Journal of Fluency Disorders 70 (2021) 105846

Contents lists available at ScienceDirect

Journal of Fluency Disorders

journal homepage: www.elsevier.com/locate/jfludis

Working memory in adults who stutter using a visual N-back task

Zoi Gkalitsiou *, Courtney T. Byrd

The University of Texas at Austin, Department of Speech, Language, and Hearing Sciences, USA

ARTICLE INFO

Keywords:
Stuttering
Working memory
Central executive
N-back
Adults

ABSTRACT

Purpose: The purpose of this study was to investigate working memory in adults who do (AWS) and do not (AWNS) stutter using a visual N-back task. Processes involved in an N-back task include encoding, storing, rehearsing, inhibition, temporal ordering, and matching.

Methods: Fifteen AWS (11 males, 4 females; *M* = 23.27 years, *SD* = 5.68 years) and 15 AWNS (*M* = 23.47 years, *SD* = 6.21 years) were asked to monitor series of images and respond by pressing a "yes" button if the image they viewed was the same as the image one, two, or three trials back. Stimuli included images with phonologically similar (i.e., phonological condition) or phonologically dissimilar (i.e., neutral condition) names. Accuracy and manual reaction time (mRT) were analyzed.

Results: No difference was found between AWS and AWNS in accuracy. Furthermore, both groups were more accurate and significantly faster in 1- followed by 2- followed by 3-back trials. Finally, AWNS demonstrated faster mRT in the phonological compared to neutral condition, whereas AWS did not.

Conclusion: Results from this study suggest different processing mechanisms between AWS and AWNS for visually presented phonologically similar stimuli. Specifically, a phonological priming effect occurred in AWNS but not in AWS, potentially due to reduced spreading activation and organization in the mental lexicon of AWS. However, the lack of differences between AWS and AWNS across all N-back levels does not support deficits in AWS in aspects of working memory targeted through a visual N-back task; but, these results are preliminary and additional research is warranted.

1. Introduction

Stuttering is a complex, multifactorial disorder characterized by atypical disruptions in the forward flow of speech (Smith & Weber, 2017). Several factors contribute to its development and persistence including genetic predisposition (e.g., Drayna & Kang, 2011), neurophysiological differences (e.g., Giraud et al., 2008; Watkins et al., 2008), differences in emotional reactivity and regulation (e.g., Jones et al., 2014; Karrass et al., 2006) or variances in speech motor control (e.g., Alm, 2004; Max et al., 2004; Namasivayam & van Lieshout, 2008). Most relevant to the present study, differences in linguistic/cognitive processing of individuals who stutter have also been suggested as a contributing factor to stuttered speech (e.g., Byrd et al., 2012; Byrd et al., 2015a; Coalson & Byrd, 2015; Sasisekaran et al., 2006; Weber-Fox et al., 2004).

Speakers do not randomly select words for production, instead they carefully select and access words from their long-term memory in order to communicate their thoughts and goals. Therefore, executive control (or executive functions or executive processes) plays a

* Corresponding author at: The University of Texas at Austin, 1 University Station A1100, Austin, TX 78759, USA.
E-mail address: zoigkal@utexas.edu (Z. Gkalitsiou).

https://doi.org/10.1016/j.jfludis.2021.105846
Received 1 July 2020; Received in revised form 16 January 2021; Accepted 18 March 2021
Available online 26 March 2021

key role in language production as competing information needs to be inhibited in order for target words to be selected and produced (e.g., Miyake et al., 2000). According to (Diamond, 2013), there are three core executive functions: (a) working memory (i.e., ability to retain and manipulate information for short periods of time), (b) inhibition (i.e., inhibitory control, self-control, and interference control), and (c) cognitive flexibility (i.e., ability to shift attention and adapt to new information). Attention interacts with the three core executive processes and plays a fundamental role for an individual's successful performance in a variety of cognitive tasks that require ability to maintain information in active memory, particularly during interference (Engle & Kane, 2004).

Before producing a word, the speaker needs to choose the appropriate word to convey the intended message, access its lemma, and construct its phonetic plan, which will then be executed (e.g., Levelt, 1989). This process requires a wise selection of a target word; thus, inhibiting irrelevant distractors, which will then lead to the retrieval of its phonemes from long-term memory and their temporary storage and verbal manipulation before they become available for motor execution (e.g., Levelt, 1989). Working memory will temporarily store this information until it is ready for articulation (Baddeley, 2003, 2012).

In simple linguistic tasks, such as picture naming, researchers have found that attention (e.g., Jongman et al., 2019), inhibition (e.g., Shao et al., 2012; Shao et al., 2013), and working memory (e.g., Piai & Roelofs, 2013) affect naming latencies, providing further evidence for the influence of executive control in language production. For example, the ability to maintain focus of attention prior to speech planning initiation influences the speed of language production (e.g., Jongman et al., 2019). The ability to update and monitor information in working memory has also been linked to the speed a speaker can name pictures (e.g., Piai & Roelofs, 2013).

The purpose of the present study was to investigate visual-verbal working memory in adults who stutter (AWS) when they process phonologically similar nameable pictures via an N-back task. Briefly, in an N-back paradigm, participants are required to recognize the stimuli (e.g., Pelegrina et al., 2015), monitor series of stimuli and respond by pressing a button whether the current stimulus is the same as the stimulus *N* trials back. Processes involved in an N-back task include: stimulus recognition, encoding, storing, rehearsing, monitoring, matching, inhibiting, and updating (Jonides et al., 1997). Stimuli in the present study included nameable line drawings; therefore, participants were expected to initially recognize the images visually, thus engage in nonverbal/nonlinguistic processing, and access the images' names (at least to some degree given that labels were available for the common images that were used in the study), thus engage in verbal/linguistic processing (Kelley et al., 1998).

1.1. Working memory and stuttering

According to Baddeley and colleagues (Baddeley, 2003, 2012; Baddeley et al., 2011), working memory is a set of interacting processes that involve temporary storage and manipulation of information and includes two slave systems: a) the phonological loop, which includes a phonological short-term store responsible for storing verbal information, and an articulatory rehearsal mechanism responsible for refreshing information to prevent decay, and b) the visuospatial sketchpad, responsible for storage and manipulation of visual and spatial information. Regarding verbal information processed through the phonological loop, auditory-verbal information has direct access to the phonological store whereas visual-verbal information (e.g., printed words) needs to be visually analyzed, phonologically recoded and gain access to the phonological store through articulatory rehearsal (e.g., Baddeley et al., 1998; Repovs & Baddeley, 2006; Vallar & Baddeley, 1984; Vallar & Papagno, 2002). Regarding nonverbal visual information processed by the visuospatial sketchpad, a store component and a rehearsal mechanism, analogous to the phonological loop, have been suggested (e.g., Papagno, 2002).

Working memory also includes a central executive component, which is considered to be the most complex component as it is responsible for attention focus, dividing attention, switching between tasks, decision making, and providing the interface between long-term memory and all subsystems of working memory (i.e., phonological loop, visuospatial sketchpad, episodic buffer). Finally, the episodic buffer is a limited capacity multidimensional store that allows all other components (i.e., phonological loop, visuospatial sketchpad, central executive) to interact with each other and with long-term memory by binding their separate codes into integrated segments (Baddeley, 2003, 2012).

The phonological loop and central executive are the two components of working memory that have received the most attention in stuttering research due to their critical role in spoken word planning and production; thus, these are the two components that will be further discussed.

1.1.1. The role of phonological loop in stuttering

The phonological loop comprises a phonological store, which stores information for a short period of time, and an articulatory rehearsal mechanism, which has a dual role. Articulatory or subvocal rehearsal is a required step for visually presented verbal stimuli to access the phonological store after being coded into phonological codes (Vallar & Papagno, 2002). In addition, information in the store decays but can be refreshed via the process of articulatory/subvocal rehearsal (Baddeley, 2012). A distinction between a separate auditory short-term store and a separate visual short-term store is supported by studies wherein patients demonstrated better memory span performance with visual-verbal compared to auditory-verbal stimulus presentation (Vallar & Baddeley, 1984; Warrington et al., 1971). Under auditory-verbal stimulus presentation, access to the phonological store is obligatory. In contrast, under visual-verbal stimulus presentation, as in the present study, access is provided via a recoding process and subvocal rehearsal.

A variety of experimental paradigms (e.g., nonword repetition, phoneme monitoring, rhyme judgment, list recall, etc.) that assess phonological working memory (i.e., phonological loop) support a link between stuttering and phonological working memory deficits (e.g., Byrd et al., 2012; Byrd et al., 2015a; Byrd et al., 2015b; Sasisekaran & Weisberg, 2014; Weber-Fox et al., 2004, cf., Nippold 2002). Poorer performance during word (e.g., Bosshardt, 1990) and nonword reading tasks (e.g., Sasisekaran, 2013) has been observed in AWS compared to adults who do not stutter (AWNS), supporting a relationship between stuttering and slower phonemic encoding and/or speech planning and motor execution. Deficits in subvocal rehearsal have also been demonstrated in AWS via list recall tasks (e.g., Byrd et al., 2015b). Furthermore, results from nonword repetition tasks and/or additional metalinguistic tasks including phoneme manipulation (e.g., phoneme elision) indicate slower and less accurate performance in AWS (e.g., Byrd et al., 2012; Byrd et al., 2015a; Sasisekaran & Weisberg, 2014), providing further evidence for aberrant function of the phonological loop in AWS as opposed to AWNS.

1.1.2. The role of central executive in stuttering

The central executive is a critical and complex component of working memory, which is responsible for focusing attention, dividing attention, switching between tasks, and providing the interface between long-term memory and all the subsystems of working memory (Baddeley, 2012). Tasks used to investigate central executive function in AWS typically required participants to engage in two tasks concurrently (i.e., dual-task paradigms) (e.g., Bajaj, 2007).

During dual-tasks, which usually involve a primary and a secondary task, individuals who stutter have been reported to demonstrate reduced attentional resources in primary (Bosshardt et al., 2002) as well as in secondary (Maxfield et al., 2016) tasks. Results on the impact of dual-tasks on speech fluency in people who stutter are mixed. Some studies have found decreased fluency abilities (e.g., Bosshardt, 2002), others increased fluency abilities (e.g., Eichorn et al., 2016), while others found no difference in AWS' speech fluency during dual-task performance (e.g., Bosshardt et al., 2002). These results indicate a complex relationship between dual-task performance and stuttering frequency, with interpretations suggesting more vulnerable phonological and articulatory systems in AWS when attention resources are also allocated in a concurrent verbal task (Bosshardt, 2002). Alternatively, increased fluency in AWS may be the result of distribution of working memory and attentional resources to tasks other than speaking (e.g., Eichorn et al., 2016).

Taken together, these data suggest individuals who stutter demonstrate deficits in working memory, specifically with regard to phonological loop (i.e., phonological encoding and subvocal rehearsal) *and* the function of the central executive.

1.2. Purpose of the present study

To date, the vast majority of studies that investigated phonological loop and/or central executive function in individuals who stutter employed tasks that required overt speech production, making it difficult to disentangle the influence of speech planning on the required overt production (e.g., Bosshardt et al., 2002; Byrd et al., 2012; Byrd et al., 2015b, Maxfield et al., 2016). For example, in a nonword repetition task, AWS may have difficulty rehearsing the letter sequences comprising the nonwords; alternatively, they could have difficulty programming and executing the motor plan of a nonword. Thus, the present study employed a visual N-back task, a novel experimental paradigm that focused on visual-verbal working memory without requiring overt speech production.

The N-back task has been used to assess working memory ability in typical as well as clinical populations (e.g., aphasia in Wright et al., 2007; thalamic lesions in Kubat-Silman et al., 2002). Successful performance on the N-back task requires recognition, encoding, storing, rehearsing, monitoring, matching, inhibiting, and updating information (Jonides et al., 1997; Pelegrina et al., 2015); therefore, the task places demands on phonological loop and central executive. During this task, participants are required to monitor series of stimuli (e.g., letters, objects) and respond each time whether the most recently presented stimulus is the same as the stimulus *N* trials back. The working memory load can be manipulated by changing the N-back level (e.g., 0-back, 1-back, 2-back etc.).

In the present study, line drawings of actual, nameable objects were used; therefore, participants were expected to engage in both nonlinguistic and linguistic processing of the stimuli (e.g., Kelley et al., 1998). Specifically, participants needed to decode the image visually, access its lemma and generate a name for each image (e.g., Indefrey & Levelt, 2000), maintain and rehearse the images' names and match whether each current image was the same as the image 1-, 2-, or 3- trials back; thus, recruiting phonological loop and central executive resources. Subvocal rehearsal was assumed to rehearse and prevent the verbal information from decay as well as to provide access of the visual-verbal information to the phonological store (e.g., Vallar & Baddeley, 1984; Vallar & Papagno, 2002).

Three different N-back levels (i.e., 1-back, 2-back, 3-back) and two linguistic conditions (i.e., phonologically similar versus phonologically dissimilar items) were incorporated in the study, which allowed investigation of the impact of increasing cognitive demands on participant performance.

1.2.1. Predictions regarding working memory load

Incorporation of different N-back levels is a typical way to manipulate the task's difficulty and demands placed on working memory. Participants perform worse as the N-back level increases due to the increased working memory demands and cognitive resources (e.g., attention) placed in the higher levels. This tendency has been reported in both neurotypical adults (e.g., Jonides et al., 1997) and in individuals with brain injury (e.g., Kubat-Silman et al., 2002). Therefore, we expected that both AWS and AWNS would be more accurate and faster in 1-back, followed by 2-back, followed by 3-back. Since working memory demands in 1-back are minimal, no differences were expected between the two groups in that level. But, based on the reported difficulties of AWS in tasks tapping on phonological loop and central executive function (for a review see Bajaj, 2007), we anticipated poorer performance in AWS compared to AWNS in 2- and 3-back trials.

1.2.2. Predictions regarding linguistic condition

Incorporating phonologically similar stimuli in the study was another way to manipulate the task's cognitive demands. Phonologically similar stimuli have been reported to affect working memory performance in healthy adults (e.g., Mueller et al., 2003; Sweet et al., 2008) and AWS (e.g., Byrd et al., 2015b). Specifically, during 2-back tasks, accuracy has been reported to be lower when visually presented consonants have names that rhymed compared to consonants that did not rhyme (e.g., Sweet et al., 2008). In word serial recall tasks, fewer words are recalled from phonologically similar sets than dissimilar ones, and this effect has been observed in both healthy adults (e.g., Mueller et al., 2003) and AWS (e.g., Byrd et al., 2015b).

One reason that may contribute to decreased participant performance under phonologically similar items includes the increased difficulty of those items to be remembered because they are “harder to discriminate in terms of the articulatory code in which they are stored” (Vallar & Baddeley, 1984, p. 152). Another possibility may be that phonological codes may decay faster when they are phonologically similar than dissimilar (Mueller et al., 2003). Alternatively, phonological codes may decay at the same rate regardless of their phonological similarity, but during recall, the degraded codes of phonologically similar items may be more difficult to reconstruct (Mueller et al., 2003). Finally, attention-related accounts have also been suggested, wherein participants are required to implement more “efficient” attentional strategies when processing phonologically similar items (i.e., more taxing), by disengaging their attention from other destructing/unrelated processes in order to focus on the task at hand and improve their concentration to perform well on the task (e.g., Sweet et al., 2008).

In this study, we anticipated that both groups would perform better in the neutral compared to the phonological condition (e.g., Byrd et al., 2015b; Mueller et al., 2003; Sweet et al., 2008; Vallar & Baddeley, 1984). Furthermore, based on evidence suggesting that AWS demonstrate deficits in lexical access, phonological loop, and central executive (e.g., Byrd et al., 2015a, 2015b; Maxfield et al., 2012; McGill et al., 2016), we predicted that AWS would be significantly less accurate and slower than AWNS in the phonological condition at the 2- and 3-back levels, the most cognitively demanding of the required tasks.

2. Method

2.1. Participants

Approval of the study was provided by the authors’ university institutional review board and written informed consent was obtained for each participant. Participants were recruited from the authors’ university, city and surrounding areas. Participants in the study included 15 AWS ($M = 23.27$ years, $SD = 5.68$ years, $n = 11$ males, $n = 4$ females) and 15 AWNS ($M = 23.47$ years, $SD = 6.21$ years, $n = 11$ males, $n = 4$ females). The two groups did not differ statistically in age, $t(27.8) = .09$, $p = .931$. The package *SIMR* (Green & MacLeod, 2016) was used to determine the study’s required sample size for power of at least 80 % when using mixed-effects model analyses. Data from the first 12 participants, 6 participants per group, were used to estimate an initial power. This initial *SIMR* model was extended to n = 30 participants, using Monte Carlo simulations, and yielded power of 94 % to observe a 2-way interaction (i.e., Group x Working memory load) and 95 % to detect a 3- way interaction (i.e., Group x Linguistic condition x Working memory load).

To be eligible to participate in the study, participants needed to be native speakers of American English, have typical or corrected-to-normal vision and report no past or present history of speech or language disorders (with the exception of stuttering for AWS). All participants in the study were required to pass a binaural, pure-tone hearing screening at 20 dB HL at 1000 Hz, 2000 Hz and 4000 Hz (ASHA, 1997). In addition, all participants completed a near-vision acuity screening, using an Early Treatment Diabetic Retinopathy Study (ETDRS) chart testing at 40 cm, with a score of at least 20/30 or better, which is considered within the range of “normal vision” according to ICD-10-CM: United States Department of Health and Human Services (2016). Handedness was determined by the revised version of the Edinburgh Handedness Inventory (Dragovic, 2004) and only right-handed participants were included in the study.

In addition, each participant completed a case history form and reported no prior or current neurological, social, emotional, or psychiatric diagnoses or treatments. Information regarding the participants’ age, gender, race/ethnicity, educational level and primary spoken language were included in the case history form. All participants spoke primarily English and about half of the participants in each group reported speaking a language other than English (AWS: $n = 7$, AWNS: $n = 8$). In the AWS group, four participants reported speaking Spanish, one Mandarin, one Vietnamese and one Hindi. In the AWNS group, five participants reported speaking Spanish, one Mandarin, one ASL, and one French. Participants who reported speaking another language were also asked to report how proficient they were in that language on a 1–10 scale (1= not proficient, 10- highly proficient). In the AWS group, four participants reported high proficiency (scores 8–10), two participants reported being moderately proficient (scores 5–7) and one participant reported being not proficient (score of 1). In the AWNS group, four participants reported high proficiency (scores 8–10), two participants reported being moderately proficient (scores 5–7) and two participants reported being not proficient (score of 1). Finally, participants were asked to report current use of any medication that could potentially influence their performance in the study. Apart from one AWS (medication for acne) and one AWNS (medication for Hypothyroidism), no other participants reported use of any prescribed medication at the time of the study.

Participants in the two groups (i.e., AWS and AWNS) were matched in age (±3 years), gender, handedness, and educational level (i.e., highest degree obtained). Ten participants in each group reported high-school degree as their highest degree obtained and were attending college at the time of the study, one participant per group had a bachelor’s degree, and four participants in each group had a graduate degree.

2.1.1. Stuttering identification and talker group classification

Stuttering status was determined by a licensed speech-language pathologist based on the following criteria: a) self-identification as an individual who stutters by the participant; b) a score of 11 or higher on the *Stuttering Severity Instrument- 4th Edition* (SSI-4; Riley, 2009); c) confirmation of the presence of stuttering by the first author, a licensed speech-language pathologist. Three speech samples were collected and analyzed in terms of disfluencies for each participant: a reading sample, a narrative sample (i.e., picture description) and a conversational sample. In the present study, seven AWS received a severity SSI-4 score of very mild, four a severity score of mild, two a severity score of moderate, and two a severity score of severe.

2.2. Pre-testing and baseline measures

In an attempt to ensure that no participants in either group demonstrated vocabulary or cognitive deficits, each participant was required to perform within one standard deviation from the mean on the following pre-testing measures: a) *Test of Nonverbal Intelligence- Fourth Edition* (TONI-4; Brown et al., 2010), b) *NIH Toolbox for Assessment of Neurological and Behavioral Function* (2013) including the tests of *Picture Vocabulary Test* (receptive vocabulary), *List Sorting Working Memory Test* (working memory), *Oral Reading Recognition test* (oral reading), *Flanker Inhibitory Control and Attention Test* (inhibition), *Dimensional Change Card Sorting Test* (cognitive flexibility), and *Picture Sequence Memory Test* (episodic memory).

Participants' baseline manual reaction time was assessed using a visual task that resembled the experimental procedures used in the present study. Participants were asked to manually respond to a visual symbol (i.e., a black square) presented on the left or right of the computer screen, as quickly as possible, by pressing the left button with their left index finger or their right button with the right index finger respectively on a SuperLab response pad. The two buttons on the response pad represented the two sides on the screen. The order of the possible locations was randomized for each interval; however, participants received an equal number of left and right location trials. Stimuli were randomly presented at 500, 1000, 1500, and 2000 millisecond intervals, in 4 blocks of 10 trials (N = 40 total trials), wherein each time-interval was randomly presented 10 times across the 4 blocks.

To ensure that there were no performance differences between the two groups, independent samples t-tests were performed in the pre-testing and baseline measures, with alpha level set to $\alpha = .01$ due to multiple comparisons (Pituch & Stevens, 2016). Independent t-tests yielded no statistically significant differences between the two groups in *TONI,* $t(27.9) = .72$, $p = .477$, or any measures of the *NIH Toolbox* between AWS and AWNS; Picture Vocabulary Test, $t(27.3) = .48$, $p = .637$; List Sorting Working Memory Test, $t(27.7) = .81$, $p = .427$; Flanker Inhibitory Control Test, $t(26.5) = 1.50$, $p = .144$; Dimensional Change Card Sorting Test, $t(25.0) = 1.57$, $p = .128$; Picture Sequence Memory Test, $t(27.6) = 1.91$, $p = .067$; and Oral Reading Recognition Test, $t(27.8) = .02$, $p = .986$. In addition, statistically significant differences were not found between AWS and AWNS in the accuracy ($t(14) = 2.26$, $p = .041$) or manual reaction time ($t(27.8) = .04$, $p = .968$) in the baseline task. Table 2 provides information about the two groups' means in the baseline and pre-testing measures.

2.3. Experimental tasks and stimuli

2.3.1. Stimuli

Sixteen line drawings were included in the study, nine of which were obtained from Snodgrass and Vanderwart (1980), three from Szekely et al. (2004) and four from the Internet. The following ratings have been obtained for all stimuli based on their influence on picture recognition or picture naming tasks: a) visual complexity (i.e., the number of lines and the amount of detail in a picture, which affects the visual recognition system; Alario et al., 2004), b) image agreement (i.e., the degree to which mental images generated by individuals based on a picture's name agree with a presented picture, a parameter that impacts the visual recognition system; Alario

Table 1
List of stimuli included in the phonological and neutral conditions and information regarding the properties that stimuli were controlled for.

	Stimulus	VC	IA	NA	Imageab.	Freq.	ND	Phon. Prob-pos	Phon. Prob-bigr
Phono	block[1]	2.81	3.18	89	5.7	40.53	13	0.04162434	0.00163694
	clock[3]	2.81	3.91	95	6.8	58.63	17	0.04976693	0.0018563
	lock[2]	2.07	4.10	100	5.6	56.57	38	0.0494651	0.00146983
	rock[3]	3.38	3.82	95	6.4	86.16	41	0.06101901	0.00175076
	sock[2]	1.93	4.18	100	6.7	8.98	42	0.07432213	0.00109422
	black[1]	1.20	3.00	11	6.4	167.94	14	0.04432759	0.00264476
	log[3]	4.54	4.10	100	6.3	11.96	16	0.02746027	0.00060842
	sack[1]	2.50	3.30	32	5.5	12.92	38	0.07619447	0.00212016
	Mean	2.66	3.70	78	6.18	55.46	27.4	0.05302248	0.001647674
Neutral	tree[2]	4.00	4.36	100	6.8	65.00	16	0.05010673	0.00833318
	lips[2]	1.20	4.18	95	6.8	31.18	23	0.0565887	0.00344461
	cup[2]	1.47	3.00	63	6.7	51.65	19	0.06976891	0.00550372
	fox[2]	3.27	3.36	89	6.4	21.61	32	0.05455606	0.00272133
	doll[1]	4.06	3.36	100	6.4	24.76	26	0.06447224	0.00147759
	nail[2]	2.00	4.00	89	6.6	18.65	46	0.03911539	0.0012634
	bus[2]	4.20	3.40	95	6.8	74.18	25	0.07291384	0.00685404
	star[2]	1.00	4.60	100	6.8	81.35	9	0.04877823	0.00691405
	Mean	2.65	3.78	91	6.7	46.05	24.5	0.05703751	0.00456399

VC = Visual Complexity, IA = Image Agreement, NA = Name agreement, Imageab = Imageability, Freq = word frequency per million words, ND = Neighborhood Density, Phon. Prob-pos = positional phonotactic probability, Phon. Prob-bigr = bigram phonotactic probability, Phono = Phonological condition, Neutral = Neutral condition.
Note. Mean ratings for VC, IA, and Imageab. and the percentage of NA are reported. WF is reported as frequency per million words and ND is reported as the number of phonological neighbors of each stimulus. The word-average positional and bigram probabilities are reported.

[1] Images from the Internet.
[2] Images from Snodgrass and Vanderwart (1980).
[3] Images from the International Naming Project (Szekely et al., 2004).

et al., 2004), c) name agreement (i.e., the degree to which participants agree on the name of a picture, which affects naming difficulty and speed as well as recall ability; Alario et al., 2004), d) imageability (i.e., the ease with which an object name evokes few or many different mental images for a particular object and affects processing at a semantic level; Cortese & Fugett, 2004), e) frequency (i.e., how often a given word is used, which usually relies on counts of written corpora and affects picture naming latency; Brysbaert & New, 2009), f) neighborhood density (i.e., number of phonological or orthographic neighbors a word has by changing one sound or letter in the word, which affects accuracy and speed of picture naming; Balota et al., 2007; Vitevitch, 2002), and g) phonotactic probability (i.e., the likelihood of sound sequences that are present in a given word; Vaden et al., 2009).

Ratings for visual complexity, image agreement and name agreement were obtained from three different cohorts of undergraduate and graduate students ($n = 11$, $n = 11$, $n = 19$ respectively) following the guidelines described in Snodgrass and Vanderwart (1980). Mean ratings for visual complexity, image agreement, and imageability and the percentage of name agreement are reported for the aforementioned variables in Table 1. Furthermore, word frequency is reported as frequency per million words and neighborhood density is reported as the number of phonological neighbors of each stimulus. Two measures of phonotactic probability were obtained, positional and bigram. Positional probability refers to the probability of each phoneme occurring in a specific position in a given word (Vaden et al., 2009) and bigram probability refers to the probability of a sequence of two phonemes occurring in a specific position in a given word (Vaden et al., 2009). The word-average positional and bigram probabilities are reported. Table 1 includes detailed information about the study's stimuli.

Independent samples *t*-tests were performed to compare the two linguistic conditions' average scores for each accounted variable. Alpha level was determined to $\alpha = .01$ due to multiple comparisons (e.g., Pituch & Stevens, 2016). Due to unequal variances in the samples, Welch's *t*-tests are reported with fractional degrees of freedom. None of the pre-identified variables were significantly different between the two linguistic conditions (i.e., visual complexity, $t(12.8) = .01$, $p = .994$; image agreement, $t(13.4) = -.32$, $p = .754$; name agreement, $t(8.7) = -1.03$, $p = .331$; imageability, $t(8.7) = -2.57$, $p = .031$; frequency, $t(10.0) = .45$, $p = .660$; neighborhood density, $t(13.6) = .47$, $p = .647$; positional probability, $t(12.4) = .56$, $p = .584$; biphoneme probability, $t(7.7) = -2.98$, $p = .020$).

2.3.2. *Linguistic conditions*

Two different linguistic conditions were developed for the purpose of the present study, one phonological (i.e., phonologically similar stimuli) and one neutral (i.e., phonologically dissimilar stimuli). The phonological condition included images with names that were phonologically maximally similar and semantically maximally dissimilar. Eight line drawings were used in this condition (i.e., black, sack, sock, block, rock, clock, lock, log). The neutral condition included eight pictures with names that were semantically and phonologically maximally dissimilar and comprised the control condition (i.e., tree, lips, cup, fox, doll, nail, bus, star).

2.3.3. *N-back levels*

In order to manipulate working memory load, three different levels of difficulty and working memory load were incorporated, including 1-back, 2-back, and 3-back. Participants were instructed to identify whether each presented image was the same as the image 1-, 2-, or 3- trials back.

2.3.4. *Trial description*

All N-back levels and linguistic conditions included two practice blocks, one containing 12 trials and 3 targets (i.e., a "yes" response was required) and another containing 15 trials with 5 targets, followed by two experimental blocks, each containing 20 targets. The 1-back included two blocks of 60 trials, the 2-back included two blocks of 62 trials, and the 3-back included two blocks of 63 trials each. Therefore, the percentage of trials that included a target was 31 %–32 %, which is consistent with prior literature that suggests about one third of the trials to include targets (e.g., Braver et al., 1997). Each stimulus was used as target in each N-back level 5 times in the experimental trials and one time in the practice trials. The number of times each stimulus was used as a filler item (i.e., non-target) was also accounted for (e.g., each stimulus was used about 30 times as a non-target item in the experimental trials and about 6 times in the practice trials across all N-back levels).

Table 2
Summary of participant pre-testing measures, with means and standard deviations on parenthesis (*p* value of .01 was used to determine significance).

Measure	AWS	AWNS	*p*
Test of Nonverbal Intelligence (TONI)	107.00(9.84)	109.67(10.42)	.48
NIH-Picture Vocabulary Test	117.73(12.78)	115.67(10.89)	.64
NIH-List Sorting Working Memory Test	98.60(10.97)	102.00(12.10)	.43
NIH-Flanker Inhibitory Control Test	108.73(14.71)	101.47(11.54)	.14
NIH-Dimensional Change Card Sorting Test	108.33(9.29)	114.93(13.32)	.13
NIH-Picture Sequence Memory Test	99.73(15.44)	111.20(17.38)	.07
NIH-Oral Reading Recognition Test	121.40(11.01)	121.33(10.12)	.99
Visual Baseline Accuracy (%)	99.33(1.14)	100	.04
Visual Baseline Manual Reaction Time (ms)	335.67(41.72)	225.07(37.99)	.97

2.4. Procedure

Visual stimuli were presented centrally in the screen, on a 20-inch Dell computer monitor, using SuperLab Pro 4.5. All pictures were black line drawings presented on a white background. The dimensions of each image on the screen were approximately 12 cm in width and 12 cm in height and they sustained a maximum visual angle of 10.55° horizontally and a maximum visual angle of 10.40° vertically. Participants sat on a chair and the distance between the participant and the screen was approximately 65 cm. Participants' responses included the push of a button and were recorded via a SuperLab TB-844 response pad. No overt speech response was required for task completion.

During the experimental tasks, participants were required to monitor series of back-to-back presented line drawings and to respond by pressing the "yes" (i.e., right) button each time the most recently presented image was the same as the image *N* trials back with their right index finger. If the current image was not the same as the image *N* trials back, participants were instructed to press the "no" button (i.e., left button) on the response pad with their left index finger. Participants were instructed to respond as quickly as possible.

The stimuli remained on the screen for 1000 ms, based on picture naming latencies of monosyllabic names (e.g., Szekely et al., 2004), in order to allow enough time for participants to access the drawings' names. Participants had an additional 2500 ms white screen with a fixation cross in the center of the screen in order to provide their response by the push of a button, before the next stimulus appeared on the screen. Reaction time was measured from stimulus onset.

2.5. Data preparation and screening

Participant responses were recorded and scored as either correct or incorrect automatically via the SuperLab software. A response was marked as correct if the participant pressed the "yes" button on a target trial (i.e., the current image matched the image *N* trials back) and the "no" button on a nontarget trial (i.e., the current image did not match the image *N* trials back). SuperLab also recorded manual reaction times (mRT) in milliseconds. Analysis was performed in the target responses only (i.e., a "yes" response was required). Therefore, 240 trials were included from each participant in the analysis (120 trials per linguistic condition).

Trials that included mRTs that were less or more than 2 standard deviations from the participant's group mean in each set of conditions were excluded from the analysis (Ratcliff, 1993). Three hundred and fifteen of 7200 total trials (4.4 %) were excluded, 145 from AWS and 170 trials from AWNS. Reaction time latencies that are more or less than 2 standard deviations from the mean in a condition are most likely associated with inattention or "fast guessing", and thus, not reflective of the processes involved in completing the task (Ollman, 1966). Therefore, outlier data were removed from both the accuracy and the reaction time analyses. Regarding reaction time data, another 2021 trials were excluded (AWS: 1062 trials; AWNS: 959 trials) because they were inaccurate, leaving a total of 4844 accurate trials to be analyzed (AWS: 2,384 trials; AWNS: 2460 trials). In summary, 95.4 % of all trials (6865 of 7200) were analyzed for accuracy and 67.3 % of all trials (4844 of 7200) were analyzed for reaction time.

2.6. Data analyses

To analyze the participants' mRT latencies, a mixed-effects ANOVA type model using the Kenward-Roger (KR) approximation for degrees-of-freedom (i.e., fractional degrees of freedom will be reported) using Type III sums of squares was used. To analyze response accuracy, a mixed-effects Likelihood Ratio Test (LRT) using a binomial family was used. A mixed-effects model instead of a traditional

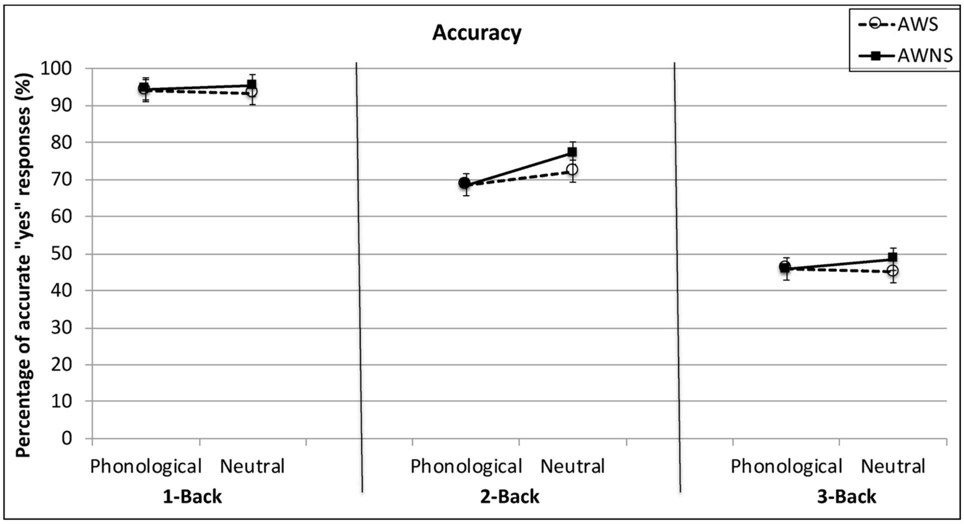


Fig. 1. Accuracy performance for AWS and AWNS per linguistic condition (i.e., phonological vs. neutral) and N-back level (i.e., 1-back, 2-back, 3-back). Error bars represent standard error of the mean.

ANOVA was used to analyze the data due to the flexibility of these models to handle missing data, their ability to capture individual variability in the data, and for allowing for the incorporation of random factors in addition to fixed factors, which are the only factors used in traditional models (for more details on mixed-effects models see Harel & McAllister, 2019). The *afex* package, version 0.27–2 (Singmann et al., 2020) in R Studio, version 3.5.1 (RStudio Team, 2018) was used to run both models via the function "mixed". The package *emmeans*, version 1.4.6 (i.e., Least-Squares Means, Lenth et al., 2020) was used to calculate post hoc pairwise comparisons for significant interactions, with Holm correction applied. Fixed factors in the model (i.e., between-subjects factors) included: Group (i.e., AWS vs. AWNS), Linguistic condition (i.e., phonological vs. neutral), Working memory load (i.e., 1-back vs. 2-back vs. 3-back), and their interactions. Participant was used as a random intercept in the model (i.e., within-subjects factor). The following pre-testing measures were tested as individual covariates and as interaction terms with Group, due to their relevance to N-back performance (e.g., Jaeggi, Buschkuehl, Perrig, & Meier, 2010): a) TONI (intelligence), b) NIH-Flanker (inhibition), c) NIH- List sorting working memory (working memory), and d) manual reaction time (speed). TONI was the only pre-testing measure that turned out significant in both mRT and accuracy; therefore, it was used as a covariate in the following analyses.

3. Results

3.1. Participant accuracy

Both AWS and AWNS were significantly more accurate in 1-back ($M = 94.3$ %, $SE = .6$%) followed by 2-back ($M = 72.8$ %, $SE = 1.8$ %) followed by 3-back ($M = 46.5$ %, $SE = 2.2$ %), as indicated by a significant main effect for Working memory load, $\chi^2(2) = 1390.29$, $p < .001$. A significant main effect of TONI, $\chi^2(1) = 11.45$, $p < .001$, was also observed. Differences were not observed between AWS and AWNS in accuracy, $\chi^2(1) = .40$, $p = .530$ (AWS: $M = 76.2$ %, $SE = 2.1$ %; AWNS: $M = 78.1$ %, $SE = 2.1$ %) and no other main effects or interactions were significant in the model (see Fig. 1 for details regarding participant accuracy results per linguistic condition and working memory load).

3.2. Participant manual reaction time latencies

3.2.1. Participant manual reaction latencies and working memory load

A significant main effect of Working memory load, $F(2,4804.5) = 980.74$, $p < .001$ indicated that participants were faster in 1-back ($M = 465.4$ ms, $SE = 15.4$ ms), followed by 2-back ($M = 599.0$ ms, $SE = 15.5$ ms), followed by 3-back ($M = 650.0$ ms, $SE = 15.7$ ms).

Furthermore, both AWS and AWNS demonstrated similar patterns in their mRT in the three N-back levels, that is, longer mRT in 3-back, followed by 2-back, followed by 1-back. However, different statistical effects were observed within each group, indicative of a significant Group x Working memory load interaction, $F(2,4805.0) = 6.54$, $p = .001$. Post hoc pairwise comparisons demonstrated that both groups demonstrated longer mRT in 2- compared to 1-back (AWS: $z = -25.61$, $p < .001$; AWNS: $z = -21.01$, $p < .001$), in 3- compared to 1-back (AWS: $z = -29.01$, $p < .001$; AWNS: $z = -27.29$, $p < .001$), and in 3- compared to 2-back (AWS: $z = -6.11$, $p < .001$; AWNS: $z = -8.73$, $p < .001$). Based on the z-ratio tests, both AWS and AWNS demonstrated larger statistical effects in 1- versus 3-back, followed by 1- versus 2-back, followed by 2- versus 3-back. Also, AWS demonstrated a larger statistical effect than AWNS in 1- versus 2-back and in 1- versus 3-back while AWNS demonstrated a larger statistical effect than AWS in 2- versus 3-back. In sum, the same trend

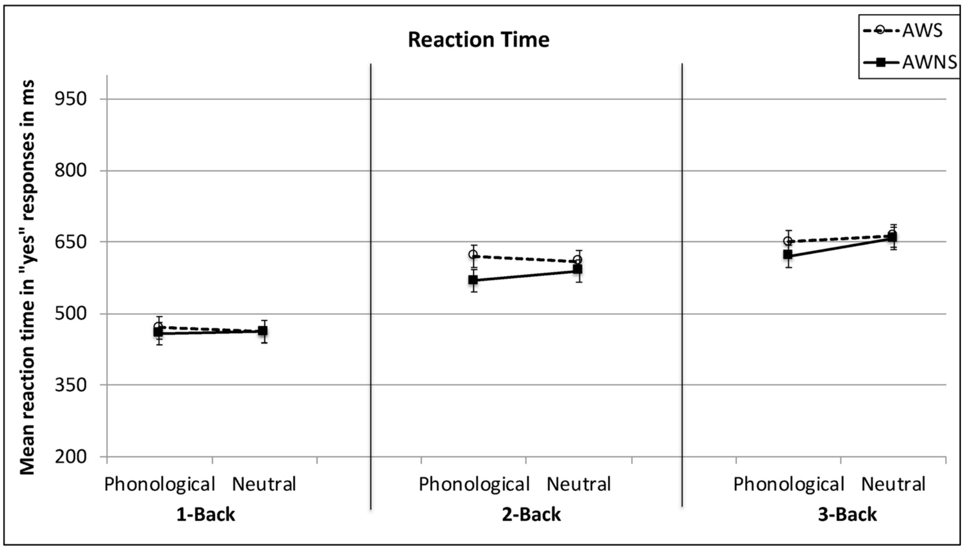


Fig. 2. Mean manual reaction time latencies for AWS and AWNS for the phonological and neutral linguistic conditions in visual input modality. Error bars represent standard error of the mean.

was observed between AWS and AWNS in the three N-back levels (i.e., longer mRT in 3-, followed by 2-, followed by 1-back), but magnitude differences were observed among the N-back level comparisons within each group, representative of the significant Group x Working memory load interaction.

3.2.2. *Participant manual reaction latencies and linguistic condition*

A significant main effect for Linguistic condition, $F(1,4804.5) = 6.30$, $p = .010$, revealed shorter mRT in the phonological ($M = 566.8$ ms, $SE = 15.4$ ms) compared to the neutral linguistic condition ($M = 576.1$ ms, $SE = 15.4$ ms).

In addition, a significant Linguistic condition x Working memory load interaction, $F(2,4804.4) = 4.22$, $p = .010$, and post hoc pairwise comparisons showed that participants were significantly slower in the neutral ($M = 662.5$ ms, $SE = 16.1$ ms) compared to the phonological linguistic condition ($M = 637.6$ ms, $SE = 16.1$ ms) in 3-back, $z = -3.30$, $p = .003$, but not in 1-back, $z = .32$, $p = .919$ (phonological: $M = 466.2$ ms, $SE = 15.7$ ms; neutral: $M = 464.5$ ms, $SE = 15.7$ ms) or 2-back, $z = -.74$, $p = .919$ (phonological: $M = 596.8$ ms, $SE = 15.8$ ms; neutral: $M = 601.3$ ms, $SE = 15.8$ ms). Furthermore, within both linguistic conditions, the differences in mRT latencies between 1- versus 2-back (phonological: $z = -22.57$, $p < .001$; neutral: $z = -24.17$, $p < .001$), 1- versus 3-back (phonological: $z = -26.05$, $p < .001$; neutral: $z = -30.37$, $p < .001$) and 2- versus 3-back (phonological: $z = -5.89$, $p < .001$; neutral: $z = -8.98$, $p < .001$) were all significant. The difference in mRT was larger between 1- versus 3-back, followed by 1- versus 2-back, followed by 2- versus 3-back in both AWS and AWNS. In sum, participants in both groups demonstrated significantly longer mRT in the neutral compared to the phonological linguistic condition in the 3-back level only and they also demonstrated significantly longer mRT in their 3- compared to 2- compared to 1-back levels in both linguistic conditions (see Fig. 2 for details).

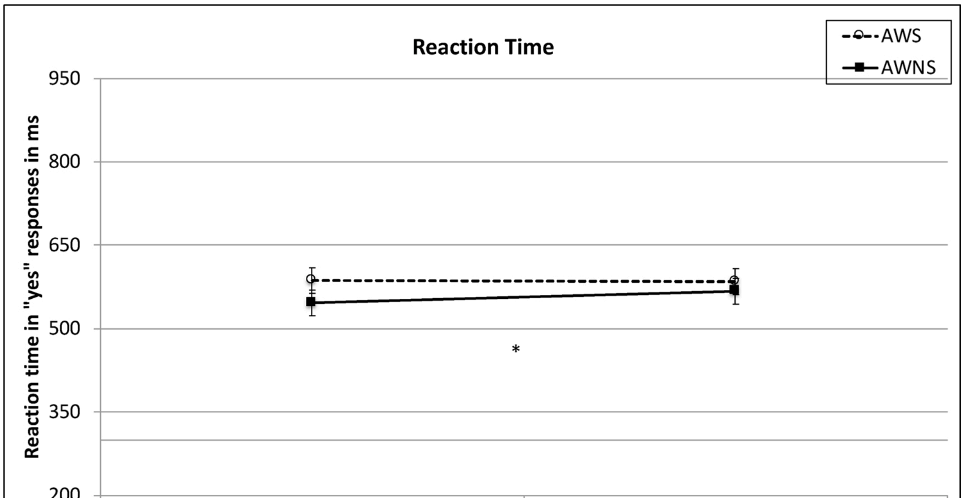
Finally, although no differences were found in mRT latencies between AWS and AWNS in the phonological or the neutral linguistic condition, within group comparisons revealed a difference between the two linguistic conditions for AWNS but not for AWS (i.e., Group x Linguistic condition interaction), $F(1,4804.6) = 9.54$, $p = .002$. Specifically, post hoc pairwise comparisons revealed that AWNS demonstrated significantly shorter mRT in the phonological compared to the neutral linguistic condition, $z = -3.99$, $p < .001$ (phonological: $M = 546.7$ ms, $SE = 21.8$ ms; neutral: $M = 567.4$ ms, $SE = 21.8$ ms), whereas AWS did not, $z = .41$, $p = .684$ (phonological: $M = 586.9$ ms, $SE = 22.0$ ms; neutral: $M = 584.8$ ms, $SE = 22.0$ ms) (see Fig. 3 for details).

3.2.3. *Participant manual reaction latencies and TONI*

A significant main effect was observed for TONI, $F(126.9) = 4.69$, $p = .040$. No other main effects or interactions were significant in the model.

4. Discussion

The purpose of the present study was to investigate visual-verbal working memory in AWS and AWNS using an N-back task. Use of this task permitted exploration of working memory without requiring a spoken response, thereby minimizing production influences on participant performance. Accuracy and manual reaction time (mRT) latencies were analyzed to assess the influence of increasing cognitive demands on working memory performance. To achieve that, we modified the linguistic nature of the stimuli (i.e.,



	Phonological	Neutral
AWS	586.94	584.8
AWNS	546.73	567.41

Fig. 3. Manual reaction time performance for AWS and AWNS per linguistic condition (i.e., phonological vs. neutral) and N-back level (i.e., 1-back, 2-back, and 3-back). Error bars represent standard error of the mean.

phonologically related [phonological linguistic condition] vs. phonologically unrelated stimuli [neutral linguistic condition]) and incorporated three different working memory loads (i.e., 1-back, 2-back, 3-back).

To review, we assumed that participants would engage in both nonverbal/nonlinguistic processing, to visually recognize the images, and verbal/linguistic processing (i.e., involvement of phonological loop and lemma selection) to access the images' names in order to complete the task.

Due to the increasing cognitive demands of the task, it was expected that both AWS and AWNS would demonstrate better performance in accuracy and mRT in 1- compared to 2- compared to 3-back across both linguistic conditions. No differences were expected between AWS and AWNS in 1-back in any linguistic condition. We further hypothesized that both groups would perform worse in terms of accuracy and mRT in the phonological as opposed to neutral linguistic condition, since worse WM performance has been observed with phonologically related versus unrelated stimuli (e.g., Byrd et al., 2015b; Mueller et al., 2003; Sweet et al., 2008; Vallar & Baddeley, 1984). Finally, we predicted that AWS would be less accurate and slower compared to AWNS in the phonological linguistic condition at the 2- and 3-back levels, the most cognitively demanding of the required tasks.

4.1. Participant performance and working memory load

Our first hypothesis was met, as both groups were slower and less accurate in 3- versus 2- versus 1-back levels across both linguistic conditions. Based on the nature of the N-back task, participants were not required to rely on working memory at 1-back, since they only had to match two images presented back-to-back. However, at 2-back and 3-back, participants were required to maintain two and three verbal items in their working memory, update each set of items in their working memory every time a new item was presented, and then provide their response based on matched or nonmatched stimuli. Studies incorporating the N-back task in neurotypical adults (e.g., Jonides et al., 1997) or adults with brain injury (e.g., Kubat-Silman et al., 2002) have also documented decreased performance as the N-back level increased. This is because the 2- and 3-back levels place significantly higher demands on working memory and require additional resources from participants. Specifically, subvocal rehearsal is needed to rehearse and maintain sets of verbal stimuli in working memory. In addition, the central executive (Baddeley, 2003, 2012) is also recruited, because with each new trial, a new item is added in working memory, one item is removed from working memory, and the list of items maintained in working memory needs to be refreshed, sequenced and matched (e.g., Jonides et al., 1997).

4.2. Group performance differences across N-back levels

Our hypothesis of poorer performance in AWS compared to AWNS in the higher N-back levels was not confirmed, as no differences were found between AWS and AWNS in any N-back level. A few reasons may explain why we did not see differences between the two groups in our study whereas several other studies suggest that AWS demonstrate deficits in verbal working memory (e.g., list recall tasks, NWR etc.).

First, incorporation of images instead of letter/words (as used in other studies, e.g., Jonides et al., 1997) allowed participants to engage in both verbal and nonverbal processing of the stimuli (Kelley et al., 1998). Even though our results suggested that AWNS mostly relied on verbal processing to complete the task, based on the phonological priming effect we saw in the study, we cannot be certain about the degree to which participants in each group engaged in verbal and nonverbal processing. The latter, in combination with the relatively small number of participants in each group, which could be inadequate to capture individual processing patterns within AWS and AWNS, might have hindered a between group-difference in the task.

Second, and most important, different tasks may capture different working memory processes. The N-back task includes several processes, some of which are also found in "traditional" working memory tasks, such as encoding, storage, maintenance, updating etc. But, the task also involves processes that are not typically found in other working memory tasks. One of the most prominent differences between the N-back task and other working memory tasks is that it requires information to be recognized instead of recalled (e.g., Pelegrina et al., 2015). That is, participants in an N-back task are required to continuously recognize stimuli and match whether the most recent stimulus is the same as the stimulus *N* trials back. Therefore, it has been suggested that participants must rely on the processes of familiarity and recollection (e.g., Jaseggi et al., 2010; Kane et al., 2007). For example, stimuli that have been recently presented (e.g., 1-back) would have higher levels of activation in working memory and would elicit higher levels of familiarity; thus, they could be recognized based on familiarity. On the other hand, at higher N-back levels, the item *N* trials back may not receive the highest levels of activation and familiarity since additional items intervene between the current item and the item *N* trails back. In the latter case, recollection is required to process any other additional information, such as order of stimulus presentation, to achieve recognition. In addition, inhibition and interference resolution are required in higher N-back levels to resolve conflicts when the current stimulus does not match the stimulus *N* trials back but matches the item N-1 or N-2 trials back (e.g., Jaeggi et al., 2010; Kane et al., 2007).

If recognition and familiarity are two key processes involved in the N-back task, which are not typically involved in more "traditional" measures of working memory that have been used in the past with AWS, we can assume that our results suggest that AWS do not seem to have difficulties with item recognition and familiarity-based interference. Our results do not exclude AWS of having deficits on other processes of working memory, as has been supported in other studies (e.g., subvocal rehearsal). Instead, we suggest that it is possible that only some processes involved in working memory may be different in AWS but not others. This suggestion about selective differences in people who stutter has also been suggested in other executive functions, such as inhibition (e.g., Eggers et al., 2018; Gkalitsiou et al., 2020). For example, Eggers et al. (2018) suggested that there may be different types of inhibitory control, and individuals who stutter may present with deficits only in one type but not in another.

Alternatively, recall the multiple processes that are involved in an N-back task. If AWS present with deficits in only a few of those processes, statistically significant differences between the two groups may be difficult to detect if the study and its stimuli are not specifically designed to disentangle the different processes involved in the task. The primary purpose of our study was not to differentiate among the different processes involved in an N-back task; thus, potential differences between the two groups in a subset of processes would be less likely to emerge.

4.3. Participant performance and linguistic condition

Regarding accuracy, phonological similarity had no effect in participants' responses, with both AWS and AWNS being equally accurate in the phonological and neutral linguistic conditions. Regarding mRT, participants were significantly faster in the phonological compared to neutral linguistic condition in the 3-back level (significant Linguistic condition x Working memory load interaction), an opposite pattern of what was expected. Our hypothesis for poorer performance by both groups in the phonological compared to neutral linguistic condition, was based on the effects phonological similarity has on working memory; that is, performance is worse when items are phonologically similar (e.g., Mueller et al., 2003; Sweet et al., 2008; Vallar and Baddeley, 1984). However, in the present study we did not find such a pattern.

One possible explanation for the difference between our results and those presented in other studies may lie in the nature of the stimuli used in the present study. In our study, we presented nameable line-drawing images, which require both verbal and nonverbal processing (e.g., Kelley et al., 1998), whereas Mueller et al. (2003) and Vallar and Baddeley (1984) used letters of the alphabet, which require verbal processing only (e.g., Braver et al., 1997; Kelley et al., 1998). On a similar note, Byrd et al. (2015b) incorporated auditory verbal stimuli (i.e., words), wherein participants recalled fewer words from phonological lists compared to semantic lists. In our study, it was unclear how much verbal and nonverbal processing each participant engaged into when processing the stimuli. Therefore, the combination of verbal and nonverbal processing involved in our stimuli, which could differ among participants, might have prevented the phonological similarity effect we initially anticipated to emerge.

The nature of the stimuli included in the present study might have also influenced the lexical access demands required in the study. When participants viewed the nameable line-drawings, they would access the images' names and they would rehearse the images' names in order to complete the task. In that case, the present N-back task would resemble a silent picture naming task. Evidence suggests a phonological faciliatory effect during picture naming, with shorter picture naming latencies when phonologically related distractor stimuli (i.e., primes) are available (e.g., Jescheniak & Schriefers, 2001; Vitevitch, 2002). The premise is that the word forms of the primes, which are also phonologically related to those of the targets, facilitate the activation of the targets' word forms due to their shared phonological segments (e.g., Vitevitch, 2002). In our study, every image name would serve as a phonological prime for the upcoming stimulus; thus, participants would access the stimuli's names faster in the phonological compared to the neutral condition, leading to the shorter mRTs in the phonological compared to the neutral condition that were observed in the present study.

4.4. Group performance differences across linguistic conditions

Our hypothesis of poorer performance in AWS compared to AWNS in the phonological linguistic condition was not confirmed, as any direct comparison between AWS and AWNS was not significant in the study. However, differences on how AWS and AWNS approached the task arose via a significant Group x Linguistic condition interaction, with differences evolving between the two conditions within each group. That is, AWNS were significantly faster in the phonological compared to the neutral condition, whereas the mRT of AWS did not differ between the two conditions.

The lack of between-group differences in the phonological condition may be related to the nameable line-drawings we employed in the present study. Recall that participants would engage in both verbal and nonverbal processing of the line drawings. Since we cannot be certain as to the extent to which each type of processing was implemented by each participant in each group, we could assume that the varying degrees of verbal processing and lexical access the participants engaged into might have influenced their access to the phonological store and subvocal rehearsal (Baddeley, 2003). The latter, in combination to the relatively small sample size of the study, make it challenging for potential between-group differences to emerge.

However, a difference was observed within the two groups' mRT latency patterns. Specifically, phonological similarity did not influence AWS' mRT latencies and had a facilitation effect in AWNS. Specifically, AWNS demonstrated shorter mRT latencies in the phonological compared to the neutral condition. Thus, given the mRT differences between the two linguistic conditions in AWNS, we can presume that AWNS accessed and rehearsed the images' names during the 2-and 3-back trials. Therefore, a phonological priming effect was observed in AWNS, indicative of the faster mRTs in the phonological compared to the neutral condition.

The lack of a phonological facilitation effect in AWS may suggest that AWS did not demonstrate a phonological priming effect or their priming effect was very weak to reach significance. This account can be supported by the reported difficulties of AWS with lexical access, phonological encoding (e.g., Byrd et al., 2015a; McGill et al., 2016; Sasisekaran & Weisberg, 2014; Weber-Fox et al., 2004) and weaker spreading activation abilities (e.g., Byrd et al., 2015b; Maxfield et al., 2012),

For example, several studies have supported differences in phonological priming between AWS and AWNS (e.g., Maxfield et al., 2012; Maxfield et al., 2015; Wijnen & Boers, 1994; cf. Hennessey et al., 2008), which suggest that, compared to AWNS, AWS may require more overlapping phonological information between the prime and the target in order to demonstrate a phonological priming effect (e.g., Wijnen and Boers, 1994). In the present study, we used a combination of overlapping onsets and rimes in the phonological condition; therefore, the degree of overlapping phonological information varied among the stimuli. This might have prevented AWS from demonstrating a priming effect or their priming effect was too weak to reach significance.

Alternatively, unstable semantic and phonological representations and weaker lemma/word form connections compared to AWNS have also been suggested as factors influencing phonological priming in AWS (e.g., Maxfield et al., 2012; Maxfield et al., 2015). Taken together, the reported difficulties observed in AWS when processing phonological information might have withheld them from demonstrating a phonological priming effect.

On a final note, we also need to consider that our relatively small size of 15 participants per group, even though adequate to capture 3-way interactions based on our SIMR analyses, may not be large enough to possibly capture all the individual variability that might have been employed by the participants in the task.

4.5. Limitations and future directions

One potential limitation in the present study includes the relatively restricted diversity of stimuli (i.e., 8 items per linguistic condition) and the multiple repetitions of each stimulus in the task (i.e., each image was used approximately 30 times). Both factors have been linked to better memory performance (e.g., Roelofs, 2001). That is, a small number of items is easier to be kept in short-term memory. Also, using the same set of stimuli multiple times during a task can improve establishing this set of stimuli in memory, and subsequently improve performance. Nonetheless, the number of stimuli used in this study is comparable to that of other N-back studies completed in healthy adults (e.g., McEvoy et al., 1998; n = 12 capital letters) or in clinical populations (e.g., aphasia, Wright et al., 2007; n = 5 stimuli per linguistic condition), suggesting that effects can be observed even with a smaller set of stimuli.

Another limitation of the study was that hand response for target and no-target trails was not counterbalanced. Even though we only included right-handed participants in the study, a response hand bias might have been introduced in the study that could potentially influence participant performance,

Additionally, even though a power analysis was implemented in order to determine a satisfactory sample size to detect potential performance differences between the two groups, ideally, a larger participant pool would allow for a better representation of participant individual variability. Individual variability in participant mRT data and the varying degree of nonverbal and verbal processing each participant engaged into could have influenced the results. However, it needs to be noted that our sample size of 30 participants (i.e., 15 per group) is a common sample size used in other behavioral experimental paradigms investigating phonological encoding and phonological working memory in AWS (e.g., McGill et al., 2016, *n* = 13 participants per group; Weber-Fox et al., 2004, *n* = 11 participants per group), as well as in similar studies that have employed an N-back task with other clinical and typical populations (e.g., Braver et al., 2001, n = 28 healthy young adults; Wright et al., 2007, n = 9 adults with aphasia). Nevertheless, a larger sample size would better allow us to capture individual variability within each group.

Another variable that might have also affected participant performance in our task is the participants' bilingual status. Even though all participants in the study reported native American English proficiency and English was their dominant language, the fact that almost half of AWS and AWNS also reported speaking languages other than English might have contributed to participant performance variability. Teubner-Rhodes et al. (2016) found that bilingual speakers outperformed monolingual speakers in a 3-back task, when printed words were visually presented to participants. Bilingual status appeared to influence performance only in the high-conflict condition (i.e., nontarget trials included lures that were matches for N-2, N-4 and N-5 back) but not in the low-conflict condition (i.e., nontarget trials could not be a possible match for any other N-back level). It needs to be noted though that there are also studies that either find a very small effect of a bilingual advantage or they fail to find an advantage when different statistical analyses are performed (e.g., Bayesian analyses) and when potential biases, such as SES or educational level, are being addressed (e.g., Lukasik et al., 2018). In the present study, both groups included a similar number of bilingual speakers (i.e., about half of the participants in each group), whose proficiency levels were also equally distributed among the two groups and their educational level was matched in the two groups. Therefore, if bilingualism would influence participant performance, we would expect it to have the same effect in both groups. Future studies should incorporate various types of N-back tasks (e.g., verbal, spatial) and include monolingual and bilingual participants of varying levels of proficiency as distinct groups in order to obtain a clearer picture of how bilingual status may affect performance in an N-back task.

Finally, in the present study we analyzed only target responses (i.e., "yes" responses). Even though analyzing only target responses in an N-back task is common, future studies should also include lures in the non-target trials (i.e., matches on a different N-back level than the one currently assessed) and analyze both target and non-target trials in order to obtain a more comprehensive picture of the participants' processing strategies in the task.

5. Conclusions

The present study investigated visual-verbal working memory in AWS and AWNS via an N-back task. Results of the study indicated that both groups were more accurate and significantly faster in 1-back, followed by 2-back, followed by 3-back. Even though statistically significant differences were not found when directly comparing AWS to AWNS, our preliminary results indicate different processing patterns between AWS and AWNS in the task. Specifically, a priming effect for phonologically similar stimuli was observed in AWNS but not in AWS. That is, AWNS demonstrated faster mRT in the phonological compared to neutral condition, suggesting that their access of the images' names and phonological similarity among the stimuli in the phonological condition facilitated their performance in that condition. Our results did not reveal a phonological facilitation effect in AWS, suggesting that AWS did not demonstrate a phonological priming effect possibly due to their reported difficulties in lexical access and/or phonological encoding. Finally, the lack of differences between AWS and AWNS in any N-back level does not support deficits in AWS in processes involved in the N-back task, such as recognition and/or familiarity, suggesting that individuals who stutter may have difficulties with only some processes involved in working memory but not with others. However, these results are preliminary and considering the multiple different processes involved in an N-back task, additional research is warranted to elucidate the relationship between specific aspects of working memory and the speech breakdowns unique to stuttering.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank the adults who do and do not stutter, who were willing to give their time to participate in this study and help us further our knowledge of the underlying nature of stuttering. We would also like to acknowledge the participant compensation funding from the Michael and Tami Lang Stuttering Institute Endowment awarded to the second author.

References

Alario, F., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods Instruments & Computers, 36*, 140–155.
Alm, P. A. (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders, 37*, 325–369.
American Speech-Language and Hearing Association. (1997). *). Guidelines for audiologic screening [Guidelines]*. Available from http://www.asha.org/policy.
Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews, 4*, 829–839.
Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.
Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of episodic buffer. *Neuropsychologia, 49*, 1393–1400.
Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.
Bajaj, A. (2007). Working memory involvement in stuttering: Exploring the evidence and research implications. *Journal of Fluency Disorders, 32*, 218–238.
Balota, D., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.
Bosshardt, H. (1990). Subvocalization and reading rate differences between stuttering and nonstuttering children and adults. *Journal of Speech and Hearing Research, 33*, 776–785.
Bosshardt, H. (2002). Effects of concurrent cognitive processing on the fluency of word repetition: Comparison between persons who do and do not stutter. *Journal of Fluency Disorders, 27*, 93–114.
Bosshardt, H., Ballmer, W., & De Nill, L. (2002). Effects of category and rhyme decisions on sentence production. *Journal of Speech Language and Hearing Research, 45*, 844–857.
Braver, T., Barch, D., Kelley, W., Buckner, R., Cohen, N., Miezin, F., Snyder, A., Ollinger, J., Akbudak, E., Conturo, T., & Petersen, S. (2001). Direct comparison of prefrontal cortex regions engaged by working memory and long-term memory tasks. *NeuroImage, 14*, 48–59.
Braver, T., Cohen, J., Nystrom, L., Jonides, J., Smith, E., & Noll, D. (1997). Parametric study in prefrontal cortex involvement in human working memory. *Neuroimage, 5*, 49–62.
Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of nonverbal intelligence* (4th ed.). Austin, TX: PRO-ED.
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.
Byrd, C. T., Vallely, M., Anderson, J. D., & Sussman, H. (2012). Nonword repetition and phoneme elision in adults who do and do not stutter. *Journal of Fluency Disorders, 37*, 188–201.
Byrd, C. T., McGill, M., & Usler, E. (2015a). Nonword repetition and phoneme elision in adults who do and do not stutter: Vocal versus nonvocal performance differences. *Journal of Fluency Disorders, 44*, 17–31.
Byrd, C. T., Sheng, L., Berstein Ratner, N., & Gkalitsiou, Z. (2015b). Veridical and false recall in adults who stutter. *Journal of Speech Language and Hearing Research, 58*, 28–42.
Coalson, G., & Byrd, C. T. (2015). Metrical encoding in adults who do and do not stutter. *Journal of Speech Language and Hearing Research, 58*, 601–621.
Cortese, M., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods Instruments & Computers, 36*, 384–387.
Dragovic, M. (2004). Towards an improved measure of the Edinburgh handednessinventory: A one-factor congeneric measurement model using confirmatory factor analysis. *Laterality, 9*, 411–419.
Drayna, D., & Kang, C. (2011). Genetic approaches to understanding the causes of stuttering. *Journal of Neurodevelopmental Disorders, 3*, 374–380.
Eggers, K., De Nil, L. F., & Van den Bergh, B. R. H. (2018). Exogenously triggered response inhibition in developmental stuttering. *Journal of Fluency Disorders, 56*, 33–44.
Eichorn, N., Marton, K., Schwartz, R. G., Melara, R. D., & Pirutinsky, S. (2016). Does working memory enhance or interfere with speech fluency in adults who do and do not stutter? Evidence from a dual-task paradigm. *Journal of Speech Language and Hearing Research, 59*, 415–429.
Engle, R., W., & Kane, M., J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *In B.H. Ross (Ed.) The psychology of learning and motivation: Advances in research and theory* (pp. 145–149). Elsevier Science.
Giraud, A., Neumann, K., Bachoud-Levi, A., von Grudenberg, A. W., Euler, H. A., Lanfermann, H., & Preibisch, C. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain and Language, 104*, 190–199.
Gkalitsiou, Z., Byrd, C. T., & Griffin, Z. (2020). Executive control in adults who stutter: The antisaccade task. *Journal of Speech Language and Hearing Research, 63*, 1688–1699.
Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*, 493–498.
Harel, D., & McAllister, T. (2019). Multilevel models for communication sciences and disorders. *Journal of Speech Language and Hearing Research, 62*, 783–801.
Hennessey, N. W., Nang, C. Y., & Beilby, J. M. (2008). Speeded verbal responding in adults who stutter: Are there deficits in linguistic encoding? *Journal of Fluency Disorders, 33*, 180–202.
Indefrey, P., & Levelt, W. (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed, pp. 845–865). Cambridge, MA: MIT Press.
Jaeggi, S., Buschkuehl, M., Perrig, W., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory, 18*(4), 394–412.
Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture-word interference. *Quarterly Journal of Experimental Psychology: Human Experimental psychology, 54*, 371–382.
Jones, R. M., Conture, E. G., & Walden, T. A. (2014). Emotional reactivity and regulation associated with fluent and stuttered utterances of preschool-age children who stutter. *Journal of Communication Disorders, 48*, 38–51.
Jongman, S. R., Roelofs, A., & Lewis, A. G. (2019). Attention for speaking: Prestimulus motor-cortical alpha power predicts picture naming latencies. *Journal of Cognitive Neuroscience, 32*, 747–761.

Jonides, J., Schumacher, E., Smith, E., Lauber, E., Awh, E., Minoshima, S., & Koeppe, R. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9*(4), 462–475.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology Learning, Memory, and Cognition, 33*, 615–622.

Karrass, J., Walden, T. A., Conture, E. G., Graham, C. G., Arnold, H. S., Hartfield, K. N., & Schwenk, K. A. (2006). Relation of emotional reactivity and regulation to childhood stuttering. *Journal of Communication Disorders, 39*, 402–423.

Kelley, W. M., Miezin, F. M., McDermott, K. B., Buckner, R. L., Raichle, M. E., Cohen, N. J., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., & Petersen, S. E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron, 20*, 927–936.

Kubat-Silman, A., Dagenbach, D., & Absner, J. (2002). Patterns of impaired verbal, spatial, and object working memory after thalamic lesions. *Brain and Cognition, 50*, 178–193.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *Estimated marginal means, aka Least-Squares Means, R package version 1.4.6*. https://cran.r-project.org/web/packages/emmeans/index.html.

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

Lukasik, K. M., Lehtonen, M., Soveri, A., Waris, O., Jylkka, J., & Laine, M. (2018). Bilingualism and working memory performance: Evidence from a large scale online study. *PloS One, 13*(11), Article e0205916.

Max, L., Guenther, F., Gracco, V., Ghosh, S., & Wallace, M. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Sciences and Disorders, 31*, 105–122.

Maxfield, N. D., Pizon-Moore, A. A., Frisch, S. A., & Constantine, J. L. (2012). Exploring semantic and phononological picture-word priming in adults who stutter using event-related-potentials. *Clinical Neurophysiology, 123*, 1131–1146.

Maxfield, N. D., Morris, K., Frisch, S. A., Morphew, K., & Constantine, J. L. (2015). Real-time processing in picture naming in adults who stutter: ERP evidence. *Clinical Neurophysiology, 126*, 284–296.

Maxfield, N., Olsen, W., Kleinman, D., Frisch, S., Ferreira, V., & Lister, J. (2016). Attention demands of language production in adults who stutter. *Clinical Neurophysiology, 127*, 1942–1960.

McEvoy, L. K., Smith, M. E., & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: Effects of memory load and task practice. *Cerebral Cortex, 8*(563-574), 1047–3211.

McGill, M., Sussman, H., & Byrd, C. T. (2016). From grapheme to phonological output: Performance of adults who stutter on a word jumble task. *PloS One, 11*(3). e0151107.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological working complexity in verbal working memory. *Journal of Experimental Psychology Learning, Memory, and Cognition, 29*, 1353–1380.

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

Nippold, M. A. (2002). Stuttering and phonology: Is there an interaction? *American Journal of Speech Language Pathology, 11*, 99–110.

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science, 6*, 155–156.

Papagno, C. (2002). Progressive impairment of constructional abilities: A visuospatial sketchpad deficit? *Neuropsychologia, 40*, 1858–1867.

Pelegrina, S., Lechuga, M. T., Garcia-Madruga, J. A., Elosua, M. R., Macizo, P., Carreiras, M., Fuentes, L. J., & Bajo, T. (2015). Normative data on the n-back task for children and young adolescents. *Frontiers in Psychology, 6*, 1544.

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica, 142*, 332–342.

Pituch, K., & Stevens, J. (2016). *Applied multivariate statistics for the behavioral sciences. Analyses with SAS and IBM's SPSS* (6th ed.). New York, NY: Routledge.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.

Repovs, G., & Baddeley, A. (2006). The multicomponent model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139*, 5–26.

Riley, G. D. (2009). *Stuttering severity instrument* (4th ed). Austin, TX: PRO-ED.

Roelofs, A. (2001). Set size and repetition matter: Comment on Caramazza and Costa (2000). *Cognition, 80*, 283–290.

RStudio Team. (2018). *RStudio: Integrated developmental for r*. Boston, MA: RStudio, Inc.. https://rstudio.com/

Sasisekaran, J. (2013). Nonword repetition and nonword reading abilities in adults who do and do not stutter. *Journal of Fluency Disorders, 38*, 275–289.

Sasisekaran, J., & Weisberg, S. (2014). Practice and retention of nonwords in adults who stutter. *Journal of Fluency Disorders, 41*, 55–71.

Sasisekaran, J., De Nil, L. F., Smyth, R., & Johnson, C. (2006). Phonological encoding in the silent speech of persons who stutter. *Journal of Fluency Disorders, 31*, 1–21.

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

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

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., Fox, J., Lawrence, M., Mertens, U., Love, J., Lenth, R., & Christensen, R. H. B. (2020). *afex: Analysis of factorial experiments, R package version 0.27-2*. https://cran.r-project.org/web/packages/afex/index.html.

Smith, A., & Weber, C. (2017). How stuttering develops: The multifactorial dynamic pathways theory. *Journal of Speech Language and Hearing Research, 60*(9), 2483–2505.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name Agreement. Image Agreement, Familiarity and Visual Complexity. *Journal of Experimental Psychology Human Learning and Memory, 6*, 174–215.

Sweet, L. H., Paskavitz, J. F., Hasley, A. P., Gunstad, J. J., Mulligan, R. C., Nyalakanti, P. K., & Cohen, R. A. (2008). Imaging phonological similarity effects on verbal working memory. *Neuropsychologia, 46*, 1114–1123. et al., 2008.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikoova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., … Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51*, 247–250.

Teubner-Rhodes, S. E., Mishler, A., Corbett, R., Andreu, L., Sanz-Torrent, M., Trueswell, J. C., & Novick, J. M. (2016). The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery. *Cognition, 150*, 213–231.

Vaden, K., Halpin, H., & Hickok, G. (2009). *Irvine phonotactic online dictionary, version 2.0.*. [Data file]. Available from http://iphod.com.

Vallar, G., & Baddeley, A. (1984). Fractionation of working memory: Neurophysiological evidence for a phonological short-term store. *Journal of Verbal Learning and Behavior, 23*, 151–161.

Vallar, G., & Papagno, C. (2002). Neuropsychological impairments if verbal short-term memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (pp. 249–270). Chickester, UK: Wiley.

Vitevitch, M. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology Learning, Memory, and Cognition, 28*, 735–747.

Warrington, E., Logue, V., & Pratt, R. (1971). Anatomical localisation of selective impairment of auditory verbal short-term memory. *Neuropsychologia, 9*, 377–387.

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain, 131*, 50–59.

Weber-Fox, C., Spencer, R. M. C., Spruill, J. E., & Smith, A. (2004). Phonological processing in adults who stutter: Electrophysiological and behavioral evidence. *Journal of Speech Language and Hearing Research, 47*, 1244–1258.

Wijnen, F., & Boers, I. (1994). Phonological priming effects in stutterers. *Journal of Fluency Disorders, 19*, 1–20.

Wright, H., Downey, R., Gravier, M., Love, T., & Shapiro, L. (2007). Processing distinct linguistic information types in working memory in aphasia. *Aphasiology, 21*, 802–813.

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

Zoi Gkalitsiou is an Assistant Professor in the Department of Communication Sciences and Disorders at the University of Texas at Austin and the Director of the Stuttering and Linguistic Processing lab at UT. Her primary research interests include the contribution of cognitive and linguistic factors to the development and persistence of stuttering and the manifestation of stuttering in bilinguals.

Courtney T. Byrd is a Professor and Associate Chair in the Department of Speech, Language and Hearing Sciences at The University of Texas at Austin and Founding Director of the Michael and Tami Lang Stuttering Institute. Her research interests include the study of speech-language contributions to childhood stuttering and the development of innovative treatment and clinical training tools.