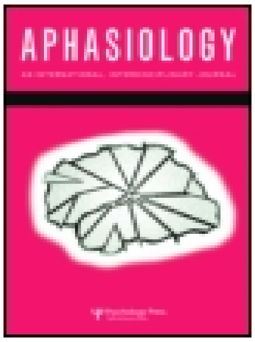

Aphasiology

Routledge
Taylor & Francis Group

ISSN: 0268-7038 (Print) 1464-5041 (Online) Journal homepage: https://www.tandfonline.com/loi/paph20

How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia

Jean K. Gordon & Sharice Clough

To cite this article: Jean K. Gordon & Sharice Clough (2020): How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia, Aphasiology, DOI: 10.1080/02687038.2020.1712586

To link to this article: https://doi.org/10.1080/02687038.2020.1712586

APHASIOLOGY
https://doi.org/10.1080/02687038.2020.1712586

Routledge
Taylor & Francis Group

How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia

Jean K. Gordon[a] and Sharice Clough[b]

[a]Department of Communication Sciences & Disorders, University of Iowa, Iowa City, IA, USA; [b]Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

ABSTRACT

Background: While persons with aphasia (PwA) are often dichotomised as fluent or nonfluent, agreement that fluency is not an all-or-nothing construct has led to the use of continuous variables as a way to quantify fluency, such as multi-dimensional rating scales, speech rate, and utterance length. Though these measures are often used in research, they provide little information about the underlying fluency deficit.

Aim: The aim of the study was to identify how well commonly used continuous measures of fluency capture variability in spontaneous speech variables at lexical, grammatical, and speech production levels.

Methods & Procedures: Speech samples of 254 English-speaking PwA from the AphasiaBank database were analyzed to examine the distributions of four continuous measures of fluency: the WAB-R fluency scale, utterance length, retracing, and speech rate. Linear regression was used to identify spontaneous speech predictors contributing to each fluency outcome measure.

Outcomes & Results: All the outcome measures reflected the influence of multiple underlying dimensions, although the predictors varied. The WAB-R fluency scale, speech rate, and retracing were influenced by measures of grammatical competence, lexical retrieval, and speech production, whereas utterance length was influenced only by measures of grammatical competence and lexical retrieval. The strongest predictor of WAB-R fluency was aphasia severity, whereas the strongest predictor for all other fluency proxy measures was grammatical complexity.

Conclusions: Continuous measures allow a variety of ways to objectively quantify speech fluency; however, they reflect superficial manifestations of fluency that may be affected by multiple underlying deficits. Furthermore, the deficits underlying different measures vary, which may reduce the reliability of fluency diagnoses. Capturing these differences at the individual level is critical to accurate diagnosis and appropriately targeted therapy.

ARTICLE HISTORY
Received 12 July 2019
Accepted 2 January 2020

KEYWORDS
Aphasia; assessment; fluency; measurement; spontaneous speech

CONTACT Jean K. Gordon jean-k-gordon@uiowa.edu Department of Communication Sciences & Disorders, University of Iowa, Iowa City, IA, USA

Supplemental data for this article can be accessed here.

Introduction

What distinguishes fluent speech from nonfluent speech? Fluency, which refers to the ease and efficiency with which ideas can be formulated into words and sentences, can be disrupted at any stage from message formulation to articulation. The concept is relevant to issues in normal language production, such as the evaluation of teaching effectiveness (e.g., Carpenter, Wilford, Kornell, & Mullaney, 2013) or the understanding of psycholinguistic planning processes (e.g., Fraundorf & Watson, 2014) and in second language learning (e.g., Chambers, 1997; De Jong, 2018), as well as in communication disorders such as stuttering, apraxia of speech, and aphasia. In describing oral expressive deficits in aphasia, fluency incorporates linguistic dimensions of word retrieval and sentence formulation, as well as more peripheral aspects of speech production such as phonological and phonetic encoding, articulation, and melodic line. This is evident in fluency rating scales such as those in the *Boston Diagnostic Aphasia Exam* (BDAE, Goodglass, Kaplan, & Barresi, 2001b), and the *Western Aphasia Battery* (WAB-R, Kertesz, 2006).

An additional potential source of discrepancy is that the term is used both diagnostically and descriptively (Feyereisen, Pillon, & De Partz, 1991). In traditional neuroanatomic approaches to aphasia syndrome diagnosis (used in the BDAE, WAB-R, and *Aphasia Diagnostic Profiles* (ADP, Helm-Estabrooks, 1992)), syndromes are classified as “nonfluent” (e.g., Broca’s aphasia, transcortical motor aphasia) or “fluent” (e.g., Wernicke’s aphasia, anomic aphasia). Descriptively, the fluency concept is used to characterise how smoothly and efficiently language is produced by a given speaker in a given task at a given point in time. Both uses are clinically useful and serve different purposes. However, these different uses can result in disagreements in the characterization of spontaneous speech in a PwA. For example, an individual with anomic aphasia who pauses frequently and abandons sentences in the search for words may be described as “fluent” according to diagnostic categories, but as nonfluent in a descriptive sense. Our goal is to understand how aspects of spontaneous speech and language influence diagnostic and descriptive designations of fluency. In a companion paper (Clough & Gordon, 2020), we examined spontaneous speech features contributing to dichotomous classifications of fluency. In the current paper, we consider contributors to continuous measures of aphasia fluency.

Because fluency is multidimensional, it has been measured in various ways, such as the rate at which speech is produced (e.g., Howes, 1964; Nozari & Faroqi-Shah, 2017), the amount of pause time relative to speech time (e.g., Feyereisen, Verbeke-Dewitte, & Seron, 1986; Park et al., 2011), the average or typical length of utterances (e.g., Goodglass, Quadfasel, & Timberlake, 1964; Helm-Estabrooks, 1992), or the complexity of grammatical forms (Goodglass, Christiansen, & Gallagher, 1993; Rochon, Saffran, Berndt, & Schwartz, 2000). Studies have demonstrated that these dimensions vary in the extent to which they co-vary (Vermeulen, Bastiaanse, & Van Wageningen, 1989; Wagenaar, Snow, & Prins, 1975), and may even dissociate, giving rise to sources of disagreement in fluency classifications (Deloche, Jean-Louis, & Seron, 1979; Feyereisen et al., 1991; Gordon, 1998; Marshall, 1986). To maximize the reliability and validity of fluency measures, some researchers have recommended cautious consideration of both quantitative and qualitative dimensions (Feyereisen et al., 1991; Goodglass et al., 1993; see also Chambers, 1997). Other researchers have responded by calling for a moratorium on fluency classifications because of their lack of psychometric reliability (Poeck, 1989) and theoretical relevance (Caplan, 1987).

However, the fluency dimension has weathered the hail of criticism and remains in common usage today, albeit with a somewhat battered reputation. Many speech-language pathologists find the fluency distinction to be clinically relevant, what Goodglass and colleagues refer to as "an effective 'first cut' in diagnostic classification" (Goodglass, Kaplan, & Barresi, 2001a, p. 75). In addition to diagnostic use, fluency affects psychosocial interactions. It has been demonstrated that people with aphasia (PwA) are perceived more negatively overall – even by their spouses – on cognitive (e.g., competence, intelligence) and personality characteristics (e.g., extroversion, likeability), and that listeners report being less comfortable and less likely to engage with them (Croteau & Le Dorze, 2001; Lasker & Beukelman, 1999; Zraick & Boone, 1991). Furthermore, perceptions are particularly negative for those with nonfluent aphasia (Duffy, Boyle, & Plattner, 1980; Harmon, Jacks, Haley, & Faldowski, 2016; Khvalabov, 2019).

The use of rating scales allows consideration of multiple variables, including more qualitative aspects of fluency (see Casilio, Rising, Beeson, Bunton, & Wilson, 2019 for a helpful discussion). This approach has a long history and was advocated by such aphasiology pioneers as Arthur Benson (1967) and Harold Goodglass (Goodglass et al., 1964). Today, the *Western Aphasia Battery* (WAB-R, Kertesz, 2006) and the *Boston Diagnostic Aphasia Exam* (BDAE-3, Goodglass et al., 2001b) take this approach, combining ratings of multiple dimensions in the diagnosis of fluency. In the BDAE, syndrome diagnoses (and their corresponding fluency categories) are made on the basis of scores on a set of separate rating scales (e.g., melodic line, word finding relative to fluency, grammatical form). In the WAB-R, fluency diagnoses arise from a single rating scale that incorporates descriptions of multiple dimensions (e.g., paraphasias, telegraphic speech, hesitations). However, there is ample evidence that such ratings also lack reliability or may not accord with clinical impressions (Clough & Gordon, 2020; Gordon, 1998; John et al., 2017; Swindell, Holland, & Fromm, 1984; Trupe, 1984; Wertz, Deal, & Robinson, 1984).

In aphasia research, the advent of increasingly sensitive neuroimaging techniques has revived interest in identifying the neural substrates of expressive impairments in aphasia. For example, Borovsky, Saygin, Bates, and Dronkers (2007) used voxel-based lesion analysis to identify networks underlying aspects of speech production. Fluency was related to damage in a broad area centering in the left inferior frontal gyrus and insula. Diffusion tensor imaging has implicated the arcuate fasciculus in fluency of output (Wang, Marchina, Norton, Wan, & Schlaug, 2013). Using MRI, Fridrikkson and colleagues also identified the anterior arcuate fasciculus (Fridrikkson, Guo, Fillmore, Holland, & Rorden, 2013), as well the aslant tract and the uncinate fasciculus (Basilakos et al., 2014), as important regions underlying speech fluency. Two recent MRI studies have aimed to identify broad language factors and their neural substrates. In one, Lacey and colleagues conducted a factor analysis and showed that the "Word Finding/Fluency" factor was related to inferior frontal lobe regions and a small area of dorsal parietal white matter (Lacey, Skipper-Kallal, Xing, Fama, & Turkeltaub, 2017). In the other, a principal component analysis related components of fluency to damage in the left precentral gyrus, superior insula, putamen, and neighboring white matter (Halai, Woollams, & Lambon Ralph, 2017). Similar approaches have been used to examine brain areas corresponding to fluent speech production in primary progressive aphasia (e.g., Wilson et al., 2010).

Together, these studies provide an emerging picture of the neural substrates underlying the fluent production of spontaneous speech. However, progress is hindered by the same lack of specificity in definitions of fluency that occurs in clinical practice and research. As in prior work, these imaging studies vary widely in the indices of fluency used. Borovsky et al. (2007) used the number of tokens as a proxy for fluency. In both of Fridrikkson's studies (Fridrikkson et al., 2013, 2012), fluency was defined by the WAB-R Spontaneous Speech fluency rating scale. Wang et al. (2013) used words per minute to capture articulatory agility, and Correct Information Units per minute (Nicholas & Brookshire, 1993) to reflect efficiency in a way that also considered the informativeness of the output. Halai et al. (2017) measured the number of words, MLU, and type-token ratio, which were found to reflect two principal components of fluency: quantity of speech produced (including MLU and number of words), and variety of words (TTR). This diversity of measures makes it difficult to find converging evidence across studies or to draw any firm conclusions about what the identified brain areas are actually responsible for doing. In addition, most of the measures used capture superficial manifestations rather than underlying causes of dysfluency.

The current study

Given the continued relevance of the fluency dimension in both clinical and research domains, we aim to improve the reliability and validity of fluency measurements. As the first step in that direction, we examined the spontaneous speech characteristics that contribute to dichotomous (Part A, Clough & Gordon, 2020) and continuous (Part B, the current paper) measures of fluency in aphasia. To do so, we made use of a large and fairly representative sample of participants with aphasia from AphasiaBank (https://aphasia.talkbank.org/) to maximise the generalizability of the findings. We investigated how characteristics of the speakers' spoken language production affected the relative importance of factors contributing to fluency. For our analyses, we examined the WAB-R fluency scale (a subjective multidimensional scale) and three more objective quantitative measures that are commonly used to index fluency, but which are hypothesised to capture different components of fluency: mean length of utterance, speech rate, and retracing. Because the objective measures reflect fluency indirectly and incompletely, they should be considered as proxies for fluency. Each of these measures may be affected by a combination of the three primary dimensions of spoken language production: word retrieval, grammatical formulation, and speech production abilities.

The WAB-R fluency scale, as discussed above, explicitly combines qualitative descriptions of multiple speech and language dimensions. Fridrikkson and colleagues (2013) note that the WAB-R fluency scale is widely accepted by clinical and scientific communities "as a flawed but usable shorthand" (p. 3452). As it relies on matching PwA to qualitatively different categories of performance, the continuous nature of the scale is questionable; however, it is used as such. Scores on the scale are taken to reflect the degree to which output is fluent. We expected from our own analyses in Part A (Clough & Gordon, 2020) that the WAB-R fluency scale would reflect the severity of aphasia most strongly, as it did in when predicting dichotomous classifications of aphasia, but also be influenced by word retrieval and grammatical formulation abilities.

The three objective measures reflect different characteristics of spoken language behaviour. Utterance length reflects the ability to connect speech at the sentence level. Thus, it is frequently used as a proxy for grammatical competence, a common problem in nonfluent aphasia (Goodglass, 1993; Helm-Estabrooks, 1992). Although a long utterance does not necessarily ensure an accurate or complex utterance, it is strongly associated with these dimensions but easier to calculate, and is thus an attractive index of fluency for both clinical and research purposes. Speech rate has frequently been identified as an important contributor to impressions of fluency (e.g., Gordon, 2006; Park et al., 2011). This measure is broader in scope than MLU and can reflect the efficiency of connecting ideas in discourse, connecting words to form sentences, or combining sounds to form words. We expected this measure to reflect lower-level aspects of speech production ability as well as higher linguistic levels.

Retracing, in which an utterance is stopped and restarted, is also an intuitive and salient marker of dysfluency. Interruptions and repairs may occur because of message-level difficulties (e.g., losing a train of thought), lexical retrieval problems, difficulty structuring a grammatical sentence, or difficulty formulating an accurate phonological representation. Repetition and repair processes are frequently analyzed in the assessment of second language fluency (e.g., Bosker, Pinget, Quené, Sanders, & De Jong, 2012) and stuttering (Howell, 2004), but there has been relatively little focus on these processes in aphasia. In describing the Quantitative Production Analysis procedures, Rochon and colleagues note that false starts are “frequent manifestations of nonfluent aphasic speech” (2000, p. 196). Likewise, Feyereisen and colleagues (1991) describe false starts and repetitions among the behaviours that give rise to impressions of effortful speech. However, the relationship of repair behaviours to fluency is far from clear. Marshall and Tompkins (1982) noted that self-correction behaviours were related more to aphasia severity than to fluency. Wilson and colleagues (2010) included repaired sequences as a measure contributing to impressions of nonfluent speech in their examination of different types of primary progressive aphasia (PPA), but found that repairs were more frequent relative to control speakers only in logopenic PPA, but not nonfluent PPA. In the current study, we expected to find that repairs, because they were measured at the utterance level, would be most strongly predicted by grammatical factors, but perhaps also by aspects of lexical retrieval and/or speech production.

Despite their clinical utility, these characteristics of spontaneous speech have been described as “shallow measures” that are “inadequate when the assessment is aimed at the identification of the defective mechanism” (Feyereisen et al., 1991, p. 14). The same might be said of the WAB-R fluency scale, given its multidimensionality (e.g., see Trupe, 1984). Thus, our goal is to unpack these superficial representations of fluency, providing a window into the underlying dimensions that affect them.

Methods

This study is part of a larger project funded by the American Speech Language Hearing Foundation. It was approved by the Institutional Review Board (IRB) of the University of Iowa.

Participants

Connected speech samples were downloaded from AphasiaBank for the current study. The selection process is described in a companion paper (Clough & Gordon, 2020). To summarise, we analysed Cinderella story samples from 254 unique individuals with aphasia, representing a range of aphasia types and severity. The most commonly occurring types of aphasia in this set were Broca's aphasia and anomic aphasia. Aphasia Quotients from the WAB-R ranged from 10.8 to 99.6. Their ages ranged from 25 to 90 years, with an average of 61 years. Time post-onset ranged from 4 months to 27 years, with an average of just over 5 years.

For some of the analyses below, we divided the participants into fluent (n = 139) and nonfluent (n = 115) subgroups according to the impressions of the clinician who tested them. The guidelines for making these judgements were intentionally open-ended so that impressions would reflect typical clinical practice, with all its inherent variability. The aphasia data in this sample were contributed from 21 different sites, from clinicians whose experience averaged 15 years (range: 6 months to 40 years). In addition to capitalising on their ecological validity, relying on clinical impression was a matter of practicality, as there is no single established *objective* measure that reliably captures the multidimensionality of fluency.

Outcome measures

We examined four continuous measures, all of which have been widely used to index fluency in aphasia. The WAB-R has two Spontaneous Speech scales, one focusing on the information content of spontaneous speech and the other on the manner of its production (grammaticality, lexical retrieval, timing). Here we use the latter, commonly referred to as the *WAB-R "fluency" scale. Utterance length* was measured as the mean number of words per utterance (MLU), where utterances are defined on the basis of syntactic, semantic, and intonational cues, following the Quantitative Production Analysis (Berndt, Wayland, Rochon, Saffran, & Schwartz, 2000; as explained by MacWhinney, Fromm, Forbes, & Holland, 2011). *Speech rate* was measured by counting the number of words produced per minute. *Retracing* occurs when a speaker interrupts an utterance and reformulates it, changing the structure but not the meaning (MacWhinney, 2000). In determining dichotomous fluency classifications, retracing was included as a predictor (Part A, Clough & Gordon, 2020). However, because this measure is itself multiply determined (as described above), its contribution to other proxy fluency measures would not significantly clarify the underlying dimensions of fluency in the current analysis. Rather, we felt it would be instructive to identify the speech-language components underlying this more superficial manifestation of fluency.

Speech-language predictors

Speech and language characteristics of the Cinderella story retelling task were analysed. We selected this spontaneous speech task from the AphasiaBank protocol because it balances the trade-off between naturalness (as in conversational speech) and control of the target content (as in picture description). Standardising the content facilitates comparison across speakers, relative to more open-ended tasks like conversation. Compared to single-picture description, storytelling promotes a more lexically diverse sample, likely because it reduces the tendency to simply list items and events (Fergadiotis & Wright, 2011; Thompson et al., 2012).

Measures of spontaneous speech were selected to represent the major dimensions of spoken language production, any of which might contribute to fluency disruption: *grammatical competence* (accuracy and complexity of sentences, morphological errors, and degree of verb inflection), *lexical retrieval* (accuracy, specificity, efficiency, and diversity of words used), and *facility of speech production* (phonological and neologistic errors, presence of apraxia and dysarthria, and pitch variation). More details about the variables, and their selection and measurement, are provided in the companion paper (Clough & Gordon, 2020). Briefly, we selected a set of variables representing the dimensions described above that could be extracted using CLAN (MacWhinney, 2000) or other automated measures, and that showed relatively low intercorrelations (rs $<$ $|.500|$). An intercorrelation table is also available in the companion paper (Clough & Gordon, 2020). As described above, one of the predictor variables analysed in Part A – retracing – was analyzed as an outcome here, because it is potentially influenced by multiple underlying causes. Another measure (the ratio of content to function words) may also be affected by either lexical or grammatical abilities but was retained here as a predictor because it is more transparent to interpret. The ratio may be affected by a decline in content words (a lexical retrieval problem) or a decline in function words (a grammatical formulation problem). Thus, the direction of the effect provides clues as to the underlying cause. We also included covariates to control for overall sample length (the total number of utterances) and aphasia severity, measured by the WAB-R Aphasia Quotient (AQ). For reference, a list of the variables, along with a brief description of each and the fluency component it is intended to represent, is shown in Table 1.

Analyses

As a preliminary analysis, we compared the means of each proxy fluency measure for fluent and nonfluent PwA (as classified by clinical impression), to illustrate the ability of the four proxy fluency measures to distinguish between fluency categories. Next, we plotted the distributions of each fluency measure for all PwA, and for the fluent and

Table 1. Predictor variables categorised by linguistic dimension (*adapted from* Clough & Gordon, 2020).

Underlying component	Dimension	Predictor Variable
Grammatical competence	Grammatical accuracy	Presence of ungrammatical utterances: Y/N (Gram Err)
	Grammatical complexity	Proportion of complex grammatical relations (Gram Comp)
	Morphological accuracy	Presence of morphological errors: Y/N (Morph Errs)
	Morphological complexity	Proportion of verbs inflected (Inflect Vbs)
Lexical retrieval ability	Lexical accuracy	Presence of semantic errors: Y/N (Sem Errs)
	Lexical specificity	Presence of empty utterances: Y/N (Empty)
	Lexical efficiency	Presence of circumlocutory utterances: Y/N (Circum)
		Propositional density (Prop Dens)
	Lexical diversity	Moving Average Type-Token Ratio (MATTR)
Facility of speech production	Phonological encoding	Presence of phonological errors: Y/N (Phon Errs)
		Presence of neologistic errors: Y/N (Neo Errs)
	Motor speech	Apraxia of Speech: Y/N (AoS)
		Dysarthria: Y/N (Dys)
	Melodic line	Pitch variation: SD of F_0 (Pitch Var)
Combined measure	Grammatical & lexical	Content:function word ratio (ConFun)
Covariates	Global	Total Utterances (Total Utts)
		Aphasia Severity (WAB-R AQ)

nonfluent subgroups, to illustrate the extent to which they represent truly continuous measures, both within and across subgroups. Examining the distributions separately for fluent and nonfluent PwA allows us to investigate whether, for example, there might be floor or ceiling effects on a given measure within a subgroup, which would carry clinical implications for that measure's utility.

In our main analysis, we examined the contributors to each fluency outcome measure by conducting linear regression models. We followed the same model-building procedure as in our investigation of dichotomous measures of fluency in the companion paper (Clough & Gordon, 2020). For each model, the 16 spontaneous speech measures and aphasia severity (AQ) were entered into the model together and were removed one by one according to their *p*-values (highest *p*-value first), until only significant predictors remained. To ensure that we were capturing only robust results, we set a conservative alpha criterion of $p < .01$. Each final model was assessed to see if it met model assumptions for linear regression; if not, we used model diagnostics[1] to identify any problems, then re-ran the model. Below, we also report exploratory analyses of the extent to which predictors of fluency differed for fluent and nonfluent subgroups. To do this, we added interaction terms to the final models between clinical fluency category (F *vs* NF) and each significant predictor.

Results

The Cinderella story samples in this study ranged from 5 s to 20 min, with a mean of 3.4 min in length. Data were available for all PwA on all fluency proxy measures, but a few speakers were missing data for some of the speech-language predictors. MATTR scores were missing for 3 (0.8%) individuals; content:function ratios for 4 (1.2%) individuals; and dysarthria and apraxia diagnoses for, respectively, 20 (7.9%) and 18 (7.1%) PwA. In all, 1.0% of data points were missing. Missing data points were left out of the analyses, i.e., we did not use data imputation.

Table 2 demonstrates the ability of each proxy fluency measure to differentiate between fluent and nonfluent aphasia subtypes. The WAB-R AQ is also shown and illustrates that people with nonfluent aphasia are generally more severely aphasic than people with fluent aphasia. Differences between fluent and nonfluent groups were robustly significant ($p < .001$) on each measure, although the WAB-R fluency scale, MLU, and WPM clearly distinguished between fluent and nonfluent speakers better than retracing did. In addition, it is noteworthy that retracing was more common among fluent than nonfluent speakers, a finding we explore in more depth in the analyses below.

Table 2. Means (in raw units), standard deviations, and *t*-test results for aphasia severity and each fluency outcome measure for fluent and nonfluent subgroups.

	Clinician fluency classification		
Outcome measure	Fluent (*n* = 139) M (SD)	Nonfluent (*n* = 115) M (SD)	F vs NF *p*-value
WAB-R AQ (severity)	80.2 (15.37)	62.7 (16.93)	<.001
WAB-R fluency scale	7.9 (1.55)	4.5 (2.01)	<.001
Utterance length (MLU)	7.6 (2.08)	4.3 (1.91)	<.001
Speech rate (WPM)	75.3 (34.11)	37.2 (20.07)	<.001
Percent retraced utterances	.279 (.134)	.190 (.148)	<.001

Distributions

Fluency scale

WAB-R scores on the Spontaneous Speech fluency scale ranged across the full scale from 0 to 10, as shown in Figure 1(a). The distribution is negatively skewed, which may in part be an artefact of the AphasiaBank protocol – many individuals with more severe aphasia cannot complete the protocol so are not well represented in the database – and may also reflect the chronicity of many of the PwA. The mean fluency score is 6.3, and the median is 6.0. The distribution is roughly bimodal, with most of the sample clustering either at the high end or in the middle of the range (i.e., at the high end of the nonfluent range). Thus, the two modes correspond to the respective modes for each (negatively skewed) subgroup distribution. The scale distinguishes between fluent and nonfluent aphasia for the majority of PwA (as indicated by the *t*-tests reported above). However, nonfluent aphasia ranges from 0 to 9, and fluent aphasia from 4 to 10, so there is considerable overlap of subgroups on the scale (see Part A for further discussion, Clough & Gordon, 2020).

Although the two modes seem to have clinical validity, there are relative gaps in the scale at scores 3 and 7. This “lumpy” distribution arises because the scale consists of qualitative descriptions that fit certain aphasia syndromes varying in the frequency of occurrence. For example, the description for rating 4 (“halting, telegraphic speech … characteristic of agrammatic nonfluent aphasia”) corresponds closely to Broca’s aphasia, and rating 9 fits anomic aphasia. These are the most commonly occurring syndromes, at least among more chronic PwA. By contrast, the descriptions for rating 3 (“Longer, recurrent stereotypic or automatic utterances”) and rating 7 (“Phonemic jargon … characteristic of severe Wernicke’s aphasia”) represent less commonly observed patterns.

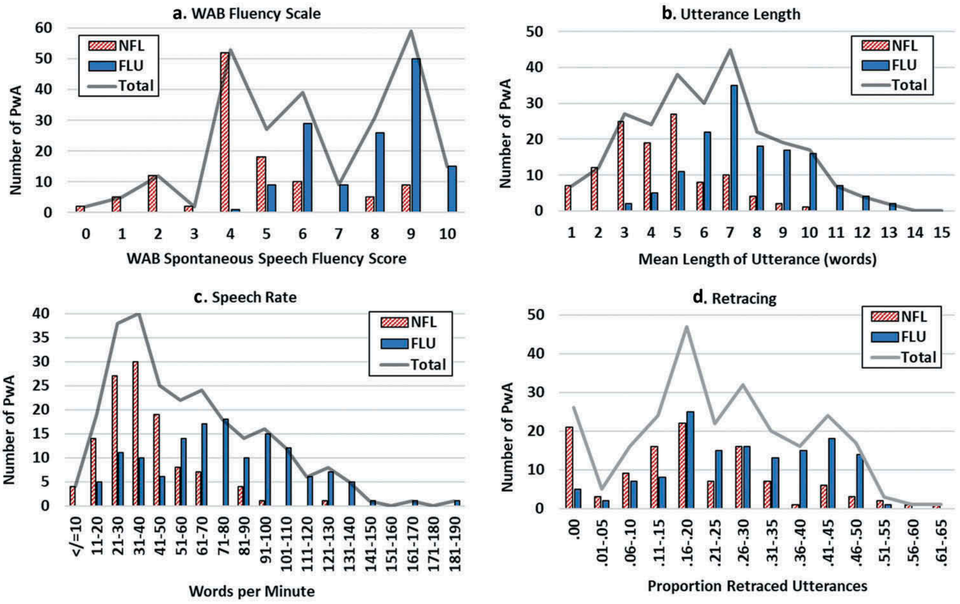

Figure 1. Distributions of continuous fluency measures ((a) WAB-R fluency scale; (b) MLU, (c) Speech rate; (d) Retracing) by clinician fluency diagnosis (NFL=nonfluent, FLU=fluent).
[To view this figure in color, please see the online version of this journal.]

Utterance length

Compared to the fluency scale, MLU shows a relatively normal distribution with a mode of 7 words per utterance, and an equivalent mean and median of 6.1 words (Figure 1(b)). Each subgroup also shows a normal distribution, with nonfluent aphasia showing a modal value of 3 words and fluent aphasia 7 words. Although there is considerable overlap in the middle of the distribution, utterance length appears to distinguish fluent and nonfluent syndromes more cleanly than the WAB-R fluency scale.

Speech rate

The distribution of the rate of speech (Figure 1(c)) is positively skewed and unimodal, with most values clustering towards the slower end (10–40 words per minute), and a few individuals showing more rapid rates of speech. The modal value is between 30 and 40 wpm, with a higher median (51 wpm) and mean (58 wpm), as these values are influenced by the positive skewness of the distribution. Both the clinically defined subgroups are positively skewed as well; however, the fluent subgroup distribution is multimodal, with peaks at 20–30, 70–80, and 90–100 wpm, suggesting that there may be factors other than aphasia fluency subtype that contribute to speech rate. There is considerable overlap between fluent and nonfluent subgroups, suggesting that speech rate does not distinguish these syndromes well.

Retracing

The distribution of retracing proportions (Figure 1(d)) is irregular, with a mean of 24%, median of 22%, and a modal value between 15% and 20% across the full group. Both subgroups also had modes at 15-20%, but the nonfluent subgroup had another mode at 0% and the fluent subgroup had another mode at 40-45%. Although the subgroups differ in mean value, their distributions overlap almost completely across the range.

Linear regressions

Before conducting the regression analyses, we examined the relationships between each predictor variable and each of the fluency outcome measures (see Supplementary Table). We also generated histograms of our raw predictor variables to look for highly skewed distributions and obvious outliers (those separated from the rest of the distribution by a gap). Because many of the error variables showed high proportions of zeroes (Gram Err: 13%, Morph Err: 72%, Circum: 77%, Empty: 61%, Sem Err: 35%, Neo Err: 56%, Phon Err: 40%), we transformed all the error proportions into dichotomous variables – either they produced some instances or none of the given error behavior. From the remaining continuous variables, 7 outlying data points were removed: 1 each from the variables MATTR and Pitch Var; 2 from Total Utt; and 3 from ConFun.

In all the initial regression models except the Retracing analysis, violations were noted. The WAB-R and WPM model violations were addressed by deleting a few influential outlying cases (3 in each model). In addition, the outcome variable WPM was transformed using a square root transformation to correct its positive skew. For the MLU model, we also log-transformed the content:function ratio, and deleted additional outlying cases (14 in total). The total number of cases included following model correction is reported in the regression tables below and ranges from 88% to over 99% of the original sample. It should be noted that the outcomes of the original and corrected models did not differ substantially in their findings; for each outcome measure, the same set of predictors was identified with similar coefficient estimates and effect sizes.

Fluency scale

Table 3 shows the final model predicting scores on the continuous WAB-R fluency scale for all PwA. Predictors are ordered from the strongest to the weakest effect size of each predictor. WAB-R fluency scale scores were most strongly influenced by aphasia severity ($p < .001$, $\eta^2_p = .556$), reflecting the strong tendency for nonfluent PwA to have more severe aphasia. Fluency scores were also predicted by greater lexical diversity ($p < .001$, $\eta^2_p = .100$), more complex grammar ($p < .001$, $\eta^2_p = .097$), fewer cases of dysarthria ($p < .001$, $\eta^2_p = .062$), and more semantic errors ($p < .001$, $\eta^2_p = .057$). Together, these variables accounted for 74% of the variance in WAB-R fluency scores.

MLU

Table 4 displays the final model for the outcome measure of mean utterance length. Consistent with our predictions, grammatical complexity was the strongest predictor ($p < .001$, $\eta^2_p = .590$). Also positively related to utterance length were WAB-R AQ ($p < .001$, $\eta^2_p = .159$), propositional density ($p < .001$, $\eta^2_p = .096$), and lexical diversity ($p < .001$, $\eta^2_p = .086$). Content:function ratio was negatively associated with utterance length ($p < .001$, $\eta^2_p = .043$), indicating that utterances were longer for those speakers who could produce relatively more function words. The final model accounted for 82% of the variance in MLU.

Table 3. Final model predicting WAB-R fluency scale scores.

Predictors of WAB-R fluency scale	Overall model fit (*n* =229) $F(5,223) = 130.5$, $p < .001$ *Adjusted* $R^2 = 0.74$		
Predictor	*t*-value	*p*-value	η^2_p
(Intercept)	−6.26	< .001	0.150
WAB-R Aphasia Quotient	16.71	< .001	0.556
Lexical diversity (MATTR)	4.96	< .001	0.100
Grammatical complexity	4.89	< .001	0.097
Dysarthria (yes)	−3.83	< .001	0.062
Semantic errors (yes)	3.67	< .001	0.057

Table 4. Final model predicting utterance length.

Predictors of MLU	Overall model fit (*n* =229) $F(5,223) = 210.6$, $p < .001$ *Adjusted* $R^2 = 0.82$		
Predictor	*t*-value	*p*-value	η^2_p
(Intercept)	−5.51	< .001	0.120
Grammatical complexity	17.90	< .001	0.590
WAB-R Aphasia Quotient	6.50	< .001	0.159
Propositional density	4.85	< .001	0.096
Lexical diversity (MATTR)	4.57	< .001	0.086
Content:function ratio (log)	−3.18	< .001	0.043

Speech rate

The final model predicting speech rate is shown in Table 5. Speech rate was not predicted by severity. As for utterance length, the most important predictor was grammatical complexity ($p < .001$, $\eta^2_p = .229$), showing that speakers with more complex grammar also spoke more quickly. Consistent with this, content:function word ratios were negatively associated with speech rate ($p < .001$, $\eta^2_p = .114$), illustrating that the speech rate was facilitated by retrieval of relatively more function words. Speech rate was also positively associated with lexical diversity ($p < .001$, $\eta^2_p = .063$) and propositional density ($p < .001$, $\eta^2_p = .059$) indicating that word retrieval skills facilitate faster speech. A positive association was also found for pitch variation ($p < .001$, $\eta^2_p = .060$), and a negative association with speech apraxia ($p = .004$, $\eta^2_p = .039$), implicating speech production processes in the rate of production. A small association was also noted between speech rate and the total number of utterances produced ($p = .008$, $\eta^2_p = .032$). The final model accounted for 56% of the variance in speech rate.

Retracing

The final model predicting proportions of retraced utterances is shown in Table 6. As for MLU and speech rate, the strongest predictor was grammatical complexity ($p < .001$, $\eta^2_p = .092$), but with more complex grammar associated with more retracing. Speakers who produced more circumlocutions also retraced more often ($p = .001$, $\eta^2_p = .040$). These associations are consistent with the greater incidence of retracing on average in those with fluent aphasia that was indicated by our preliminary *t*-tests. However, retracing was also positively associated with grammatical error production ($p = .001$, $\eta^2_p = .041$), and speakers with less variable pitch produced more retracing ($p = .001$, $\eta^2_p = .041$). This is what one would expect if

Table 5. Final model predicting speech rate.

Predictors of speech rate (WPM-SqRt)	Overall model fit (*n* =223) *F*(7,217) = 40.8, *p* < .001 *Adjusted* R^2 = 0.56		
Predictor	*t*-value	*p*-value	η^2_p
(Intercept)	−1.70	.091	0.013
Grammatical complexity	7.99	< .001	0.229
Content:function ratio	−5.27	< .001	0.114
Lexical diversity (MATTR)	3.80	< .001	0.063
Pitch variation (SD)	3.70	< .001	0.060
Propositional density	3.67	< .001	0.059
Apraxia of speech (yes)	−2.94	.004	0.039
Total utterances	2.68	.008	0.032

Table 6. Final model predicting retracing.

Predictors of retracing	Overall model fit (*n* =253) *F*(4,248) = 19.0, *p* < .001 *Adjusted* R^2 = 0.22		
Predictor	*t*-value	*p*-value	η^2_p
(Intercept)	4.17	< .001	0.066
Grammatical complexity	5.00	< .001	0.092
Grammatical errors (yes)	3.27	.001	0.041
Pitch variation (SD)	−3.25	.001	0.041
Circumlocution (yes)	3.22	.001	0.040

less fluent speakers tend to repair their utterances more often, suggesting that retracing occurs in both fluent and nonfluent aphasia, but for different reasons. These predictors collectively account for a relatively small amount of the variance (22%) in retracing.

Fluency interactions

Post-hoc regression analyses of each fluency outcome measure were conducted using the predictors of each final model and their interactions with clinical fluency impression. (As exploratory analyses, we report here all interactions significant at $p < .05$ and present two sample graphs as supplementary materials to illustrate the interactions.) In predicting WAB-R fluency scale scores, significant interactions were found for semantic errors ($p = .015$), grammatical complexity ($p = .017$), and aphasia severity ($p = .022$). Examination of the interactions illustrated that semantic errors (Supplementary Figure A) and aphasia severity were more predictive of WAB-R fluency for those in the nonfluent than the fluent category. On the other hand, grammatical complexity was a better predictor for fluent than nonfluent individuals. Similarly, aphasia severity was a stronger predictor of MLU for nonfluent PwA ($p = .034$), whereas propositional density was a stronger predictor of MLU for fluent PwA ($p = .027$). Propositional density was also a significantly stronger contributor to speech rate for fluent than nonfluent PwA ($p < .001$, Supplementary Figure B). None of the predictors interacted with fluency category in accounting for retracing.

Discussion

We examined spontaneous speech predictors of several common continuous measures of fluency in a large sample of individuals with aphasia. Most fluency measures were affected by aspects of word retrieval, grammatical formulation, and speech production, although there were differences in which predictors dominated for the different outcome measures. For the WAB-R fluency scale, the strongest predictor by far was aphasia severity; for utterance length, speech rate, and retracing, it was a grammatical complexity. To make the findings across the analyses more accessible, our results are summarised in Table 7. Here we show the significant predictors (*not* including covariates) for each outcome variable, color-coded by predictor speech-language dimension (grammatical competence, lexical retrieval, speech production) and indicating the size of each effect. The R^2 values in Table 7 are different from those shown in Tables 3–6, because they reflect the variance accounted for in each outcome measure by just the specific speech-language predictors, excluding the covariates (severity and total utterances).

To summarise, of the four fluency measures assessed, three (WAB-R fluency, retracing, and speech rate) were affected by a combination of variables representing all three underlying components of spoken language production. MLU was significantly predicted only by grammatical and lexical measures. This measure of fluency can be conceptualised, then, as reflecting linguistic aspects of fluency. By contrast, speech rate showed a stronger influence of variables at the speech production level, including pitch variation and apraxia of speech, although grammatical and lexical variables contributed as well. Examining the variance accounted for in each model (see Table 7) illustrates that MLU was predicted best by the speech-language variables (79% of the variance), followed by speech rate (54%) and WAB-R fluency (42%). The variance in WAB-R fluency accounted for by the speech-language

Table 7. Speech-language predictors of each outcome measure (not including covariates), shown by predictor category[a] and effect size[b].

Predictor	Effect sizes
WAB-R fluency scale	**R^2 = .42**
Lexical diversity (MATTR)	Medium (LEX)
Grammatical complexity	Medium (GRAM)
Dysarthria	Medium (SPCH)
Semantic errors	Small (LEX)
Utterance length (MLU)	**R^2 = .79**
Grammatical complexity	Very Large (GRAM)
Propositional density	Medium (LEX)
Lexical diversity (MATTR)	Medium (LEX)
Content:function ratio	Small (GRAM)
Speech rate (WPM)	**R^2 = .54**
Grammatical complexity	Large (GRAM)
Content:function ratio	Medium (LEX)
Lexical diversity (MATTR)	Medium (LEX)
Pitch variation	Medium (SPCH)
Propositional density	Small (LEX)
Apraxia of speech	Small (SPCH)
Retracing	**R^2 =.22**
Grammatical complexity	Medium (GRAM)
Grammatical errors	Small (GRAM)
Pitch variation	Small (SPCH)
Circumlocution	Small (LEX)

[a] GRAM indicates grammatical characteristics; LEX indicates lexical characteristics; SPCH indicates speech production characteristics.

[b] Effect size benchmarks: small = .01-.05; medium = .06-.13; large = .14-.29; very large > .30 (Cohen, 1988).

predictors alone is considerably less than the original model (Table 3) because the strongest predictor – overall severity – is omitted here. As we discuss in the companion paper (Clough & Gordon, 2020), the fact that the WAB-R fluency scale is dominated by severity limits its utility as a measure of fluency. A relatively small proportion of the variance in retracing (22%) was accounted for by the speech-language predictors, suggesting that this behaviour is affected by other factors. For example, the tendency to retrace may reflect conceptual-level processing difficulties or premorbid characteristics affecting speech style.

In general, the finding that predictors from multiple linguistic and speech levels affect most measures of fluency supports previous observations that fluency is multiply determined (Casilio et al., 2019; Feyereisen et al., 1991; Goodglass et al., 2001a; Gordon, 1998; Trupe, 1984; Vermeulen et al., 1989; Wagenaar et al., 1975). Our findings are consistent, for example, with those of Casilio and colleagues (2019), who also examined a wide range of speech-language variables and identified 4 factors underlying connected speech, which they labelled paraphasia, logopenia, agrammatism, and motor speech. The latter three were interpreted as reflecting different components of fluency, although the factors did not map onto the aphasia subtypes of the 24 participants they examined with any consistency. Thus, it is unclear how these factors relate to fluency as a diagnostic term. Nevertheless, our study supports their conclusion that word retrieval, grammatical competence, and motor speech all contribute to fluency, and extends these findings to a much larger sample.

Our results also illustrate that different measures of fluency weight the contributions of these underlying sources of variability differently. Across the analyses, an important role is evident for grammatical complexity, which affects all of the fluency measures to some degree and exerts large effects on utterance length and speech rate. This variable was calculated by dividing the number of grammatical relations that encode embeddings (as identified in the CLAN manual, MacWhinney, 2000) by the total number of grammatical relations. Among these, the most frequently occurring were finite clausal complements to a verb (e.g., "She doesn't want to tell him *why she has to go*" or "She goes cryin' *she can't go ball at all*") and clausal modifiers (e.g., "They have her go to the party and be with the prince *who is the heavenly host*" or "They were probably hoping that *they were gonna be ones that would somebody this handsome man was gonna marry her*"). Notably, this measure captures embeddings regardless of the accuracy with which they are implemented. The importance of grammatical complexity reflects the frequent and well-established co-occurrence of agrammatism with nonfluency, and is consistent with the results of Nozari and Faroqi-Shah (2017). They examined the influence of measures of word retrieval, syntactic production, comprehension, working memory, and self-monitoring on fluency, and found that only syntactic measures had a direct effect on fluency. However, their word retrieval measures may have been non-significant contributors because they were obtained in single-word picture-naming tasks rather than connected speech. (It is unsurprising that comprehension and working memory were not predictive of fluency in that study.)

It is noteworthy that one of the measures of fluency – retracing – was revealed to be more strongly related to *fluent* characteristics, such as more complex grammar and circumlocution, than to *nonfluent* language characteristics (although retracing was also more common in speakers with less variable pitch and more grammatical errors). The coexistence of these seemingly contradictory effects suggests that fluent and nonfluent speakers retrace for different reasons, but that it is more common in fluent speakers. Fraundorf and Watson (2014) proposed that retracing is a production strategy that is more available to speakers who are more fluent. After examining the frequency and location of three types of dysfluencies (fillers, silent pauses, and repeats/repairs [similar to retraces here]) in undergraduates during a story retelling task, they concluded that repetition is a "preferred" strategy. Repetitions maintain production once the utterance has been initiated and when the material is available to be repeated or revised, whereas pauses indicate grammatical or lexical planning difficulties and are used "when other strategies are not available" (p. 1092). Although the language production of undergraduates is clearly different from PwA, parallels can be drawn: fluent PwA tend to have access to a wider range of lexical and grammatical options than nonfluent PwA (hence the well-known contrast between substitution and omission behaviors [e.g., Kleist, 1916; Kolk & Heeschen, 1992]).

These findings have implications for methods of assessing fluency. For example, if the goal is to provide a global picture of fluency, particularly across a group of PwA who vary in their underlying deficits, then a multidimensional measure such as the WAB-R fluency scale or speech rate is more likely to capture the broad range of ways fluency can be disrupted. However, the distribution of scores across the WAB-R scale indicates that it is better viewed as a categorical than a continuous measure, reflecting subtypes of aphasia approximately arrayed by aphasia severity. Although speech rate captures several ways in which PwA vary continuously on fluency dimensions, the distribution of speech rates indicates that it is not sufficiently specific to differentiate between fluency categories. Furthermore, such multidimensional measures will be less useful in diagnosing the reasons for fluency disruptions. MLU may be useful to focus on linguistic (*vs* speech) deficits but is clearly affected by both grammatical and lexical factors. To identify underlying causes of dysfluency at an individual level would require a set of specific measures, such as grammatical complexity and lexical diversity, in combination with indicators of fluency. The multidimensionality of fluency is also problematic for efforts to identify neural substrates of fluency (e.g., Basilakos et al., 2014; Fridrikkson et al., 2013; Lacey et al., 2017; Wang et al., 2013; Wilson et al., 2010). Vastly different findings are likely attributable, at least in part, to the widely varying measures and combinations of measures used. It would be more meaningful to identify neural substrates of the components of fluency than of "fluency" as a global construct.

Examining interactions between fluency category and the predictors suggests that the importance of different variables might also be different at different levels of fluency. That is, what interferes with fluency in those who are already relatively fluent may be different from what affects fluency for less fluent speakers. Our findings indicate, for example, that higher WAB-R fluency scale scores are obtained for speakers who can produce more complex grammatical constructions, but that this association is stronger for fluent speakers. Presumably, the production of nonfluent speakers is limited by other factors. Similarly, greater propositional density (the ability to formulate propositions by connecting nouns with verbs and adjectives) facilitates speech rate among fluent speakers more than nonfluent speakers. Although nonfluent speakers show an even greater range of propositional density than fluent speakers in our sample (see Supplementary Figure B), it is not as strongly associated with speech rate. It may be that the effort to formulate propositions entails a trade-off for some nonfluent speakers, such that more propositions can be produced, but this entails a slowing of speech output.

The finding that semantic errors are associated with higher WAB-R fluency scores, but only for nonfluent PwA, may seem counter-intuitive, but probably also reflects a trade-off. The ability to produce words, even in error, allows the flow of speech to continue compared to the speaker who stops to search for the correct word. (This resonates with Kolk and Heeschen's [1992] characterization of the difference between agrammatism and paragrammatism as a strategic choice between nonfluent elliptic or fluent but errorful production.) The association of semantic errors with higher fluency also highlights the fact that fluency is influenced by the combined effects of both positive and negative symptoms of aphasia. Thus, greater fluency may reflect productive aspects such as more complex grammar, but also counter-productive aspects such as empty speech and semantic errors. Although these analyses were exploratory, they suggest how consideration of the underlying causes of dysfluency at an individual level can help identify clinical targets for therapy.

Other sources of variability that might influence outcomes include the nature of the task used and the heterogeneity of participants studied. In particular, studies using conversational tasks (e.g., Borovsky et al., 2007; Casilio et al., 2019) are likely to identify different spontaneous speech characteristics than studies examining expression in monologue tasks like picture description or story retelling (e.g., Gordon, 2006; Kong & Wong, 2018; Park et al., 2011). For example, conversation often consists of shorter and more elliptical utterances. In terms of participant sampling, the distribution of syndromes, or the average and range of severity level, may vary across studies. For example, even in studies examining relatively unselected groups, the proportion of the sample consisting of individuals with anomic aphasia varies widely, from less than a third (e.g., Casilio et al., 2019; Halai et al., 2017) to almost half (e.g., Borovsky et al., 2007; Lacey et al., 2017). In the current sample, just over a third of the PwA were characterized as having anomic aphasia. Some studies include only individuals in the chronic phase (Fridrikkson et al., 2013; Halai et al., 2017; Wang et al., 2013); others examine acute aphasia (John et al., 2017). Somewhat atypically, Park and colleagues (2011) combined stroke participants with a variety of other populations, such as semantic dementia, Alzheimer's disease, and non-brain-damaged participants, in their investigation of fluency. Wang and colleagues (2013) excluded participants with "pure anomia", whereas other studies (like the current study) included participants whose aphasia had resolved according to the WAB-R criterion of an AQ score greater than 93.8 (Basilakos et al., 2014; Borovsky et al., 2007; Lacey et al., 2017).

Differences such as these may influence which dimensions are identified as being most salient or which underlying brain areas are identified as most important for fluent expression. The smaller the subject sample, the greater the impact of such sources of variability. Thus, one of the strengths of the current study is the large sample of PwA included, thanks to the dataset provided in AphasiaBank.

Limitations and caveats

There are several limitations to the current study. Foremost is the lack of a continuous *perceptual* measure that focuses specifically on fluency. Instead, we rely on a variety of quantitative measures which have been widely used, in both clinical and research domains, to reflect fluency. Clinical impressions were available as dichotomous perceptions of fluency (analyzed in Part A and used here to investigate fluency within aphasia subtypes), but do not provide information on *how fluent* the speakers were perceived to be. The WAB-R fluency scale score provides the closest approximation to a perceptual scale, but it is not continuous, and its interpretation is complicated by the fact that it incorporates aspects of grammaticality, word retrieval, prosody, and articulatory effort (Trupe, 1984). Although these dimensions are probabilistically related to fluency (as our results reinforce), they may not be directly related to the judgement of fluency. For future study of this issue, we are currently conducting surveys to collect perceptions of fluency, along with perceptions of other aspects of production.

We also acknowledge that there are aspects of spontaneous speech production that are not well captured by our set of variables. Most notably, AphasiaBank provides few cues to articulatory agility beyond the presence or absence of motor speech disorders. The occurrence of errors of phonological encoding and the variation of pitch provide some additional information relevant to lower-level production processes, but these are no substitute for the impression of "effort" that is so important to fluency judgements (e.g., Goodglass et al., 2001a; Park et al., 2011). We also have minimal information to distinguish dysfluencies at grammatical and lexical levels from dysfluencies that might have their origin at the message level. Such factors might explain some of the considerable variance left unaccounted for in some of our models, particularly retracing.

As mentioned above, a limitation of the sample in the current study, although quite large, is the relatively low representation of global aphasia. This may account for the skewing of the nonfluent subsample towards the upper end of the nonfluent scale (see Figure 1(a)). With more individuals with global aphasia and a better measure of motor speech, we might have obtained different results, such as a stronger influence of co-occurring motor speech disorders.

Conclusions

In this study, we examined four continuous measures of fluency and the speech-language factors affecting these measures. Making use of a large and relatively unselected sample of individuals with aphasia allowed us to analyze a large number of potential predictors and helped to overcome some of the variability that can come from sampling differences. Results showed that, although most of the measures were multiply determined, they each reflected different aspects of fluency. Thus, they may be appropriate for different purposes in clinical and research settings. The discontinuous distribution of scores on the WAB-R fluency scale indicates that it is not well suited as a continuous measure of fluency. Rather, it consists of a set of qualitative patterns of expression corresponding to different aphasia syndromes that are more representative of severity than fluency. Speech rate is a multidimensional measure of fluency that shows a more continuous (although skewed) distribution but does not distinguish between fluent and nonfluent categories well. Mean length of utterance shows a more normal distribution of scores and captures a relatively large amount of variance in connected speech measures at grammatical and lexical levels, while speech rate reflects the influence of lower-level aspects of speech production better than MLU does. Retracing reflects more specific aspects of production and, importantly, is more closely associated with fluent than nonfluent speech production.

Particularly when extrapolating results for clinical purposes, it is important to consider that these findings reflect general patterns, and the utility of any given measure depends on the profile of the individual speaker. For example, pitch variation, as the only measure of melodic line, played a relatively small role in predicting measures of fluency at the group level, but there are some PwA for whom disrupted melodic line may be the most prominent feature of dysfluency. The disruption of fluency in different speakers reflects the potential contributions of different deficits in varying combinations.

Variation in fluency measurement arises from speakers with different types of aphasia, from different measures of fluency, and from different conceptions of what fluency is (Clough & Gordon, 2020; Gordon, 1998; the current study). Although this variation reduces the reliability of categorical fluency diagnoses, Feyereisen and colleagues (1991) highlight the utility of fluency measurement for descriptive purposes: "For the same reasons that made them inappropriate in the classification of aphasia … fluency measures display a versatility and an ecological relevance that allow for several adequate uses in the study and treatment of single cases of aphasia" (p. 17). In our view, the solution to the fluency reliability problem lies in clearly distinguishing between the surface manifestations of fluency and the underlying deficits which contribute to fluency measures, whether they be difficulties with lexical retrieval, sentence formulation, phonological encoding, articulatory implementation, or a combination of these. In our ongoing project, we are developing an assessment tool that relates fluency to a set of potential speech and language contributors, to improve the accuracy of diagnosis and facilitate treatment planning.

Note

1. Package *gvlma* in R was used to assess linearity, skewness, kurtosis, and homoscedasticity of residuals. To diagnose problems with the data, we used the commands *vif* to assess for variance inflation; *skewness* to assess normality; *ncvTest* to assess homoscedasticity; *Durbin–*

Watson test to assess the independence of errors; *OutlierTest, outliers,* and *hist* to identify outliers; and plotted Cook's Distance to identify influential cases.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was generously supported by a New Century Scholars Grant from the American Speech-Language-Hearing Foundation.

ORCID

Jean K. Gordon http://orcid.org/0000-0003-0073-8016

References

Basilakos, A., Fillmore, P., Rorden, C., Guo, D., Bonhila, L., & Fridrikkson, J. (2014, October). Regional white matter damage predicts speech fluency in chronic post-stroke aphasia. *Frontiers in Human Neuroscience*, 1–9. doi:10.1044/jshr.3602.338.

Benson, D. F. (1967). Fluency in aphasia: Correlation with radioactive scan localization. *Cortex*, *3*, 373–394. doi:10.1016/S0010-9452(67)80025-X

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *Quantitative production analysis: A training manual for the analysis of aphasic sentence production*. Hove, UK: Psychology Press.

Borovsky, A., Saygin, A. P., Bates, E., & Dronkers, N. (2007). Lesion correlates of conversational speech production deficits. *Neuropsychologia*, *45*, 2525–2533. doi:10.1016/j.neuropsychologia.2007.03.023

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*, 159–175. doi:10.1177/0265532212455394

Caplan, D. (1987). *Neurolinguistics and linguistic aphasiology: An introduction*. Cambridge, UK: Cambridge University Press.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychometric Bulletin & Review*, *20*, 1350–1356. doi:10.3758/s13423-013-0442-z

Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology,*1–19. doi:10.1044/jshr.3602.338.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*, 535–544. doi:10.1016/S0346-251X(97)00046-8

Clough, S., & Gordon, J. K. (2020). Fluent or nonfluent? Part A. Underlying contributors to categorical diagnoses of fluency in aphasia. *Aphasiology*.

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

Croteau, C., & Le Dorze, G. (2001). Spouses' perceptions of persons with aphasia. *Aphasiology*, *15*, 811–825. doi:10.1080/02687040143000221

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*, 237–254. doi:10.1080/15434303.2018.1477780

Deloche, G., Jean-Louis, J., & Seron, X. (1979). Study of the temporal variables in the spontaneous speech of five aphasic patients in two situations, interview and description. *Brain and Language*, *8*, 241–250. doi:10.1016/0093-934X(79)90052-X

Duffy, J. R., Boyle, M., & Plattner, L. (1980). Listener reactions to personal characteristics of fluent and nonfluent aphasic speakers. *Clinical Aphasiology Conference Proceedings, 1980*, Bar Harbor, ME, 117–126.

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, *25*, 1414–1430. doi:10.1080/02687038.2011.603898

Feyereisen, P., Pillon, A., & De Partz, M.-P. (1991). On the measures of fluency in the assessment of spontaneous speech production by aphasic subjects. *Aphasiology*, *5*, 1–21. doi:10.1080/02687039108248516

Feyereisen, P., Verbeke-Dewitte, C., & Seron, X. (1986). On fluency measures in aphasic speech. *Journal of Clinical and Experimental Neuropsychology*, *8*, 393–404. doi:10.1080/01688638608401329

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, *29*, 1083–1096. doi:10.1080/01690965.2013.832785

Fridrikkson, J., Guo, D., Fillmore, P., Holland, A., & Rorden, C. (2013). Damage to the anterior arcuate fasciculus predicts non-fluent speech production in aphasia. *Brain*, *136*, 3451–3460. doi:10.1093/brain/awt267

Fridrikkson, J., Hubbard, H. I., Hudspeth, S. G., Holland, A., Bonhila, L., Fromm, D., & Rorden, C. (2012). Speech entrainment enables patients with Broca's aphasia to produce fluent speech. *Brain*, *135*, 3815–3829. doi:10.1093/brain/aws301

Goodglass, H. (1993). *Understanding aphasia*. Boston, MA: Academic Press.

Goodglass, H., Christiansen, J. A., & Gallagher, R. E. (1993). Comparison of morphology and syntax in free narrative and structured tests: Fluent vs. nonfluent aphasics. *Cortex*, *29*, 377–407. doi:10.1016/S0010-9452(13)80250-X

Goodglass, H., Kaplan, E., & Barresi, B. (2001a). *The assessment of aphasia and related disorders* (3rd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Goodglass, H., Kaplan, E., & Barresi, B. (2001b). *Boston diagnostic aphasia examination* (3rd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Goodglass, H., Quadfasel, F. A., & Timberlake, W. H. (1964). Phrase length and type and severity of aphasia. *Cortex*, *1*, 133–153. doi:10.1016/S0010-9452(64)80018-6

Gordon, J. K. (1998). The fluency dimension in aphasia. *Aphasiology*, *12*, 673–688. doi:10.1080/02687039808249565

Gordon, J. K. (2006). A quantitative production analysis of picture description. *Aphasiology*, *20*, 188–204. doi:10.1080/02687030500472777

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2017). Using principal components analysis to capture individual differences with a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. *Cortex*, *86*, 275–289. doi:10.1016/j.cortex.2016.04.016

Harmon, T. G., Jacks, A., Haley, K. L., & Faldowski, R. A. (2016). Listener perceptions of simulated fluent speech in nonfluent aphasia. *Aphasiology*, *30*, 922–942. doi:10.1080/02687038.2015.1077925

Helm-Estabrooks, N. (1992). *Aphasia diagnostic profiles*. Austin, TX: PRO-ED.

Howell, P. (2004). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communication Sciences & Disorders*, *31*, 122–139. doi:10.1044/jshr.3602.338.

Howes, D. (1964). Application of the word-frequency concept to aphasia. In A. V. S. DeReuck & M. O'Connor (Eds.), *Disorders of language* (pp. 47–78). London, Eng.: J.A. Churchill.

John, A. A., Javali, M., Mahale, R., Mehta, A., Acharya, P. T., & Srinivasa, R. (2017). Clinical impression and Western aphasia battery classification of aphasia in acute ischemic stroke: Is there a discrepancy? *Journal of Neurosciences in Rural Practice*, *8*, 74–78. doi:10.4103/0976-3147.193531

Kertesz, A. (2006). *Western aphasia battery-revised*. San Antonio, TX: Pearson.

Khvalabov, N. (2019). *Listener judgements of fluency and perceptions of aphasia*. (BA), University of Iowa. Retrieved from https://ir.uiowa.edu/honors_theses/.

Kleist, K. (1916). Uber leitungsaphasie und grammatische storungen. *Monatschrift Fur Psychiatrie Und Neurologie*, *40*, 118–199. doi:10.1159/000190892

Kolk, H., & Heeschen, C. (1992). Agrammatism, paragrammatism and the management of language. *Language and Cognitive Processes*, *7*, 89–129. doi:10.1080/01690969208409381

Kong, A. P.-H., & Wong, C. W.-Y. (2018). An integrative analysis of spontaneous storytelling discourse in aphasia: Relationship with listeners' rating and prediction of severity and fluency status of aphasia. *American Journal of Speech-Language Pathology*, *27*, 1491–1505. doi:10.1044/2018_AJSLP-18-0015

Lacey, E. H., Skipper-Kallal, L. M., Xing, S., Fama, M. E., & Turkeltaub, P. E. (2017). Mapping common aphasia assessments to underlying cognitive processes and their neural substrates. *Neurorehabilitation and Neural Repair*, *3*, 442–450. doi:10.1177/1545968316688797

Lasker, J. P., & Beukelman, D. R. (1999). Peers' perceptions of storytelling by an adult with aphasia. *Aphasiology*, *13*, 857–869. doi:10.1080/026870399401920

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). New Jersey: Lawrence Erlbaum Associates.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, *25*, 1286–1307. doi:10.1080/02687038.2011.589893

Marshall, J. C. (1986). The Description and Interpretation of aphasic language disorder. *Neuropsychologia*, *24*, 5–24. doi:10.1016/0028-3932(86)90040-0

Marshall, R. C., & Tompkins, C. A. (1982). Verbal self-correction behaviors of fluent and nonfluent aphasic subjects. *Brain and Language*, *15*, 292–306. doi:10.1016/0093-934X(82)90061-X

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech & Hearing Research*, *36*, 338–350 doi:10.1044/jshr.3602.338.

Nozari, N., & Faroqi-Shah, Y. (2017). Investigating the origin of nonfluency in aphasia: A path modeling approach to neuropsychology. *Cortex*, *95*, 119–135. doi:10.1016/j.cortex.2017.08.003

Park, H., Rogalski, Y., Rodriguez, A. D., Zlatar, Z., Benjamin, M., Harnish, S., … Reilly, J. (2011). Perceptual cues used by listeners to discriminate fluent from nonfluent narrative discourse. *Aphasiology*, *25*, 998–1015. doi:10.1080/02687038.2011.570770

Poeck, K. (1989). Fluency. In C. Code (Ed.), *The characteristics of aphasia* (pp. 23–32). Philadelphia, PA: Taylor & Francis.

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, *72*, 193–218. doi:10.1006/brln.1999.2285

Swindell, C. S., Holland, A., & Fromm, D. (1984). *Classification of aphasia: WAB-R type versus clinical impression*. Clinical Aphasiology Conference Proceedings, 1984, Seabrook Island, SC.

Thompson, C. K., Cho, S., Chien-Ju, H., Wieneke, C., Rademaker, A., Weitner, B. B., … Weintraub, S. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology*, *26*, 20–43. doi:10.1080/02687038.2011.584691

Trupe, E. H. (1984). *Reliability of rating spontaneous speech in the western aphasia battery: Implications for classification*. Clinical Aphasiology Conference Proceedings, 1984, Seabrook Island, SC.

Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language*, *36*, 252–274. doi:10.1016/0093-934X(89)90064-3

Wagenaar, E., Snow, C., & Prins, R. (1975). Spontaneous speech of aphasic patient: A psycholinguistic analysis. *Brain and Language*, *2*, 281–303. doi:10.1016/S0093-934X(75)80071-X

Wang, J., Marchina, S., Norton, A. C., Wan, C. Y., & Schlaug, G. (2013, December). Predicting speech fluency and naming abilities in aphasic patients. *Frontiers in Human Neuroscience*, 1–13. doi: 10.3389/fnhum.2013.00831.

Wertz, R. T., Deal, J. L., & Robinson, A. J. (1984). *Classifying the aphasias: A comparison of the Boston diagnostic aphasia examination and the western aphasia battery*. Clinical Aphasiology Conference Proceedings, 1984, Seabrook Island, SC.

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., … Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain*, *133*, 2069–2088. doi:10.1093/brain/awq129

Zraick, R. I., & Boone, D. R. (1991). Spouse attitudes toward the person with aphasia. *Journal of Speech, Language, & Hearing Research*, *34*, 123–128. doi:10.1044/jshr.3401.123