USING DISCOURSE RATING SCALES TO MEASURE EFFECTIVENESS OF TREATMENT IN PEOPLE WITH APHASIA

A Thesis
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
MASTER OF ARTS OF SPEECH, LANGUAGE, HEARING SCIENCE

---

by
Jaclyn S. Navarro
Diploma Date August 2020

Thesis Approvals:

Gayle DeDe, Thesis Advisor, Department of Communication Sciences and Disorders
Nadine Martin, Department of Communication Sciences and Disorders
Francine Kohen, Department of Communication Sciences and Disorders

ABSTRACT

Improving discourse is often targeted in aphasia treatment because it is an important skill for meaningful conversation and interaction. The aphasia literature offers a variety of methods to analyze and treat discourse impairments in aphasia, however, there is no true consensus on what the best method is for discourse analysis. Very few studies have utilized listener perception as a method to capture discourse-related changes in aphasia. However, many researchers in other areas of speech-language pathology (e.g. dysarthria, fluency) use listener perceptions and rating scales as a valid measure to assess connected speech. The overarching goal of this study is to determine whether people with aphasia (PWA) and naïve listeners perceive changes in discourse associated with conversational treatment. A questionnaire, the Discourse Rating Scale for Aphasia, was created based on three constructs of discourse analysis in aphasia: macrolinguistic, microlinguistic, and functional features. Six PWA and nine naïve listeners listened to 30-35 second speech samples obtained before and after conversational treatment and rated their judgments on the questionnaire. We examined the relationship between the ratings on the Discourse Rating Scale for Aphasia (DRSA) and standardized language tests to validate the items and rating scale. Additionally, we looked for descriptive pre and post differences within the data to determine whether the DRSA was sensitive to treatment. It was found that each item and total DRSA score correlated highly with standardized tests of language in aphasia. We did not find strong evidence for the DRSA's sensitivity to treatment; however, we discuss the clinical implications of utilizing listener perception in the assessment of discourse in aphasia.

Keywords: aphasia, discourse analysis, listener perception, conversation treatment

This thesis is dedicated to all the individuals with aphasia,
communication partners, speech-language pathologists,
and clinical researchers who dedicate their lives to
learning and advocating for others.

ACKNOWLEDGMENTS

First, I would like to thank Dr. Gayle DeDe for her support and mentorship throughout the thesis process. This work would not have been possible without Dr. DeDe's guidance and expertise. Secondly, I would like to thank my committee members, Professor Francine Kohen and Dr. Nadine Martin for their help, advice, and mentorship throughout the creation and execution of this study. I would also like to thank Chiara Mancuso and Gabrielle Caminiti for collecting data and administering the questionnaire to the participants. I would also like to thank the people with aphasia and naïve listeners who participated in this study. Lastly, I would like to thank my family and friends for supporting me throughout this process. This work was supported by grant R21DC015859 from NIDCD.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Outcome measures are crucial in assessing the effectiveness of intervention in aphasia. They provide information about the client's progress in therapy, as well as insight into the effects of therapy in the individual's everyday life. An aspect of language that captures how an individual with aphasia communicates with others is discourse production. Improvements in discourse can be assessed objectively by examining changes at the linguistic level or can be measured subjectively based on how the individual perceives their communication in everyday situations. Measuring outcomes in discourse intervention is crucial in assessing meaningful change in response to aphasia treatment (Wright, 2011). The purpose of the present study is to evaluate a new tool for evaluating discourse production in people with aphasia.

Discourse intervention is crucial in the treatment of aphasia because it targets connected speech, which is important for everyday communication and interaction. Effective discourse relies on cohesively linking units of information together in order to convey a meaningful message (Wright, 2011). Discourse production demonstrates language processing in "real time," combining the syntactic, lexical, and phonological levels of language processing together to communicate with others (Armstrong, 2000). Discourse is important for people with aphasia to engage in meaningful interactions and form connections. Given the impairments at the lexical and grammatical levels in aphasia, discourse production presents as one of the main challenges in aphasia (Armstrong, 2000).

When discourse is impaired, it becomes challenging for a person with aphasia (PWA) to participate in daily life activities. It is common for daily communicative interactions to become difficult for PWA, leading to communication breakdown and social isolation (Parr, 2007). PWA may have fewer communicative opportunities because others do not know how to communicate with them (Kagan, 1998). PWA often lose connections with their spouses because communication becomes inaccessible and their partners may not be trained on how to incorporate them into conversations. Croteau and Le Dorze (2001) found that spouses of PWA thought of their partner as more dependent and not as likable compared to individuals without aphasia. Consequently, listener perception significantly influences communicative interactions for individuals with aphasia (Harmon, Jacks, Haley, & Faldowski, 2016). People with aphasia may become apprehensive about how their spoken language is perceived by others, which creates many communication barriers in daily life activities. Previous research has supported PWA's negative views of their communication abilities, both for familiar and unfamiliar communication partners. In a qualitative study surveying 50 PWA, Parr (2001) found that many social barriers resulted from reported negative attitudes such as ignorance and pity of the communication partner. Additionally, Simmons-Mackie and Elman (2015) suggested that unfamiliar listeners believe that PWA are unintelligent because of their communication impairment. These studies demonstrate the importance of assessing discourse and the need for discourse intervention in aphasia treatment.

Methods of Analyzing Discourse in People with Aphasia (PWA)

Discourse production is one component of the assessment process for people with aphasia. It allows speech-language pathologists to gain insight into how the individual

with aphasia conveys meaning in everyday situations. In addition to evaluating communicative effectiveness, assessing discourse is important for correctly diagnosing type of aphasia, planning treatment, and analyzing outcomes for intervention (Kong & Wong, 2018). Two perspectives have shaped the way clinicians assess and analyze discourse in aphasia: the structuralist and functionalist frameworks. They both provide a window into the complexity of discourse production and the overall intervention process, but they differ in their views of what is pertinent to study about discourse in aphasia.

The structuralist framework characterizes discourse based on individual linguistic components at the word and phrase level. Structuralists hone in on areas such as phonological processing, lexical selection, grammatical construction, and cohesion (Marini, Andreetta, Tin, & Carlomagno, 2011). Thus, the structuralist paradigm is associated with a focus on the microlinguistic aspects of discourse (Saffran, Berndt, & Schwartz, 1989; Prins, Snow, & Wagenaar, 1978; Mayer & Murray, 1988). For example, Prins and colleagues (1978) assessed spontaneous speech in adults with aphasia in terms of six variables: fluency, telegraphic speech, grammatical errors, articulation, paraphasias, and empty speech. This method has been used by clinicians to evaluate severity of aphasia and to differentiate fluent versus nonfluent types of aphasia through evaluating level of impairment for each of these factors. Saffran and colleagues (1989) devised Quantitative Production Analysis (QPA) to quantitatively measure structural and morphological aspects of spontaneous speech in a sample of agrammatic speakers. QPA aims to analyze the syntactic accuracy and complexity of discourse in aphasia based on proportion of verb inflections, closed-class words, and well-formed sentences. Instead of assessing errors in spontaneous speech, QPA evaluates the number of correct grammatical

structures and the extent in which phrases are expanded beyond a subject-verb utterance. QPA is effective in objectively comparing linguistic impairments across discourse in aphasia, as well as differentially diagnosing type of aphasia. Clinicians can also use QPA to determine changes in discourse in response to intervention over time (Rochon et al., 2000). Using these structuralist measures to assess spontaneous speech at the microlinguistic level provides clinicians with a detailed profile of an individual's discourse production and gives an idea of where the breakdown in communication happens. This is important for clinicians to identify in order to intervene in treatment. However, these methods can be time consuming and require considerable knowledge about linguistics and aphasia. Therefore, this method of analysis is not frequently used by clinicians to assess discourse in aphasia.

On the other hand, the functionalist framework characterizes discourse based on a gestalt view of language—mainly focused on an overall ability to convey meaning. From the functionalist perspective, language is a semantic unit, where purpose and context are the main factors and linguistic variables are simply interwoven (Marini et al., 2011). Analyzing discourse based on the functionalist paradigm involves assessing the meaning and efficiency of an individual's message (Nicholas & Brookshire, 1993; Nicholas & Brookshire, 1995). For example, in an effort to capture informativeness of discourse in people with aphasia, Nicholas and Brookshire (1993) created Correct Information Units (CIUs). CIUs refer to the words in a speech sample that are correct, appropriate, and informative to the task at hand. This method combines both quantitative and functional measures of discourse, assessing number of words and CIUs in a sample, words per minute, CIUs per minute, and percentage of words that are informative (%CIUs). CIU

analysis is consistently used in practice, has high diagnostic sensitivity, and can be used with a range of language samples (Marini et al., 2011). CIUs are beneficial in quantifying informativeness and efficiency of discourse, which can be applied in evaluating progress over the course of intervention (Nicholas & Brookshire, 1993).

Another functionalist measure is main concepts. In a study analyzing main concepts, or the "gist," of spontaneous speech, Nicholas and Brookshire (1995) found that individuals with aphasia had more inaccurate, incomplete, and omitted main concepts from their message compared to the non-brain damaged group. Main concepts analysis is a reliable measure used for assessing the informativeness and efficiency of discourse. Additionally, main concept analysis serves as a valuable outcome measurement tool in assessing the overall detail and effectiveness of an individual's message (Nicholas & Brookshire, 1995). Using these functionalist influenced measures provides insight into how efficient and informative an individual's discourse is, which gives a glimpse into overall communicative effectiveness. These measures focus on quantifying improvements in spontaneous speech that may be relatively subtle, but are nonetheless important for the everyday communication of individuals with aphasia.

Both structuralist and functionalist measures can be used to objectively describe discourse in aphasia. However, approaches for analyzing discourse have either focused on the linguistic nature of discourse or the overall ability to convey meaning within a message (Marini et al., 2011). Although both clinicians and researchers have demonstrated awareness regarding the importance of discourse production in aphasia, there is a lack of consensus on the best method to characterize discourse impairments in PWA (Dietz & Boyle, 2018). The range of available measures poses challenges in both

clinical and research settings. A critical issue is identifying the best method to characterize treatment-induced changes in discourse ability.

Treatment Outcome Measures for Aphasia

Treatment outcome measures should reflect the priorities of both the individual with aphasia and their conversation partners, which means that treatment should target aspects of communication that are functional and meaningful for daily routines (Wallace, Worrall, Rose, & Dorze, 2014). Discourse and functional communication using the World Health Organization (WHO) International Classification of Functioning, Disability and Health (ICF) framework (World Health Organization, 2001) is an important goal in aphasia treatment. Here, the word "functional" refers to the effectiveness of communicative interactions in daily situations and how the PWA feels about their skills as a communicator. Functional measures in this sense are more subjective and may evaluate discourse production abilities on the basis of the PWA's feelings about their communication in everyday situations.

Currently, there are a variety of functional communication outcome measures for aphasia. Some use rating scales to quantify feelings about communication (e.g., Lomas et al., 1989; Frattali et al., 1995; Hilari et al., 2003; Hula et al., 2015) and some use role play to assess functionality of communication (Holland, 1980). For example, the Communicative Effectiveness Index (CETI) aims to measure the individual's performance in a variety of communication situations using a self-rating scale. The CETI is a sensitive outcome measure for evaluating improvement in functional communication over time from both the caregiver and PWA perspective (Lomas et al., 1989). The Aphasia Communication Outcome Measure (ACOM) is a patient reported outcome

measure that also assesses functional communication in everyday communication tasks. The ACOM evaluates how effectively the PWA feels they can communicate with family and strangers in a variety of situations (Hula et al., 2015). The ACOM is a reliable and valid outcome measure for functional communication. Since functional communication measures often elicit the perspective of the individual, rating scale procedures can be relatively helpful in simplifying communicative scenarios by defining feelings through specific points on a scale, allowing for more adequate reflection about language. However, feelings about specific situations can sometimes be difficult to quantify and may require further explanation than a simplified number on a scale. In order to address this issue, treatment outcomes may also be measured objectively to evaluate improvement at the linguistic level of discourse. Speech-language pathologists often measure comprehensive treatment effects using standardized assessments such as the Western Aphasia Battery or the Boston Naming Test. However, these assessments do not provide a fine-grained assessment of discourse production.

The lack of consensus and the wide discrepancy between discourse outcome measures remains an issue in practice today. As discussed, some discourse measures focus on communication in everyday situations, whereas others focus on objectively measuring discourse at the linguistic level. It would be valuable to combine both of these methods and incorporate listener perspectives when assessing outcomes of treatment for discourse since it is a form of spoken language used in everyday situations. It would be helpful to have a measure that targets various levels of discourse and includes the individual's perception of their own discourse. This measure should evaluate the linguistic level of discourse in a less time-consuming manner than current available

methods through the use of rating scales. Such a measure would be valuable in providing clinicians with information about how the individual feels as a communicator.

Listener Perceptions of PWA's Discourse Samples

Previous studies have investigated listener perceptions of discourse in aphasia. Lasker and Buekelman (1999) compared listener perceptions of healthy adults for PWA using either unaided or aided communication during story tell. Each sample was video recorded. The researchers found that most of the listeners rated PWA using unaided communication more favorably compared to those using aided communication. Harmon and colleagues (2016) conducted a study to investigate negative perceptions of discourse produced by PWA and evaluate the effects that fluency has on perceptions of communication ability. The researchers asked naïve listeners to judge speech samples produced by people with aphasia. The perceptual rating scale was created based on questionnaires developed in the stuttering literature. It consisted of nine statements on a seven point Likert scale divided into three categories—speech output (behavioral), speaker attributes (cognitive), and listener feelings (affective). They found that listeners perceive PWA less favorably compared to the neurologically healthy individuals.

Using the same rating scale as Harmon et al. (2016), Kong and Wong (2018) examined the relationship between three different discourse analysis approaches (proposition-based, linguistically based, and story grammar analysis) and both aphasia severity scores and listener perceptions. The researchers collected audio files from 68 PWA and 68 healthy controls narrating a story retell of "The Boy Who Cried Wolf." Each language sample was coded using three types of discourse analyses. Severity scores were assessed using the Cantonese Aphasia Battery (CAB). Naïve listeners completed the same

perceptual rating scale from the Harmon and colleagues (2016) study regarding behavioral, cognitive, and affective aspects of the speech sample. They found that all three approaches highly correlated with listener ratings. However, the linguistic approach to analyzing discourse showed the strongest correlations with both aphasia severity and naïve listener ratings. This demonstrates that listener ratings may provide a similar measure of discourse compared to objective and quantitative approaches to discourse analysis.

Aside from the mentioned studies, there is little research on how listeners perceive discourse produced by PWA. However, rating scales have been used to examine listener perception of spontaneous speech in other communication disorders. For example, in the dysarthria literature, Yorkston and Beukelman (1978) explored the relationship between intelligibility as assessed by objective measures and rating scale measures. For the objective measures, they had naive listeners transcribe speech samples from individuals with dysarthria to measure intelligibility. For the rating scale measures, the researchers had the naïve listeners estimate percentage of intelligibility and rate overall intelligibility of the speech sample. The researchers found that the rating scale procedures ranked the clients with dysarthria similarly to the objective measures in terms of severity. They also found that the percentage estimates reflected intelligibility scores from the transcription tasks. The results of this study demonstrate the usefulness of analyzing listener perception as a method in assessing spontaneous speech.

Additionally, rating scales are used commonly as assessment and outcome instruments in both stuttering and voice treatment. One of the most prominent stuttering instruments used to assess how stuttering affects an individual's life is the Overall

Assessment of the Speaker's Experience of Stuttering (OASES; Yaruss & Quesal, 2006). The OASES consists of a variety of questions scored on a Likert scale from one to five. It is helpful in determining severity of impact, as well as serving as an outcome measure for treatment. Rating scales are also used in the assessment of voice disorders. The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) is used to describe the severity of auditory-perceptual voice characteristics that can be communicated among other clinicians (Kempster et al., 2009). Each voice attribute is rated on a visual analog scale (VAS) by indicating the degree of voice deviation from mild to severe. The CAPE-V is valuable in determining severity of a voice problem, quantifying perceptual characteristics, and providing consensus amongst clinicians. Since voice and fluency disorders impair communication at the level of connected speech and have benefited from the use of rating scales, rating scales may also be beneficial in the assessment of discourse in aphasia.

Rating scales have the potential to provide information about how the speaker or listener perceives connected speech samples of people with aphasia. Currently, rating scales for people with aphasia are mainly used to capture how the individual feels about their ability to communicate in a variety of situations. However, there is no measure that assesses what the individual notices about their discourse. That is, there is no measure in which the PWA listen to their own speech sample and rate their discourse. In this study, we were focused on how people with aphasia perceive their connected speech—both functionally and linguistically. Existing rating scales for PWA typically focus on assessing functional communication, which is only one aspect of discourse analysis. Therefore, the Discourse Rating Scale for Aphasia (DRSA) was created to target

microlinguistic (e.g., grammatical errors), macrolinguistic (e.g., relevance), and functional (e.g., confidence) aspects of the connected speech samples using a visual analog scale. In addition, naïve listeners rated the speech samples to evaluate how they compare with the PWA's own judgments of their discourse as well as objective measures of aphasia severity.

CHAPTER 2

AIMS

The purpose of this study was to validate the DRSA using two methods. First, we examined whether the DRSA correlated with measures of language ability. Second, we examined whether the DRSA was sensitive to changes in discourse associated with conversation treatment. We predicted a positive correlation between aphasia severity scores on standardized measures of language ability (Comprehensive Aphasia Test, Aphasia Communication Outcome Measure, and the Verb Naming Test) and DRSA scores, and that DRSA scores would improve following conversation treatment.

CHAPTER 3

GENERAL METHODS

Naïve listeners and PWA listened to speech samples collected at different time points within a conversation treatment study and rated their perceptions using the DRSA. General study methods, including (1) an overview of the conversation treatment study and (2) the structure and rationale of the DRSA are presented first. Next, methods and results are presented for each group separately, beginning with the naïve listeners and followed by PWA. Finally, results across the two groups are compared.

Conversation Treatment Study

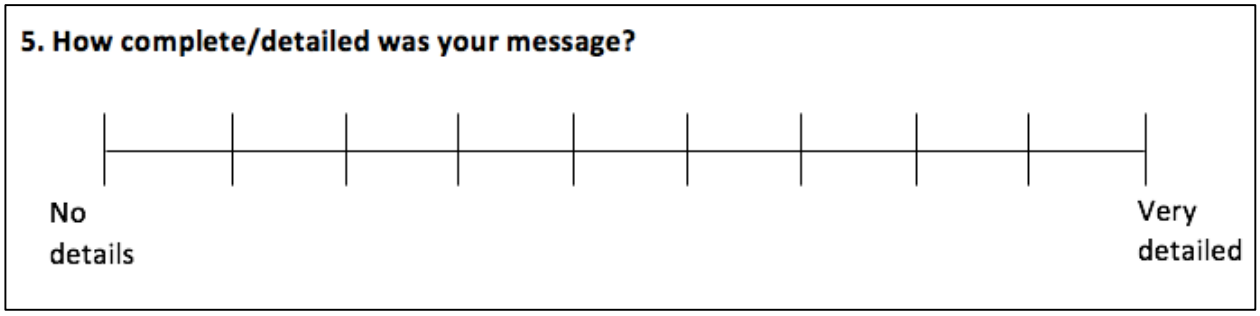
Data from a study of conversation treatment were used to examine whether self-reported and naïve listener-reported ratings demonstrate a positive change from pre to post conversational treatment. In this study, DeDe, Hoover, and Maas (2019) explored the ideal number of clients in a conversation group by comparing two hypotheses: the group dynamic hypothesis (larger group > smaller group) and the treatment dosage hypothesis (smaller group > larger group). Forty-eight PWA from Boston University and Temple University were randomly assigned to one of three conditions: the large group (six to eight PWA), the dyad condition (two PWA), and the delay control group. The present study collapsed across treatment conditions; that is, treatment compared to no-treatment groups. Treatment sessions were one hour and took place twice a week for 10 weeks. Each conversation group followed a socially oriented approach and was generally unstructured, consisting of topics that are common in daily peer interaction and communicative events. Each participant had two goals that were identified based on the initial assessment and were targeted naturally within conversation. Treatment sessions

were led by trained graduate SLP students under the supervision of a licensed SLP. Each participant was assessed at four different time points using standardized assessments, patient reported outcome measures, and connected speech sample analyses (e.g. CIUs). See the participant section in *Study 2: People with Aphasia* for more information about the assessments and inclusion criteria.

The Discourse Rating Scale for Aphasia (DRSA)

The DRSA consisted of nine questions probing both linguistic and functional aspects of the speech samples. One version of the DRSA was specific to the naïve listener group and the other was specific for the PWA group; however, each question measured the same construct for each group's questionnaire. Each response was rated on a visual analog interval scale (VAIS), which is a specific type of visual analog scale (VAS) that is measured on a 100 mm line and consists of eight undefined tick marks and two defined anchors at the end representing opposite extremes (e.g. not confident- very confident). Overall, VASs aim to capture feelings or attitudes along a continuum without defining intervals between the two anchors (See *APPENDIX A and B* for the DRSA). Figure 1 shows an example of the visual analog interval scale used for each item on the DRSA.

Figure 1: Example of a Visual Analog Interval Scale (VAIS)



The VAIS was chosen in order to maximize reliability of the measure and its accessibility for people with aphasia. Likert scales are commonly used to capture

individuals' feelings and opinions in rating studies. However, since Likert scales define each of the intervals along a continuum, the overall design requires individuals to fit their perceptions into a specific category. With VASs, individuals can interpret their own ideas and feelings along a continuum without the persuasion of defined intervals. Importantly, Wewers and Lowe (1990) found that VASs are much easier to understand than Likert scales when the rater's reading ability is impaired. Thus, VASs might be preferred for PWA, who often have reading impairments. VASs also show strong reliability. In one study, researchers found that VASs are highly reliable as measured by intra-class correlations (Bijur, Silver, & Gallagher, 2001). Ninety percent of pain ratings were reproducible within 9 mm, which is comparable to breaking up the line in 10 mm intervals. Thus, including undefined intervals to a VAS may increase the meaning of the points of the scale without binding the measurement to a specific definition. When comparing the different type of VASs, Sriwatanakul and colleagues (1983) found that most of the subjects preferred a horizontal (rather than vertical) scale containing gradations. The Communication Confidence Rating Scale for Aphasia (CCRSA) is a seven-point VAS used to measure PWA's confidence in communicating in everyday situations. Cherney and colleagues (2011) found that the CCRSA showed statistically significant changes from pre to post-treatment, demonstrating the reliability and sensitivity of the VAS scale on the CCRSA in analyzing participants' self-report. Additionally, VASs increase the ease of measurement and comparison when analyzing treatment outcomes, as well as the overall responsiveness of the participant (Dias et al., 2001). Therefore, a VAS was used in this study.

Questions were categorized as reflecting one of three underlying constructs: microlinguistic, macrolinguistic, and functional features in order to identify whether certain features are more sensitive to perceivable change. Questions coded as microlinguistic examined individual elements of language processing such as syntax, morphology, and semantics. Questions coded as macrolinguistic examined the overall message and amount of details within the speech sample. Questions coded as functional assessed communicative effectiveness and feelings of communication in the speech sample. In order to capture speech fluency, a question about rate of speech was created. This item was not categorized into a specific construct because fluency is a multidimensional concept that is not clearly macrolinguistic or microlinguistic. Table 1 shows how each question on the DRSA was categorized.

The DRSA was developed based on the different methods clinicians use to measure discourse in aphasia (e.g. macrolinguistic, microlinguistic, and functional constructs). To validate the DRSA, each question was correlated with a corresponding standardized measure of language ability. Standardized measures included subtests of the Comprehensive Aphasia Test (CAT) and the Aphasia Communication Outcome Measure (ACOM). The standardized tests were selected based on the purpose of each question on the DRSA. Question coded as macrolinguistic assessed amount of details, relevance, and repetitiveness of the sample. Questions coded as microlinguistic assessed word finding difficulty, paraphasias, and overall grammaticality. Questions coded as functional assessed perceived communicative effectiveness and amount of the message understood. Thus, the purpose of each question on the DRSA was carefully examined, and based on the measure, a corresponding CAT subtest or ACOM score was correlated with that item

to validate the measure. Table 1 shows which CAT subtest or ACOM score each item was correlated with.

Table 1: *Naïve Listener Version of the DRSA with the Coded Constructs and the Corresponding Assessments*

Question Item	Construct	Test
How complete/detailed was the person's message?	Macrolinguistic	• CAT Total Picture Description Score
How much of what the person said was relevant to the main point?	Macrolinguistic	• CAT Total Picture Description Score
How repetitive was the message?	Macrolinguistic	• CAT Total Picture Description Score
How often did the person have trouble thinking of the word they wanted to say?	Microlinguistic	• CAT Total Naming Score
How often did the person use the wrong word?	Microlinguistic	• CAT Total Naming Score • CAT Repetition of Words and Nonwords
How much of the grammar was correct?	Microlinguistic	• CAT ICWs: well formedness • CAT Repetition of Sentences
How easy would it be to have a conversation with this person?	Functional	• Aphasia Communication Outcome Measure
What percent of the overall message was understood?	Functional	• CAT Total Picture Description Score
How was the person's rate of speech?	Not coded	• CAT ICWs: Speed Score

Each macrolinguistic item was correlated with the CAT Total Picture Description score because this subtest provides an overall measure of appropriate and inappropriate

words, syntactic variety, well-formedness, and speed ratings for a picture description task. Therefore, the CAT Total Picture Description assesses the overall message, accounting for amount of details, relevance, and repetitiveness. Each microlinguistic item was correlated with a different subtest depending on the purpose of the question. The word finding question (*How often did the person have trouble thinking of the word they wanted to say?*) was correlated the CAT Total Naming Score because this subtest provides a glimpse into word finding difficulty by assessing naming errors during confrontation naming tasks (e.g. actions and objects). The naming error question (*How often did the person use the wrong word?*) aimed to capture presence of paraphasias, therefore, it was also correlated with the CAT Total Naming Score. This item was also correlated with CAT Repetition of Words and Nonwords to determine whether phonological processing was related to production of paraphasic errors. The grammar question (*How much of the grammar was correct?)* was correlated with CAT ICWs: Well Formedness because this subtest provides an overall measure of grammaticality and syntactic accuracy for a picture description task. This question was also correlated with CAT Repetition of Sentences, because morphosyntactic knowledge may support repetition of sentences. Each functional question was correlated with a measure to either reflect communicative effectiveness or understanding of the overall message. The listener feelings question (*How easy would it be to have a conversation with this person?*) was correlated with the ACOM because this provides an overall measure of perceived communicative effectiveness in specific communication situations. The percent understood question (*What percent of the overall message was understood?*) aimed to assess the gist of the individual's message within one rating, and was correlated with the

CAT Total Picture Description Score because this subtest takes into account a variety of linguistic variables (syntactic variety, appropriate words, well-formedness, and speed) during a connected speech sample, providing a glimpse into the overall message of one's sample. The speed question was not categorized as a specific construct and was correlated with CAT ICWs: Speed Score because this subtest determines rate of speech, which is reflective of speech fluency.

In addition to correlating each item on the DRSA with a specific subtest, we also correlated the total DRSA score with the total score on three different assessments that correspond to the three identified constructs. These assessments included the CAT, the ACOM, and the Verb Naming Test (VNT). We chose the CAT Total Picture Description Score as a macrolinguistic measure because it provides a total score of connected speech accounting for relevance and details during a picture description task. The ACOM was chosen as a functional measure since it provides a measure of communicative effectiveness during specific communication scenarios. The VNT was chosen as a microlinguistic measure because it examines the production of a variety of verbs in different contexts. Correlating the total DRSA score with these standardized measures contributes further information regarding the validation of the DRSA.

CHAPTER 4

STUDY 1: NAÏVE LISTENERS

Participants

Nine neurotypical naïve listeners participated in this present study. This group was sampled from undergraduates at Temple University who had minimal experience interacting with people with aphasia. This was determined through self-report and a simple screening form evaluating amount of education about aphasia. Each listener was a native English speaker as defined by learning English before the age of six and not in a classroom setting. Naïve listeners were between the ages of 18 and 30 years old in order to sample a wide range of individuals obtaining their undergraduate degree. Each naïve listener rated nine PWA's speech samples. Five of the nine PWA rated by naïve listeners participated in the self-rating study (Study 2) and thus rated their own samples (see Table 4 and 5 under *Study 2: People with Aphasia* for more information on participants with aphasia).

Procedures

Thirty- to thirty-five second clips were extracted from the beginning of each individual's Cinderella retell. The thirty-second time frame was chosen because it was estimated that thirty-seconds would be an adequate amount of time to convey enough information to rate the quality of the sample based on the content, while also minimizing the listening time and maximizing the number of samples played in a session. Furthermore, since each participant spoke for different amounts of time, equating the sample lengths would reduce variation introduced by this factor.

In order to further reduce the total testing time, naïve listeners only rated the Cinderella retell and did not rate other picture descriptions or the maintenance time period. Thus, naïve listeners listened to nine different PWA's Cinderella retell samples from pre- and post-treatment. Of the nine PWA rated by the naïve listeners, three PWA were from the control group and six PWA were from the treatment group from the previous conversation treatment study. Therefore, each naïve listener listened to a total of 18 speech samples.

Each participant was tested individually in a quiet room. Each speech sample was pseudo-randomized with the constraint that participants did not hear two samples from the same PWA in a row. Participants were allowed to listen to the sample twice. Speech samples and the DRSA were presented in Qualtrics. There was a brief five to ten-minute training period prior to the study, which included a description of the questions and scale, as well as a practice item. The practice item consisted of a sample taken from Aphasia Bank. The participant was encouraged to ask any questions during the practice trial. However, the administrator did not answer specific questions that would possibly influence or bias the participants' opinions. Once it was clear that that the participant understood the procedure, they proceeded to complete the study independently. After listening to a speech sample, the participant immediately rated the sample using the appropriate rating scale and repeated this procedure for each sample presented. Testing lasted about one hour.

Data Analysis Plan

Construct Validity

In order to validate each item on the DRSA and determine the sensitivity of the measure, scores for each question were averaged across the naïve listener group (n= 9) for each PWA sampled (n= 9). The average score for each question was then correlated with a corresponding subtest on the CAT or the ACOM to validate the items. Thus, for each DRSA item, there were nine data points (from the 9 PWA), each of which contained averaged data from nine naïve listeners. The average score on each question was then correlated with the PWA's performance on the corresponding assessment. Spearman's non-parametric correlations were used given the small sample sizes. Correlations with a p-value less than .05 were considered significant. Additionally, total DRSA scores were correlated with standardized assessments that measure macrolinguistic, microlinguistic, and functional aspects of discourse. These correlations were used as a measure of construct validity, demonstrating that the DRSA measured what it was intended to measure.

Sensitivity to Treatment

The total score on each DRSA was averaged across raters to calculate pre and post scores for each participant sampled. A Wilcoxon Signed Rank test was used to assess pre and post differences for the three PWA in the control group and the six PWA in the treatment group. Additionally, scores were averaged across each construct (microlinguistic, macrolinguistic, and functional) for each time point and assessed descriptively. A change in scores by ten percent (10 points) was interpreted as a

meaningful deflection given that this would be a change in one anchor point on the DRSA (also cf. Bijur, Silver, & Gallagher, 2001).

Results

Construct Validity

In order to assess construct validity, DRSA scores from the pretreatment time point were correlated with standardized test scores. Results are presented in Table 2. Overall, each test item correlated with at least one standardized test, with significant correlations ranging from .71 to .84. Next, correlations between total DRSA scores and identified macrolinguistic, microlinguistic, and functional assessments were examined. As Table 3 shows, all correlations were significant.

Table 2: *Naïve Listener Version of the DRSA and CAT Subtest Correlations*

Question Item	Test	P-Value	Rho
How complete/detailed was the person's message?	• CAT Total Picture Description Score	.009*	.81
How much of what the person said was relevant to the main point?	• CAT Total Picture Description Score	.01*	.78
How repetitive was the message?	• CAT Total Picture Description Score	.04*	.69
How often did the person have trouble thinking of the word they wanted to say?	• CAT Total Naming Score	.03*	.71
How often did the person use the wrong word?	• CAT Total Naming Score • CAT Repetition of Words and Nonwords	.06 .03*	.64 .71
How much of the grammar was correct?	• CAT ICWs: well formedness • CAT Repetition of Sentences	.005* .14	.84 .53

Table 2 Continued: *Naïve Listener Version of the DRSA and CAT Subtest Correlations*

How easy would it be to have a conversation with this person?	• Aphasia Communication Outcome Measure	.01*	.80
What percent of the overall message was understood?	• CAT Total Picture Description Score	.009*	.81
How was the person's rate of speech?	• CAT ICWs: Speed Score	.004*	.84

Table 3: *Correlations Between Total DRSA Score and Corresponding Assessment*

Test	Construct	P-value	Rho
CAT Total Picture Description Score	Macrolinguistic	.005*	.83
Verb Naming Test Total Score	Microlinguistic	.02*	.76
A-COM T-Score Estimate	Functional	.03*	.73

Sensitivity to Treatment

Wilcoxon Signed Ranks Test showed no significant differences from pre to post time points for either the treatment or control group. Next, changes in the average construct scores were assessed from pre-treatment to post-treatment, using the benchmark of a 10% change as a meaningful deflection. No participants in the control group demonstrated positive deflections from pre to post-treatment in any of the 3 constructs. In fact, each participant in this group demonstrated a negative deflection in at least one of the constructs from pre to post-treatment. In constrast, two of the six PWA in the treatment group demonstrated positive deflections greater than 10 from pre to post-treatment in at least one construct. Two of the PWA in the treatment group demonstrated negative deflections from both time points in at least one construct, and one participant

did not show any deflections from pre to post-treatment. Figures 1, 2, and 3 below highlight positive and negative deflections greater or equal to ten across the control and treatment group for the three different constructs.

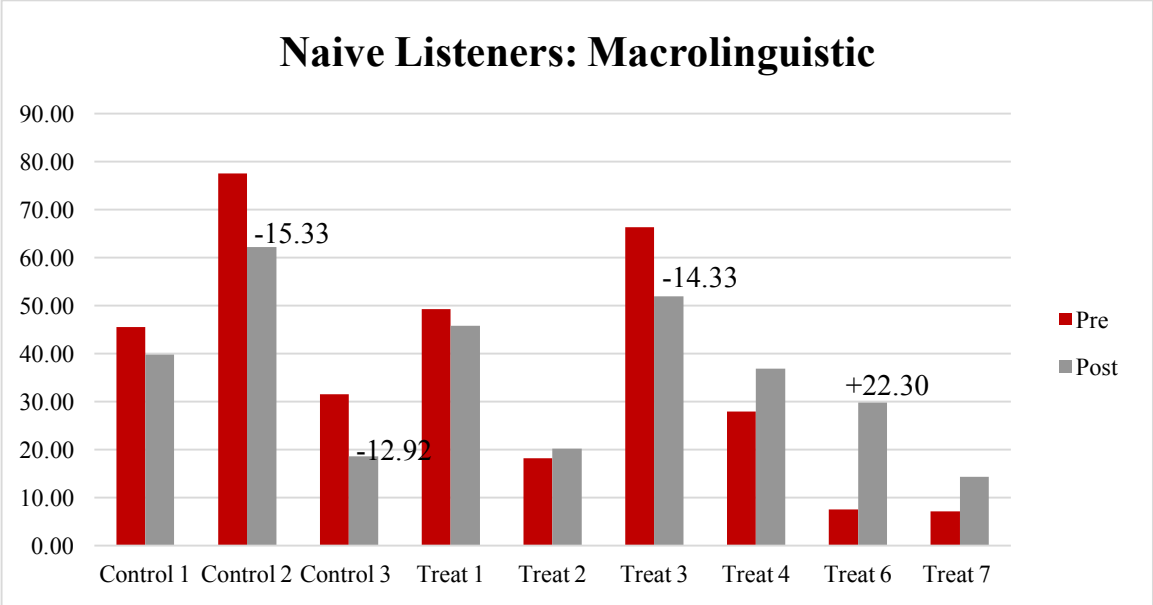


Figure 2*:* Graph representing average scores from NL group for the macrolinguistic construct for each participant. Numbers highlight +/- meaningful deflections from pre to post time points.

For the macrolinguistic construct (as seen in Figure 2), two out of the three participants in the control group showed negative deflections of greater than 10 from pre to post-treatment (Control 2 and Control 3). No controls demonstrated positive deflections for this construct. For the treatment group, one out of the six participants showed a positive deflection from pre-treatment to post-treatment (Treat 6). Treat 3 demonstrated a negative deflection from pre to post-treatment. No other participants showed meaningful deflections from pre to post-treatment.

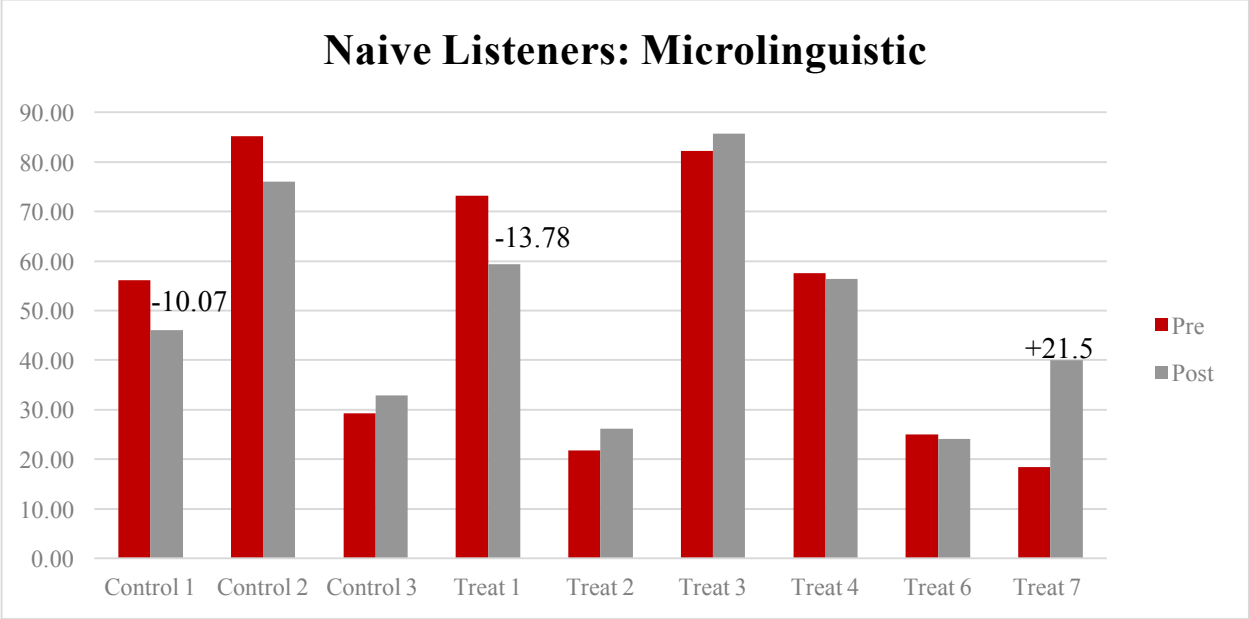


Figure 3*:* Graph representing average scores from NL group for the microlinguistic construct for each participant. Numbers highlight +/- meaningful deflections from pre to post time points.

For the microlinguistic construct (as seen in Figure 3), one out of the three participants in the control group showed negative deflections of greater than 10 from pre to post-treatment (Control 1). No controls demonstrated positive deflections for this construct. For the treatment group, one out of the six participants showed a positive deflection from pre-treatment to post-treatment (Treat 7). Treat 1 demonstrated a negative deflection from pre to post-treatment. All other participants did not show a meaningful change from pre to post-treatment.

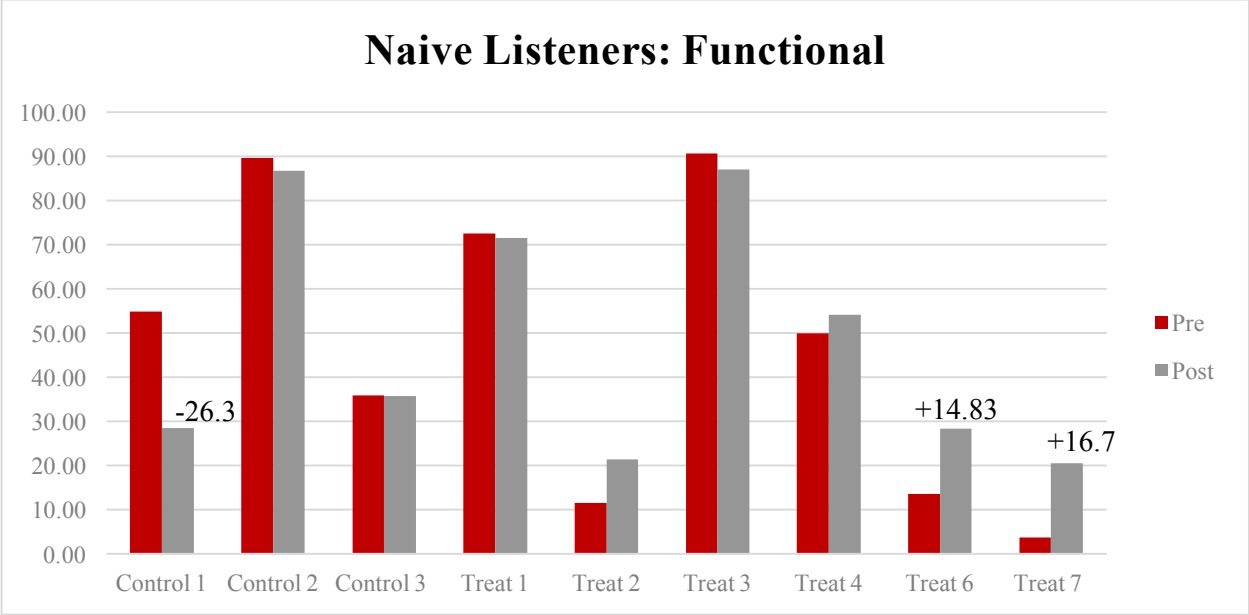


Figure 4: Graph representing average scores from NL group for the functional construct for each participant. Numbers highlight +/- meaningful deflections from pre to post time points.

For the functional construct (as seen in Figure 4), one out of the three participants in the control group showed negative deflections of greater than 10 from pre to post-treatment (Control 1). No positive deflections were observed for the control group from pre to post-treatment. For the treatment group, two out of the six participants showed a positive deflection from pre-treatment to post-treatment (Treat 6). No negative deflections were observed in the treatment group. No other participants showed a meaningful deflection from pre to post-treatment.

CHAPTER 5

STUDY 2: PEOPLE WITH APHASIA

Participants

A total of 10 PWA were recruited from a larger treatment study assessing effects of group size in aphasia conversation treatment (DeDe et al, 2019). In this follow-up study, participants were only obtained from Temple University. Inclusion criteria included a) age 18 years or older, b) at least 5-months post stroke, c) English as their first language, d) presence of aphasia as determined by CAT scores and clinician judgment. All participants with aphasia from Temple University were contacted to determine whether they were willing to participate in the study. Out of the 10 PWA who agreed to participate in the study, six PWA listened to and self-rated their own speech samples from different time points. Five of the six PWA were also rated by naïve listeners in Study 1. The exception was Treat 5. See Table 4 and Table 5 for background data on the PWA included in this study.

PWA were distributed across two different conditions of the conversation treatment study— the treatment group (large group and dyads) and the delay (control) group. Out of the six in the PWA group (self-raters), three participants were part of the control group and three participants were part of the treatment group. We recruited individuals with a range of aphasia severities to evaluate if the DRSA was appropriate and sensitive to differences in severity. Severity of aphasia was based on the CAT Total Naming T-score, given that impairments in naming are a characteristic feature of aphasia. See Table 5 for aphasia severity according to CAT naming scores. In the conversation treatment study (DeDe et al., 2019), aphasia syndrome included a variety of fluent and

nonfluent types, with mean CAT naming scores ranging between 41 (dyad group) and 50.2 (large group). They were individually assessed pre (T1), post (T2), 6 weeks after (T3), and 11 months after (T4) treatment using the CAT, *Northwestern Assessment of Verbs and Sentences* (NAVS), *Philadelphia Naming Test* (PNT), *Aphasia Communication Outcome Measure (ACOM)*, and the *Lubben Social Network Scale*.

Table 4*: Background Data on PWA*

Measure	Pre	Post
CAT section total scores		
Auditory comprehension	50.1 (5.9)	54.7 (5.9)
Reading Comprehension	45.0 (6.7)	45.7 (5.6)
Repetition	59.7 (9.6)	59.3 (10.7)
Naming	59.7 (14.7)	61.9 (14.4)
Oral Reading	47.5 (12.6)	46.3 (14.3)
Spoken Pic Description	26.6 (11.8)	26.3 (10.5)
ACOM Final T-Score	51.9 (9.0)	54.4 (9.2)
PNT Total Score	27.0 (4.7)	26.2 (2.9)

Note. Displayed are the averages of the total scores for each measure across the PWA included in this study (n= 10) from both pre-treatment and post-treatment time points. The standard deviation of total scores for each subtest across the PWA included in the study is in parentheses.

Table 5*: CAT Total Naming T-Scores and Aphasia Severity*

Participants	Age	Treatment Condition	Rating Group	CAT Total Naming T-Score	Aphasia Severity
Control 1	65	Control	Self & NL	61	Mild
Control 2	57	Control	Self & NL	69	Mild
Control 3	51	Control	Self & NL	52	Moderate
Treat 1	60	Treat	NL	61	Mild
Treat 2	54	Treat	NL	55	Moderate
Treat 3	69	Treat	NL	63	Mild
Treat 4	58	Treat	NL	60	Mild
Treat 5	62	Treat	Self	58	Moderate
Treat 6	55	Treat	Self & NL	61	Mild
Treat 7	67	Treat	Self & NL	49	Severe

Note. Aphasia severity was based off of pretreatment CAT total naming T-score by the following classification: >57 mild, 46-56 moderate, 35-45 severe. The rating condition represents whether or not participants rated their own speech samples (Self), were rated by naïve listeners (NL), or included in both (Self & NL).

Procedures

Thirty- to thirty-five second clips were extracted from the beginning of the Comprehensive Aphasia Test (CAT) picture description, the Cat Rescue picture description from Nicholas and Brookshire, and the Cinderella retell. Each PWA rated their own speech samples from three different time points throughout the study: pre-treatment, post-treatment, and 6-week maintenance. Thus, each PWA listened to a total of nine speech samples. PWA were presented with an audiotaped sample played on the computer and answered questions on the DRSA via paper. They did not answer on Qualtrics given that many of the PWA were not comfortable using the computer. Therefore, the scale was slightly different from the version presented on Qualtrics for the naïve listener group. The paper scale was measured from one to 10, where as the Qualtrics scale was measured from zero to 100. All other testing procedures were the same as the naïve listener group (see *Study 1: Naïve Listeners* for more details on procedures).

Data Collection and Analysis

Data analysis was generally similar to the naïve listener group in regards to sensitivity of treatment. Total DRSA scores from pre-treatment, post-treatment, and maintenance were qualitatively assessed for pre and post differences for the three participants in the control and the three participants in treatment group. Additionally, scores were averaged across the three constructs for each time point and assessed descriptively. A change in scores by ten percent (or a change in 1 point for the paper scale) was noted as a meaningful deflection given that this would be a change in one

anchor point on the DRSA. Construct validity was not measured in this group given the small sample size and number of samples rated.

Results

Sensitivity to Treatment

The total score for each PWA was averaged across speech samples separately for each of the three time points: pre-treatment, post-treatment, and maintenance. The total scores were compared across each time point to qualitatively look for differences. Since the scale was slightly different from the naïve listener group in terms of how it was measured, a difference of one was considered a meaningful deflection (10% change, or change in one anchor line of the scale). In the group of six PWA sampled, three were part of the control group and the other three were part of the treatment group. All three PWA in the control group were also rated by the naïve listener group, and two out of the three PWA in the treatment group were also rated by the naïve listener group. These participants are indicated with an asterisk in Figures 5, 6, 7, and 8.

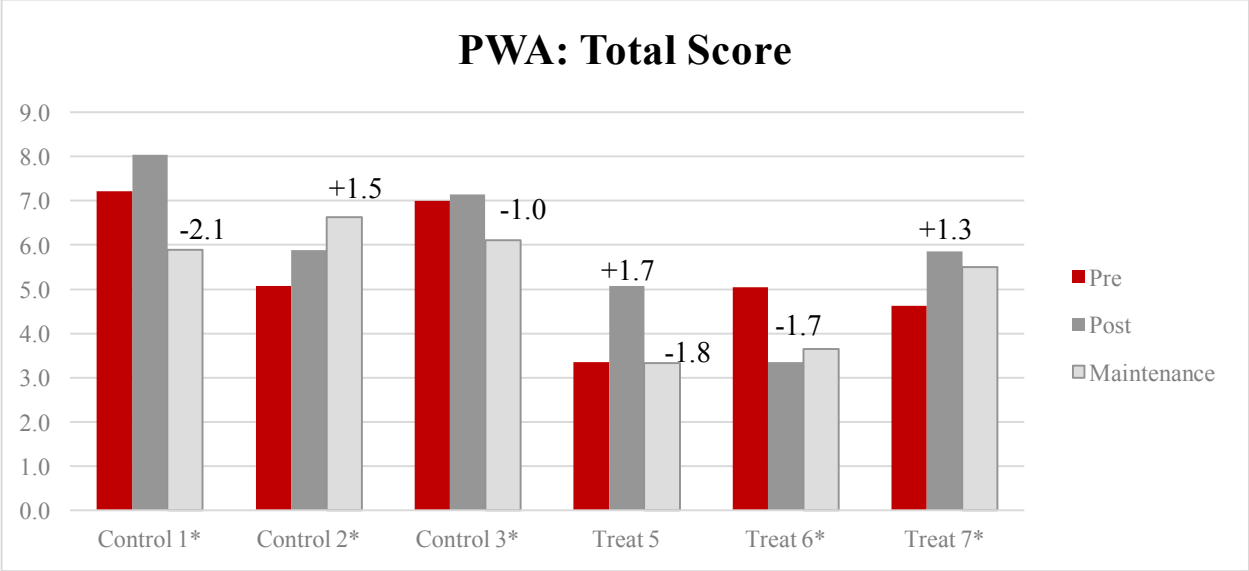


Figure 5: Graph representing total DRSA scores for each PWA from pre, post, and maintenance time points. Numbers reflect +/- meaningful deflections. *Denotes PWA who were also rated in Study 1.

No participants in the control group demonstrated meaningful deflections from pre-treatment to post-treatment in the total DRSA score. One participant in the control group (Control 2) demonstrated a positive deflection in the total score from pre-treatment to maintenance. Control 1 showed a negative deflection from pre-treatment to maintenance, as well as post-treatment to maintenance. Control 3 demonstrated a negative deflection from post-treatment to maintenance only. Within the treatment group, two out of the three PWA (Treat 5 and Treat 7) perceived positive deflections of greater than one from pre to post-treatment. Treat 6 showed a negative deflection from pre-treatment to post-treatment, as well as a negative deflection from pre-treatment to maintenance. Treat 5 perceived a negative deflection from post-treatment to maintenance only. The other changes in the data were not meaningful. Figure 4 above highlights the positive and negative deflections for the total score on the DRSA for the PWA group.

Similar to the naïve listener group, PWA's self-ratings for questions from pre, post, and maintenance were averaged across their corresponding construct. These

averages were examined across each time point and assessed for meaningful deflections of greater than one (as seen in Figure 6, 7, 8 below).

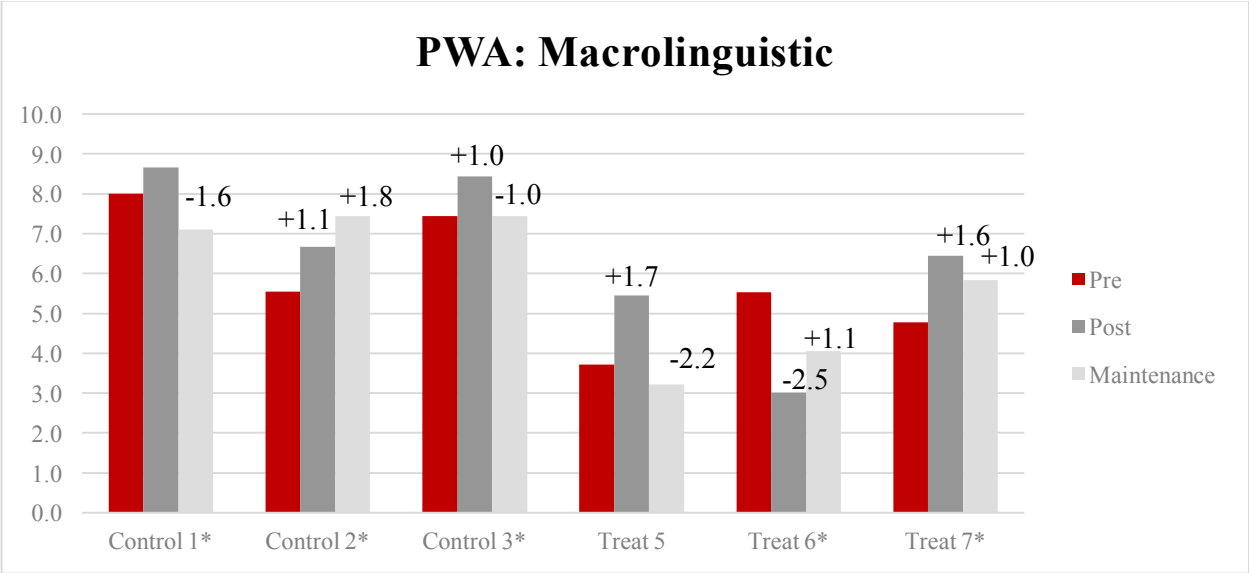


Figure 6: Graph representing PWA average scores for the macrolinguistic construct for each participant from pre, post, and maintenance time points. Numbers highlight +/- meaningful deflections. *Denotes PWA who were also rated in Study 1.

For the macrolinguistic construct (as seen in Figure 6), two out of the three participants in the control group (Control 2 and Control 3) showed positive deflections of greater than one from pre to post-treatment. Two out of the three participants (Control 1 and Control 3) showed a negative deflection from post-treatment to maintenance. For the treatment group, two out of the three participants (Treat 5 and Treat 7) showed a positive deflection from pre-treatment to post-treatment. Treat 6 showed a negative deflection from pre to post-treatment, but showed a positive deflection from post-treatment to maintenance. Treat 5 showed a negative deflection from post-treatment to maintenance.

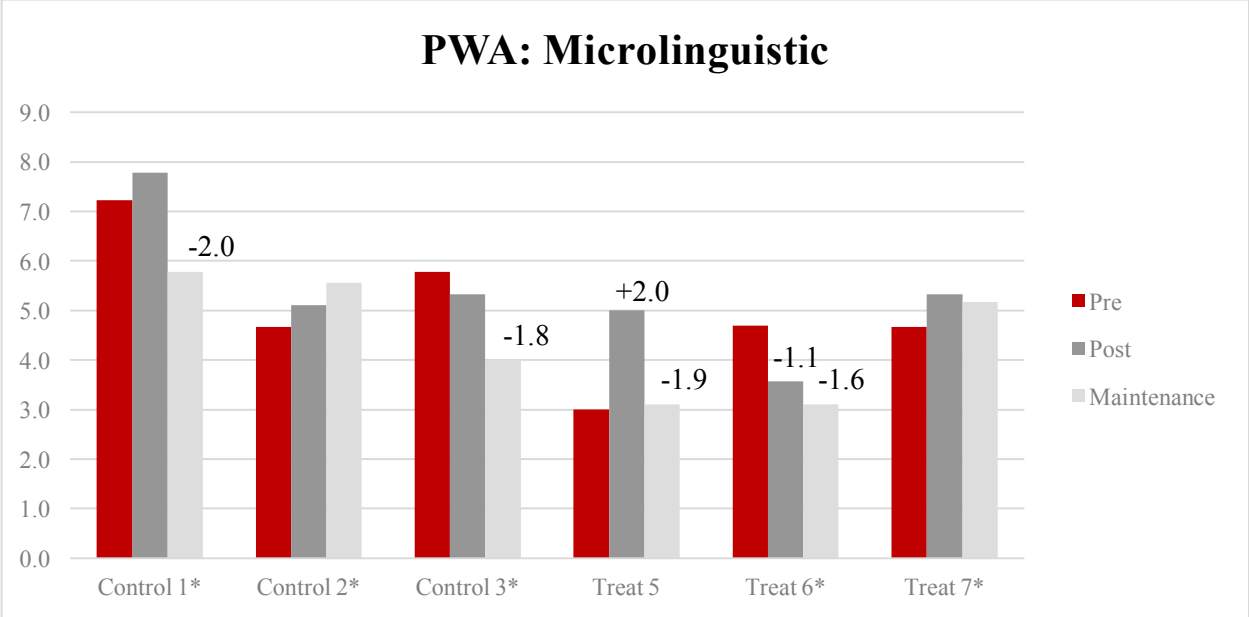


Figure 7: Graph representing PWA average scores for the microlinguistic construct for each participant from pre, post, and maintenance time points. Numbers highlight +/- meaningful deflections. *Denotes PWA who were also rated in Study 1.

For the microlinguistic variable (as seen in Figure 7), no positive deflections were observed for the control group. One out of the three participants (Control 1) showed a negative deflection from post-treatment to maintenance. For the treatment group, one participant out of the three showed a positive deflection from pre-treatment to post-treatment (Treat 5). Treat 6 showed a negative deflection from pre-treatment to post-treatment, as well as pre-treatment to maintenance. Treat 5 showed a negative deflection from post-treatment to maintenance.

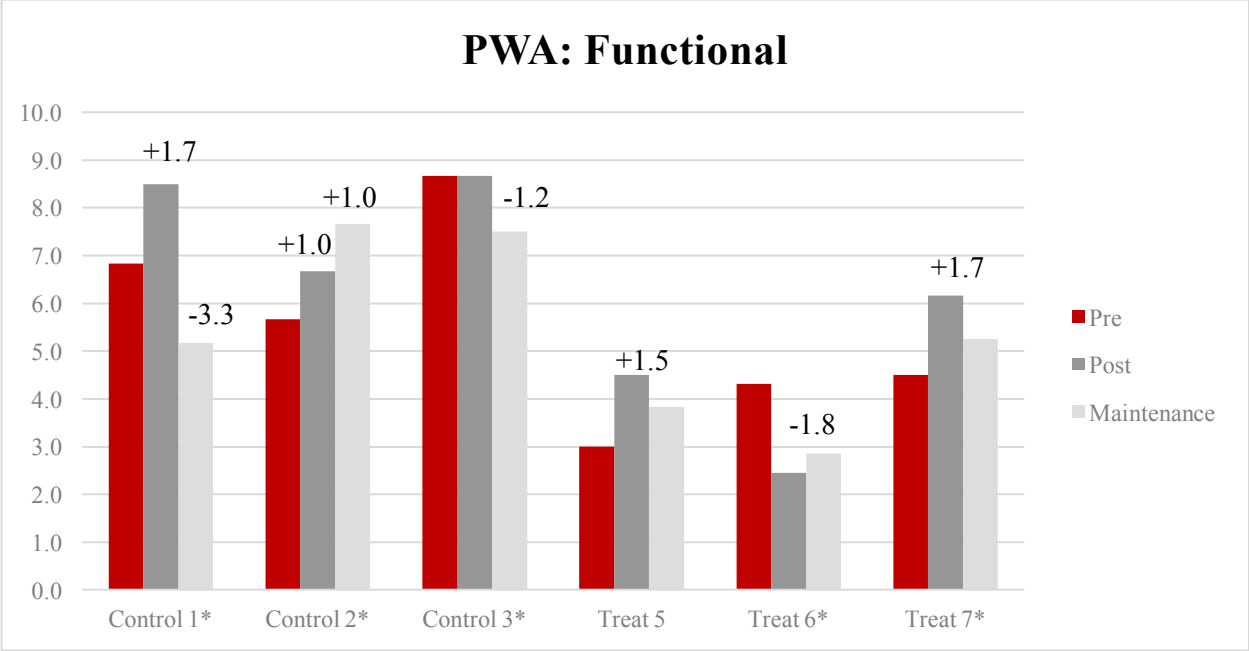


Figure 8: Graph representing PWA average scores for the functional construct for each participant from pre, post, and maintenance time points. Numbers highlight +/- meaningful deflections. *Denotes PWA who were also rated in Study 1.

For the functional construct (as seen in Figure 8), two out of the three participants in the control group (Control 1 and Control 2) showed positive deflections from pre-treatment to post-treatment, with two people (Control 1 and Control 3) showing a negative deflection from post-treatment to maintenance. Control 2 also showed a positive deflection from post-treatment to maintenance, as well as pre-treatment to maintenance. Similarly, in the treatment group, two out of the three people (Treat 5 and Treat 7) showed a positive deflection from pre-treatment to post-treatment; however, the remaining participant (Treat 6) showed a negative deflection from these two time points.

Agreement Between Groups

Given that the participants with aphasia overlapped in the PWA's self-ratings and naïve listener ratings, agreement between the two groups was also examined. Participants Treat 6 and Treat 7 were sampled in both groups. Overall, the naïve listener group

perceived more positive deflections for these participants compared to the participant's self-rating. On average, the naïve listener group perceived two positive deflections across the three constructs for Treat 6, as well as two positive deflections for Treat 7. There were no negative deflections perceived for these participants in the naïve listener group. For the PWA group's self-ratings, Treat 6 perceived negative deflections across all three constructs, with no positive deflections. Treat 7 perceived two positive deflections, with agreement in the functional construct compared to the naïve listener group. For the participants in the control group, there was no agreement between the naïve listener and PWA groups across the three constructs and two time points. Interestingly, the naïve listener group perceived only negative deflections for the control group; however, Control 1 and Control 3 in the PWA group perceived more positive deflections for themselves across the macrolinguistic and functional constructs from pre-treatment to post-treatment.

CHAPTER 6

DISCUSSION

This study focused on validating a questionnaire about listener perceptions of discourse in people with aphasia. The Discourse Rating Scale for Aphasia (DRSA) examined microlinguistic, macrolinguistic, and functional aspects of discourse, consisting of nine questions rated on a visual analog interval scale. Naïve listeners and PWA listened to thirty-second speech samples and completed the DRSA. The goals of the study were to examine the construct validity of the DRSA and to determine whether it was sensitive to changes in discourse samples elicited before and after conversational treatment.

Construct Validity of DRSA

The first aim of the study was to examine the construct validity of the DRSA by correlating average scores on the DRSA with a corresponding standardized measure of language ability. Results indicated that each item on the DRSA highly correlated with a subtest of a corresponding assessment (see Table 2). Additionally, the total DRSA score also showed significant correlations with total assessment scores (See Table 3). This provides evidence that the DRSA is measuring what it intends to measure, and is sensitive to aphasia severity.

A previous study examined the relationship between patient reported outcome measures (PROMs) and standardized measures of language ability using the data from the same conversation treatment study (Navarro, DeDe, & Hoover, 2018). The patient reported outcome measures were the Lubben Social Network Scale (Lubben et al, 2006) and the ACOM. The results showed few significant correlations between the PROMs and subtests of the CAT. However, naming correlated strongly with the ACOM. In this

present study, the DRSA highly correlated with the ACOM, suggesting that microlinguistic features (such as naming), combined with other constructs of discourse analysis, may have a relationship with how people perceive their communicative effectiveness. In fact, the total DRSA score showed strong correlations with three different assessments that measure macrolinguistic, microlinguistic, and functional features of discourse (CAT total picture description score, VNT, and the ACOM, respectively). Therefore, these findings suggest that creating a DRSA that incorporates listener perception of these three identified constructs (macrolinguistic, microlinguistic, and functional) may provide a sensitive measure of discourse. Additionally, these correlations demonstrate that probing listener perception of specific constructs in discourse analysis may provide a valid measure of verbal expression. Therefore, the DRSA may serve as a helpful supplemental tool to assess listener perception of multiple aspects of discourse in aphasia.

Although there are few studies that have utilized rating scales and listener perception for discourse analysis in aphasia, our results are consistent with findings from other researchers. Harmon et al (2016) examined listener perceptions of simulated fluency in PWA. The researchers modified the fluency of the participants' speech samples and had students rate their perceptions of the samples using a rating scale that probed behavioral, cognitive, and affective aspects of connected speech. Researchers found that listener perceptions were more positive as speech fluency improved, which was consistent across items related to speech output, speaker attributes, and listener feelings. These results demonstrated that naïve and trained listeners could reliably perceive changes in connected speech and may serve as another measure of discourse analysis. In

this present study, the DRSA contained an item designed to capture fluency, which was related to speed, and found significant correlations with the speed measure on the CAT picture description.

In another study, Kong & Wong (2018) correlated three different discourse analysis approaches (proposition-based, linguistically based, and story grammar) with listener ratings on the same questionnaire used in Harmon et al (2016). The researchers found that the aphasia severity was highly correlated with listener ratings of speech output, speaker's attributes, and the listener's comfort while listening to the sample. These results are similar to the findings in this present study, except we also found significant correlations between functional measures of discourse and the DRSA, which adds to the versatility of this questionnaire. Thus, this present study adds to a growing body of literature to support the use of listener perceptions when assessing discourse in aphasia.

Sensitivity to Treatment

Another aim of the study was to determine if the DRSA was sensitive to treatment effects in people with aphasia. In the same conversation treatment study that the existing data were drawn from (DeDe et al., 2019), researchers found that the ACOM (functional measure of connected speech) and the CAT picture description total score (macrolinguistic and functional measure of connected speech) were sensitive to treatment changes in the large treatment group.

The present study did not find any evidence of treatment changes using the DRSA. This may be due to the small sample size compared to the original study (n= 48 PWA). For the naïve listener group, there were no significant differences from pre to

post-treatment for either the control or treatment groups. When descriptively looking at the data calculated for each construct, the trends within the naïve listener group ratings suggested that there were four more positive deflections in the treatment group overall compared to the control group. Additionally, the naïve listener data showed that the functional construct may be more sensitive to treatment given that the controls showed either negative deflections or changes <10; however, the treatment group showed positive deflections or changes <10. However, we would need more data to confirm this.

For the PWA group, the positive and negative deflections were fairly similar for both the control and treatment groups across constructs and time points. Therefore, there were no definitive trends in the PWA data to suggest that the DRSA was sensitive to treatment changes. On the contrary, the trends in the naïve listener data are a positive sign that the DRSA may be sensitive to treatment changes. However, these data should be interpreted cautiously since these are merely trends within the data and not statistically significant.

A related point is that inspection of the data revealed that naïve listeners rated PWA differently than the PWA rated themselves. Interestingly, the PWA in the treatment group perceived more negative deflections in themselves compared to the naïve listeners, whereas the PWA in the control group perceived more positive deflections in themselves compared to the naïve listeners. Therefore, it is important to also interpret individual differences in perception and attitude towards one's aphasia, as well as outside supports, expectations of oneself, and environment. This could be examined qualitatively through self-report, or through the use of quality of life measure, such as the Stroke and Aphasia Quality of Life Scale (SAQOL; Hilari et al., 2003) to assess overall impact of aphasia or

the Lubben Social Network Scale (Lubben et al., 2006) to look at social support. These factors can make a noteworthy difference in how an individual rates their own discourse, and it was evident in this study that for specific participants, naïve listeners rated their speech samples more positively compared to the PWA's self-ratings.

Limitations

There were several limitations to this study. One issue was minor discrepancies in task administration across the groups. For example, the PWA group completed the DRSA on paper that was formatted on a rating scale that was ranked from one to 10. The naïve listener group completed the DRSA on Qualtrics, which was formatted as a rating scale that was ranked zero to 100. Therefore, the scales were off by one interval and average scores were computed somewhat differently. It is not clear whether these differences would affect the overall results, but it would be helpful to use more consistent procedures across groups. Additionally, the groups rated different speech samples. The PWA group rated three different samples: the CAT picture description, Nicholas and Brookshire's Cat Rescue picture, and the Cinderella Retell. The naïve listener group only rated the Cinderella retell in an effort to reduce testing time to about an hour. The Cinderella retell requires the individual to spontaneously structure a cohesive fictional narrative without the help of visual support. Given that the Cinderella retell is a more demanding task compared to the picture descriptions, the results may have differed if the naïve listener group also rated the picture descriptions. Additionally, both groups could have also rated conversational samples (e.g. stroke story) given that this would be a more accurate reflection of functional communication. Therefore, in order to have the naïve listeners rate the same samples as the PWA, we could have had them come in for two sessions and

rate more samples. Lastly, we were aiming for a larger sample size; however, due to the COVID-19 pandemic, the sample size was reduced significantly. Therefore, it was harder to detect significant changes in the data. Despite these limitations, the data trends suggest that the DRSA could be sensitive to treatment effects based on naïve listener ratings. However, the design should be replicated across a larger sample size for PWA self-ratings to further validate the measure.

CHAPTER 7

CONCLUSION

Currently, there is a wide range of tools that measure discourse in aphasia, but, each measure assesses one specific level of discourse analysis, whether it is objective or self-reported. Given that discourse production is multifaceted and crucial for meaningful interactions, discourse should be analyzed at all levels (microlinguistic, macrolinguistic, and functional), as well as incorporate the individual's perception of their own discourse. This study aimed to create a valuable tool in providing information about how an individual with aphasia feels as a communicator. However, the long term goal is that clinicians begin to integrate these objective and functional measures into their assessment of discourse, while also taking into consideration the listener's judgment of their own communication as an outcome measure.

REFERENCES

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, *14*(9), 875-892.

Bijur, P. E., Silver, W., & Gallagher, E. J. (2001). Reliability of the visual analog scale for measurement of acute pain. *Academic Emergency Medicine*, *8*(12), 1153-1157.

Cherney, L. R., Babbitt, E. M., Semik, P., & Heinemann, A. W. (2011). Psychometric properties of the communication Confidence Rating Scale for Aphasia (CCRSA): phase 1. *Topics in stroke rehabilitation*, *18*(4), 352-360.

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

DeDe, G., Hoover, E., & Maas, E. (2019). Two to tango or the more the merrier? A randomized controlled trial of the effects of group size in aphasia conversation treatment on standardized tests. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1437-1451.

Dias, J. J., Bhowal, B., Wildin, C. J., & Thompson, J. R. (2001). Assessing the outcome of disorders of the hand: is the patient evaluation measure reliable, valid, responsive and without bias?. *The Journal of Bone and Joint Surgery. British volume*, *83*(2), 235-240.

Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: Have we reached the tipping point?. *Aphasiology*, *32*(4), 459-464.

Frattali, C., Thompson, C. K., Holland, A. L., Wohl, C. B., & Ferketic, M. (1995). Functional Assessment of Communication Skills for Adults.

Harmon, T. G., Jacks, A., Haley, K. L., & Faldowski, R. A. (2016). Listener perceptions of simulated fluent speech in nonfluent aphasia. *Aphasiology*, *30*(8), 922-942.

Hilari, K., Byng, S., Lamping, D. L., & Smith, S. C. (2003). Stroke and aphasia quality of life scale-39 (SAQOL-39) evaluation of acceptability, reliability, and validity. *Stroke*, *34*(8), 1944-1950.

Holland, A. L. (1980). *Communicative abilities in daily living (CADL): A test of functional communication for aphasic adults*. Baltimore, MD: University Park Press.

Hula, W. D., Doyle, P. J., Stone, C. A., Austermann Hula, S. N., Kellough, S., Wambaugh, J. L., ... & St. Jacque, A. (2015). The Aphasia Communication Outcome Measure (ACOM): Dimensionality, item bank calibration, and initial validation. *Journal of Speech, Language, and Hearing Research*, *58*(3), 906-919.

Kagan, A. (1998). Supported conversation for adults with aphasia: Methods and resources for training conversation partners. *Aphasiology*, *12*(9), 816-830.

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18(*2), 124-132.

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*(4), 1491-1505.

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

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Development and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*(1), 113-124.

Lubben, J., Blozik, E., Gillmann, G., Iliffe, S., von Renteln Kruse, W., Beck, J., & Stuck, A.(2006). Performance of an abbreviated version of the Lubben Social Network Scale Among three European community-dwelling older adult populations. *The Gerontologist, 46*(4), 503-513.

Marini, A., Andreetta, S., Del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, *25*(11), 1372-1392.

Murray, P. L., & Mayer, R. E. (1988). Preschool children's judgments of number magnitude. *Journal of educational psychology*, *80*(2), 206.

Navarro, J., DeDe, G., & Hoover, E. (2018). Patient reported outcomes and language ability in aphasia. Poster presented at Clinical Aphasiology Conference, Austin, Texas.

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, Language, and Hearing Research*, *36*(2), 338-350.

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults

with aphasia. *Journal of Speech, Language, and Hearing Research*, *38*(1), 145-156.

Parr, S. (2001). Psychosocial aspects of aphasia: Whose perspectives?. *Folia Phoniatrica et Logopaedica*, *53*(5), 266-288.

Parr, S. (2007). Living with severe aphasia: Tracking social exclusion. *Aphasiology*, *21*(1), 98-123.

Prins, R. S., Snow, C. E., & Wagenaar, E. (1978). Recovery from aphasia: Spontaneous speech versus language comprehension. *Brain and Language*, *6*(2), 192-211.

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*(3), 193-218.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, *37*(3), 440-479.

Simmons-Mackie, N., & Elman, R. J. (2015). Negotiation of identity in group therapy for aphasia: the Aphasia Café. *International Journal of Language & Communication Disorders*, 1-12.

Sriwatanakul, K., Kelvie, W., Lasagna, L., Calimlim, J. F., Weis, O. F., & Mehta, G. (1983). Studies with different types of visual analog scales for measurement of pain. *Clinical Pharmacology & Therapeutics*, *34*(2), 234-239.

Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2014). Measuring outcomes in aphasia research: A review of current practice and an agenda for standardisation. *Aphasiology*, *28*(11), 1364-1384.

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, *13*(4), 227-236.

WHO. International Classification of Functioning, Disability and Health (ICF). Geneva, Switzerland: World Health Organization; 2001.

Wright, H. H. (2011). Discourse in aphasia: An introduction to current research and future directions. *Aphasiology*, *25*(11), 1283-1285.

Yaruss, J. S., & Quesal, R. W. (2006). Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders*, *31*(2), 90-115.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, *11*(6), 499-512.

APPENDIX A

DISCOURSE RATING SCALE FOR APHASIA (NL)

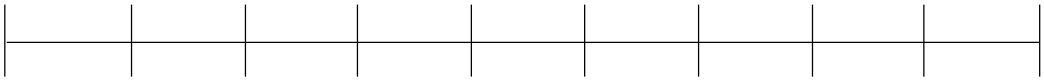
1. How complete/detailed was the person's message?

No details — Very detailed

2. How easy would it be to have a conversation with this person?

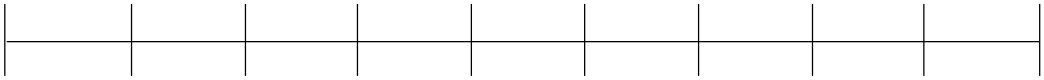
Very difficult — Very easy

3. How much of the grammar was correct?



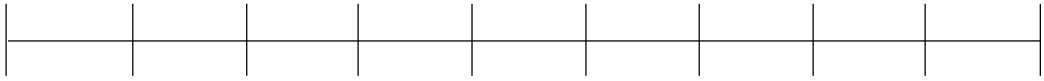
A lot of mistakes — All correct

4. How much of what the person said was relevant to the main point?



Not relevant — Very relevant

5. How often did the person have difficulty thinking of the word they wanted to say? (Even if they thought of it eventually)



A lot — Never

6. How often did the person use the wrong word? (Even if they used the right word eventually)

A lot Never

7. How repetitive was the person's message?

Very repetitive Not repetitive

8. How was the person's rate of speech?

Too slow Too fast

9. What percent of the overall message was understood?

0% 100%

APPENDIX B

DISCOURSE RATING SCALE FOR APHASIA (PWA)

1. How complete/detailed was your message?

No details — Very detailed

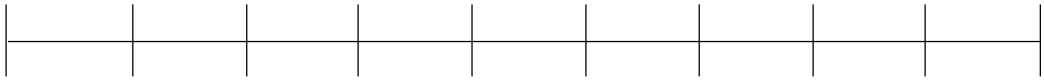
2. How much of the grammar was correct?

A lot of mistakes — All correct

3. How much of what you said was relevant to the main point?

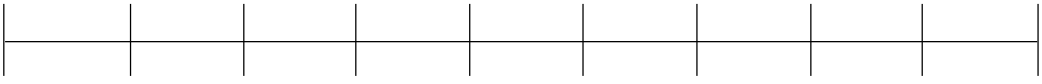
Not relevant — Very relevant

4. How often did you have difficulty thinking of the word you wanted to say? (Even if you thought of it eventually)



A lot — Never

6. How often did you use the wrong word? (Even if you used the right word eventually)

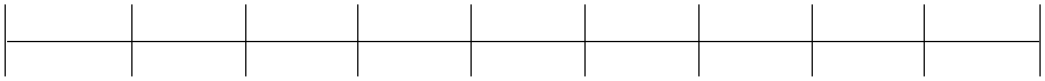


A lot — Never

7. How repetitive was your message?

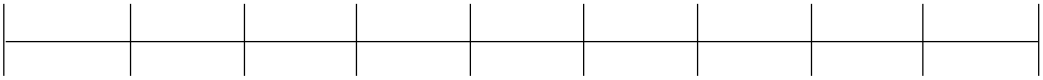
Very repetitive — Not repetitive

8. How was your rate of speech?



Too slow — Too fast

9. How confident do you feel about your speech in this sample?



Not confident — Very confident