Journal of Fluency Disorders 49 (2016) 1–12

Contents lists available at ScienceDirect

Journal of Fluency Disorders

Linguistic and emotional-valence characteristics of reading passages for clinical use and research

Boaz M. Ben-David[a,b,c,*], Maroof I. Moral[d,e], Aravind K. Namasivayam[b], Hadas Erel[a,e], Pascal H.H.M. van Lieshout[b,c,d,f,g]

[a] Communication, Aging and Neuropsychology Lab (CANlab), Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC) Herzliya, Herzliya, Israel
[b] Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, Toronto, Ontario, Canada
[c] Toronto Rehabilitation Institute, Toronto, Ontario, Canada
[d] Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada
[e] Faculty of Information, University of Toronto, Toronto, Ontario, Canada
[f] Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada
[g] Rehabilitation Sciences Institute, University of Toronto, Toronto, Ontario, Canada

ARTICLE INFO

Article history:
Received 13 June 2016
Received in revised form 27 June 2016
Accepted 29 June 2016
Available online 30 June 2016

Keywords:
Reading passage
Fluency disorders
Emotion
Threat
Arousal
Lexical characteristics
Readability
Phonetic properties
Word familiarity

1. Introduction

Fluency assessments in people who stutter (PWS) typically rely on reading aloud a text and/or producing spontaneous speech (e.g., Riley, 2008). Reading materials used for fluency assessment are as ubiquitous as they are varied in the stuttering literature, ranging from standard reading passages (e.g. Rainbow passage, Fairbanks, 1960), Newsweek magazine pieces, junior high level texts to scientific papers. Notably, even the American Speech and Hearing Association's (ASHA) presidential address has been utilized (Adams & Hutchinson, 1974; Armson & Stuart, 1998; Duchin & Mysak, 1987; Freeman & Armson, 1998; Hedge, 1982).

This wide assortment of reading materials enriches the assessment process for gauging fluency. However, these materials may differ significantly in the cognitive, emotional, linguistic and speech-motor demands that they place upon the speaker,

* Corresponding author at: Communication, Aging and Neuropsychology Lab (CANlab), Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC) Herzliya, Herzliya, Israel.
E-mail address: boaz.ben.david@idc.ac.il (B.M. Ben-David).

http://dx.doi.org/10.1016/j.jfludis.2016.06.003

as well as the adequacy of the phonetic sample (Powell, 2006). Indeed, in a systematic factor analysis of 15 reading passages for speech assessment, Powell (2006) found differences in three main dimensions that relate to the readability of the words (lexical complexity), sentences (structural complexity) and the passages as a whole (passage length). Powell (2006) concluded that one must consider the adequacy of the passage for a given clinical population (that may differ in terms of literacy levels, cognitive capacity, production effort/stamina), given the substantial differences in these three dimensions. An examination of the literature reveals that these dimensions have been related to disfluencies. For example, the level of reading difficulty of a passage was associated with the frequency of disfluencies in children (for children who stutter, Blood and Hood, 1978; as well as for children who do not stutter, Cecconi, Hood, & Tucker, 1977). Word familiarity, frequency and length have also been directly related to stuttering, with higher disfluencies noted for unfamiliar and longer utterances for people who do not stutter (e.g., Hubbard and Prins, 1994). Readability and linguistic complexity are not the only factors that may affect reading fluency. In this context, it is important to consider that reading aloud can become an emotional experience related to the content of the words, specifically with respect to their valence and arousal (see Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Dresler, Mériau, Heekeren & van der Meer, 2009). Negative high-arousal words have been found to slow-down reading, as compared to neutral low-arousal words (see, Algom, Chajut & Lev, 2004; Arntz, Appels & Sieswerda, 2000). These words were also found to disrupt semantic processing (lexical decision task, MacLeod, Tata, & Mathews, 1987), motor-related responses (Chajut, Mama, Levy & Algom, 2010) and even sensory processes (Ben-David, Chajut & Algom, 2012) in comparison to neutral words. Anxiety disorders generally amplify these effects, inflating the attentional bias for such words (Amir, Freshman, & Foa, 2002; Andersson, Westöö, Johansson, & Carlbring, 2006; Chajut, Schupak, & Algom, 2010; Öhman, Flykt, & Esteves, 2001). This notion is of special importance as anxiety has been mentioned as being more prevalent among PWS (Blood, Blood, Bennett, Simpson, & Susman, 1994; Craig, 1990; Fitzgerald, Djurdjic, & Maguinet, 1992; Iverach, Menzies, O'Brian, Packman, & Onslow, 2011; Iverach and Rapee, 2014). PWS typically score higher on both trait anxiety (task-independent anxiety; Alm and Risberg, 2007; Blumgart, Tran, & Craig, 2014; Craig, 1990; Craig & Tran, 2014; Manning and Beck, 2013), state anxiety including task-related anxiety (Blumgart et al., 2014; Davis, Shisca, & Howell, 2007; Lovett, 1988) and social-related anxiety (fear of humiliation and negative assessment in performance based situations; Craig and Tran, 2014; Iverach & Rapee, 2014; Manning and Beck, 2013). How anxiety relates to stuttering has been debated in the literature (Ezrati-Vinacour and Levin, 2004; Menzies, Onslow, and Packman, 1999), with early proponents suggesting a causal role (Sheehan, 1970; Wischner, 1952), but over time, the perspective has changed from anxiety being a mediating factor (Brutten and Shoemaker, 1967; Gregory, 1991; van Riper, 1973) to the more recent view that it is a consequence of being a person who stutters (Craig, 1990; Ezrati-Vinacour & Levin, 2004; Kraaimaat, Vanryckeghem, & Van Dam-Baggen, 2002; Perkins, 1979; Ryan, 1974).

Iverach and Rapee (2014) in their review of the literature stress the need to account for anxiety influences in a fluency assessment procedure. Such a strategy is supported by indications in the literature that word specific threat-related effects are inflated for people with anxiety (Phaf and Kan, 2007; Watts, McKenna, Sharrock, & Trezise, 1986; Williams, Mathews, & MacLeod, 1996; Yiend, 2010) and clinical populations in general (Williams et al., 1996). This highlights that the relation between anxiety and the influence of threat specific words on reading performance is not specific to PWS. In fact, one of the most common tests for performance under threat involves printed threat-related words, also known as the Emotional Stroop paradigm (Algom et al., 2004; Williams et al., 1996). In this paradigm, printed threat and neutral words (e.g., FAILURE and FURNITURE) are presented to participants, and they are asked to respond by either naming their font color, deciding if the string of letters forms a word, or read them out loud (see, Algom et al., 2004). Typically, responses to threat words are slower than to neutral words, and anxiety has been found to inflate this threat effect for color naming as well as for reading aloud (Asmundson and Stein, 1994; Bar-Haim et al., 2007; Rutherford, MacLeod, & Campbell, 2004; Williams et al., 1996; Yiend, 2010).

Several theories have been suggested to explain the threatening effect of the content of words. According to the Threat Theory, the menacing content of the words activates a general-purpose defense mechanism that responds to threat by momentarily freezing all activity that is not directly related to the threat (Algom et al., 2004; Fox, Russo, Bowles, & Dutton, 2001), even at the perceptual level (Ben-David et al., 2012). In the presence of semantic threat-words, the system prioritizes resources for efficient action (Öhman et al., 2001), akin to automatic vigilance (Larsen, Mercer, & Balota, 2006), leaving less resources available for the task at hand (e.g., reading the word aloud). Another theory argues that it is the additional attention that threat words grab that induces this cognitive bias to threat (e.g., Williams et al., 1996). Alternatively, the effect was suggested to be the outcome of both a distraction generated by the threat word and a freeze effect induced by its content (Frings, Englert, Wentura, & Bermeitinger, 2010).

As stated earlier, there is evidence to suggest a close relation between stuttering and anxiety, implying that PWS might show a possible interference when responding to threatening words. Two recent studies have indeed demonstrated reduced performance in the face of threat words with PWS using the emotional Stroop paradigm. Hennessey et al. (Hennessey, Dourado, & Beilby, 2014) found that PWS were slower to respond to threat (than to neutral) words when verbal responding was required. Similar results were obtained by van Lieshout, Ben-David, Lipski, & Namasivayam (2014), where threat words influenced not only latency but also speech production mechanisms in PWS. Taken together, these two studies suggest that automatic responses to the threatening content of printed words (likely exacerbated by the presence of anxiety) tap into the same resource pool (capacity) as the speech motor system (Alm and Risberg, 2007; Craig & Tran, 2014; Neilson and Neilson, 1987). Hence, it may interfere with the ability of PWS to control their speech motor system (Alm and Risberg, 2007; Craig & Tran, 2014; Neilson and Neilson, 1987), given their limitations in speech motor skills (Namasivayam and van

Lieshout, 2011; van Lieshout, Hulstijn, & Peters, 2004). Thus, Hennessey et al. (2014) conclude that "...PWS who showed larger slowing of color naming of threat words tend to have less capacity to maintain fluency when speaking. Therefore, the emotionality effect appears to be related to a key characteristic of the speech motor system of PWS" (p. 53). The way emotion and other factors (linguistic, motor) can potentially impact on speech-motor control is detailed in the Speech Motor Skills theory (Namasivayam and van Lieshout, 2011; van Lieshout et al., 2004). In this theory, it is argued that PWS are at the low end of the speech motor skill continuum and this makes them less efficient and less flexible in handling changes in motor control parameters, which could destabilize motor output. To counteract such destabilizing influences, PWS may resort to specific motor control strategies such as increasing movement amplitude or slowing down movements. Evidence in support of these claims, in the context of semantic threat content, was provided in a recent study (van Lieshout et al., 2014). The authors showed a group difference for especially upper lip movements between control speakers and PWS when naming the color of threat stimuli. Given the predisposition of PWS towards higher levels of anxiety, the use of emotional content words should be considered during the assessment process (Guitar, 2013) as it could bias the severity levels due to an increased likelihood of speech motor instability when encountering such words (Hennessey et al., 2014; van Lieshout et al., 2014). Being aware of this additional factor can enable a better understanding of the measurement outcomes as well as choosing the most appropriate tool for the purpose of the assessment. For example, one may chose a more neutral passage for assessing stuttering severity to avoid the impact of pre-existing anxiety. Alternatively, one may choose a more emotional passage to elicit stuttering in a case where the client does not demonstrate any stuttering during assessment but says that he/she does stutter in daily situations, where anxiety may play a stronger role (Guitar, 2013).

In sum, the emotional valence of printed words can affect reading and speech performance, especially for populations where anxiety may be a significant factor. As stuttering assessment protocols involve reading aloud printed text, the valence of the words in these texts might bias performance (and severity scores), given the prevalence of higher trait anxiety and speech-related state anxiety in PWS (Hennessey et al., 2014; van Lieshout et al., 2014).

1.1. The current study

An overview of the literature on stuttering assessments (see Baken and Orlikoff, 2000) shows that the three most widely used passages are the "Rainbow" (Fairbanks, 1960), "The North Wind and the Sun" (IPA, 1999), and the "Grandfather" (Darley, Aronson, & Brown, 1975); see Appendix A for details. The benefit of having compatible passages allows for easier comparison of speech features across individuals and/or at different points in time. However, to date, there are only limited data available on factors like readability and phonetic features of these "traditional" passages that might influence reading performance, irrespective of stuttering status (Powell, 2006; Deterding, 2006; Patel et al., 2013), and there is no information on the semantic content of these passages. The current study aims to rectify this. We provide detailed content, linguistic, readability and phonetic information on the three traditional passages commonly used for assessment of speech function in PWS and other populations. This type of information is useful for comparing performance on the different passages, as well as for evaluating speech assessments provided by each passage as a function of its unique features. A secondary goal was to compare the existing passages with a novel passage, titled "My First Day." The new passage was designed to minimize variables that may influence reading in general and to minimize the potential impact of threat words on the speech of PWS. These variables include: low familiarity and frequency of words, high arousal and valence, word length, the structure of the sentences, word predictability, passage readability, as well as certain phonetic properties (Howell, Au-Yeung, Yaruss, & Eldridge, 2006; Jayaram, 1983; Kuchinke, Vo, Hofmann, & Jacobs, 2007; Mahesh and Geetha, 2013; Smith, Sadagopan, Walsh, & Weber-Fox, 2010; Strijkers, Costa, & Thierry, 2009). Apart from quantifying these properties, we also conducted a study to compare the stuttering frequency for one of the standard passages (Rainbow) with the newly developed passage. We hope that the information presented here can assist researchers and clinicians in making informed decisions about the usage of these passages, and provide possible explanations for variability in reading performance for PWS and non-stuttering individuals, when using these (and potentially other) passages.

2. Experiment 1

In this experiment, we compared the linguistic and semantic characteristics of the words across the three traditional passages. Participants were asked to rate the arousal, valence and familiarity of 714 content words (that serve more than just a grammatical function).

2.1. Method

2.1.1. Participants

A group of 122 native English speakers (average age 18.5 years old; range 18–25 years old) undergraduates at the University of Toronto Mississauga, participated in the study for partial course credit. All participants signed an informed consent and the study was approved by the Health Sciences Research Ethics Board at the University of Toronto. We excluded participants that were not native English speakers, as assessed by a self-report, and that did not achieve a minimum Mill-Hill

Table 1
Content-words features and passage-characteristics for the four passages.

	Content-words features								Passage characteristics, based on the full text		
		Rating scales	English lexicon project data-base	N of sentence	Words per sentence	Complex words					
	Content words	Non-neutral words	High-arousal words	Less-familiar words	Word frequency (log HAL)	Word length (letters)	N. of phonemes per word	N. of syllables per word			
Rainbow	108	24%	17.5%	5.5%	10.0 (2.2)	6.2 (2.3)	5.1 (2.2)	1.8 (1.0)	19	17.3	8.8%
North Wind	25	33.5%	16.5%	8%	9.5 (1.7)	6.2 (2.2)	5.2 (1.9)	1.75 (.9)	5	22.6	7.1%
Grandfather	63	22%	14.5%	8%	9.7 (2.3)	5.6 (1.7)	4.5 (1.6)	1.5 (.6)	8	16.5	5.3%
My 1st Day	75	0%	0%	0%	10.2 (1.5)	5.7 (1.6)	4.6 (1.5)	1.6 (.8)	14	11.6	3.1%

The percentages presented here for the three rating scales were based on the distribution of responses. Lexical features of the content words were taken from the English lexicon project data-base (Balota et al., 2007). Data represents averages (and standard deviations). Passage characteristics are based on all of the words in the passage (and not only content ones). Complex words are taken as polysyllabic words, with 3 or more syllables.

Vocabulary Test score of 9/20, corresponding to normal vocabulary levels for native-English speakers (Ben-David, Goy, Erel & Schneider, 2015).

2.1.2. Apparatus and procedure

Seven-hundred and fourteen content words were rated on arousal, valence and familiarity. These words include all content words in the three passages and content words gathered from several relevant databases of emotional and neutral words, used for constructing the novel passage (e.g., Larsen et al., 2006; Phaf & Kan, 2007; Algom et al., 2004). The ratings were performed in separate blocks using eight-point Likert (1936) scales. Word order within each block was randomized for each participant, such that no word order was repeated between and within participants. In the first block, participants were asked to rate the words on their emotional valence (from 1, extremely negative, to 8, extremely positive). On the second task, they were asked to rate how much they were personally affected by the words as an indicator for arousal (from 1, extremely unaffected, to 8, extremely affected; see Dresler et al., 2009 for details). Participants were encouraged to base their ratings on their immediate impression of the word. On the third block, participants were asked to rate the same words on familiarity ranging from 1 (extremely unfamiliar) to 8 (extremely familiar). Given the large number of words tested, different word subsets and rating tasks were presented to different participants, randomly.[1]

2.2. Results and discussion

Based on the distribution of responses in each scale, cutoff points were selected. Table 1 presents the percentage of content words in each passage that were rated as non-neutral (words rated lower than 4.0 or higher than 6.0 on the valence scale), high-arousal (as words were generally low on arousal, relatively high-arousal words were taken as words that received an average rating above 3.0 on the arousal scale) and less-familiar (95th percentile score as a cut-off). The data show that 14–18% of the content words in the traditional passages are relatively high on arousal, for example, WAR (Rainbow) and SUCCEEDED (North Wind). Similarly, 22–32% of the words in these passages were rated as negative, for example, WAR and FLOOD (Rainbow), or positive, for example, SUCCEEDED and MIRACLE (North Wind and Grandfather, respectively). Table 1 also shows that about 6–8% of the content words used in the traditional passages were rated as less-familiar, or even unfamiliar, for example, NORSEMAN and SUPERIMPOSITION (Rainbow).

In sum, the three traditional passages include several words that are more extreme on valence and arousal, and a few words that are relatively low on familiarity. These specific features have the potential to impact reading performance, especially when anxiety is a factor (Chajut, Mama et al., 2010; Chajut, Schupak et al., 2010; Craig, 2014; Craig, Hancock, Tran, & Craig, 2003; Hennessey et al., 2014; Hubbard & Prins, 1994; Lang, Bradley, & Cuthbert, 1990; Mogg, Millar, & Bradley, 2000; van Lieshout et al., 2014). Hence, these findings call for an examination of the traditional passages to ensure that readability and content are not affecting the quality of speech fluency assessment, especially given the variability in levels of speech-related anxiety present in the PWS population (Craig and Tran, 2014; Alm, 2014). In the following section, we

[1] Note, we were not able to use the recent database on valence and arousal ratings provided by Warriner and colleges (Warriner, Kuperman, & Brysbaert, 2013), as it does not include all of the words used in the traditional passages (88 words are missing from their corpus, 12.3%: 26 were absent and 62 were presented in a different form). However, in an examination of a subset of words that were tested in both our and the Warriner et al. (2013) dataset (158 matching and 79 different-form words), no significant differences were found in valence and arousal ratings ($F < 1$ for both, using Z scores, as original scales differed).

Table 2
Phonetic properties of the phonemes in the four passages.

	Phonemes	Vowels				Consonants	
		Frontal	Back	Central	Diphthongs	Voiceless	Voiced
Rainbow	1096	16%	6%	2%	9%	20%	47%
North Wind	361	14%	7%	4%	5%	22%	48%
Grandfather	457	18%	6%	4%	6%	25%	41%
My First Day	507	15%	4%	3%	9%	22%	47%

Ratings are based on coding by a trained linguist using North American enunciation. As can be expected, the four types of vowels were distributed unequally with the highest percentage found for frontal vowels ($M = 16\%$). However, this difference was independent of passage type ($\chi^2(9) = 0.26$ $p = 0.9$).

describe the construction of a novel passage (designed to increase readability) and compare the readability of the three traditional passages and the novel one.

3. Experiment 2

As the traditional passages were found to include several non-neutral words and a few words that are relatively low on familiarity we constructed a new reading passage entitled "My New Day". This passage was designed to minimize the influence of both content and structure factors that may affect reading performance, above and beyond stuttering. Next, we compared the traditional passages and the novel one on readability.

3.1. Constructing "My First Day" passage

The ratings of the 714 words (see above) were used to create the new reading passage, based on the following six guidelines: (a) maintain semantic cohesiveness of the text and readability; (b) avoid any negative, extremely positive and/or arousing words (based on ratings); (c) choose extremely familiar words (based on ratings); (d) present each content word only once in the text; (e) match main lexical characteristics to the average values found in the three traditional passages (obtained from Balota et al., 2007; for a similar analysis, see Ben-David, van Lieshout & Leszcz, 2011); and (f) use relatively short sentences, with a minimum number of complex words (that have more than three syllables) to enhance readability (Kincaid, Fishburne, & Rogers, 1975). A final set of 74 content words (underlined) was chosen to construct the following novel passage.

My First Day

I walked around a long fence to reach the man on patrol. I buttoned my coat and wrapped my scarf because the wind was blowing. He switched a lever to allow me through the front gate of the building. Inside, I sat on a chair at a round table with the manager of my division. His associate and I signed some forms and stapled them together. They showed me a folder with my first project. I observed closely and took notes. Then, they opened a small cardboard box with their main product inside. It looked like a modern type of engine with a small motor. Its front panel held a considerable number of plain black and white square buttons. With a paper and pencil, I designed them in various shapes and colors. At five, someone in an apron came in with some bread and butter with jelly. We raised our glasses in the air. We toasted to the end of my first day.

3.2. Passage features and readability

Table 1 presents the main passage-features used to calculate readability on all four passages (based on both content and non-content words). The table presents a complex picture, where several features do not differ between the four passages, while others do. The four passages do not differ in terms of average word length (as indicated by a nonsignificant effect of passage, $F(3,260) = 1.5$, $p = 0.2$; note, there are 264 content words that differ between the four passages), number of phonemes per word ($F(3,260) = 2.1$, $p = 0.1$), number of syllables per word ($F(3,260) = 1.8$, $p = 0.14$), and the average word frequency (log HAL, $F(3,260) < 1$). However, the passages differ in length (number of words and sentences), average words-per-sentence, and percentage of complex (polysyllabic) words. The Rainbow passage is the longest one, with the highest percentage of complex words (8.8%). In contrast, our novel passage, constructed to maximize readability, has a significantly lower words-per-sentence average and lower percentage of complex words than the traditional ones (compared to the average and 95% confidence interval, CI, calculated for all four passages).

3.2.1. Phonetic properties

Table 2 presents the phonetic properties of the four passages (types of vowels and consonants), based on coding by a trained linguist using North American enunciation. Although the influence of individual sounds on stuttering is debatable, some studies have indicated that stuttering could be higher for consonants than vowels and that there is some variety in stuttering across different vowels (Mahesh and Geetha, 2013). Within the category of consonants, voiceless fricatives may be more stutter prone (Jayaram, 1983). Furthermore, a study by Howell et al. (2006), looking at phonetic difficulty

Table 3
Readability Indices for the four passages.

	Word-Based Formulas						Word Predictability: Cloze score	
	Flesch Kincaid:		Gunning Fog	SMOG	Coleman Liau	Automated Reading Index	Full passage	First 23 fill-in
	Reading Ease	Grade Level						
Rainbow	71.2	7.6	10.5	10.2	9.8	7.7	53.7%	62.1%
North Wind	76.8	8.2	11.9	10.4	9	9.7	68.0%	68.0%
Grandfather	77.9	6.5	8.7	8.5	9.6	7.2	50.9%	53.7%
My First Day	84.5	4.4	5.9	6.5	8.6	3.9	62.7%	65.1%

Note: A lower score indicates better readability in all word-based indices except for Flesch Kincaid: Reading Ease (where the scale is reversed). In the Cloze indicator, scores represent accuracy percent, with a higher score indicating better readability.

and stuttering, indicated that manner and place of consonant productions influenced level of disfluency in their group of participants, which the authors associated with developmental aspects of motor difficulty. Although the latter study was based on spontaneous speech in PWS (children, teenagers, & adults), the influence of phonetic features could also affect speech when reading aloud words, as similar processes of phonetic encoding and speech motor planning and execution are involved. Some PWS may also display a fear for specific sounds, which may lead to a stronger expectation for stuttering as indexed by an increase in anticipatory autonomous arousal (Bowers, Saltuklaroglu, & Kalinowski, 2012). Finally, individual vowels may also affect the variability of specific articulator movements that could trigger the occurrence of speech errors (normal disfluencies; Slis and van Lieshout, 2013). In sum, phonetic features need to be addressed when comparing different passages for different purposes (e.g., Deterding, 2006; Patel et al., 2013). We did not aim for a detailed analysis of all possible phonetic features for the four passages, as this would be too onerous given the relatively limited reported impact of phonetic features on stuttering. Instead, we focused on a more global distribution of type of vowels (Mahesh and Geetha, 2013) and consonants (based on voicing characteristics; Jayaram, 1983). A visual inspection of Table 2 suggests that the distribution of types of vowels (front, back, central and diphthong) and the two types of consonants (voiced and voiceless) did not differ across the four passages. This was confirmed by finding a non-significant chi square analyses (χ^2 (9) = 0.26 p = 0.9 and χ^2 (3)=0.68, p = 0.9 for the type of vowel and the type of consonant, respectively).

3.2.2. *Passage readability based on readability indices*

Readability scores, as calculated by six commonly used word-based readability indices, are presented in Table 3 for all passages (see Appendix B for the specific formulae used to calculate these indices). All of the passages show good readability, however, our novel passage scored better on all of the indices (based on descriptive measures, using the average value and CI across all four passages) as follows: (a) The Flesch Reading Ease Index (Flesch, 1949; higher scores indicate better readability). The three traditional passages fall within the recommended range of 60–80 (appropriate for 12–15 years old, see D'Alessandro, Kingsley, & Johnson-West, 2001), yet our novel passage scored significantly higher (84.5); (b) The Flesch-Kincaid Grade Level indicator (Kincaid et al., 1975). The novel passage's grade level, 4th grade, is significantly lower than the traditional passages (i.e., it might potentially be used with individuals who attained lower education levels or school age children); (c) The Automated Readability Index, also a grade level indicator, confirms the Flesch-Kincaid index result; (d) The Gunning Fog (Gunning, 1968; better readability with lower scores), typically used for evaluating business texts. Only two passages, Grandfather and My New Day, met the recommended readability level (7–8; see Kasabwala et al., 2013). On the rest of the scores, (e) The SMOG index (McLaughlin, 1969) used for assessing the years of education needed to completely understand a text and (f) The Coleman-Liau (Coleman, & Liau, 1975) index (with lower scores indicating better readability on both), indicate that our novel passage is nominally better than the other passages.

To produce a readability score that is based on word predictability in the context of the passage as a whole, we also performed a CLOZE (Taylor, 1953) procedure.

3.3. *Method*

3.3.1. *Participants*

A group of 31 native English speakers (average age 22 years old; range 18–25 years old), taken from the same population as in Section 2.1.1, participated in the study for partial course credit. All participants signed an informed consent and the study was approved by the Health Sciences Research Ethics Board at the University of Toronto.

3.3.2. *Apparatus and procedure*

Every fifth word (randomizing the location of the first word) was deleted and replaced by a fill-in blank space (_____). Participants were asked to write the word that best fits the blank space. The order of the passages was counterbalanced across participants. Scoring was conducted by four trained raters (scoring agreement, Cronbach's Alpha = 0.95). They followed the conceptual rating process (O'Toole & King, 2011), where a response that maintains the meaning of the sentence was

considered accurate. For example, in the sentence "I buttoned my coat and _____ (wrapped) my scarf" (My New Day), accepted responses were "tightened", "tied", and "fixed", even though they are not exact synonyms.

3.4. Results and discussion

Cloze-scores for the full passages (a total of 148 Cloze fill-in blanks) and the first 23 Cloze fill-in blanks in each passage (a total of 92; to control for the differences in passage length) are presented in the rightmost columns of Table 3. No significant difference was found between the four passages on both tests (one-way ANOVA; $F(3, 144) = 2.1$, $p = 0.1$ and $F(3, 88) < 1$, respectively). It is noteworthy that the traditional passages include repetition of content words, a condition that should facilitate correct filling-in-the-blanks in a CLOZE test (e.g., 23% of the content words in the "North Wind"; Patel et al., 2013). Even though our novel passage includes *no repetition* of content words, it was still rated highly readable in the Cloze score.

4. Experiment 3

In order to test the possible influences of passage's features on fluency assessment, we compared stuttering percent for a group of PWS on the most widely used traditional passage (Rainbow) and our novel passage. A difference in fluency levels between the two passages with the same set of participants would indicate that the passage construct and the semantic features may bias the stuttering assessment process. Alternatively, if these differences will not be apparent, this will support a validation of the use of the traditional passage with PWS. Idiosyncratic differences between the two passages may still imply the need to use a variety of passages to ensure the accuracy of clinical assessment, given individual responses to arousal, valence, familiarity and passage structure features. Hence, we directly compared the Rainbow passage with our new balanced passage, My First Day, on the percentage of disfluencies elicited for both passages on the same group of individuals who stutter. The Rainbow passage was chosen as it is the most commonly used passage, as reported in the literature (see Appendix A). Note, examining Tables 1–2, this passage also represents the average of the other traditional passages in terms of general word and passage properties.

4.1. Method

4.1.1. Participants

Twenty-six PWSs participated in the study (average age, 23 years old, range 14–44 years old). At the time of testing, they were enrolled for fluency treatment at the Speech and Stuttering Institute, Toronto. All PWS participants were approached based on a list of PWS who expressed their interest in participating in a study in the past. Recruitment was conducted at the University of Toronto Oral Dynamics lab to avoid any notion that a person's decision to decline participation could affect his/her treatment in any way.

PWS participants ranged in stuttering severity from very mild to severe on the Stuttering Severity Instrument for Children and Adults (SSI-4, Riley, 2008; Riley & Bakker, 2009). The SSI includes a reading and speaking task to assess the severity of stuttering based on frequency of stuttering events (repetitions or prolongations of sounds or syllables), duration of longest blocks and physical concomitants (e.g. distracting sounds, facial grimaces, head movements and movements of the extremities; for an excellent overview on the use of SSI-4 see Guitar, 2014). PWS participants did not report and/or demonstrate any clinical signs related to cognitive, speech, language, hearing or neurological deficits, other than stuttering (see Table 4 for details). All participants signed an informed consent and the study was approved by the Health Sciences Research Ethics Board at the University of Toronto.

4.1.2. Apparatus and procedure

We audio-recorded each participant's reading of both the Rainbow and My First Day passages (order was counterbalanced across participants), using a digital recorder (Olympus WS–133M) placed at a 45-degree angle, 6–8 in from the participant's mouth. The participants had no previous experience with the reading passages, nor did they have an opportunity to rehearse them orally or silently prior to the recording. Audio recordings were analyzed off-line using PRAAT (Boersma and Weenink, 2009). A trained rater scored the number of stuttering disfluencies (pauses/hesitations, repetitions and blocks). Two raters independently re-scored a random selection of 30% of the data set (agreement reached 84.4%, K-alpha = 0.992).

4.2. Results and discussion

The average percentage of disfluencies (e.g. prolongations, repetitions etc.) on the Rainbow passage was 4.36 (SD = 4.52) and on My First Day it was 4.69 (SD = 4.95). The disfluencies percent across individuals on the two passages were highly correlated ($r_p = 0.816$, $p < 0.001$). We conducted the same correlation analysis with 16 participants that demonstrated at least 3% disfluencies at one of the passages (see Table 4), and reached a similar yet smaller correlation ($r_p = 0.667$, $p = 0.007$). Although the average disfluencies percent for the two passages were not significantly different ($t(14) = 0.7$, $p = 0.5$), half of the individuals in this sample showed a difference between the two passages that ranged from 1.6–11% of disfluencies. Such

Table 4
PWS participant demographics.

Participant ID	Age(years; months)	Gender	Severity Scale(SSI-4)	>2% disfluenciesin the current study
1	20; 06	M	severe	+
2	18; 11	M	mild	+
3	44; 11	M	severe	+
4	18; 06	M	very mild	
5	18; 10	M	very mild	
6	19; 01	M	very mild	
7	25; 10	M	moderate	+
8	25; 10	M	mild	
9	17; 11	M	very mild	
10	17; 03	M	moderate	+
11	23; 06	M	very mild	
12	14; 0	F	very mild	+
13	26; 0	M	severe	
14	14; 11	M	mild	+
15	18; 0	M	very mild	
16	21; 04	F	severe	+
17	21; 02	F	mild	+
18	24; 09	M	mild	+
19	23; 07	M	moderate	+
20	20; 03	M	severe	+
21	32; 05	M	mild	+
22	38; 0	M	>very mild	+
23	23; 0	F	moderate	+
24	18; 08	M	mild	+
25	25; 11	M	severe	+
26	23; 04	F	>very mild	

a difference may have clinical implications. We thus suggest that clinicians and researchers should consider using more than one passage to assess stuttering in individual cases.

5. General discussion

The first goal of this study was to evaluate the properties of the three most common reading passages ("Rainbow", "The North Wind and the Sun" and "Grandfather" passages) used to assess speech function in English speaking individuals, including PWS. We analyzed specific properties of these passages, such as, emotional valence, arousal, word familiarity and frequency, as well as passage-readability characteristics. To the best of knowledge, this is the first study that provides this kind of information for all three passages. The analyses show that the three traditional passages contain a share of emotional (22–34%), high arousal (15–18%), lower familiarity (6–8%) and polysyllabic (three or more syllables, 5–9%) content words. These features have been noted in the literature to impact on reading performance (slowing down) and other cognitive processes, as detailed at the outset of this paper. Readability estimates for these passages (both formula-based and CLOZE tests) were not as high as one may wish for use with a clinical population. The detailed analyses of the traditional passages, as provided in this study, point to potential strengths and weaknesses of these passages (see also Powell, 2006; Patel et al., 2013). This information can guide clinicians and researchers in choosing the appropriate passage for their clients to obtain a reliable fluency assessment. Apart from PWS, this could be of special interest when testing other populations, such as young children, older adults and people with cognitive deficits and/or anxiety (for a discussion, see Lowe et al., 2015). For example, when testing a person with anxiety, performance on the 'North Wind' passage might be affected by the large percentage of non-neutral (33.5% of 25 words) and high arousal words (16.5%). Clearly, this is a factor that needs to be accounted for in stuttering, specifically, given the high prevalence of anxiety in this population (Craig and Tran, 2014; Alm, 2014). Moreover, age can have an effect on the familiarity of words in the text (Ben-David et al., 2015) and its readability. Further research should examine the effect of these features on such populations in more detail.

The second goal of the current paper was to construct a new passage ("My First Day") designed to minimize the influence of these properties, generating a highly readable and well-balanced passage, made of neutral, low-arousal and highly-familiar words. This neutral passage would be especially suitable for populations diagnosed with high levels of anxiety. Our novel passage was also designed to score high on readability scales, while being compatible with the traditional ones in terms of structural properties. As valence, arousal and familiarity have been shown to influence reading performances (Bar-Haim et al., 2007), the new balanced passage was compared to one of the most widely used reading passages, "Rainbow." We tested 26 PWS adults, without clinical signs of cognitive deficits or anxiety. The percentage of disfluencies on the two passages was highly correlated across individuals, validating the use of both passages with this type of population. In this context, the results provide some indication that the Rainbow traditional passage although different in some ways from the new passage (in terms of general properties such as emotional valence, arousal, word familiarity and word frequency), may not necessarily lead to more disfluencies on a group basis, at least not in the sample tested here. However, when testing PWS

who have been identified with anxiety issues, the result could be different, as the presence of threat words (Hennessey et al., 2014) or words with an emotional link to stuttering and speech (van Lieshout et al., 2014) have been found to influence speech preparation and execution processes in this population.

Given that passages used for clinical assessment (or research purposes) may differ in the properties assessed in the current study, it might be advisable to use at least two different reading passages to obtain a more robust estimate of stuttering severity. Further testing of the new passage with a larger sample of PWS, a wider range of severity and different English-speaking cultures (for a discussion see, Icht and Ben-David, 2014) will be necessary to support this recommendation. Note, our "My First Day" passage also meets the recommendation of minimum sample length (>200 syllables) for obtaining stable and reliable speech fluency assessment data (Riley, 1994). For samples less than 150 syllables, the reliability of stuttering severity scores decreases significantly and thus passages such as "The Grandfather" (177 syllables; as indicated in Patel et al., 2013) and The Northwind and the Sun (125 syllables) may not be suitable for fluency assessments (Riley, 1994; Todd et al., 2014).

Finally, in the past decade, the need to tailor a reading passage to a specific population has been raised by several researchers (on tailoring speech tests to special populations, see Icht and Ben-David, 2015). For example, "The Boy who Cried Wolf" passage by Deterding (2006) was constructed to include the largest variety of English sounds. "The Caterpillar" passage (Patel et al., 2013) was designed to examine speech motor deficits within and across speech subsystems by manipulating several segmental and prosodic variables. Both studies are highly important in addressing these specific speech variables. Our study supplements them, adding the emotional and attentional purview that was missing in previous examinations. Note, the theme of "The Boy who Cried Wolf" passage, as many of its content words, is highly emotional, describing a common fairy tale of fear and morality. As such, it may not be very suitable for use with patients with high levels of anxiety.

6. Conclusion

This study presents a detailed evaluation of three popular readings passages on a wide variety of characteristics related to emotional valence, arousal, familiarity, word and sentence complexity and readability. The data from this study can provide clinicians and researchers detailed information on the potential sources of variance in reading performance for the traditional passages. In addition, we present a novel well-balanced and carefully controlled passage that can supplement and complement reading performance indices based on the traditional passages. The literature shows that it is important to verify that the passages used for assessing stuttering are balanced in terms of valance and arousal characteristics to avoid biases associated with anxiety. Moreover, different passages may yield different stuttering scores due to imbalanced word characteristics (valence and arousal). We further suggest that controlling for the possible link between stuttering severity and anxiety can help targeting individually tailored interventions. Knowledge about the characteristics of testing materials is crucial to inform clinical practice (and research). We hope that this study has been able to contribute to this process.

Acknowledgments

B. M. Ben-David was partially supported by a grant from the Ontario Neurotrauma Foundation (2008-ABI-PDF-659), and a Marie Curie Career Integration Grant (FP7-PEOPLE-2012-CIG) from the European Commission. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program (303712CRC) awarded to P. H. H. M. van Lieshout.

Appendix A.

A search in Scopus electronic database (conducted on 2015), using entire collection with search terms in abstract, title and keywords, showed that the "Rainbow" passage was referred to most often (72 citations), compared to the "Grandfather" passage (26 citations), and "The North Wind and Sun" (20 citations).

Appendix B.

Overview of formulas used for the Readability indices

Readability indices	Formulas
Flesch Kincaid: Reading Ease	$206.835 - 1.015 \times (\text{words/sentences}) - 84.6 \times (\text{syllables/words})$
Grade Level	$0.39 \times (\text{words/sentences}) + 11.8 \times (\text{syllables/words}) - 15.59$
Gunning:Fog Score	$0.4 \times ((\text{words/sentences}) + 100 \times (\text{complex Words/words}))$
SMOG Index	$1.0430 \times \text{sqrt}(30 \times \text{complex Words/sentences}) + 3.1291$
Coleman Liau Index	$5.89 \times (\text{characters/words}) - 0.3 \times (\text{sentences/words}) - 15.8$
Automated Readability Index	$4.71 \times (\text{characters/words}) + 0.5 \times (\text{words/sentences}) - 21.43$

References

Adams, M. R., & Hutchinson, J. (1974). The effects of three levels of auditory masking on selected vocal characteristics and the frequency of disfluency of adult stutterers. *Journal of Speech, Language, and Hearing Research*, *17*, 682–688.

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional stroop phenomenon: a generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, *133*, 323–338.

Alm, P. A. (2014). Stuttering in relation to anxiety, temperament, and personality: review and analysis with focus on causality. *Journal of Fluency Disorders*, *40*, 5–21.

Alm, P. A., & Risberg, J. (2007). Stuttering in adults: the acoustic startle response, temperamental traits, and biological factors. *Journal of Communication Disorders*, *40*, 1–41.

Andersson, G., Westöö, J., Johansson, L., & Carlbring, P. (2006). Cognitive bias via the internet: a comparison of Web-based and standard emotional Stroop tasks in social phobia. *Cognitive Behaviour Therapy*, *35*, 55–62.

Armson, J., & Stuart, A. (1998). Effect of extended exposure to frequency-altered feedback on stuttering during reading and monologue. *Journal of Speech, Language, and Hearing Research*, *41*, 479–490.

Arntz, A., Appels, C., & Sieswerda, S. (2000). Hypervigilance in borderline disorder: a test with the emotional Stroop paradigm. *Journal of Personality Disorders*, *14*, 366–373.

Asmundson, G. J., & Stein, M. B. (1994). Selective processing of social threat in patients with generalized social phobia: evaluation using a dot-probe paradigm. *Journal of Anxiety Disorders*, *8*, 107–117.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice* (2nd ed.). San Diego, CA: Singular Publishing Group.

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. *Behavioral Research Methods*, *39*, 445–459.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological Bulletin*, *133*, 1–24.

Ben-David, B. M., Chajut, E., & Algom, D. (2012). The pale shades of emotion: a signal detection theory analysis of the emotional Stroop task. *Psychology*, *3*, 537–541.

Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). Older is always better: Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, *30*, 856–862.

Ben-David, B. M., van Lieshout, P. H. H. M., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, *25*, 206–220.

Blood, G. W., Blood, I. M., Bennett, S., Simpson, K. C., & Susman, E. J. (1994). Subjective anxiety measurements and cortisol responses in adults who stutter. *Journal of Speech, Language, and Hearing Research*, *37*, 760–768.

Blood, G. W., & Hood, S. B. (1978). Elementary school-aged stutterers' disfluencies during oral reading and spontaneous speech. *Journal of Fluency Disorders*, *3*, 155–165.

Blumgart, E., Tran, Y., & Craig, A. (2014). Social support and its association with negative affect in adults who stutter. *Journal of Fluency Disorders*, *40*, 83–92.

Boersma, P., & Weenink, D. (2009). *PRAAT: doing phonetics by computer (version 5.1.07) (computer program)*. http://www.praat.org

Bowers, A., Saltuklaroglu, T., & Kalinowski, J. (2012). Autonomic arousal in adults who stutter prior to various reading tasks intended to elicit changes in stuttering frequency. *International Journal of Psychophysiology*, *83*, 45–55.

Brutten, E. J., & Shoemaker, D. J. (1967). *The modification of stuttering*. Englewood Cliffs, NJ: Prentice-Hall.

Chajut, E., Mama, Y., Levy, L., & Algom, D. (2010). Avoiding the approach trap: a response bias theory of the emotional Stroop effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1567–1572.

Chajut, E., Schupak, A., & Algom, D. (2010). Emotional dilution of the Stroop effect: a new tool for assessing attention under emotion. *Emotion*, *10*, 944–948.

Cecconi, C. P., Hood, S. B., & Tucker, R. K. (1977). Influence of reading level difficulty on the disfluencies of normal children. *Journal of Speech & Hearing Research*, *20*, 475–484.

Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, *60*, 283–284.

Craig, A. (1990). An investigation into the relationship between anxiety and stuttering. *Journal of Speech and Hearing Disorders*, *55*, 290–294.

Craig, A. (2014). Major controversies in fluency disorders: clarifying the relationship between anxiety and stuttering. *Journal of Fluency Disorders*, *40*, 1–3.

Craig, A., Hancock, K., Tran, Y., & Craig, M. (2003). Anxiety levels in people who stutter: a randomized population study. *Journal of Speech Language and Hearing Research*, *46*, 1197–1206.

Craig, A., & Tran, Y. (2014). Trait and social anxiety in adults with chronic stuttering: conclusions following meta-analysis. *Journal of Fluency Disorders*, *40*, 35–43.

D'Alessandro, D. M., Kingsley, P., & Johnson-West, J. (2001). The readability of pediatric patient education materials on the World Wide Web. *Archives of Pediatrics & Adolescent Medicine*, *155*, 807–812.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Audio seminars in speech pathology: motor speech disorders*. Philadelphia: WB Saunders.

Davis, S., Shisca, D., & Howell, P. (2007). Anxiety in speakers who persist and recover from stuttering. *Journal of Communication Disorders*, *40*, 398–417.

Deterding, D. (2006). The north wind versus a wolf: short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, *36*, 187–196.

Dresler, T., Mériau, K., Heekeren, H. R., & van d, M. (2009). Emotional stroop task: rffect of word arousal and subject anxiety on emotional interference. *Psychological Research*, *73*, 364–371.

Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders*, *20*, 245–257.

Ezrati-Vinacour, R., & Levin, I. (2004). The relationship between anxiety and stuttering: a multidimensional approach. *Journal of Fluency Disorders*, *29*, 135–148.

Fairbanks, G. (1960). *Voice and articulation drill book* (2nd ed.). New York: Harper & Row.

Fitzgerald, H. E., Djurdjic, S. D., & Maguin, E. (1992). Assessment of sensitivity to interpersonal stress in stutterers. *Journal of Communication Disorders*, *25*, 31–42.

Flesch, R. F. (1949). *Art of readable writing*. New York: Harper.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, *130*, 681.

Frings, C., Englert, J., Wentura, D., & Bermeitinger, C. (2010). Decomposing the emotional Stroop effect. *The Quarterly Journal of Experimental Psychology*, *63*, 42–449.

Gregory, H. (1991). Therapy for elementary school-age children. *Seminars in Speech and Language*, *12*, 323–335.

Guitar, B. (2013). *Stuttering: an integrated approach to its nature and treatment* (2nd ed.). Baltimore: Williams & Wilkins.

Guitar, B. (2014). *Stuttering: an integrated approach to its nature and treatment* (4th ed.). Baltimore: Williams & Wilkins.

Gunning, R. (1968). *The technique of clear writing*. New York: McGraw Hill Book Co.

Hedge, M. N. (1982). Antecedents of fluent and disfluent oral reading: a descriptive analysis. *Journal of Fluency Disorders*, *7*, 323–341.

Hennessey, N. W., Dourado, E., & Beilby, J. M. (2014). Anxiety and speaking in people who stutter: an investigation using the emotional Stroop task. *Journal of Fluency Disorders*, *40*, 44–57.

Howell, P., Au-Yeung, J., Yaruss, S. J., & Eldridge, K. (2006). Phonetic difficulty and stuttering in English. *Clinical Linguistics & Phonetics*, *20*, 703–716.

Hubbard, C. P., & Prins, D. (1994). Word familiarity, syllabic stress pattern, and stuttering. *Journal of Speech, Language, and Hearing Research*, *37*, 564–571.

Icht, M., & Ben-David, B. M. (2014). Oral-diadochokinesis rates across languages: English and Hebrew norms. *Journal of Communication Disorders*, *48*, 27–37.
Icht, M., & Ben-David, B. M. (2015). Oral-diadochokinesis rates for Hebrew-speaking school-age children: eeal words vs: non-words repetition. *Clinical Linguistics & Phonetics*, *29*, 102–114.
IPA. (1999). *Handbook of the international phonetic association: a guide to the use of the international phonetic alphabet*. New York: Cambridge University Press.
Iverach, L., Menzies, R. G., O'Brian, S., Packman, A., & Onslow, M. (2011). Anxiety and stuttering: continuing to explore a complex relationship. *American Journal of Speech-Language Pathology*, *20*, 221–232.
Iverach, L., & Rapee, R. M. (2014). Social anxiety disorder and stuttering: current status and future directions. *Journal of Fluency Disorders*, *40*, 69–82.
Jayaram, M. (1983). Phonetic influences on stuttering in monolingual and bilingual stutterers. *Journal of Communication Disorders*, *16*, 287–297.
Kasabwala, K., Misra, P., David, R., Agarwal, N. F., Baredes, S. F., Setzen, M. F., & Eloy, J. A. (2013). Readability assessment of the american rhinologic society patient education materials. *International Forum of Allergy & Rhinology*, *3*, 325–333.
Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count, and Flesch reading rase formula) for navy enlisted personnel research branch report*. pp. 8–75. Chief of Naval Technical Training: Naval Air Station Memphis.8–75.
Kraaimaat, F. W., Vanryckeghem, M., & Van Dam-Baggen, R. (2002). Stuttering and social anxiety. *Journal of Fluency Disorders*, *27*, 319–331.
Kuchinke, L., Võ, M. L., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*, 132–140.
Lang, P., Bradley, M., & Cuthbert, B. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*, 377–395.
Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, *6*, 62–72.
Likert, R. (1936). A method for measuring the sales influence of a radio program. *Journal of Applied Psychology*, *20*, 175.
Lowe, R., Menzies, R., Packman, A., O'Brian, S., Jones, M., & Onslow, M. (2015). Assessing attentional biases with stuttering. *International Journal of Language & Communication Disorders*, *51*(1), 84–94.
Lovett, J. W. (1988). Stuttering and anxiety. *The British Journal of Psychiatry*, *153*, 844–844.
Mahesh, M. S., & Geetha, Y. (2013). Phonetic context in disfluencies of children with stuttering. *Language in India*, *13*, 545–565.
Manning, W., & Beck, J. G. (2013). The role of psychological processes in estimates of stuttering severity. *Journal of Fluency Disorders*, *38*, 356–367.
MacLeod, C., Tata, P., & Mathews, A. (1987). Perception of emotionally valenced information in depression. *British Journal of Clinical Psychology*, *26*, 67–68.
McLaughlin, G. H. (1969). SMOG grading—a new readability formula. *Journal of Reading*, *12*, 639–646.
Menzies, R. G., Onslow, M., & Packman, A. (1999). Anxiety and stuttering: exploring a complex relationship. *American Journal of Speech-Language Pathology*, *8*, 3–10.
Mogg, K., Millar, N., & Bradley, B. P. (2000). Biases in eye movements to threatening facial expressions in generalized anxiety disorder and depressive disorder. *Journal of Abnormal Psychology*, *109*, 695–704.
Namasivayam, A. K., & van Lieshout, P. (2011). Speech motor skill and stuttering. *Journal of Motor Behavior*, *43*, 477–489.
Neilson, M. D., & Neilson, P. D. (1987). Speech motor control and stuttering: a computational model of adaptive sensory-motor processing. *Speech Communication*, *6*, 325–333.
Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*, 466–478.
O'Toole, J., & King, R. (2011). The deceptive mean conceptual scoring of cloze entries differentially advantages more able readers. *Language Testing*, *28*, 127–144.
Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., . . . & Russell, S. (2013). The caterpillar: a novel reading passage for assessment of motor speech disorders. *American Journal of Speech-Language Pathology*, *22*, 1–9.
Perkins, W. (1979). From psychoanalysis to discoordination. *Controversies about Stuttering Therapy*, *5*, 97–127.
Phaf, R. H., & Kan, K. J. (2007). The automaticity of emotional Stroop: a meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 184–199.
Powell, T. W. (2006). A comparison of English reading passages for elicitation of speech samples from clinical populations. *Clinical Linguistics & Phonetics*, *20*, 91–97.
Riley, G. (1994). *SSI-3: stuttering severity instrument* (3rd. ed.). Austin, TX: Pro-Ed.
Riley, G. D. (2008). *SSI-4: stuttering severity instrument* (4rd ed.). Austin, TX: Pro-Ed.
Riley, G. D., & Bakker, K. (2009). *Stuttering severity instrument: SSI-4*. Pro–Ed.
Rutherford, E., MacLeod, C., & Campbell, L. (2004). Negative selectivity effects and emotional selectivity effects in anxiety: differential attentional correlates of state and trait variables. *Cognition and Emotion*, *18*, 711–720.
Ryan, B. (1974). *Programmed stuttering therapy for children and adults*. Springfield: CC Thomas.
Sheehan, J. G. (1970). *Stuttering: research and therapy*. New York: Harper & Row.
Slis, A., & van Lieshout, P. (2013). The effect of phonetic context on speech movements in repetitive speech. *The Journal of the Acoustical Society of America*, *134*, 4496–4507.
Smith, A., Sadagopan, N., Walsh, B., & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *Journal of Fluency Disorders*, *35*, 1–18.
Strijkers, K., Costa, A., & Thierry, G. (2009). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*, 912–928.
Todd, H., Mirawdeli, A., Costelloe, S., Cavenagh, P., Davis, S., & Howell, P. (2014). Scores on Riley's stuttering severity instrument versions three and four for samples of different length and for different types of speech material. *Clinical Linguistics & Phonetics*, *28*, 912–926.
Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.
Van Lieshout, P., Ben-David, B., Lipski, M., & Namasivayam, A. (2014). The impact of threat and cognitive stress on speech motor control in people who stutter. *Journal of Fluency Disorders*, *40*, 93–109.
van Lieshout, P., Hulstijn, W., & Peters, H. F. (2004). Searching for the weak link in the speech production chain of people who stutter: a motor skill approach. *Speech Motor Control in Normal and Disordered Speech*, *4*, 313–355.
van Riper, C. (1973). *The treatment of stuttering*. Prentice Hall.
Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191–1207.
Watts, F. N., McKenna, F. P., Sharrock, R., & Trezise, L. (1986). Colour naming of phobia–related words. *British Journal of Psychology*, *77*, 97–108.
Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional stroop task and psychopathology. *Psychological Bulletin*, *120*, 3.
Wischner, G. J. (1952). An experimental approach to expectancy and anxiety in stuttering behavior. *Journal of Speech & Hearing Disorders*, *17*, 139–152.
Yiend, J. (2010). The effects of emotion on attention: a review of attentional processing of emotional information. *Cognition and Emotion*, *24*, 3–47.

Boaz M. Ben-David Ph.D. is an Assistant Professor at the School of Psychology, the Interdisciplinary Center (IDC), Herzliya, head of the Communication Aging and Neuropsychology lab (CANlab) and adjunct faculty at the University of Toronto. His research focuses on the interaction between sensory and cognitive aging, speech perception and the validity of paradigms.

Maroof I. Moral M.A. is a scientist-turned-consultant with academic interests in cognitive sciences and industry interests in cognitive ergonomics. As an information technology consultant, he creates or improves digital experiences by blueprinting user-friendly web and mobile user interfaces.

Aravind K. Namasivayam Ph.D. is a scientist at the PROMPT Institute (Santa Fe) and an adjunct faculty at the University of Toronto. He serves as an editorial consultant/reviewer for several peer-reviewed journals in the field of Speech-Language Pathology and conducts treatment research projects which will lead to furthering evidence-based practices.

Hadas Erel Ph.D. is a Post-Doctoral Fellow at the Baruch Ivcher School of Psychology, the Interdisciplinary Center (IDC), Herzliya. Her research interests include cognitive sciences with a focus on higher cognitive functions, attention, language, intelligence and aging

Pascal H. H. M. van Lieshout Ph.D. is a Professor and Chair of the Department of Speech-Language Pathology at the University of Toronto and director of the Oral Dynamics Lab. He studies dynamical principles in oral motor control of individuals with and without speech motor problems, including stuttering.