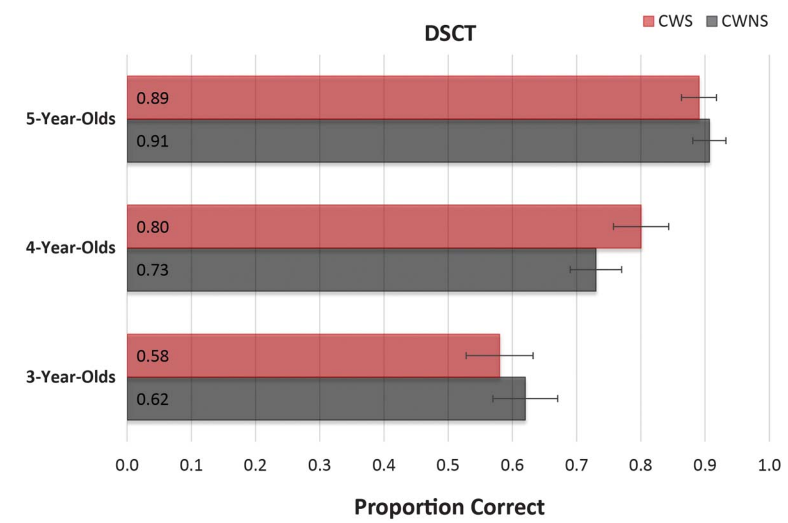
Simple contrasts across age groups revealed a significant difference in response accuracy between 3- and 4-year-old children ($p < .001$) and 4- and 5-year-old children ($p = .001$; 3-year-olds: adjusted *M* = .60, *SE* = .04; 4-year-olds: adjusted *M* = .77, *SE* = .03; 5-year-olds: adjusted *M* = .90, *SE* = .02). However, pairwise comparisons between the CWS and CWNS groups indicated no significant difference at each age group (p = .59, .26, and .72 for the 3-, 4-, and 5-year-olds, respectively). These results suggest that the CWS and CWNS groups were comparable in their semantic flexibility skills and that these skills are stronger for older children and those with higher receptive vocabulary scores.

RT

Across both groups of children, 4.35% of the RT data were considered outliers (3.09% CWS, 1.26% CWNS; all were > 2 *SD*s above the mean) and, thus, removed from the final data corpus. Following the removal of outliers, children who had fewer than six (< 46%) usable RT responses were excluded from the analyses, along with their matched

[6]For the DSCT, there were eight 3-year-olds, eleven 4-year-olds, and twelve 5-year-olds in each group of children. For the DPCT, there were five 3-year-olds, eight 4-year-olds, and eight 5-year-olds in each group.

Figure 3. Adjusted mean (and standard error of the mean) proportion correct for 3-, 4-, and 5-year-old children who stutter (CWS; n = 44) and children who do not stutter (CWNS; n = 44) in the Double Semantic Categorization Task (DSCT).



pairs. This resulted in the exclusion of 10 CWS and three CWNS (all were between 3;0 and 4;0), leaving a total of 31 children (23 boys and eight girls) in each participant group (N = 62). A chi-square goodness-of-fit test indicated that significantly more CWS were excluded from the analysis than CWNS, $\chi^2(1, N = 13) = 3.77$, $p = .05$.

As with the larger group of children, there was no significant difference between 31 CWS and 31 CWNS in chronological age (CWS: M = 54.84, SD = 10.36; CWNS: M = 55.61, SD = 10.15), $t(60) = -0.30$, $p = .77$; SES (CWS: Mdn = 53.00; CWNS: Mdn = 51.00), U = 391.00, $z = -1.26$, $p = .21$; and their performance on the four speech and language tests, $F(4, 57) = 0.77$, $p = .55$, $\eta_p^2 = .05$. There were also no significant between-groups differences in chronological age for the younger (CWS: M = 46.50, SD = 6.12; CWNS: M = 46.67, SD = 5.90), $t(29) = -0.08$, $p = .94$, and older (CWS: M = 63.73, SD = 5.08; CWNS: M = 64.00, SD = 4.27), $t(29) = -0.16$, $p = .88$, children.

Between-groups differences in mean RT were analyzed using an ANCOVA, with SADT RT serving as a covariate (see Figure 4). Findings revealed a significant main effect of group, $F(1, 57) = 5.90$, $p = .02$, $\eta_p^2 = .09$, with the CWS (adjusted M = 5,728.54 ms, n = 31) performing more slowly than the CWNS (adjusted M = 4,393.16 ms, n = 31). The covariate, SADT RT, was also significant, $F(1, 57) = 10.42$, $p = .002$, $\eta_p^2 = .16$. However, neither the main effect of age group (younger: adjusted M = 5,485.77 ms, n = 31; older: adjusted M = 4,635.94 ms, n = 31), $F(1, 57) = 1.91$, $p = .17$, $\eta_p^2 = .03$, nor the Group × Age Group interaction, $F(1, 57) = 0.09$, $p = .77$, $\eta_p^2 = .001$, were statistically significant. These findings indicate the CWS were less efficient than their typically fluent peers in their ability to switch from one semantic representation to another.

Given the significant main effect of group, the relationship between RT and overall response accuracy (i.e., total number correct) was evaluated for each group of children to determine the potential for speed–accuracy trade-offs (i.e., the tendency for slower responses to be associated with fewer errors or increased accuracy and vice versa) using Spearman's rank partial correlation coefficients, with chronological age serving as the covariate. There was no significant correlation between RT and response accuracy for the CWS ($r = -.02$, $p = .89$), suggesting that these children were not sacrificing speed for accuracy. There was, however, a significant correlation between RT and response accuracy for the CWNS ($r = -.41$, $p = .02$), but the direction of the correlation was negative. Thus, the speed and accuracy with which the CWNS responded to the task were consistent: Those who performed more slowly tended to be less accurate and vice versa (no speed–accuracy trade-off).

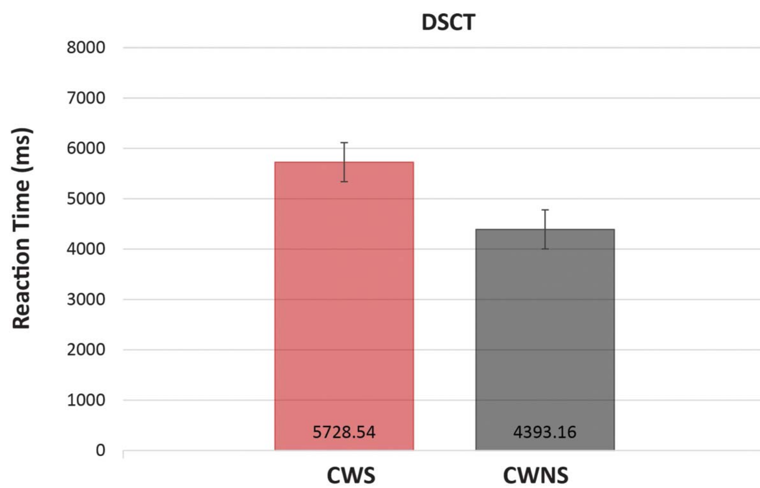
While the CWS, as a group, performed more slowly on the DSCT than the CWNS, not all CWS did. As is typical of many studies, not just in stuttering but in other disorder groups as well, there is often much variability across individual participants. The RT data for each participant on the DSCT are shown at the left side of Figure 5. As can be seen, only a few CWNS performed slower than 1 SD above the CWNS mean, whereas twice as many CWS did. Furthermore, there were more CWNS who scored below the CWNS mean, with 6 times as many CWNS than CWS performing faster than 1 SD below the mean. Thus, while there is clearly an overlap in the performance of individual CWS and CWNS, there were more individual CWS who performed more slowly and, conversely, more individual CWNS who performed faster.

DPCT

Accuracy

Results of the GEE analysis for binary outcomes (correct or incorrect), with participant group (CWS, CWNS)

Figure 4. Adjusted mean (and standard error of the mean) reaction time (ms) for children who stutter (CWS; n = 31) and children who do not stutter (CWNS; n = 31) in the Double Semantic Categorization Task (DSCT).



and age group (3-, 4-, and 5-year-olds) as between-subjects factors (see Figure 6), revealed that, while the CWS were descriptively less likely than the CWNS to produce an accurate response (CWS: M = .53, SE = .03; CWNS: M = .59, SE = .03), this difference was not statistically significant, $\chi^2(1, N = 88) = 3.14$, $p = .07$. The main effect of age group was statistically significant, $\chi^2(1, N = 88) = 16.97$, $p < .001$, but not the Participant Group × Age Group interaction effect, $\chi^2(1, N = 88) = 3.54$, $p = .17$. Given that the main effect of group approached significance, pairwise comparisons were conducted between participant groups for each age group. Findings revealed no significant differences between the 3-year-old CWS and CWNS (p = .37) and 4-year-old CWS and CWNS (p = .52). While there was a tendency for the 5-year-old CWS to perform more poorly than the 5-year-old CWNS, this difference was not statistically significant (p = .08).

Simple contrasts across age groups further indicated that the probability of producing an accurate response was similar between the 3- and 4-year-old children (p = .47), but

Figure 5. Individual reaction time (ms) data for children who stutter (CWS; red diamonds) and children who do not stutter (CWNS; black circles) in the Double Semantic Categorization Task (DSCT; n = 62) and the Double Perceptual Categorization Task (DPCT; n = 42). The upper and lower brackets indicate 1 SD above and below the adjusted mean of the CWNS in each experimental task.

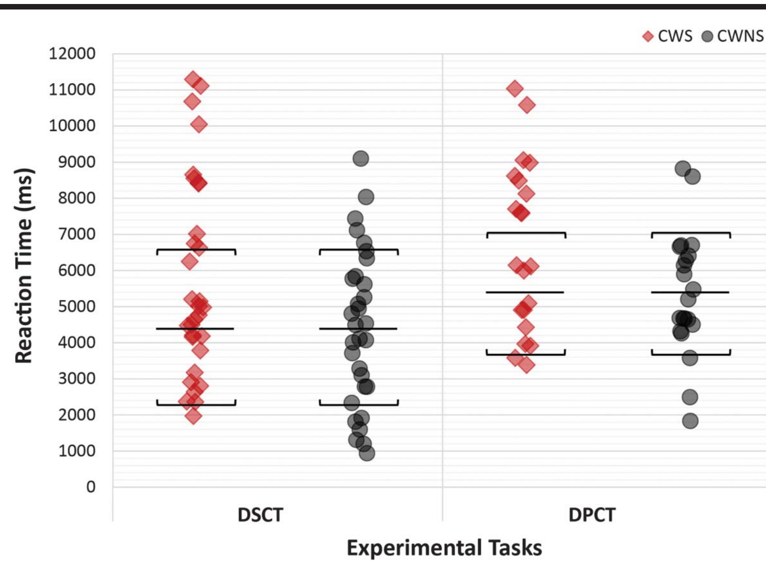
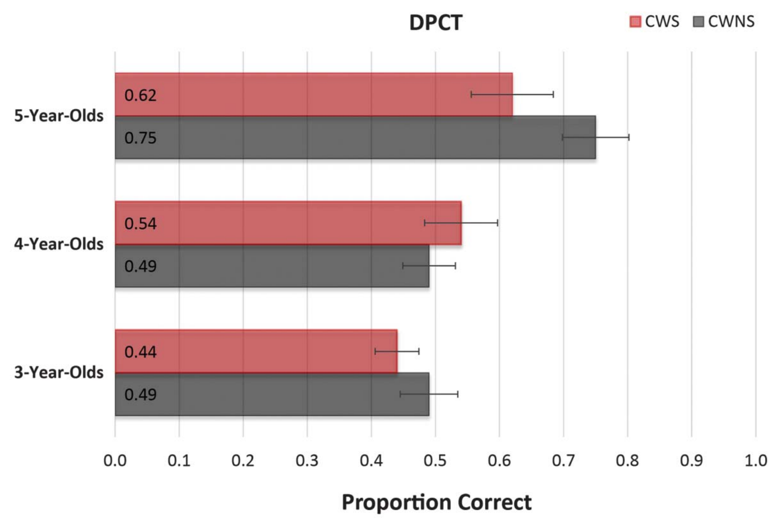


Figure 6. Adjusted mean (and standard error of the mean) proportion correct for 3-, 4-, and 5-year-old children who stutter (CWS; *n* = 44) and children who do not stutter (CWNS; *n* = 44) in the Double Perceptual Categorization Task (DPCT).



the 4-year-old children were significantly less accurate than the 5-year-old children (p = .002; 3-year-olds: M = .46, SE = .03; 4-year-olds: M = .51, SE = .04; 5-year-olds: M = .69, SE = .05). Taken together, these results indicate that the perceptual flexibility skills of CWS and CWNS were similar and that these skills are stronger among older children, especially between children 4 and 5 years of age.

RT

For both groups of children combined, 5.51% of the data (3.15% CWS, 2.36% CWNS; all were > 2 *SD*s above the mean) were removed from the final data corpus as outliers. Following the removal of outliers, any child who had less than six (< 50%) usable RT values for correct responses was excluded from the analysis, along with the child's matched pair. This resulted in the exclusion of 14 CWS (ten 3-year-olds, one 4-year-old, and three 5-year-olds) and nine CWNS (seven 3-year-olds and two 4-year-olds), and thus, the final sample size was 21 children (18 boys and three girls) in each group (N = 42). A chi-square goodness-of-fit test indicated no significant difference in the number of CWS and CWNS who were excluded from the analysis, $\chi^2(1, N = 23) = 1.09$, p = .30.

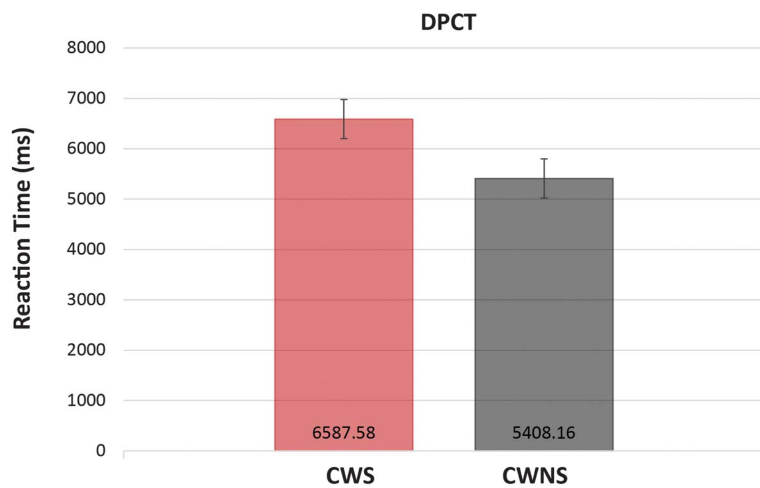
The 21 CWS and 21 CWNS did not significantly differ in chronological age (CWS: M = 55.38, SD = 10.49; CWNS: M = 56.33, SD = 9.82), $t(40) = -0.30$, p = .76; SES (CWS: *Mdn* = 57.00; CWNS: *Mdn* = 52.00), U = 234.50, $z = -1.94$, p = .11; and their performance on the four speech and language tests, $F(4, 37) = 1.14$, p = .35, η_p^2 = .11. There were also no significant between-groups differences in chronological age for the younger (CWS: M = 47.18, SD = 6.52; CWNS: M = 47.60, SD = 5.84), $t(19) = -0.15$, p = .88, and older (CWS: M = 64.40, SD = 4.97; CWNS: M = 64.27, SD = 4.08), $t(19) = -0.06$, p = .95, children.

Findings from the ANCOVA revealed a significant main effect of group, $F(1, 37) = 4.59$, p = .04, η_p^2 = .11, with the CWS (adjusted M = 6,587.59 ms, n = 21) responding more slowly than the CWNS (adjusted M = 5,408.16 ms, n = 21; see Figure 7). The main effect of age group was also significant; the younger children (adjusted M = 6,911.30 ms, n = 21) were slower than the older children (adjusted M = 5,084.44 ms, n = 21), $F(1, 37) = 9.77$, p = .003, η_p^2 = .21. However, there was no significant Group × Age Group interaction, $F(1, 37) = 1.76$, p = .19, η_p^2 = .045, or main effect of the covariate, SADT RT, $F(1, 37) = 0.68$, p = .42, η_p^2 = .018. These findings suggest that the CWS were less efficient than the CWNS in flexibly switching from one perceptual representation to another and the younger children (both CWS and CWNS) were slower than the older children.

Given that CWS performed more slowly on the DPCT than the CWNS, the relationship between RT and overall response accuracy was examined for potential speed–accuracy trade-offs using Spearman's rank partial correlation coefficients, with chronological age serving as the covariate. Neither the CWS (r = .22, p = .35) nor CWNS ($r = -.08$, p = .73) exhibited a significant correlation between RT and response accuracy, providing no evidence of speed–accuracy trade-offs.

The individual RT data for the CWS and CWNS groups on the DPCT are shown at the right side of Figure 5. This figure clearly shows some overlap in the individual participant data. However, there were 5 times as many CWS who performed slower than 1 *SD* above the CWNS mean than the CWNS. In fact, most (81%) of the CWNS participants scored within 1 *SD* of the CWNS mean compared to approximately half (52%) of the CWS.

Figure 7. Adjusted mean (and standard error of the mean) reaction time (ms) for children who stutter (CWS; n = 21) and children who do not stutter (CWNS; n = 21) in the Double Perceptual Categorization Task (DPCT).



DSCT Versus DPCT

Accuracy

Within-group differences in proportion correct between the DSCT and DPCT were analyzed for each group of children using the Wilcoxon signed-ranks test. Results revealed a significant difference in proportion correct between the DSCT and DPCT for both the CWS, T = 133.00, p < .001, r = −.45, and CWNS, T = 135.50, p < .001, r = −.45, with both groups performing less accurately on the DPCT (CWS: Mdn = 50.00; CWNS: Mdn = 58.33) than the DSCT (CWS: Mdn = 76.92; CWNS: Mdn = 73.08). An ANCOVA, with chronological age added as a covariate, further revealed no significant difference between the CWS (adjusted M = 19.54, SE = 3.53) and CWNS (adjusted M = 17.36, SD = 3.53) in difference scores, $F(1, 15) = 0.19$, $p = .66$, $\eta_p^2 = .002$. These results indicate that both groups of children were less accurate on the DPCT than the DSCT, and this difference was similar across groups.

RT

Although both groups of children exhibited slower mean RT values on the DPCT on a descriptive basis, findings from the repeated-measures ANCOVA revealed no significant difference in RT between the DSCT and DPCT for the 36 CWS (adjusted M = 4,647.09 and 6,275.34 ms, SE = 380.49 and 483.42, respectively), $F(1, 15) = 0.00$, $p = .98$, $\eta_p^2 = .00$, and CWNS (adjusted M = 3,960.31 and 5,227.22 ms, SE = 409.60 and 367.55, respectively), $F(1, 15) = 0.17$, $p = .69$, $\eta_p^2 = .01$, who had usable RT values for both experimental tasks. The main effect of the covariate, chronological age, was significant for CWS, $F(1, 15) = 10.71$, $p = .005$, $\eta_p^2 = .42$, but not CWNS, $F(1, 15) = 2.19$, $p = .16$, $\eta_p^2 = .13$. Conversely, the main effect of the SADT RT covariate was significant for CWNS, $F(1, 15) = 5.50$, $p = .03$, $\eta_p^2 = .27$, but not CWS, $F(1, 15) = 0.77$, $p = .40$, $\eta_p^2 = .42$. There were no significant interaction effects between RT and chronological age and SADT RT for either group of children, with p values ranging from .57 to .98. The between-groups difference in the RT difference scores (i.e., RT in the DSCT − RT in the DPCT = difference) for the CWS (M = 1,628.25 ms, SE = 497.76) and CWNS (M = 1,266.91 ms, SE = 431.50) also were not significant, $F(1, 32) = 0.30$, $p = .59$, $\eta_p^2 = .009$.

Discussion

The primary purpose of this study was to assess CF in CWS and CWNS by examining the accuracy and speed of children's responses on two computerized tasks: the DSCT, which measures semantic flexibility, and the DPCT, which measures perceptual flexibility. By including two measures of CF—semantic and perceptual—one can determine if preschool CWS have difficulty with CF and whether shifting in thinking about semantic information versus perceptual information presents a greater or lesser challenge in accuracy and efficiency of response. Two main findings shed light on this topic. First and most importantly, while the accuracy with which the CWS and CWNS groups were able to shift between representations (both semantic and perceptual) was similar, the CWS did so much more slowly, as a group, than the CWNS. Second, while both tasks revealed a clear developmental trajectory, children were less accurate when switching between perceptual representations than semantic ones. Taken together, these results provide further insight into the role of CF in developmental stuttering. We now consider these findings in further detail below.

Do CWS Have Weaknesses in CF?

Even though there was good reason to hypothesize that the children in the CWNS group would outperform the children in the CWS group on the semantic and perceptual flexibility tasks, the CWS in this study, as a group, differed from their peers only in the speed with which they responded: The two groups of children were otherwise comparable in response accuracy. Furthermore, the slower performance of the CWS, compared to the CWNS, on the two tasks could not be attributed to speed–accuracy trade-off effects—that is, the CWS did not respond more slowly to "increase" their accuracy. These findings are commensurate with those of Eichorn et al. (2018), who reported that, when compared to CWNS, preschool CWS were able to switch from one rule to another with the same degree of accuracy, but they did so significantly more slowly.

While the findings from this study are similar to those of Eichorn et al. (2018), they differ from those of Eggers and Jansson-Verkasalo (2017). These latter investigators did not find a significant difference in RT between school-age CWS and CWNS during Part III of the Shifting Attentional Set subtest of the ANT, but they did find a difference in accuracy, with CWS producing significantly more errors than CWNS. While there are some obvious methodological differences between this study and that of Eggers and Jansson-Verkasalo (e.g., preschool- vs. school-age children, visual vs. auditory stimuli, inductive vs. deductive tasks) that could potentially account for the difference in findings, the precise reason for these differences is not altogether clear.

When encountering the difficult task of switching from one semantic or perceptual representation to another, children likely respond to the task difficulty in different ways. They also may process the importance of being correct differently. Therefore, if the tasks were more challenging for some CWS, irrespective of their accuracy, they may have slowed their response rate. Note that slowing response rate irrespective of accuracy, for the sake of coping with the difficulty of a task, is different from the idea of the speed–accuracy trade-off, in which a child reduces or increases rate in response to perceived accuracy. We did not observe any evidence of speed–accuracy trade-offs for either group, but this does not rule out the possibility that the groups differed in the way they responded to the difficulty of the tasks. We might infer, based on children's excitement and attentiveness in completing the tasks, that both groups of children were equally motivated to respond accurately, but if some of the CWS are experiencing more processing demands, then they may cope with that challenge by slowing down.

Eichorn et al. (2018) proffered a similar explanation for their findings: They suggested that when switching from one rule to another, CWS may have been able to maintain their perceived accuracy by responding more slowly. They speculated that the CWS adopted this strategy of slowing down because they were concerned about the prospect of making errors, perhaps as a result of perfectionistic tendencies. While this is a possibility, given that there is some evidence to suggest that adults who stutter may be more perfectionistic (Amster, 1995) and concerned about the prospect of making mistakes (Brocklehurst et al., 2015) than adults who do not stutter, there is no evidence to our knowledge to suggest that the same is true of young CWS. Therefore, especially given the fact that our study did not suggest the presence of speed–accuracy trade-offs, we offer the more general explanation that perhaps the CWS, experiencing a greater processing challenge than the CWNS, responded to the challenge by reducing rate of response.

If the CWS, as a group, were slower than the CWNS because the tasks were truly more challenging for them despite similar performance in accuracy, there are several possible explanations from the cognitive development literature for why switching from one representation to another might be harder for this group. First, it is well documented that, unlike 4- and 5-year-olds, 3-year-old children tend to have marked difficulty switching from one dimension or rule to another (e.g., Blaye & Jacques, 2009; Mennetrey & Angeard, 2018; Perner & Lang, 2002). For example, young children typically have no difficulty sorting a set of cards based on an initial dimension (e.g., color), but when they must then sort the same cards according to another dimension (e.g., shape), they tend to perseverate (i.e., they continue to sort the cards based on the initial dimension). A variety of hypotheses have been put forth to explain why young children (i.e., 3-year-olds) and older preschool children with lower CF skills have difficulty with these tasks, some of which can be applied to this study.

One possibility, which is akin to the redescription account developed by Perner and Lang (2002), is that children might not perform well on CF tasks if they have difficulty thinking about the target pictures in more than one way (Jacques & Zelazo, 2001). Thus, as pertains to this study, some CWS may have had difficulty with the abstraction component of the task—identifying a second salient dimension on which two objects could be matched. Children's knowledge of thematic, taxonomic, and perceptual relations is associated with language development (Blaye & Bonthoux, 2001; Pitchford & Mullen, 2001), as is CF in general (Deák, 2003). Given that some studies have revealed that CWS, as a group, have subtly weaker language skills than CWNS (Ntourou et al., 2011), it may be that these weaknesses result in difficulties with abstraction, accounting for increased processing demands of the CWS.

A second possibility, taken from the attentional inertia account in the cognitive development literature (Diamond et al., 2005; Kirkham et al., 2003), is that some CWS may have had difficulty disengaging their attentional focus from the first match to concentrate instead on the second match. In other words, CWS may have difficulty inhibiting the initial, most salient match. Alternatively, the activation-deficit account (Chevalier & Blaye, 2008; Müller et al., 2006) suggests that second matches that have previously been inhibited while attending to the first match may be difficult to reactivate. Therefore, it is possible that some CWS may have difficulty in reactivating the previously inhibited second match. Taken together, applying these two accounts to explain why the tasks may have been more challenging for some CWS, it could be because the CWS had difficulty shifting away from the representation they had initially attended to (attentional inertia) and/or shifting to a representation that they had initially ignored (activation deficit). Both of these possibilities warrant consideration, given that some studies (e.g., Anderson & Wagovich, 2017; Piispala et al., 2017, 2018) have reported inhibition difficulties among CWS.

It is important to keep in mind that, like most studies in the developmental stuttering literature, this is a group study, and as such, we would fully expect there to be some overlap between the two groups of children in their individual performance, even though the CWS, as a group, performed more slowly than the group of CWNS on the tasks. Indeed, the individual participant data shown in Figure 5 bear out this expectation. Not all CWS performed slowly relative to the average performance on these tasks, just as not all CWNS performed more quickly. Individual differences in the performance of CWS are not surprising; in fact, individual variability appears to be a hallmark of developmental stuttering. One consequence of this is inconsistency in findings across research groups (cf. MacPherson & Smith, 2013). What the present findings do suggest, however, is that there may be a subgroup of CWS who have difficulty with CF, and future studies should endeavor to explore this possibility further.

Semantic Versus Perceptual Flexibility

Both groups of children were significantly less accurate on the DPCT than the DSCT, indicating that they had more difficulty switching from one perceptual representation to another than switching between different semantic representations. Nevertheless, even though the DPCT was more demanding than the DSCT, flexibility scores (proportion of items for which both correct responses were selected) for the DPCT were still well above chance (higher than .33),[7] even for the 3-year-olds. Children also tended to respond more slowly on the DPCT compared to the DSCT, but these differences were not statistically significant. We caution against overinterpreting the lack of significance for this particular analysis, however, as the sample size was considerably smaller, since children had to have enough usable RT values in both tasks to be included in the analysis, and the standard error was higher, making it more difficult to achieve statistical significance.

There are two possible explanations for why switching between perceptual representations may have been more challenging than switching between semantic ones. The first has to do with a difference in item construction. While both tasks require categorization, in the DSCT, there were two main categorical differences (thematic or taxonomic), and none of the stimuli were used more than once in constructing the items—that is, there were 12 items, each consisting of four distinct picture stimuli (the target and three possible responses), and within each item, one correct response was thematically related to the target and the other was taxonomically related (with the third being an unrelated foil). In contrast, on the DPCT, there were three categorical differences (color, shape, and size) reflected across the entire set of items, with only two presented within an item (counterbalanced), along with an unrelated foil. In addition, each stimulus picture was used more than once across items on the DPCT. Thus, one possible reason why children were less accurate on the DPCT than the DSCT is that the inhibition demands were greater in the DPCT. Both tasks have inhibition demands; when children are presented with a stimulus set, children must inhibit the other associates when selecting the first and second matches, respectively. However, because the DPCT used picture stimuli more than once, it had an additional demand: Children had to inhibit previously relevant information from the preceding stimulus sets. That is, they had to resist proactive interference. For example, as shown in Appendix B, the medium yellow boat appears in Items 2 and 4. The relevant dimension of the medium yellow boat in the second item is its shape, whereas the relevant dimension in the fourth item is its size. Thus, when considering the medium yellow boat in the fourth item, children must suppress the previously relevant shape information, which is no longer relevant, and focus instead on its size.

Related to this explanation is the possibility that children may have found it easier to shift between two dimensions that were the same across items (i.e., the two dimensions of the DSCT) than between two out of three possible dimensions across items (i.e., the three dimensions of the DPCT). That is, they may have benefited from the fact that the items in the DSCT were more predictable in that they could expect one response always to be a thematic relationship and the other to be a taxonomic relationship.

Another potential explanation for why children had more difficulty with the DPCT than the DSCT, again, has to do with the nature of the stimuli. In particular, unlike the thematic and taxonomic relations depicted in the DSCT, the perceptual relations (color, shape, and size) among objects in the DPCT represent abstract property attributes, which must be conceptualized independently of the context in which the object is perceived (see Kowalski & Zimiles, 2006; Pitchford & Mullen, 2001). For example, the color yellow is an abstract property that can be equally applied to a wide variety of objects, such as a rubber duck, school bus, and lemon. Thus, when children see an object with different attributes, they must analyze and break the visual image down into its component parts, with color, shape, or size being one of those parts (Kowalski & Zimiles, 2006). Perceptual attributes are also unique in that they provide little information about how objects function (e.g., the yellowness of a lemon doesn't tell you what a lemon is or what you do with it), and consequentially, they are more difficult to acquire (Pitchford & Mullen, 2001). Thus, since perceptual attributes represent abstract properties that lack functional significance, children likely expend more cognitive effort when processing the relationships among these stimuli relative to semantic ones, making the DPCT more challenging.

---

[7]The probability of being correct based on chance alone is .67 ($2 \div 3 = .67$) for the first match and .50 ($1 \div 2 = .50$) for the second match. Thus, the proportion of correct responses expected by chance alone for both matches combined is .33 ($\frac{2}{3} \times \frac{1}{2} = .33$).

While the decrease in accuracy on the DPCT relative to the DSCT was similar for both groups of children, the CWS were significantly slower, as a group, than the CWNS on both tasks. This finding is noteworthy because, as will be recalled, one goal of this study was to determine if the weaknesses in CF for perceptual information observed in previous studies (Eggers & Jansson-Verkasalo, 2017; Eichorn et al., 2018) also extend to the verbal domain. Indeed, using similar measures, we found that CWS have reduced CF not only in the nonverbal domain but also in the verbal domain. Thus, regardless of the nature of the stimuli (perceptual or semantic), the CWS, as a group, had more difficulty flexibly switching from one representation to another, suggesting generalized (i.e., not domain-specific) weaknesses in CF.

Limitations and Conclusions

This study is the first to explore CF for both semantic and perceptual information in CWS and their typically fluent peers, furthering our understanding of the executive function skills of preschool CWS. This study has several potential limiting factors. First, children were given prompts to select an associate or press a button during the tasks, if needed. This was done because the computerized tasks proved to be challenging for some children, particularly the 3-year-olds. Note, however, that children did not receive any feedback about their performance. Nevertheless, while it would have been ideal if no prompts had been given at all, the fact that children in both groups received prompts attenuates this concern.

Second, children who were unable to produce enough usable RT values had to be excluded from the analyses because mean RT cannot be meaningfully calculated based on just a few values. Of note, most of the children who were excluded from these analyses were 3 years of age, which was not unexpected given that previous research has consistently shown that CF is more challenging for 3-year-old children than 4- and 5-year-old children (e.g., Blaye & Jacques, 2009; Mennetrey & Angeard, 2018). Nevertheless, by excluding these participants, the sample size was reduced to 31 per group in the DSCT and 21 per group in the DPCT. Though not ideal, compared to other studies in the field, whose sample sizes are often in the range of 10–20 per group, the sample sizes are reasonable.

Third, as previously indicated, the DSCT and DPCT were not indistinguishable. As will be recalled, the DSCT was modeled after the DCT (Blaye & Jacques, 2009). The DSCT and DCT differed only in that the DSCT contained different stimulus items than the DCT and was administered via the computer. The DPCT, on the other hand, was developed to be in keeping with the general format and procedures of the DSCT, but the stimulus items were comparable to that of the FIST (Jacques & Zelazo, 2001). The DPCT differed from the FIST in that the former used the primary colors of blue, red, and yellow (the FIST used orange, purple, and pink); the shapes of a teddy bear, boat, and duck (the FIST used fish, phone, and socks); and three associates (color, shape, and size) instead of four (the FIST used color, shape, size, and number). We adopted these changes to the stimulus items to make the task easier for the youngest participants in our study. However, the choice to model the DPCT after the FIST resulted in a task that differed from the DSCT in that the stimulus items had to be used more than once. Thus, while it would have been ideal if the DSCT and DPCT had been identical, save the difference in the nature of the stimuli, it was not feasible to do so.

Fourth, we acknowledge that many of the CWS who participated in this study will likely spontaneously recover from stuttering. As a result, the extent to which CF might differ in these children relative to those who continue to stutter is unknown. Nevertheless, as we have argued elsewhere (Anderson et al., 2019), as a first step in this line of research, from a methodological as well as practical standpoint, it seems reasonable to establish whether there are differences between CWS and CWNS in processes like CF before examining the role this factor may play in stuttering persistence or recovery. Of course, any differences observed between CWS and CWNS would not necessarily be linked to persistence or recovery. That is, early in development, children could have subtle weaknesses or delays in acquiring cognitive, language, and/or motor processes that contribute to the development of stuttering without those same factors being related to stuttering persistence/recovery patterns. In short, these are separate questions, both important, but with one line of questions (i.e., exploring the nature of differences between groups early in development) extending naturally to the other (exploring how those differences might link to profiles of persistence or recovery later on).

In summary, based on previous research suggesting that CWS, as a group, may have weaknesses in components of inhibition and working memory, we speculated that the CWS would perform less accurately and more slowly than the CWNS on both the semantic and perceptual flexibility tasks. This hypothesis was partially confirmed: The CWS, as a group, performed significantly more slowly on both tasks but were otherwise similar to the CWNS in response accuracy. Of course, not all CWS performed more slowly just as not all CWNS performed more quickly; individual differences in performance were evident in both groups of children. Some CWS may have responded more slowly to the tasks due to difficulties with abstraction, inhibition, and/or activation. Both groups of children had more difficulty switching between different perceptual representations than semantic ones, and this difficulty may be attributed to increased inhibition and/or abstract property demands. Furthermore, the fact that the CWS, as a group, had difficulty flexibly switching between both semantic and perceptual representations suggests that the difficulty they have with CF is domain-general, extending to both the verbal and nonverbal domains.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01DC012517, awarded to Julie D. Anderson.

References

American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening*.

Amster, B. J. (1995). Perfectionism and stuttering. In C. Starkweather & H. Peters (Eds.), *Stuttering: Proceedings of First World Congress on Fluency Disorders* (Vol. II, pp. 540–543). Nijmegen University Press.

Anderson, J. D., & Conture, E. G. (2004). Sentence-structure priming in young children who do and do not stutter. *Journal of Speech, Language, and Hearing Research, 47*(3), 552–571. https://doi.org/10.1044/1092-4388(2004/043)

Anderson, J. D., & Ofoe, L. C. (2019). The role of executive function in developmental stuttering. *Seminars in Speech and Language, 40*(4), 305–319. https://doi.org/10.1055/s-0039-1692965

Anderson, J. D., & Wagovich, S. A. (2017). Explicit and implicit verbal response inhibition in preschool-age children who stutter. *Journal of Speech, Language, and Hearing Research, 60*(4), 836–852. https://doi.org/10.1044/2016_JSLHR-S-16-0135

Anderson, J. D., Wagovich, S. A., & Brown, B. T. (2019). Phonological and semantic contributions to verbal short-term memory in young children with developmental stuttering. *Journal of Speech, Language, and Hearing Research, 62*(3), 644–667. https://doi.org/10.1044/2018_JSLHR-S-18-0039

Anderson, J. D., Wagovich, S. A., & Hall, N. E. (2006). Nonword repetition skills in young children who do and do not stutter. *Journal of Fluency Disorders, 31*(3), 177–199. https://doi.org/10.1016/j.jfludis.2006.05.001

Archibald, L. M. D. (2017). Working memory and language learning: A review. *Child Language Teaching and Therapy, 33*(1), 5–17. https://doi.org/10.1177/0265659016654206

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baldwin, D. A. (1992). Clarifying the role of shape in children's taxonomic assumption. *Journal of Experimental Child Psychology, 54*(3), 392–416. https://doi.org/10.1016/0022-0965(92)90027-4

Bennett, J., & Müller, U. (2010). The development of flexibility and abstraction in preschool children. *Merrill-Palmer Quarterly, 56*(4), 455–473. https://doi.org/10.1353/mpq.2010.0004

Blaye, A., & Bonthoux, F. (2001). Thematic and taxonomic relations in preschoolers: The development of flexibility in categorization choices. *British Journal of Developmental Psychology, 19*(3), 395–412. https://doi.org/10.1348/026151001166173

Blaye, A., & Jacques, S. (2009). Categorical flexibility in preschoolers: Contributions of conceptual knowledge and executive control. *Developmental Science, 12*(6), 863–873. https://doi.org/10.1111/j.1467-7687.2009.00832.x

Brocklehurst, P. H., Drake, E., & Corley, M. (2015). Perfectionism and stuttering: Findings of an online survey. *Journal of Fluency Disorders, 44,* 46–62. https://doi.org/10.1016/j.jfludis.2015.02.002

Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological priming in young children who stutter: Holistic versus incremental processing. *American Journal of Speech-Language Pathology, 16*(1), 43–53. https://doi.org/10.1044/1058-0360(2007/006)

Carroll, D. J., Blakey, E., & FitzGibbon, L. (2016). Cognitive flexibility in young children: Beyond perseveration. *Child Development Perspectives, 10*(4), 211–215. https://doi.org/10.1111/cdep.12192

Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities, 35*(6), 510–525. https://doi.org/10.1177/00222194020350060301

Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science, 11*(3), 339–353. https://doi.org/10.1111/j.1467-7687.2008.00679.x

Chevalier, N., Wiebe, S. A., Huber, K. L., & Espy, K. A. (2011). Switch detection in preschoolers' cognitive flexibility. *Journal of Experimental Child Psychology, 109*(3), 353–370. https://doi.org/10.1016/j.jecp.2011.01.006

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: A review. *International Journal of Psychological Research, 3*(1), 58–67. https://doi.org/10.21500/20112084.844

Cragg, L., & Chevalier, N. (2012). The processes underlying flexibility in childhood. *Quarterly Journal of Experimental Psychology, 65*(2), 209–232. https://doi.org/10.1080/17470210903204618

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171–237. https://doi.org/10.1006/jecp.1996.2356

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior, 31,* 271–327. https://doi.org/10.1016/S0065-2407(03)31007-9

De Sonneville, L. M. J. (2009). *Amsterdamse Neuropsychologische Taken* [Amsterdam Neuropsychological Tasks]. Boom Test Publishers.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool children's performance in task switching on the Dimensional Change Card Sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology, 28,* 689–729. https://doi.org/10.1207/s15326942dn2802_7

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. AGS. https://doi.org/10.1037/t15144-000

Eggers, K., De Nil, L. F., & Van den Bergh, B. R. H. (2010). Temperament dimensions in stuttering and typically developing children. *Journal of Fluency Disorders, 35*(4), 355–372. https://doi.org/10.1016/j.jfludis.2010.10.004

Eggers, K., De Nil, L. F., & Van den Bergh, B. R. H. (2013). Inhibitory control in childhood stuttering. *Journal of Fluency Disorders, 38*(1), 1–13. https://doi.org/10.1016/j.jfludis.2012.10.001

Eggers, K., De Nil, L. F., & Van den Bergh, B. R. H. (2018). Exogenous triggered response inhibition in developmental stuttering. *Journal of Fluency Disorders, 56,* 33–44. https://doi.org/10.1016/j.jfludis.2018.02.001

Eggers, K., & Jansson-Verkasalo, E. (2017). Auditory attentional set-shifting and inhibition in children who stutter. *Journal of Speech, Language, and Hearing Research, 60*(11), 3159–3170. https://doi.org/10.1044/2017_JSLHR-S-16-0096

Eichorn, N., Marton, K., & Pirutinsky, S. (2018). Cognitive flexibility in preschool children with and without stuttering disorders. *Journal of Fluency Disorders, 57,* 37–50. https://doi.org/10.1016/j.jfludis.2017.11.001

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 234–250. https://doi.org/10.1080/713755604

Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*(6), 966–979. https://doi.org/10.1037/0012-1649.33.6.966

Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). *BRIEF-P: Behavior Rating Inventory of Executive Function–Preschool Version: Professional manual*. Psychological Assessment Resources.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2)*. AGS. https://doi.org/10.1037/t15098-000

Hakim, H. B., & Ratner, N. B. (2004). Nonword repetition abilities of children who stutter: An exploratory study. *Journal of Fluency Disorders, 29*(3), 179–199. https://doi.org/10.1016/j.jfludis.2004.06.001

Hollingshead, A. (1975). *Four-factor index of social status* [Unpublished manuscript]. Department of Sociology, Yale University.

Hresko, W., Reid, D., & Hammill, D. (1999). *Test of Early Language Development–Third Edition (TELD-3)*. Pro-Ed.

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology, 30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology, 20*(3), 573–591. https://doi.org/10.1207/S15326942DN2003_2

Jacques, S., & Zelazo, P. D. (2005). On the possible roots of cognitive flexibility. In B. D. Homer & C. S. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 53–81). Erlbaum.

Jacques, S., Zelazo, P. D., Lourenco, S. F., & Sutherland, A. E. (2009). *The roles of labeling and abstraction in the development of cognitive flexibility* [Unpublished doctoral dissertation]. University of Toronto.

Karalunas, S. L., Huang-Pollock, C. L., & Nigg, J. T. (2013). Is reaction time variability in ADHD mainly at low frequencies? *The Journal of Child Psychology and Psychiatry, 54*(5), 536–544. https://doi.org/10.1111/jcpp.12028

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*(5), 449–476. https://doi.org/10.1111/1467-7687.00300

Kowalski, K., & Zimiles, H. (2006). The relation between children's conceptual functioning with color and color term acquisition. *Journal of Experimental Child Psychology, 94*(4), 301–321. https://doi.org/10.1016/j.jecp.2005.12.001

Kronenberger, W. G., & Pisoni, D. B. (2018). Neurocognitive functioning in deaf children with cochlear implants. In H. Knoors & M. Marschark (Eds.), *Evidence-based practice in deaf education* (pp. 363–396). Oxford University Press. https://doi.org/10.1093/oso/9780190880545.003.0016

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978–990. https://doi.org/10.3758/s13428-012-0210-4

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

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16*(1), 1–27. https://doi.org/10.1016/0010-0285(84)90002-1

Martin, N., & Gupta, P. (2004). Exploring the relationship between word processing and verbal short-term memory: Evidence from associations and dissociations. *Cognitive Neuropsychology, 21*(2–4), 213–228. https://doi.org/10.1080/02643290342000447

Mennetrey, C., & Angeard, N. (2018). Cognitive flexibility training in three-year-old children. *Cognitive Development, 48,* 125–134. https://doi.org/10.1016/j.cogdev.2018.08.004

Mohan, R., & Weber, C. (2015). Neural systems mediating processing of sound units of language distinguish recovery versus persistence in stuttering. *Journal of Neurodevelopmental Disorders, 7,* Article 28. https://doi.org/10.1186/s11689-015-9124-7

Müller, U., Dick, A. S., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the dimensional change card sort task. *Child Development, 77*(2), 395–412. https://doi.org/10.1111/j.1467-8624.2006.00878.x

Müller, U., Jacques, S., Brocki, K., & Zelazo, P. D. (2009). The executive functions of language in preschool children. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 53–68). Cambridge University Press. https://doi.org/10.1017/CBO9780511581533.005

Nardini, M., Bales, J., & Mareschal, D. (2016). Integration of audio-visual information for spatial decisions in children and adults. *Developmental Science, 19*(5), 803–816. https://doi.org/10.1111/desc.12327

Ntourou, K., Anderson, J. D., & Wagovich, S. A. (2018). Executive function and childhood stuttering: Parent ratings and evidence from a behavioral task. *Journal of Fluency Disorders, 56,* 18–32. https://doi.org/10.1016/j.jfludis.2017.12.001

Ntourou, K., Conture, E. G., & Lipsey, M. W. (2011). Language abilities of children who stutter: A meta-analytical review. *American Journal of Speech-Language Pathology, 20*(3), 163–179. https://doi.org/10.1044/1058-0360(2011/09-0102)

Ofoe, L. C., Anderson, J. D., & Ntourou, K. (2018). Short-term memory, inhibition, and attention in developmental stuttering: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 61*(7), 1626–1648. https://doi.org/10.1044/2018_JSLHR-S-17-0372

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation, 9,* Article 6. https://doi.org/10.7275/qf69-7k43

Pelczarski, K. M., & Yaruss, J. S. (2016). Phonological memory in young children who stutter. *Journal of Communication Disorders, 62,* 54–66. https://doi.org/10.1016/j.jcomdis.2016.05.006

Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task. *Infant and Child Development, 11*(2), 93–105. https://doi.org/10.1002/icd.299

Piispala, J., Kallio, M., Bloigu, R., & Jansson-Verkasalo, E. (2016). Delayed N2 response in Go condition in a visual Go/Nogo ERP study in children who stutter. *Journal of Fluency Disorders, 48,* 16–26. https://doi.org/10.1016/j.jfludis.2016.02.001

Piispala, J., Määttä, S., Pääkkönen, A., Bloigu, R., Kallio, M., & Jansson-Verkasalo, E. (2017). Atypical brain activation in children who stutter in a visual Go/Nogo task: An ERP study. *Clinical Neurophysiology, 128*(1), 194–203. https://doi.org/10.1016/j.clinph.2016.11.006

Piispala, J., Starck, T., Jansson-Verkasalo, E., & Kallio, M. (2018). Decreased occipital alpha oscillation in children who stutter during a visual Go/Nogo task. *Clinical Neurophysiology, 129*(9), 1971–1980. https://doi.org/10.1016/j.clinph.2018.06.022

Pisoni, D. B., Conway, C. M., Kronenberger, W. G., Henning, S., & Anaya, E. (2010). Executive function, cognitive control, and sequence learning in deaf children with cochlear implants. In M. Marshark & P. E. Spencer (Eds.), *The Oxford handbook of deaf studies, language, and education* (Vol. 2, pp. 439–457). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195390032.013.0029

Pisoni, D. B., Conway, C. M., Kronenberger, W., Horn, D. L., Karpicke, J., & Henning, S. C. (2008). Efficacy and effectiveness of cochlear implants in deaf children. In M. Marschark & P. C. Hauser (Eds.), *Deaf cognition: Foundations and outcomes* (pp. 52–101). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195368673.003.0003

Pitchford, N. J., & Mullen, K. T. (2001). Conceptualization of perceptual attributes: A special case for color? *Journal of Experimental Child Psychology, 80*(3), 289–314. https://doi.org/10.1006/jecp.2001.2634

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Riley, G. D. (2009). *Stuttering Severity Instrument for Children and Adults–Fourth Edition (SSI-4)*. Pro-Ed.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217–236. https://doi.org/10.1068/p5117

Samyn, V., Roeyers, H., Bijttebier, P., Rosseel, Y., & Wiersema, J. R. (2015). Assessing effortful control in typical and atypical development: Are questionnaires and neuropsychological measures interchangeable? A latent-variable analysis. *Research in Developmental Disabilities, 36,* 587–599. https://doi.org/10.1016/j.ridd.2014.10.018

Sasisekaran, J., Brady, A., & Stein, J. (2013). A preliminary investigation of phonological encoding skills in children who stutter. *Journal of Fluency Disorders, 38*(1), 45–58. https://doi.org/10.1016/j.jfludis.2012.12.003

Sasisekaran, J., & Byrd, C. (2013). Nonword repetition and phoneme elision skills in school-age children who do and do not stutter. *International Journal of Communication Disorders, 48*(6), 625–639. https://doi.org/10.1111/1460-6984.12035

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*(4), 344–358. https://doi.org/10.1016/j.jfludis.2012.06.001

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *The Journal of Child Psychology and Psychiatry, 54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Waxman, S. R., Senghas, A., & Benveniste, S. (1997). A cross-linguistic examination of the noun-category bias: Its existence and specificity in French- and Spanish-speaking preschool-aged children. *Cognitive Psychology, 32*(3), 183–218. https://doi.org/10.1006/cogp.1997.0650

Weber-Fox, C., Wray, A. H., & Arnold, H. (2013). Early childhood stuttering and electrophysiological indices of language processing. *Journal of Fluency Disorders, 38*(2), 206–221. https://doi.org/10.1016/j.jfludis.2013.01.001

Williams, K. T. (2007). *Expressive Vocabulary Test–Second Edition (EVT-2)*. AGS. https://doi.org/10.1037/t15094-000

Yairi, E., & Ambrose, N. (1992). A longitudinal study of stuttering in children: A preliminary report. *Journal of Speech and Hearing Research, 35*(4), 755–760. https://doi.org/10.1044/jshr.3504.755

Yairi, E., & Seery, C. H. (2015). *Stuttering: Foundations and clinical applications* (2nd ed.). Pearson Education.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). MIT Press.

Appendix A

Experimental Stimuli for the Double Semantic Categorization Task

Stimulus set	Target object	Thematic associate	Taxonomic associate	Unrelated associate
1	squirrel	tree	cat	chair
2	glass	straw	spoon	ball
3	bread	knife	corn	star
4	door	key	window	pear
5	cow	milk	pig	ring
6	dog	bone	bear	telephone
7	flower	bee	tree	doll
8	car	road	train	fork
9	apple	worm	banana	snowman
10	mouse	cheese	fish	pencil
11	bird	nest	butterfly	toothbrush
12	foot	shoe	arm	leaf
13	carrot	rabbit	pumpkin	book

Appendix B

Experimental Stimuli for the Double Perceptual Categorization Task

Stimulus set	Target	Relevant dimension	Shape associate	Color associate	Size associate	Unrelated associate
1	large blue teddy bear	size, color	n/a	small blue duck	large red duck	small yellow boat
2	small red boat	shape, size	medium yellow boat	n/a	small blue teddy bear	large yellow teddy bear
3	medium yellow duck	color, shape	large red duck	small yellow teddy bear	n/a	large blue boat
4	medium red teddy bear	size, color	n/a	large red duck	medium yellow boat	small blue boat
5	large blue boat	shape, size	small yellow boat	n/a	large yellow duck	medium red teddy bear
6	small red duck	color, shape	small yellow duck	small red teddy bear	n/a	small blue teddy bear
7	medium yellow boat	size, color	n/a	large yellow duck	medium blue duck	large red duck
8	medium blue teddy bear	shape, size	large red teddy bear	n/a	medium red boat	large red boat
9	small red boat	color, shape	small yellow boat	small red teddy bear	n/a	small blue duck
10	small yellow teddy bear	size, color	n/a	medium yellow teddy bear	small red teddy bear	large blue teddy bear
11	large red duck	shape, size	medium yellow duck	n/a	large blue teddy bear	medium yellow teddy bear
12	medium blue teddy bear	color, shape	small red teddy bear	large blue duck	n/a	small red boat