TEST-RETEST STABILITY OF DISCOURSE MEASURES IN INDIVIDUALS WITH APHASIA

by

Kristina Marie Fassbender

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

in

Speech-Language Pathology

Department of Communication Sciences and Disorders

The University of Utah

August 2016

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Kristina Marie Fassbender**

has been approved by the following supervisory committee members:

Julie Wambaugh	, Chair	**5/10/2016** Date Approved
Sean Redmond	, Member	**5/10/2016** Date Approved
Shannon Mauszycki	, Member	**5/10/2016** Date Approved
Jacob Kean	, Member	**5/10/2016** Date Approved

and by **Michael Blomgren** , Chair of

the Department of **Communication Sciences and Disorders**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

This investigation examined the effect of repeated sampling (i.e., test- retest) produced within the context of discourse elicited by the Nicholas and Brookshire (1993) discourse elicitation and language analysis procedures. The Nicholas and Brookshire (1993) Correct Information Units (CIUs) are extensively used in aphasic literature to gauge treatment outcomes, but few researchers have examined the temporal stability of this language measure in individuals with aphasia.

Eighteen individuals with aphasia produced language samples over three sampling times. A repeated measures, group design was used to examine the stability of language measures over repeated sampling occasions. The following measurements were calculated and compared: total number of CIUs, percent CIU productivity, number of CIU nouns, number of CIU verbs, open class CIU words, CIU closed class words, well-formed sentences, and lexical diversity.

Values for correlation coefficients were used to assess group stability of performance and standard error of measurement was used to assess individual stability of performance. Measures stable enough to use in group research included number of words, number of CIUs, percent CIUs, number of CIU nouns, number of CIU verbs, number of CIU open class words, and number of CIU closed class words. At the individual level, no participants achieved stability in performance across all measures, but 1 participant achieved stability in performance for all but CIU open class words. The majority of the

participants were not stable in performance for the majority of the measures. Researchers and clinicians using the Nicholas and Brookshire (1993) language elicitation system can expect stability in performance for the examined language measures in groups of participants. For individuals, performance for the examined language measures is expected to be not stable in performance for some and stable in performance for others.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Aphasia is an acquired language disorder caused by a focal brain lesion in the tabsence of other cognitive, motor, or sensory impairments. Aphasia may impact numerous aspects of communicative functioning and life participation (Papathanisiou, Coppens, & Potagas, 2013). Communication is often disrupted by word-finding problems, language comprehension deficits, and language production deficits. Aphashia typically affects all language modalities, with patterns of impairment being unique for each person with aphasia (PWA). Fortunately, there are evidence-based treatments for aphasia that have proven to be effective at mitigating chronic language deficits.

In the treatment of aphasia, discourse abilities have recently received increased focus (Armstrong & Fergusen, 2010; Wright, 2011). Discourse is defined as communication beyond the single sentence level (Papathanisiou, Coppens, & Potagas, 2013). As discussed by Wright (2011), discourse is pertinent to the study of aphasia for several reasons: 1) persons with aphasia (PWA) often have difficulties communicating in the context of discourse, 2) discourse level analyses can potentially be used to objectively evaluate ability to communicate with others, and 3) discourse measures may be used to document change as a response to treatment.

Various language measures have been proposed for measurement within the context of discourse produced by PWA (Berko-Gleason, Goodglass, Obler, Green, Hyde,

& Weintraub, 1980; Berndt & Schwartz, 1989; Nicholas, & Brookshire, 1993; Yorkston, & Beukelman, 1980). However, as noted by Boyle (2014), test-retest stability data are not available for the majority of such measures. The purpose of the current project is to examine the test-retest stability of select language variables produced within the context of discourse by PWA. In this introduction, an overview of the types of discourse that are frequently studied, and the types of discourse analyses that have been proposed/utilized in the study of aphasic language production, will be discussed. In addition, the need for demonstration of test-retest stability will be addressed.

Discourse Elicitation Methods and Types of Discourse

The study of discourse as it relates to aphasia has been considered from various perspectives (e.g., structuralist and functionalist) and has been defined in numerous ways with definitions often relating to the perspective (for a review, see Armstrong, 2000). For the purpose of this project, a broad definition of discourse will be utilized: discourse is any language that surpasses the boundaries of solitary sentences/utterances and is used to convey a message (Wright, 2011).

Clinical discourse elicitation has traditionally employed pictorial stimuli (Olness, 2006). Pictorial stimuli elicitation tasks have typically taken the form of requests for descriptions of single pictures and picture sequences (Goodglass & Kaplan, 1983; Kertesz, 1982; Nicholas & Brookshire, 1993). Story retell has also been frequently employed in which a familiar story, such as the "Cinderella Story," has been used with or without accompanying picture stimuli (Doyle, McNeil, Spencer, Jackson Goda, Cottrell, & Lutig 1998; Saffran, Berndt, & Schwartz, 1989).

Various other nonpictorial elicitation tasks have also be used in the study of aphasic discourse. For example, personal narratives require the PWA to describe their life experiences or personal perspectives (Hinkley, 2007). Procedural discourse elicitation entails description of activities involved in the completion of a specified procedure such as writing and sending a letter (Nicholas & Brookshire, 1993; Wambaugh, Nessler, & Wright, 2014). Structured interviews, role-playing, and conversation have also been employed to elicit discourse from PWA. In addition, spontaneous discourse production (language that has not been formally elicited) has served as a source of discourse analysis in PWA (Prins & Bastiaanse, 2004).

There are numerous ways that types of discourse have been described that often relate to the method of elicitation or the purpose of the discourse. Some researchers have described discourse in terms of genres, such as narrative discourse (e.g., descriptions of scenes) or procedural discourse (e.g., descriptions of procedures) (Armstrong, 2000). Others have used degree of spontaneity to describe the discourse sample (e.g., semispontaneous – supported by pictures, spontaneous – elicited through interview) (Prins & Bastiaanse, 2004). There is currently no agreed-upon method for describing discourse samples in aphasia, with perhaps the exception of use of the descriptive terms, monologue, and dialogue.

There is a limited, but growing, body of evidence that suggests that elicitation method and/or type of discourse may influence results of discourse-based language analyses in PWA (Armstrong, 2000; Armstrong, Ciccone, Godecke, & Kok, 2011). Due to the potential variability of discourses in PWA across context

and topic, Armstrong and Fergusen (2010) recommend that researchers and clinicians use a variety of discourse elicitation techniques to create an adequate sample size and a variety of communication behaviors when making clinical or research decisions about the language of PWA.

For the purposes of clinical application, extensive sampling of discourse is often fnot easible due to time constraints and issues with limited reimbursement. Nicholas and Brookshire (1993) developed and validated a clinically applicable discourse sampling procedure for use in the quantification of informativeness. Their procedures are used to elicit a combination of narrative, procedural, and personal recount samples of discourse. Test-retest stability data are available for the 10-item Nicholas and Brookshire sampling battery as well as for a five-item version of the battery for several measures that center around production of correct information units (CIUs) and words (not specified by word type) (Boyle, 2014; Brookshire & Nicholas, 1994; Nicholas & Brookshire, 1993). Brookshire and Nicholas (1994) reported that group stability of CIU scores was strong (*r*-correlation coefficient = 0.94) but that there was also individual variability. Unfortunately, the authors did not expand on how the discourse varied at the individual level further than minimum-maximum values or give detail on the profiles of the individual participants' scores. Information is not available concerning the test-retest stability of other language measures that may be obtained from the discourses elicited using the Nicholas and Brookshire procedures.

Types of Discourse-Based Language Analyses

Structuralist approaches. Analyses of language produced by PWA in the context of discourses has often been driven by a structuralist perspective (Armstrong, 2000). That is, analyses have been focused on the "language as a system, in and of itself" and have been directed mainly toward text microstructure (Armstrong, 2000, p. 877; Prins & Bastiaanse, 2004). Examination of discourse microstructure typically involves the analysis of individual lexical types (e.g., nouns, verbs) and syntactic units (dependent clauses, types of sentences). Lexical and syntactic structural analyses have been undertaken in the context of discourse. Lexical analyses have included analyses of instances of word-retrieval difficulties (e.g., types of paraphasias) and tallies/comparisons of different word classes (e.g., nouns, verbs, adjectives). Lexical diversity, a reflection of range of vocabulary, has also received recent attention (Fergadiotis, Wright, & Green, 2011; Fergadiotis, Wright, & West, 2013). Syntactic analyses have included measures of syntactic complexity, clause-argument structure, and syntactic errors.

Prins and Bastiaanse (2004) provided a critical review of methods of discourse analysis in aphasia (note: these authors used the term "spontaneous speech" to encompass different types of discourse). In their summarization of methods of discourse-based "linguistic" analyses, Prins and Bastiaanse described qualitative and quantitative methods.

Qualitative methods have been used to provide a characterization of linguistic skills and have usually taken the form of rating scales. Ratings have often been employed in assessment tools in which a restricted number of scales such as

phrase length, grammatical form, and paraphasias are used to depict select language characteristics (e.g., Boston Diagnostic Aphasia Examination; Goodglass & Kaplan, 1983). As discussed by Prins and Bastiaanse, there are numerous limitations associated with such rating scales. In particular, insufficient demonstrations of test-retest reliability are a weakness in the use of linguistic rating scales to measure changes in discourse.

Prins and Bastiaanse (2004) described several of the *quantitative* procedural systems that have been developed for the analysis of linguistic variables produced in aphasic discourse. Of the available analysis systems, Quantitative Production Analysis (QPA) developed by Saffran and colleagues (1989) and Prins and colleagues (1978) is relatively more comprehensive than others.

QPA is a system for quantifying linguistic variables that are pertinent to the study of agrammatic language production. Discourse samples are elicited in the form of the telling of well-known fairy tales (e.g., Cinderella Story) so that a minimum of 150 words are available for analysis (the QPA is applied to the first 150 words). The QPA is used to tally various lexical items independent of utterance type (e.g., open-class words, nouns, determiners, verbs). In addition, utterances that are designated as minimal sentences (e.g., noun plus verb) are further analyzed for structural complexity (e.g., sentence elaboration index, median length of utterance).

Rochon, Saffran, Schwartz, and Berndt (2000) provided test-retest reliability data for the QPA. Eighteen participants provided narrative samples at two separate times. The intervening time intervals were not specified nor were possible therapeutic activities during the intervals. Although the investigators indicated that test-retest reliability was calculated, utterances from each sampling occasion were conflated; “all the even numbered utterances

from the two samples were combined and compared to all the odd numbered utterances from the two samples" (p. 200). Intraclass correlations for the two samples (which the authors termed test-retest) ranged from .53 (well-formed sentences) to .92 (proportion of pronouns).

Functionalist approaches. Discourse-based language analyses have also been undertaken from a functionalist perspective wherein the interest lies in use of language in social context. That is, the focus of functionalist analyses has been on the overall meaning and organizational framework of the text. Functionalist analyses of aphasic discourse have included examination of various macrostructural, or "above the word," elements of discourse, such as discourse structure (Ulatowska, Allard, & Chapman, 1990; Ulatowska, Reyes, Santos, & Worle, 2011), interactional elements of discourse (e.g., turn-taking, conversational repairs), speech functions, nonverbal and paralinguistic behaviors, and overall discourse effectiveness (Armstrong, 2000).

Other approaches. Nicholas and Brookshire (1993) developed a discourse elicitation procedure for use in quantifying the informativeness of connected language. In addition to elicitation procedures, Nicholas and Brookshire devised operationalized procedures for measuring correct information units (CIUs) and words. As discussed by Armstrong (2000), Nicholas and Brookshire's CIU metric crosses structuralist/functionalist boundaries. CIUs are "words that are intelligible in context, accurate in relation to the picture(s) or topic, and relevant to and informative about the content of the picture(s) or the topic" (p. 348). Nicholas and Brookshire's CIU metric has been used relatively extensively in aphasia treatment

investigations as an indirect measure of word-retrieval (see Boyle, 2014, for a review).

Discourse and Aphasia Treatment

In general, difficulties at the levels of phonology, semantics, and syntax are considered to be the primary sources of communication problems in aphasia (Papathanasiou, Coppens, & Potagas, 2013). Although some macrostructural aspects of language have been shown to be occasionally disrupted in PWA when compared to the macrostructural components of discourse in non-brain-damaged individuals, such disruptions are likely related to deficiencies in language microstructure (Armstrong, 2000).

Aphasia treatments reflect the evidence that deficits in structural aspects of language are the primary source of communication difficulties in aphasia. That is, numerous treatments have been devised to target phonology, semantics, and syntax (Papathanasiou et al., 2013; www.ANCDS.org). Treatments that are derived from a functionalist perspective have most often been focused on provision of compensatory mechanisms for working around structural language difficulties (Raymer & Rothi, 2010).

Regardless of treatment approach, performance in discourse has been increasingly focused upon as a means of measuring outcomes from a potentially, ecologically valid perspective (Wright, 2011). Unfortunately, the suitability of using discourse as the context for outcome measurement has not been clearly established for many language variables (Boyle, 2014; Prins & Baastiaanse, 2004).

Importance of Test-Retest Stability of Discourse Measures

According to Boyle (2014), "Test-retest stability refers to the assessment of whether a test produces the same results during repeated applications when the participants who are being tested have not changed on the domain being measured" (i.e., are not receiving treatment). Test-retest stability is crucial for both clinicians and researchers to achieve in order to gauge treatment effect sizes and track progress. A measure that is not reasonably stable from session to session is neither a valid nor reliable measure of an individual's discourse (Boyle, 2014). In the aphasia literature, there is no requisite amount of time required between test and retest, and the interval length varies from 1 day to several weeks, although the typical amount of time is 2 to 14 days (Boyle, 2014; Fitzpatrick, Davey, Buxton & Jones, 1998).

There is a paucity of literature regarding test-retest stability of discourse in individuals with aphasia that compromises the ability to separate discourse variability as the result of day-to-day changes from treatment effects or spontaneous recovery. Little has been reported on the test-retest stability of the discourse measures mentioned previously in this paper. In particular, the Nicholas and Brookshire (1993) CIUs are extensively used in aphasic literature, but few researchers have examined the temporal stability in individuals with aphasia.

Cameron, Wambaugh, and Mauszycki (2010) replicated the Nicholas and Brookshire tasks with five fluent and six nonfluent PWA and correlated the linguistic outcomes with months post onset of aphasia and scores on standardized assessments. The authors found that there were not significant differences in the group test-retest discourse, which corroborated the findings of Nicholas and

Brookshire (1993, 1994). However, the authors found that individual variability for %CIU, WPM, and CIUs/minute was greater than reported by Nicholas and Brookshire (1993).

Boyle (2014) examined the test-retest stability of word-retrieval measures in individuals with aphasia who completed a 5-minute discourse stimulus task across three sessions. The author found that only words per minute, CIUs per minute, and per cent of T-units with time fillers or delays were stable enough from session to session to use for individual clinical decisions. Boyle (2004) and Fitzpatrick et al. (1998) recommend a reliability standard of at least 0.70 to assess groups in research studies and 0.90 to make clinical decisions about individuals. Individual clinical decision making requires greater reliability standard to minimize the confidence interval for the measurement and therefore obtain a more precise measurement of the individual's true score.

Purpose of the Research

The purpose of this project is to investigate the test-retest stability of several language variables produced within the context of discourse elicited with the protocol developed by Nicholas and Brookshire (1993). In addition to the CIU and word measures defined and studied by Nicholas and Brookshire, other linguistic variables were measured. These variables were determined by examining the discourse generalization outcomes from a compilation of aphasia treatment reports created by Dr. Pelagie Beeson (http://aphasiatx.arizona.edu/) in conjunction with the Academy of Neurologic Communication Disorders and Sciences (ANCDS) treatment guidelines project (www.ANCDS.org).

Dr. Beeson's corpus reflects extensive and continual data-based searches for

published aphasia treatment studies. The reports are grouped with respect to their primary outcome focus: overall language performance, lexical retrieval, speech production and fluency, syntax, and alternative communication. The discourse generalization outcomes from these studies were reviewed and the most common variables were considered for inclusion in the language analyses for this project.

Ideally, for the purposes of pretreatment, posttreatment outcome measurements and test-retest stability should be established for the time intervals across which the measurements would be occurring. However, aphasia treatments may require many months to complete. Extended sampling intervals (e.g., upwards of 6-9 months) were considered to be not feasible for this project. As a compromise, repeated sampling of discourse was conducted at three time intervals reflecting shorter and intermediate interval lengths: initial sample, 1 week following initial sample, and 4 weeks following initial sample.

Specific Aims

The purpose of the current investigation is to examine the stability of language production measures commonly used in aphasia language production treatment investigations. This investigation was designed to determine the effect of repeated sampling (i.e., test-retest) on the following language production measures as demonstrated within a group of 18 participants with chronic aphasia: total number of CIUs (Nicholas & Brookshire, 1993), percent CIU productivity (Nicholas & Brookshire, 1993), number of nouns, number of verbs, open class words, closed class words, well-formed sentences (Saffran et al., 1989), and lexical diversity.

The outcome of this project will inform researchers and speech language pathologists about the effect of repeated sampling on measurement of aphasic language. These measures will be clinically relevant to professionals who wish to improve aphasic language therapy outcomes.

Based on the existing relevant literature on test-retest stability, the specific experimental hypotheses regarding stability of performance were as follows:

Hypothesis 1: Group performance for the measure under study will be stable across the three sampling times. That is, for the group, scores from each sampling time will reflect a moderate to high degree of correspondence to the other sampling times.

Hypothesis 2: Individual performance for the measures under study will be variable for some participants. Specifically, obtained values will exceed score bands used to predict performance across sampling times for at least some individuals.

Table 1: Participant characteristics

ID	Sex	Age	BI location/type	MPO	Years of education	Premorbid handedness	Hemiparesis	Race/ethnicity
P1	M	30	L MCA/ischemic CVA	26	13	R	R UE, LE	White-nH/L
P2	F	84	L&R(frontal white matter older stroke) MCA	55	12	L	R UE	White-nH/L
P3	M	58	L MCA ischemic	88	11	R	R UE, LE	White-nH/L
P4	F	46	L MCA & PCA / ischemic CVA & PRES	24	14+	R	R UE	White-nH/L
P5	M	49	L MCA / ischemic CVA	111	12	R	R UE, LE	White-nH/L
P6	F	47	L MCA / ischemic CVA	90	14	R	UE, LE	White-nH/L
P7	M	38	L subarachnoid hemorrhagic in MCA	40	11	R	R UE, LE	White-nH/L
P8	M	43	L frontoparietal w/BG & small R parietal / ischemic CVA	119	14	R	UE, LE	White-nH/L
P9	M	65	L MCA, ICA thrombosis with slight hemorrhagic change	137	16+	R	R UE, LE	White-nH/L
P10	M	55	L BG hematoma, vasculitis	268	14	R	R UE, LE	White-nH/L
P11	F	29	L AVM led to hemorrhagic stroke	125	12	R	R UE, LE	White-nH/L
P12	F	64	L MCA / ischemic CVA	133	14-15	R	R UE, LE	White-nH/L
P13	F	35	intraparenchymal hemorrhage	63	12	R	UE, R LE	White-nH/L
P14	F	70	L ischemic, minor stroke 5 years before	100	13	R	R UE	White-nH/L
P15	F	54	L ischemic, posterior frontal, anterior temporal, anterior parietal, basal ganglia, deep white matter tracks	137	13	R	UE, LE	White-nH/L
P16	M	52	L MCA / ischemic CVA	37	16	L	UE, LE	White-nH/L
P17	M	61	L MCA / ischemic CVA post L temporal anterior lobectomy	66	18	R	R UE, LE	White-nH/L
P18	M	68	L MCA / ischemic CVA	17	13	R	R UE, LE	White-nH/L

Note: BI = brain injury; MPO = months post onset; CVA = cerebrovascular accident; AVM = arteriovenous malformation; BG = basal ganglia; MCA = middle cerebral artery; L= left; R = right; nH/L= non Hispanic/Latino; UE = upper extremity; LE = lower extremity

Table 2: Participant speech/language profiles

ID	WAB-R AQ	Aphasia type	TONI-4 percentile	TONI-4 descriptive term
P1	69.6	Broca's	50	average
P2	61.75	Broca's	45	average
P3	68.7	Broca's	13	below average
P4	52.4	Broca's	81	above average
P5	63	Broca's	63	average
P6	85.3	Anomic	52	average
P7	65.2	Broca's	37	average
P8	96	Anomic	45	average
P9	57.5	Broca's	58	average
P10	58.6	Broca's	27	average
P11	93.6	Anomic	27	average
P12	86.6	Anomic	52	average
P13	93.7	Anomic	34	average
P14	94	Anomic	55	average
P15	55.7	Broca's	52	average
P16	47.1	Broca's	63	average
P17	60.8	Broca's	84	above average
P18	31.7	Broca's	24	below average

Note: WAB-R AQ = Western Aphasia Battery–Revised Aphasia Quotient; TONI-4 = Test of Non-Verbal intelligence-4; n/a = not administered

METHODS

Participants

Twenty individuals with chronic aphasia who were between 17 and 268 months post onset of single, left-hemisphere stroke were recruited to participate in the study. All were native-English speakers, had completed high school, and were home-dwelling. The participants passed a pure-tone hearing screening at 40dB for 1000 kHz, 2000 kHz, and 4000 kHz for at least one ear. Participants did not receive any speech or language therapy during the time period encompassed by the study (i.e., 6-7 weeks). The participants ranged from 29 to 84 years of age. A summary of participant characteristics is provided in Table 1 (note that data from 18 participants are provided; data from 2 participants was excluded due to insufficient production of CIUs in language samples).

As shown in Table 2, the participants' Test of Non-Verbal Intelligence (TONI-4; Brown, Rita, & Johnsen, 2010) overall percentile scores ranged from 13th to the 84th and Western Aphasia Battery – Revised Aphasia Quotients (WAB-R; Kertesz, 2007) ranged from 47.1 to 93.6. Twelve participants received a diagnosis of Broca's aphasia and 6 participants received a diagnosis of Anomic aphasia according to WAB-R criteria. All had significant word-retrieval difficulties on the Test of Adolescent-Adult Word-Finding (German, 1990). In addition, all demonstrated speech characteristics consistent with acquired apraxia of speech (AOS) as described by McNeil, Robin, and Schmidt (2009).

Experimental Design

A repeated measures, within-group design was used to examine the stability of language measures over repeated sampling occasions. Discourse samples were repeatedly elicited on three occasions with each participant: 1) initial sample, 2) 1 week after the initial sample, and 3) 4 weeks after the initial sample.

Experimental Stimuli

The experimental stimuli included Set A from Brookshire and Nicholas (1994) and the procedures described by Nicholas and Brookshire (1993) were used to elicit the samples. The stimuli consisted of two single pictures, one picture sequence, one biographical narrative about one's Sunday routine, and one procedural request about how to wash dishes by hand. A single examiner elicited all samples. All samples were audio recorded using a Zoom H-4 digital recorder and orthographically transcribed. The orthographic transcriptions were then analyzed for each of the dependent variables described below.

Language Analyses

The following language measures were obtained from the discourse samples: 1) number of CIUs (Nicholas & Brookshire, 1993), 2) words by lexical type, 3) lexical diversity, and 4) sentence/phrase production. Systematic Analysis of Language Samples (SALT; Miller & Iglesias, 2012) was used as a platform for conducting the analyses other than the CIU calculations.

Content information units. CIUs were determined according to Nicholas and

Brookshire's (1993) criteria, which are defined as "words that are intelligible in context, accurate in relation to the picture(s) or topic, and relevant to and informative about the content of the picture(s) or the topic" (p. 340). The percent CIUs (number of CIUs as a percent of the total number of words), and total CIUs per testing sample will be calculated.

Words by lexical type. The following guidelines set by Nicholas and Brookshire (1993) were used to determine which words were included and excluded in the language samples.

> Words or partial words that are not in intelligible in context to someone who knows the picture(s) or topic being discussed will not be counted. Words will be excluded when they are repeated or do not add new information to the utterance, are not necessary for cohesion or grammatical correctness, and are not purposely used to intensify meaning. (Nicholas & Brookshire, 1993, p 340)

For each language sample, the number of CIU words belonging to the following lexical types were counted: nouns, verbs, open class words, and closed class words.

Lexical diversity. The diversity of words in the discourse samples was calculated. Fergadiotis et al. (2013) explain that calculating lexical diversity by type-token ratio may be biased by the length of the sample. A moving average helps control for this factor and may yield a better picture of the speaker's lexical diversity. Consequently, the Moving-Average Type Token Ratio (MATTR-2; Covington & McFall, 2007) was selected for use to calculate lexical diversity. This software program computes a moving average of type-token ratio by averaging the type-token ratios computed in multiple selections of the same text.

Sentence production. Procedures described by Saffran et al. (1989) were used to calculate numbers of complete sentences.

A transcribed and coded discourse sample is shown in Appendix A. Conventions for SALT coding are shown in Appendix B.

Reliability

All language samples were independently verified using the audio recordings by an investigator experienced in transcription of aphasic language production prior to Systematic Analysis of Language Transcripts (SALT) coding. Any discrepancies in transcription were resolved by having both listeners relisten to the audio recording of the language sample section in question, discussing the disagreement, and coming to a consensus transcription. If consensus was not able to be reached, a third listener independently transcribed the section and the corresponding transcription was utilized.

Twenty percent of the discourse samples were selected through stratified, random sampling for the purposes of recoding language behaviors. Specifically, there were three sampling times for each participant with five discourse samples per participant for a total of 15 samples; three discourse samples were randomly selected for each participant and a second investigator assigned lexical codes to each CIU word. The second investigator received training in coding from the principal investigator (PI) using nonselected samples; training continued until tallies resulting from the coding reflected less than 15% disagreement from the PI's tallies. Point-to-point interrater agreement was 92% for CIUs and 96.5% for lexical coding.

RESULTS

Descriptive statistics (means, standard deviations, and minimum - maximum values) are shown for all variables in Table 3 for the group of 18 participants for all sampling times. Because of the wide spread of values across the group, the descriptive data are also shown for the participants divided into two subgroups based upon their verbal productivity as indicated by CIU production: low-CIU = less than 100 CIUs per sample (n=10), and high-CIU = 100 or more CIUs per sample (n =8). See Table 2 for subgroup identification. Descriptive statistics for the subgroups are shown in Tables 4 and 5.

Session-to-Session Stability of Group Measures (Hypothesis 1)

In order to provide descriptive comparisons of performance across sampling times, means of difference scores between sessions and the minimum-maximum difference scores between sessions are shown in Table 6 for all participants, Table 7 for the high-CIU group, and Table 8 for the low-CIU group.

Pearson product-moment correlation coefficients were calculated to assess the degree to which scores for the different sampling sessions were associated. The following comparisons were made: Time 1-Time 2; Time 2- Time 3; and Time 1- Time 3. Pearson product-moment correlations are shown in Tables 3, 4, and 5 for the entire group, the high-CIU group, and the low-CIU group, respectively. As indicated previously, a value

of 0.7 and above was considered to reflect adequate stability (Fitzpatrick et. al 1998).

Correct information units, words, and percent CIUs. Across all sessions, number of words, number of CIUs, and percent CIUs exceeded the $r = .70$ criterion (Fitzpatrick et al., 1998) for the group of 18 participants. As seen in Table 3, the *r*-values ranged from 0.73 to 0.98. Scatterplots displaying number of CIUs for each of the sampling time comparisons (e.g., Session 1 vs. Session 2) are shown in Figures 1, 2, and 3. In addition to illustrating the association of performance between times, these figures provide visual support for the separation of participants into CIU subgroups.

Nouns, verbs, open class words, closed class words. Correlation coefficient values exceeded $r = .70$ in all comparisons for number of nouns, verbs, open class words, and closed class words. Values for correlation coefficients ranged from 0.90 to 0.98.

Lexical diversity. For the lexical diversity measure, the r-values varied across the comparisons: Time 1 – Time 2, r =.47 (weak association), Time 2 – Time 3, r = .51 (moderate association), and Time 1 – Time 3, $r = .22$ (no association).

Number of sentences. Number of sentences per session were not included in the analyses because of the low number of observations (i.e., many participants produced no sentences during one or all sessions).

Session-to-Session Stability of Individual Participants (Hypothesis 2)

Figures 4 to 19 display performance for each participant for each sampling time for all variables: Number of words = Figures 4 and 5; number of CIUs = Figures 6 and 7; percent CIUs = Figures 8 and 9; number of nouns = Figure 10 and 11; number of verbs = Figures 12 and 13; lexical diversity = Figures 14 and 15; number of open class words =

Figures 16 and 17; number of closed class words = Figures 18 and 19.

To address ***Hypothesis 2*** regarding variability of individual performance, standard error of measurement (SEM) was calculated and used to compute score bands (confidence intervals) for each participant. SEM was calculated separately for the 10 participants whose CIU production ranged from 0-99 (low-CIUs) and for 8 participants whose CIU production exceeded 100 (high-CIUs). SEM was calculated for each measure in the respective groups for each session-to-session change (i.e., three SEMs per measure for the high-CIU group and three SEMs per measure for the low-CIU group).

In order to assess the stability of individual performance, score bands were determined for each individual participant for each variable using 1 SEM. The most conservative of the three SEMs was used to derive the bands for individuals within the subgroup. For each participant, each obtained score for the three sampling times was evaluated with respect to the score band. Score differences that exceeded the SEM band were considered to represent lack of stability in performance. If a score was encompassed within the SEM band, the session-to-session change was considered to be stable in performance. A participant was judged to be stable *overall* if each session-to-session change was encompassed by the SEM (i.e., session 1 to 2, session 2 to 3, and session 1 to 3). If a participant's performance was unstable for any one of the comparisons, the participant's performance was judged to be not stable overall. Tables 9, 10, 11, 12, 13, and 14 show the scores for each participant on each language measure, the session-to-session stability decision, and the overall stability decision.

Correct information units, words, and percent CIUs. Participant 15 was the only participant to achieve overall stability across sessions for CIUs, words, and percent

CIUs.

Two participants (of the 18) achieved overall stability in performance for number of words (Participants 9 and 15). These 2 participants were in the low-CIU group. As calculated by the smallest SEM band for each group, participants were permitted a maximum change of 37.9 words for the high-CIU group and 26.49 words for the low-CIU group. Eight participants (of the 18) were stable in performance from session 1 to session 2, 6 participants for sessions 2 to 3, and 2 participants for sessions 1 to 3.

One participant (of 18) achieved overall stability in performance for number of CIUs (Participant 15). Participant 15 was in the low-CIU group. The smallest SEM permitted a maximum change of 27.5 CIUs for the high-CIU group and 8.72 CIUs between sessions for the low-CIU group. Seven participants (of the 18) were stable in performance from session 1 to session 2, 7 participants for sessions 2 to 3, and 5 for sessions 1 to 3.

Eight participants (of 18) achieved overall stability in performance for percent CIUs (Participant 1, 9, 10, 13, 15, 16, 17, 18). One of these participants was in the high-CIU group (Participant 13) and the other 7 were in the low-CIU group. The smallest SEM permitted a maximum change of 3.54 % between sessions for the high-CIU group and 9.93% for the low-CIU group. Eight of participants (of the 18) were stable in performance from session 1 to session 2, 12 participants for session 2 to 3, and 10 for session 1 to 3.

Nouns, verbs, open class words, closed class words. None of the 18 participants achieved stability in performance for number of nouns, verbs, open class words, and closed class words across all sessions.

Three participants of 18 total achieved stability in performance for CIU nouns (Participants 4, 7, 15). Participant 7 was in the high-CIU group and Participants 4 and 15 were in the low-CIU group. The smallest SEM permitted a maximum change between sessions of 9.23 nouns for the high-CIU group and 4.10 nouns for the low-CIU group. Ten participants were stable in performance from session 1 to session 2, 4 for session 2 to 3, and 7 for session 1 to 3.

Four participants of the 18 total achieved stability in performance for CIU verbs (Participants 3, 6, 12, and 15). Three of these 4 participants were in the high-CIU group (Participant 3, 6, and 12). The smallest SEM permitted a maximum change between sessions of 11.03 for the high-CIU group and 2.56 for the low-CIU group. Ten participants were stable in performance from session 1 to session 2, 8 for session 2 to 3, and 7 for session 1 to 3.

Three participants of the 18 total achieved overall stability in performance for open class CIU words (Participants 3, 7, 12). These participants were in the high-CIU group. The smallest SEM permitted a maximum change between sessions of 17.74 words for the high-CIU group and 4.56 words for the low-CIU group. Eight participants were stable in performance from session 1 to 2, 7 for session 2 to 3, and 6 for session 1 to 3.

Five participants of the total 18 achieved overall stability in performance for closed class CIU words (Participants 6, 9, 14, 16, and 17). Two of these 5 participants were in the high-CIU group. The smallest SEM permitted a maximum change between sessions of 17.05 words for the high-CIU group and 4.67 words for the low-CIU group. Eleven participants were stable in performance from session 1 to 2, 8 for session 2 to 3, and 10 for session 1 to 3.

Lexical diversity. Three participants of the 18 total achieved overall stability across sessions for lexical diversity (Participants 11, 12, and 15). Participants 11 and 12 were in the high-CIU group and Participant 15 was in the low-CIU group. The smallest SEM permitted a maximum change between sessions of 0.03 for the high-CIU group and 0.073 for the low-CIU group. Ten participants were stable in performance from session 1 to 2, 10 for session 2 to 3, and 7 for session 1 to 3.

Table 3: Mean score, standard deviation, minimum and maximum values, and SEM for session-to-session change for all participants.

Session	**Number of Words**	**Number of CIUs**	**Percent CIUs**	**Number of CIU Nouns**	**Number of CIU Verbs**	**Number of Open Class CIU Words**	**Number of Closed Class CIU Words**	**Lexical Diversity**
Session 1								
M	269.00	167.11	56.80	51.22	34.17	93.83	72.83	0.74
SD	227.45	202.63	26.45	52.16	39.10	107.45	98.43	0.093
Min-Max	52-966	31-791	13.6-96.7	12-227	4-145	21-452	2-289	0.55-0.89
Session 2								
M	273.39	165.83	52.49	48.61	36.72	92.00	74.61	0.74
SD	207.37	183.35	22.83	48.61	42.91	91.06	104.81	0.079
Min-Max	94-903	35-663	19.8-87.9	14-141	4-146	17-335	0-380	0.62-0.92
Session 3								
M	298.23	172.67	50.75	51.67	34.33	92.89	74.83	0.71
SD	181.75	167.47	22.16	38.82	32.50	77.11	90.15	0.075
Min-Max	114-743	38-614	15.2-88.8	14-164	5-104	27-264	2-282	0.55-0.83
Sessions 1 to 2								
r	0.85	0.93	0.82	.90	0.95	0.90	0.96	0.47
SEM	83.86	48.58	10.27	14.40	8.73	30.68	20.79	0.062
Sessions 2 to 3								
r	0.92	0.96	0.95	.94	0.95	0.98	0.97	0.51
SEM	54.71	33.60	4.71	9.04	8.64	11.50	17.17	0.054
Sessions 1 to 3								
r	0.87	0.98	0.73	.96	0.90	0.94	0.98	0.22
SEM	73.54	27.51	12.59	9.33	11.37	22.58	13.88	0.074

Notes: CIUs = Correct information units; *r* = coefficient of correlation; SEM = standard error of measurement

Table 4: Mean score, standard deviation, minimum and maximum values, and SEM for session-to-session change for the high-CIU group.

Session	**Number of Words**	**Number of CIUs**	**Percent CIUs**	**Number of Nouns**	**Number of Verbs**	**Number of Open Class CIU Words**	**Number of Closed Class CIU Words**	**Lexical Diversity**
Session 1								
M	413	310.88	72.65	83.88	66.25	167.88	146.25	0.70
SD	266.61	238.43	21.37	65.61	39.69	128.79	110.89	0.06
Range	173-966	109-791	21.3-96.7	30-227	21-145	74-452	45-322	0.62-0.81
Session 2								
M	421.25	304.38	69.48	77.88	72.38	161.38	153.25	0.69
SD	236.50	204.70	18.13	47.42	42.96	99.98	117.72	0.03
Range	240-903	109-663	18.1-87.9	27-159	31-146	54-335	43-380	0.62-0.72
Session 3								
M	444.25	310.75	68.71	79.13	65.25	154.88	150.25	0.68
SD	182.24	168.74	17.63	43.23	24.12	78.81	88.94	0.04
Range	197-743	156-489	17.6-88.8	31-164	30-104	75-264	46-282	0.62-0.71
Sessions 1 to 2								
r	0.76	0.89	0.63	0.86	0.90	0.83	0.92	0.40
SEM	108.45	66.6	11.12	19.21	11.25	41.04	28.42	0.03
Sessions 2 to 3								
r	0.85	0.93	0.96	0.92	0.90	0.97	.95	0.59
SEM	84.48	51.5	3.54	14.55	11.03	17.74	22.07	0.03
Sessions 1 to 3								
r	0.83	0.98	0.45	0.97	0.76	0.91	.97	-0.02
SEM	37.90	27.5	13.67	9.23	17.13	29.40	17.05	0.04

Notes: CIUs = Correct information units; *r* = coefficient of correlation; SEM = standard error of measurement

Table 5: Mean score, standard deviation, range of scores, and SEM for session-to-session change for the low-CIU group.

Session	**Number of Words**	**Number of CIUs**	**Percent CIUs**	**Number of Nouns**	**Number of Verbs**	**Number of Open Class Words**	**Number of Closed Class words**	**Lexical Diversity**
Session 1								
M	153.8	52.1	42.42	25.10	8.50	34.60	14.1	0.78
SD	96.20	16.58	22.28	9.21	3.89	11.26	10.64	0.10
Min-Max	52-375	31-86	13.6-86.8	12-40	4-17	21-52	2-33	0.55-0.59
Session 2								
M	157.1	55	32.80	25.2	8.20	36.50	11.7	0.78
SD	57.17	14.22	20.95	10.79	2.90	13.66	8.79	0.080
Min-Max	94-231	35-85	19.8-66.0	12-51	4-13	17-68	0-25	0.64-0.92
Session 3								
M	181.5	62.2	33.43	29.70	9.60	43.30	14.5	0.73
SD	49.73	18.04	13.56	13.68	3.63	16.00	9.95	0.090
Min-Max	114-264	38-86	10.9-56.7	14-60	5-16	27-78	2-32	0.55-0.89
Sessions 1 to 2								
r	0.81	0.60	0.74	0.85	0.14	0.80	0.51	0.25
SEM	30.17	10.39	10.29	4.38	3.19	6.11	6.66	0.079
Sessions 2 to 3								
r	0.85	0.72	0.76	0.87	0.44	0.89	0.76	0.36
SEM	26.49	8.72	9.93	4.10	2.56	4.56	4.67	0.073
Sessions 1 to 3								
r	0.61	0.71	0.72	0.76	-0.06	0.86	0.70	0.076
SEM	43.13	8.78	10.73	5.50	3.54	5.22	5.28	0.087

Notes: CIUs = Correct information units; *r* = coefficient of correlation; SEM = standard error of measurement

Table 6: Mean difference of scores and range of absolute difference scores for all participants.

Measure	**Sessions 1 to 2**		**Sessions 2 to 3**		**Sessions 1 to 3**	
	Mean of Difference Scores	Range of Difference Scores	Mean of Difference Scores	Range of Difference Scores	Mean of Difference Scores	Range of Difference Scores
Words	78.06	0-312	61.00	2-256	84.94	4-311
CIUs	39.06	0-222	31.17	1-174	33.89	1-177
Percent CIUs	10.84	0.51-36.80	5.58	0.94-17.37	12.39	0.39-41.82
CIU Nouns	11.06	0-86	10.28	0-38	11.67	0-63
CIU Verbs	7.44	0-46	9.17	1-44	9.39	0-65
Open Class CIU Words	23.06	1-165	14.67	1-71	20.17	1-169
Closed Class CIU Words	16.33	1-91	16.33	1-98	14.67	1-52
Lexical Diversity	0.063	0.0020-0.22	0.058	0.0030-0.215	0.084	0.0020-0.26

Note: CIUs = Correct Information Units

Table 7: High-CIU group mean difference and range of absolute difference scores.

Measure	**Sessions 1 to 2**		**Sessions 1 to 2**		**Sessions 1 to 3**	
	Mean of Difference Scores	Range of Difference Scores	Mean of Difference Scores	Range of Difference Scores	Mean of Difference Scores	Range of Difference Scores
Words	125.75	8-312	99.75	41-256	109.25	4-311
CIUs	74.75	0-222	54.88	17-174	58.63	26-177
Percent CIUs	8.95	1.30-36.81	2.99	0.35-7.18	10.08	0.75-43.8
CIU Nouns	20.5	1-86	13.75	0-38	17	0-63
CIU Verbs	12.38	1-46	17.13	1-44	15.25	0-65
Open Class CIU Words	44.50	6-165	22.00	4-71	34.00	1-169
Closed Class CIU Words	28.25	1-91	30.00	3-98	25.75	7-52
Lexical Diversity	0.039	0.0020-0.099	0.026	0.0030-0.059	0.053	0.0020-0.13

Note: CIUs = Correct Information Units

Table 8: Low-CIU Group mean difference and range of absolute difference scores.

Measure	**Sessions 1 to 2**		**Sessions 2 to 3**		**Sessions 1 to 3**	
	Mean Difference Scores	Range of Differece Scores	Mean Difference Scores	Range of Difference Scores	Mean Difference Scores	Range of Difference Scores
Words	39.90	0-128	30.00	2-83	65.50	21-138
CIUs	10.50	1-26	12.20	1-24	14.10	1-27
Percent CIUs	9.32	0.51-35.63	6.06	2.27-17.36	11.20	0.39-38.16
Nouns	3.50	0-13	7.50	4-13	3.40	0-220
Verbs	3.50	0-9	2.80	1-8	4.70	1-11
Open Class CIU	5.90	1-20	8.80	1-14	9.10	1-30
Closed Class CIU	6.8	1-24	5.4	1-13	5.8	1-16
Lexical Diversity	0.082	0.010-0.22	0.084	0.019-0.22	0.11	0.0080-0.26

Note: CIUs = Correct Information Units

Table 9: High-CIU group scores and stability of performance decisions for words, CIUs, and percent CIUs.

P	**Number of Words** SEM: 37.90					**Number of CIUs** SEM: 27.5					**Percent CIUs** SEM: 3.54				
	Values T1, T2, T3	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*
3	173, 240, 303	NSt	NSt	NSt	NSt	127, 135, 159	St	St	NSt	NSt	73, 56, 52%	NSt	NSt	NSt	NSt
6	240, 248, 197	St	NSt	NSt	NSt	114, 166, 146	NSt	St	St	NSt	48, 67, 74	NSt	NSt	NSt	NSt
7	329, 294, 385	St	NSt	NSt	NSt	109, 109, 156	St	NSt	NSt	NSt	33, 37, 41	NSt	St	NSt	NSt
8	643, 903, 647	NSt	NSt	NSt	NSt	515, 663, 489	NSt	NSt	St	NSt	80, 73, 76	NSt	St	NSt	NSt
11	966, 654, 743	NSt	NSt	NSt	NSt	791, 569, 614	NSt	NSt	NSt	NSt	82, 87, 83	NSt	NSt	St	NSt
12	216, 397, 527	NSt	NSt	NSt	NSt	209, 238, 279	St	NSt	NSt	NSt	97, 60, 53	NSt	NSt	NSt	NSt
13	396, 273, 350	NSt	NSt	NSt	NSt	343, 240, 311	NSt	NSt	NSt	NSt	87, 88, 89	St	St	St	St
14	341, 361, 402	St	St	NSt	NSt	279, 315, 332	NSt	St	NSt	NSt	82, 87, 83	NSt	St	St	NSt

Notes: P = Participant Identifier, SEM = Standard Error of Measurement; CIUs = Correct Information Unites, T = Time, St = stable, NSt = nonstable

Table 10: Low-CIU group scores and stability of performance decisions for words, CIUs, and percent CIUs.

P	**Number of Words** SEM: 26.49					**Number of CIUs** SEM: 8.72					**Percent CIUs** SEM: 9.93				
	Values T1, T2, T3	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*
1	177, 124, 141	NSt	St	NSt	NSt	86, 61, 80	NSt	NSt	St	NSt	48, 49, 56	St	St	St	St
2	52, 146, 190	NSt	NSt	NSt	NSt	31, 35, 55	St	NSt	NSt	NSt	59, 24, 28	NSt	St	NSt	NSt
4	66, 100, 142	NSt	NSt	NSt	NSt	50, 63, 70	NSt	St	NSt	NSt	75, 63, 49	NSt	NSt	NSt	NSt
5	83, 94, 177	St	NSt	NSt	NSt	72, 62, 86	NSt	NSt	NSt	NSt	87, 66, 49	NSt	NSt	NSt	NSt
9	193, 193, 216	St	St	St	St	50, 49, 62	St	NSt	NSt	NSt	26, 25, 29	St	St	St	St
10	111 , 113, 140	St	NSt	NSt	NSt	43, 52, 60	NSt	St	NSt	NSt	39, 45, 43	St	St	St	St
15	93, 116, 114	St	St	St	St	36, 38, 43	St	St	St	St	39, 33, 38	St	St	St	St
16	375, 247, 250	NSt	St	NSt	NSt	51, 49, 38	St	NSt	NSt	NSt	14, 20, 15	St	St	St	St
17	179, 231, 264	NSt	NSt	NSt	NSt	59, 85, 86	NSt	St	NSt	NSt	33, 37, 33	St	St	St	St
18	209, 207, 181	St	St	NSt	NSt	43, 56, 42	NSt	NSt	St	NSt	21, 27, 23	St	St	St	St

Notes: P = Participant Identifier, SEM = Standard Error of Measurement; CIUs = Correct Information Unites, T = Time, St = stable, NSt = nonstable

Table 11: High-CIU group scores and stability of performance decisions for nouns, verbs, and lexical diversity.

P	**Number of Nouns** SEM: 9.23					**Number of Verbs** SEM: 11.03					**Lexical Diversity** SEM: 0.03				
	Values T1, T2, T3	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*
3	30, 35, 47	St	NSt	NSt	NSt	45, 46, 45	St	St	St	St	0.750, 0.663, 0.617	NSt	NSt	NSt	NSt
6	43, 59, 59	NSt	St	NSt	NSt	21, 27, 30	St	St	St	St	0.805, 0.706, 0.703	NSt	St	NSt	NSt
7	35, 27, 31	St	St	St	St	35, 31, 46	St	NSt	St	NSt	0.671, 0.695, 0.712	St	St	NSt	NSt
8	128, 159, 121	NSt	NSt	St	NSt	100, 146, 104	NSt	NSt	St	NSt	0.682, 0.721, 0.709	NSt	St	St	NSt
11	227, 141, 164	NSt	NSt	NSt	NSt	145, 124, 80	NSt	NSt	NSt	NSt	0.623, 0.621, 0.649	St	St	St	St
12	73, 62, 73	NSt	NSt	St	NSt	52, 57, 63	St	St	St	St	0.664, 0.687, 0.662	St	St	St	St
13	74, 73, 61	St	NSt	NSt	NSt	68, 82, 72	NSt	St	St	NSt	0.671, 0.695, 0.712	St	St	NSt	NSt
14	61, 67, 77	St	NSt	NSt	NSt	64, 66, 82	St	NSt	NSt	NSt	0.713, 0.723, 0.664	St	NSt	NSt	NSt

Notes: P = Participant Identifier, SEM = Standard Error of Measurement; CIUs = Correct Information Unites, T = Time, St = stable, NSt = nonstable

Table 12: Low-CIU group scores and stability of performance decisions for nouns, verbs, and lexical diversity.

P	**Number of Nouns** SEM: 4.10					**Number of Verbs** SEM: 2.56					**Lexical Diversity** SEM: 0.073				
	Values T1, T2, T3	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*
1	40, 31, 37	NSt	NSt	St	NSt	6, 9, 10	NSt	St	NSt	NSt	0.554, 0.774, 0.811	NSt	St	NSt	NSt
2	16, 14, 21	St	NSt	NSt	NSt	8, 5, 13	NSt	NSt	NSt	NSt	0.786, 0.645, 0.778	NSt	NSt	St	NSt
4	23, 23, 27	St	St	St	St	4, 6, 9	St	NSt	NSt	NSt	0.840, 0.830, 0.753	St	NSt	NSt	NSt
5	26, 27, 40	St	NSt	NSt	NSt	17, 8, 6	St	NSt	NSt	NSt	0.688, 0.769, 0.554	NSt	NSt	NSt	NSt
9	32, 27, 34	NSt	NSt	St	NSt	8, 4, 5	NSt	St	NSt	NSt	0.931, 0.664, 0.633	NSt	St	NSt	NSt
10	21, 20, 27	St	NSt	NSt	NSt	12, 10, 8	St	Nst	Nst	Nst	0.872, 0.810, 0.791	St	NSt	NSt	NSt
15	18, 22,18	St	St	St	St	6, 7, 6	St	St	St	St	0.796, 0.763, 0.810	St	St	St	St
16	25, 25, 14	St	NSt	NSt	Nst	11, 10, 16	St	NSt	NSt	NSt	0.893, 0.924, 0.830	St	NSt	St	NSt
17	38, 51, 60	NSt	NSt	NSt	NSt	8, 13, 14	NSt	St	NSt	NSt	0.824, 0.805, 0.648	St	NSt	NSt	NSt
18	12, 12, 19	NSt	NSt	NSt	NSt	5, 11, 8	NSt	NSt	NSt	NSt	0.721, 0.798, 0.767	NSt	St	St	NSt

Notes: P = Participant Identifier, SEM = Standard Error of Measurement; CIUs = Correct Information Unites, T = Time, St = stable, NSt = nonstable

Table 13: High-CIU group scores and stability of performance decisions for open class words and closed class words.

P	**Open Class Words** SEM: 17.74					**Closed Class words** SEM: 17.05				
	Values T1, T2, T3	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*
3	86, 73, 89	St	St	St	St	45, 60, 69	St	St	NSt	NSt
6	76, 112, 100	NSt	St	NSt	NSt	57, 49, 46	St	St	St	St
7	74, 54, 75	NSt	NSt	St	St	44, 43, 78	St	NSt	NSt	NSt
8	254, 335, 264	NSt	NSt	St	NSt	289, 380, 282	NSt	NSt	St	NSt
11	452, 287, 283	NSt	St	NSt	NSt	322, 246, 270	NSt	NSt	NSt	NSt
12	120, 114, 128	St	St	St	St	80, 100, 128	NSt	NSt	NSt	NSt
13	151, 159, 132	St	NSt	NSt	NSt	188, 202, 171	St	NSt	St	NSt
14	130, 157, 168	NSt	St	NSt	NSt	145, 146, 158	St	St	St	St

Notes: P = Participant Identifier, SEM = Standard Error of Measurement; CIUs = Correct Information Unites, T = Time, St = stable, NSt = nonstable

Table 14: Low-CIU group scores and stability of performance decisions for open class words and closed class words.

P	**Open Class Words** SEM: 4.56					**Closed Class words** SEM: 4.67				
	Values T1, T2, T3	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*	*Values T1, T2, T3*	*T1-T2*	*T2-T3*	*T1-T3*	*Overall Stability*
1	52, 43, 53	NSt	NSt	St	NSt	32, 8, 16	NSt	NSt	NSt	NSt
2	21, 17, 28	St	NSt	NSt	NSt	11, 10, 23	St	NSt	NSt	NSt
4	26, 27, 35	St	NSt	NSt	NSt	15, 25, 26	NSt	St	NSt	NSt
5	43, 41, 52	St	NSt	NSt	NSt	33, 21, 32	NSt	NSt	St	NSt
9	42, 38, 52	St	NSt	NSt	NSt	7, 4, 6	St	St	St	St
10	33, 32, 45	St	NSt	NSt	NSt	10, 16, 15	NSt	St	NSt	NSt
15	24, 29, 30	NSt	St	NSt	NSt	2, 0, 5	St	NSt	St	NSt
16	35, 41, 33	NSt	NSt	St	NSt	5, 1, 2	St	St	St	St
17	48, 68, 78	NSt	NSt	NSt	NSt	9, 11, 7	St	St	St	St
18	22, 29, 27	NSt	St	NSt	NSt	17, 21, 13	St	NSt	St	NSt

Notes: P = Participant Identifier, SEM = Standard Error of Measurement; CIUs = Correct Information Unites, T = Time, St = stable, NSt = nonstable

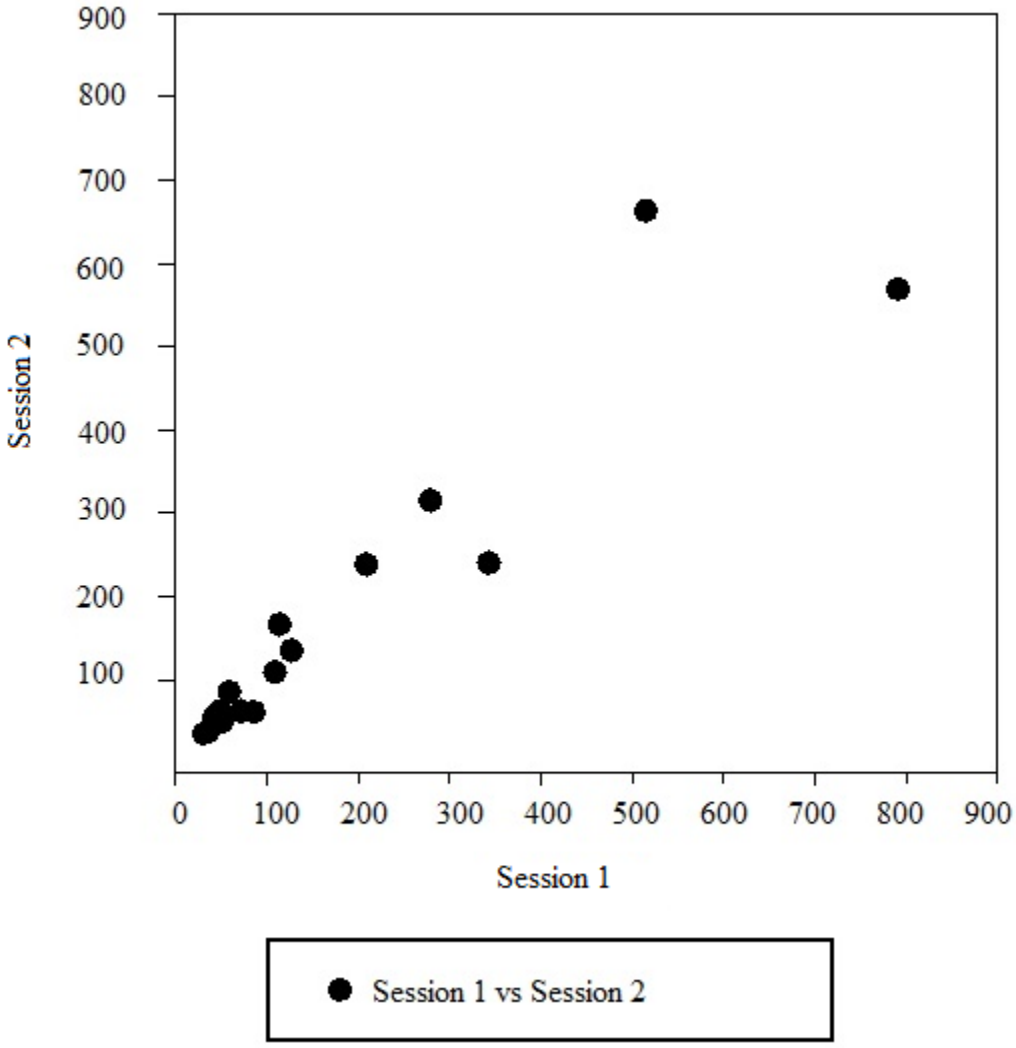

Figure 1: Scatterplot of CIU scores from Session 1 and Session 2.

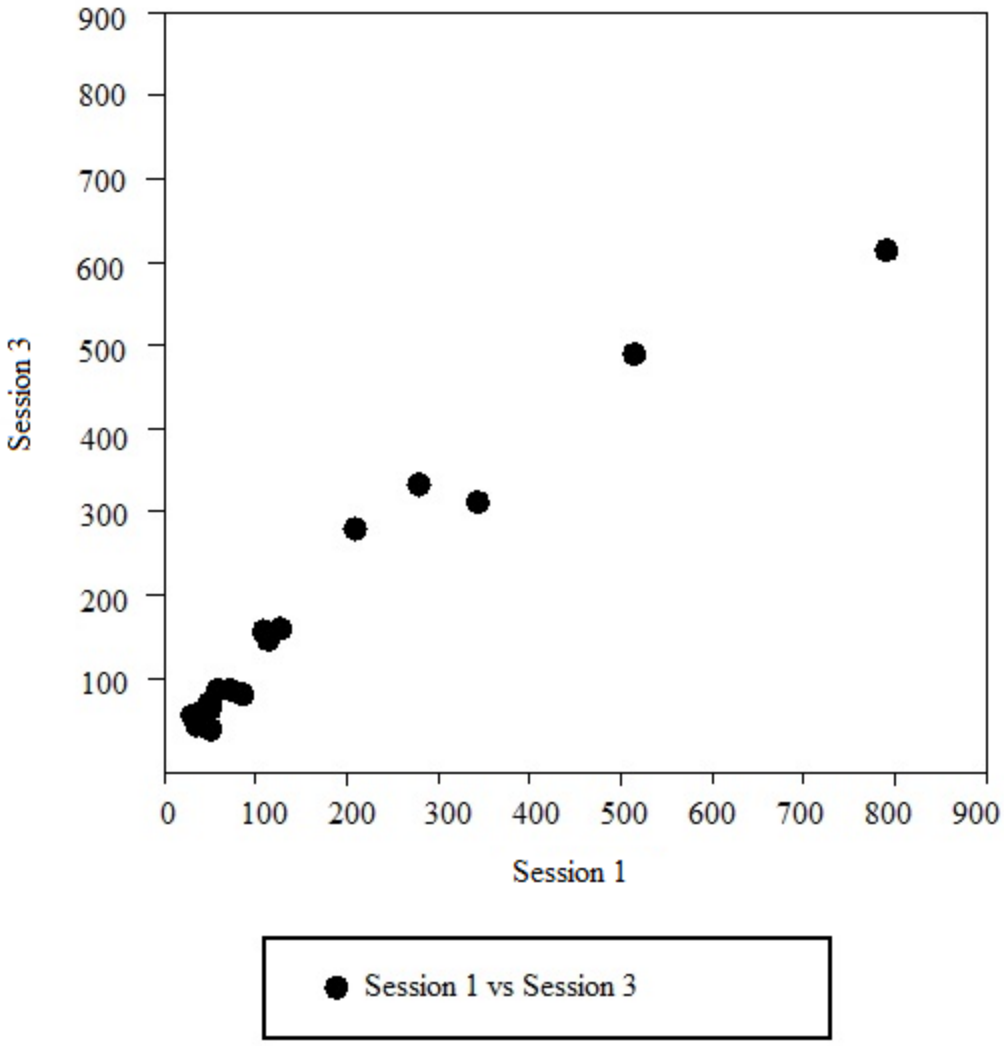

Figure 2: Scatterplot of CIU scores from Session 1 and Session 3.

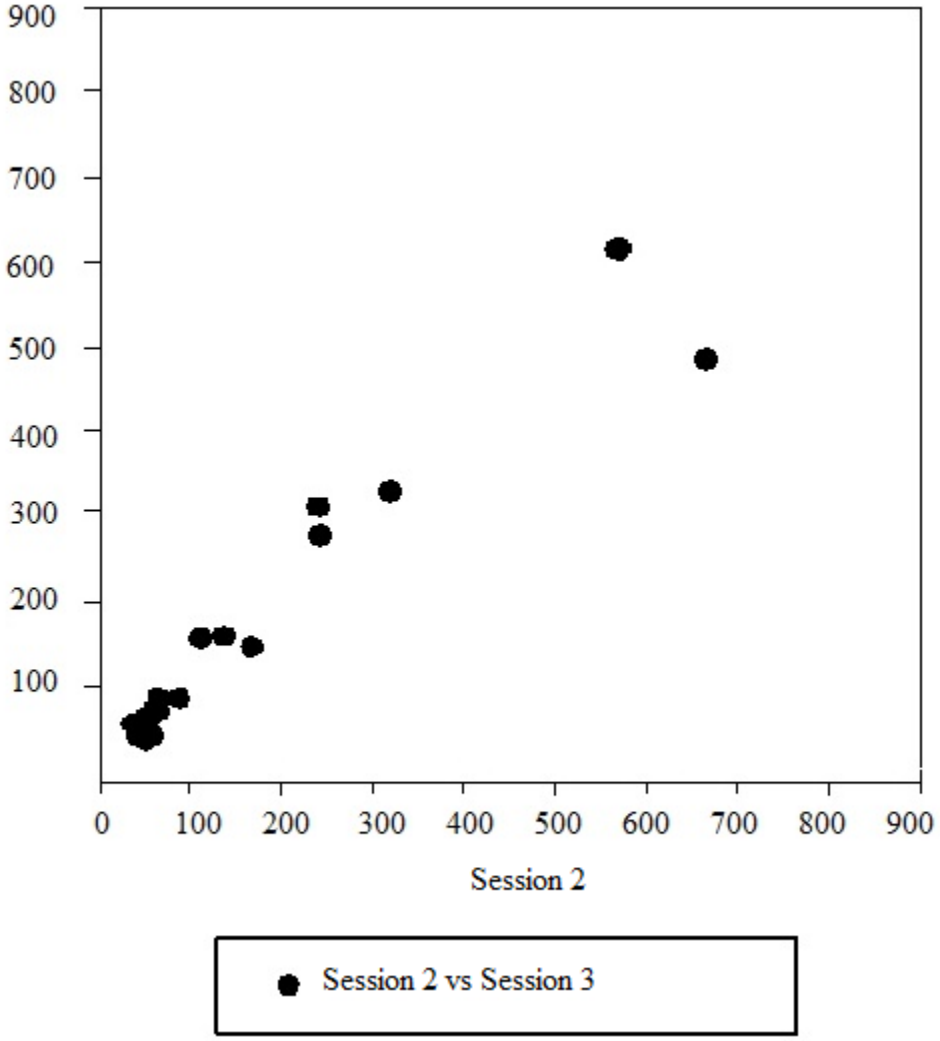

Figure 3: Scatterplot of CIU scores from Session 2 and Session 3.

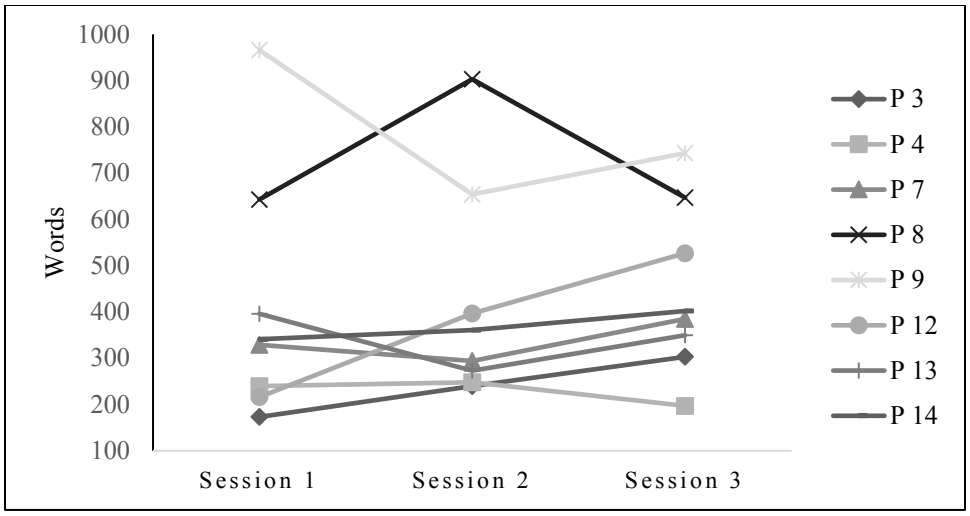

Figure 4: Number of words per session for the high-CIU Group.

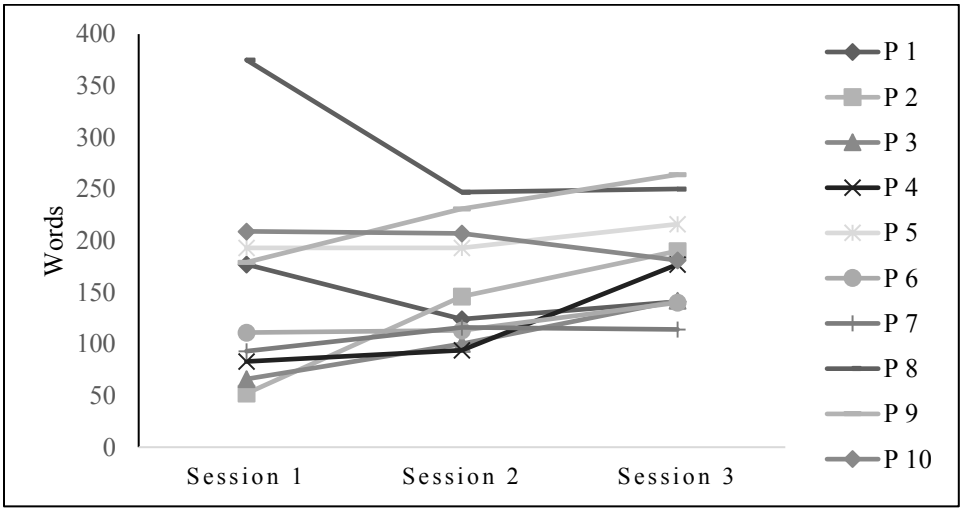

Figure 5: Number of words per session for the low-CIU group.

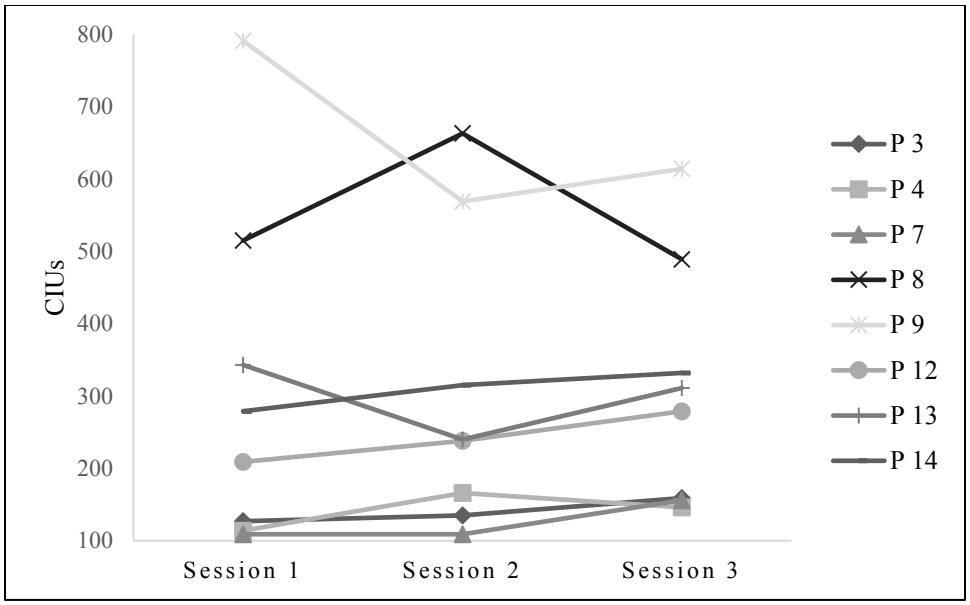

Figure 6: Number of CIUs per session for the high-CIU group.

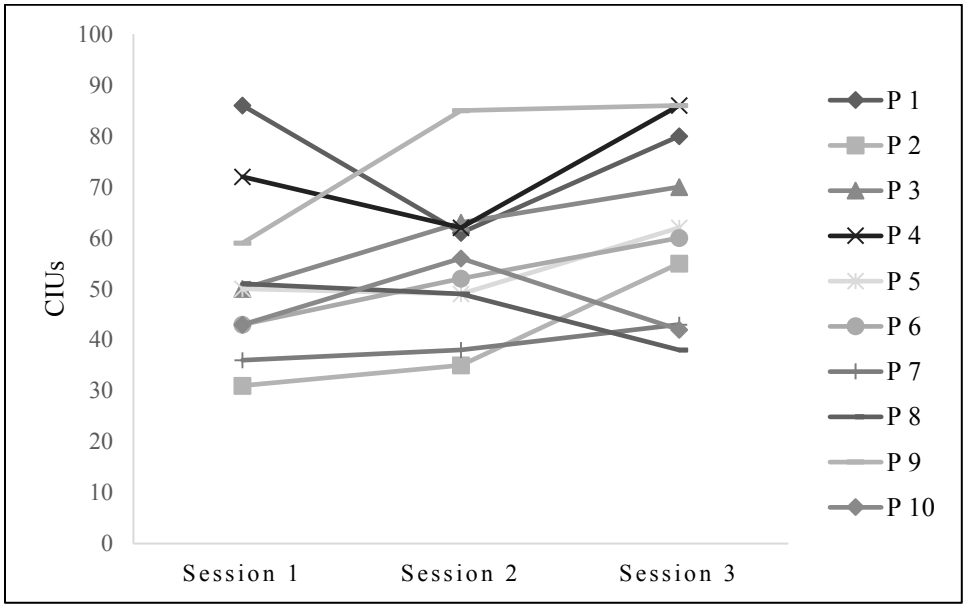

Figure 7: Number of CIUs per session for the low-CIU group.

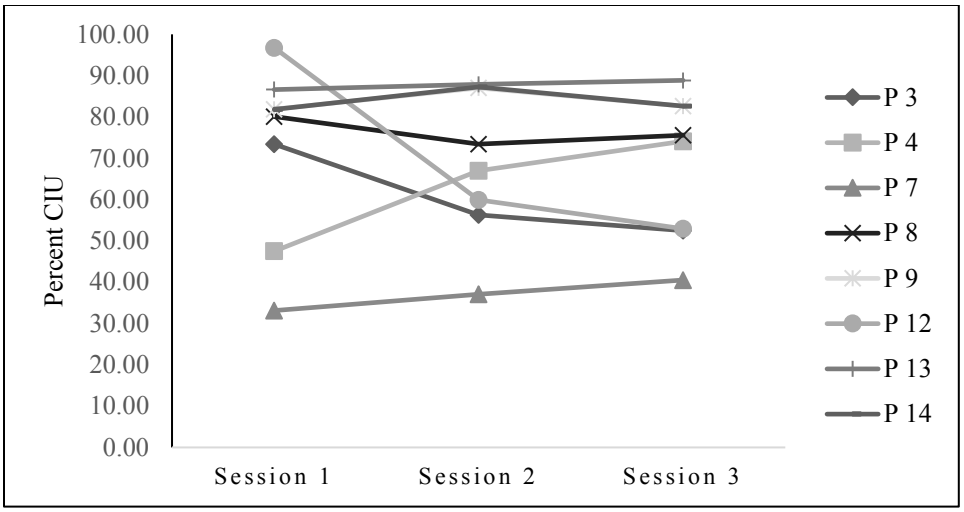

Figure 8: Percent CIUs per session for the high-CIU group.

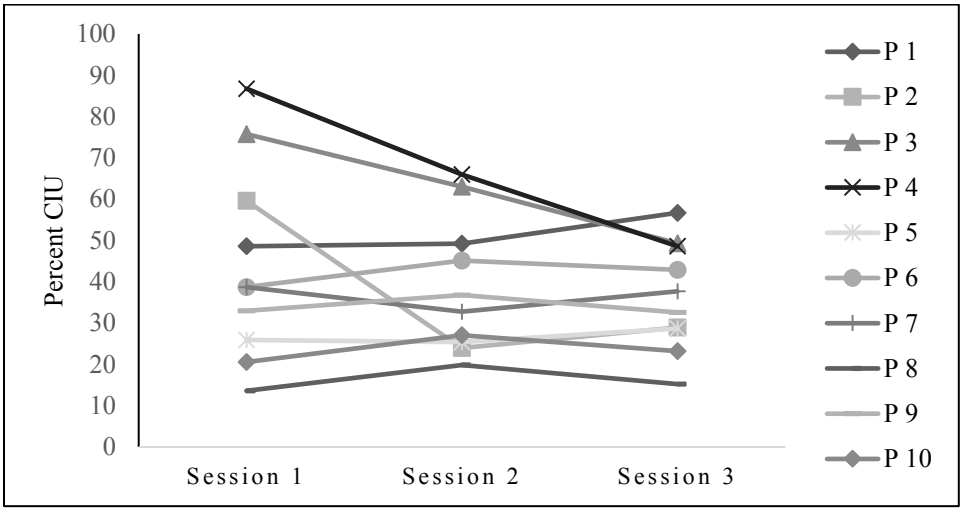

Figure 9: Percent CIUs for the low-CIU group.

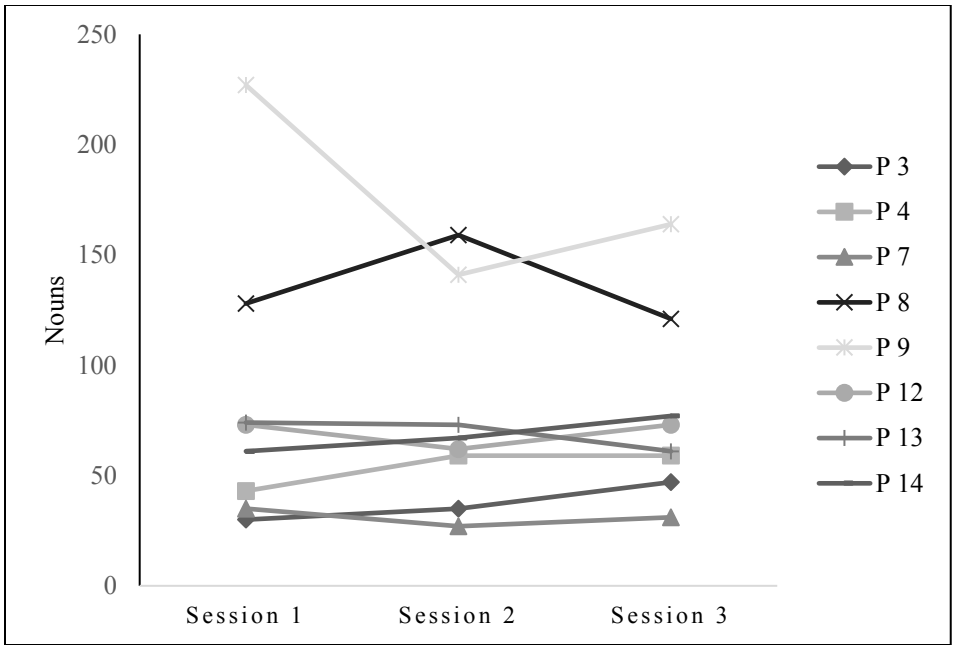

Figure 10: Number of nouns per session for the high-CIU group.

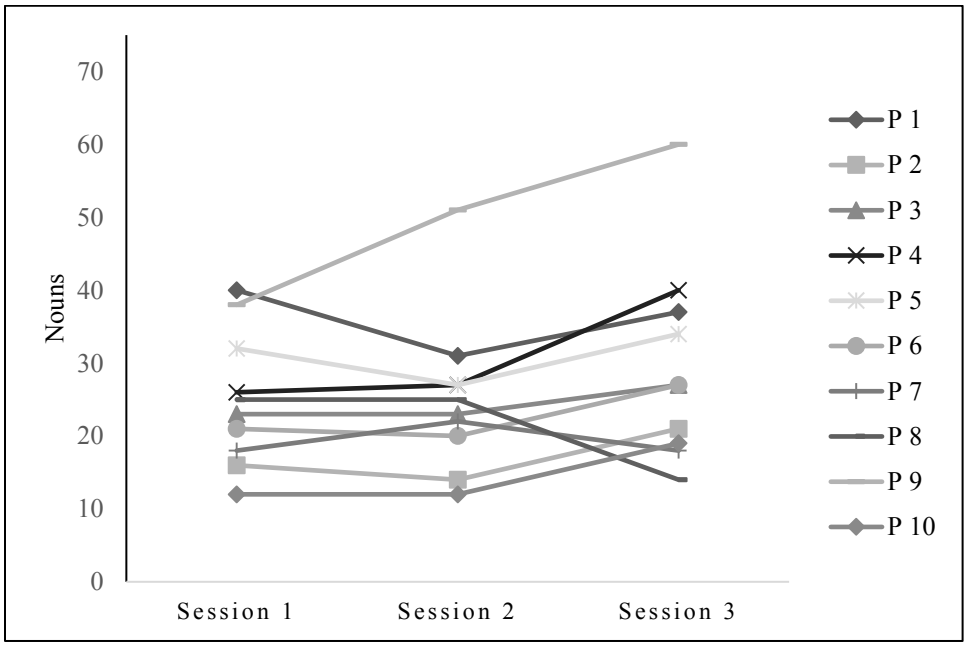

Figure 11: Number of nouns per session for the low-CIU group.

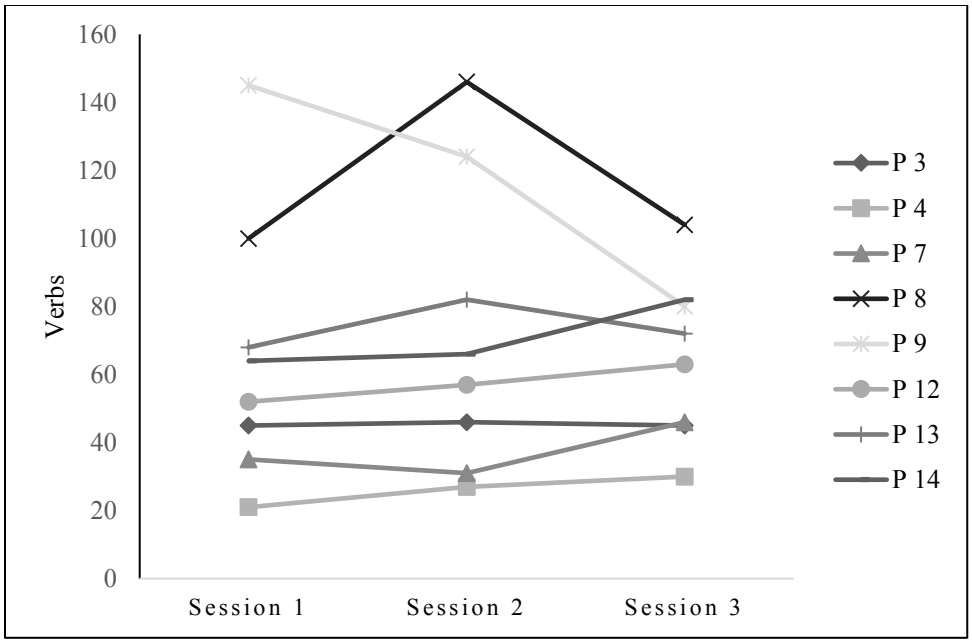

Figure 12: Number of verbs per session for the high-CIU group.

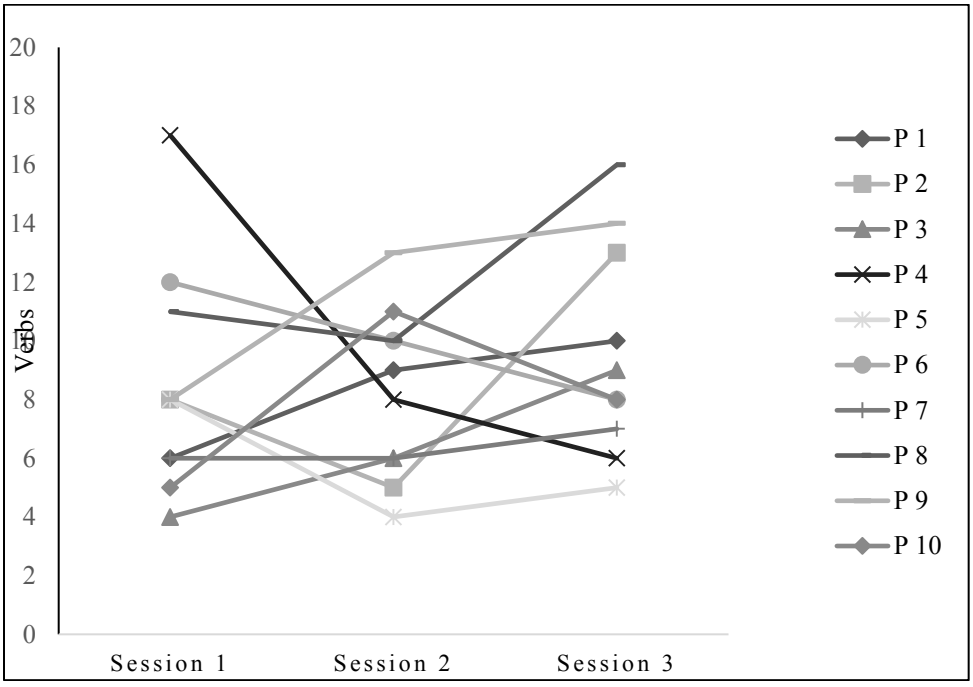

Figure 13: Number of verbs per session for the low-CIU group.

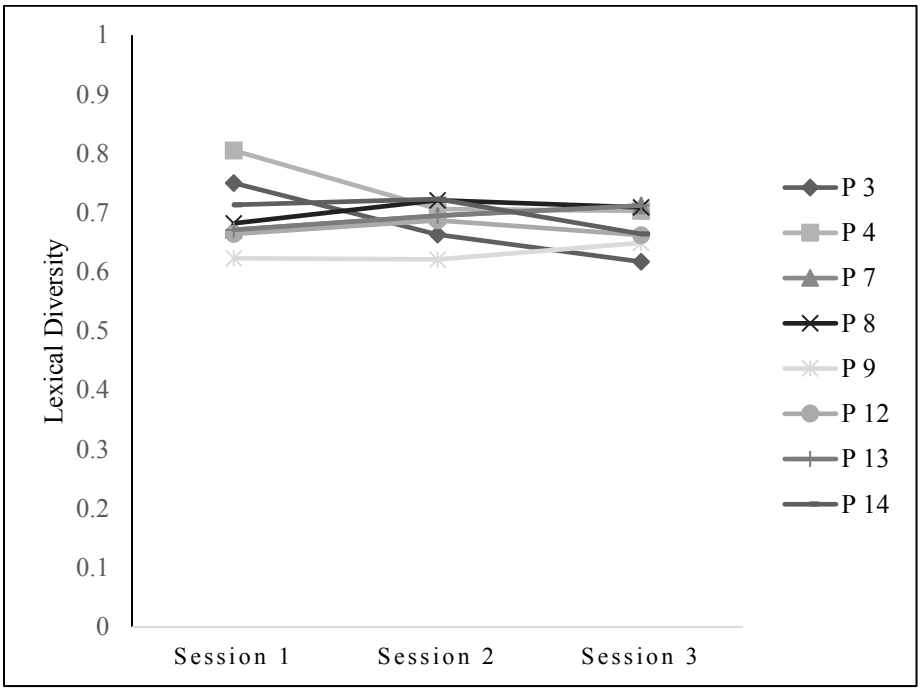

Figure 14: Lexical diversity ratio per session for the high-CIU group.

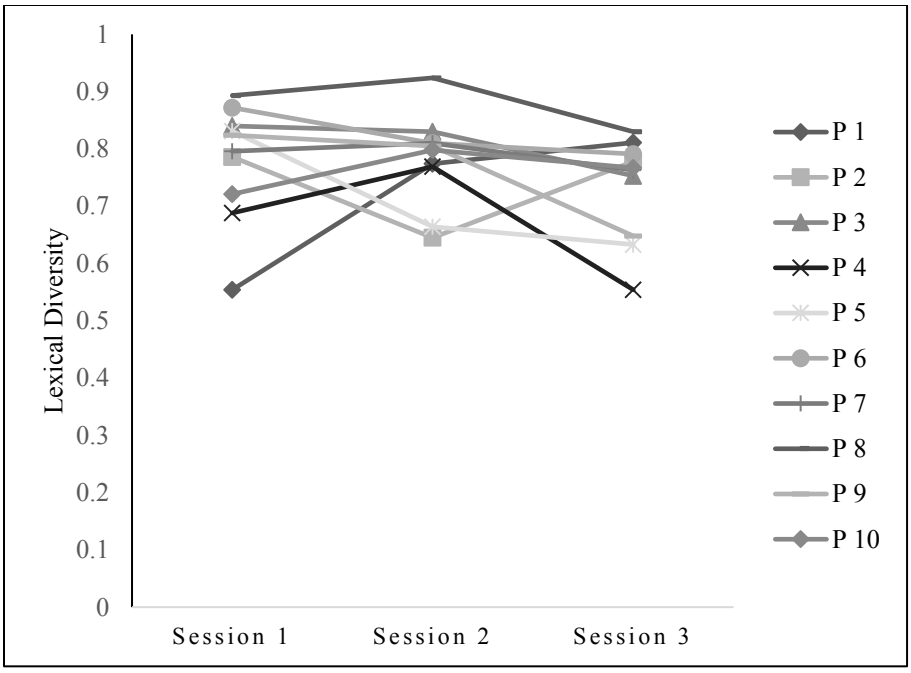

Figure 15: Lexical diversity ratio per session for the low-CIU group.

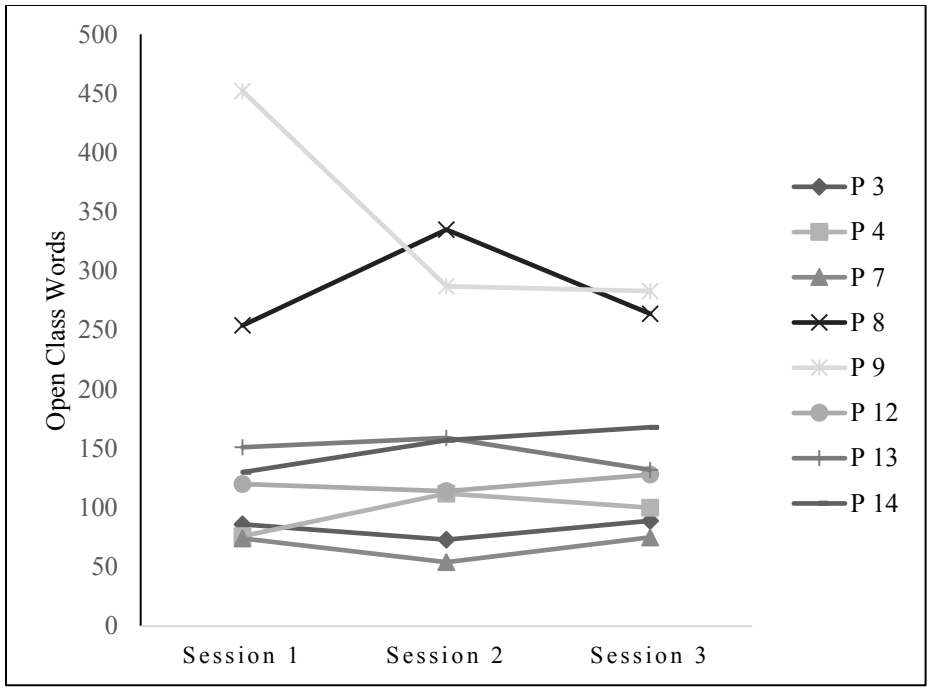

Figure 16: Number of open class words per session for the high-CIU group.

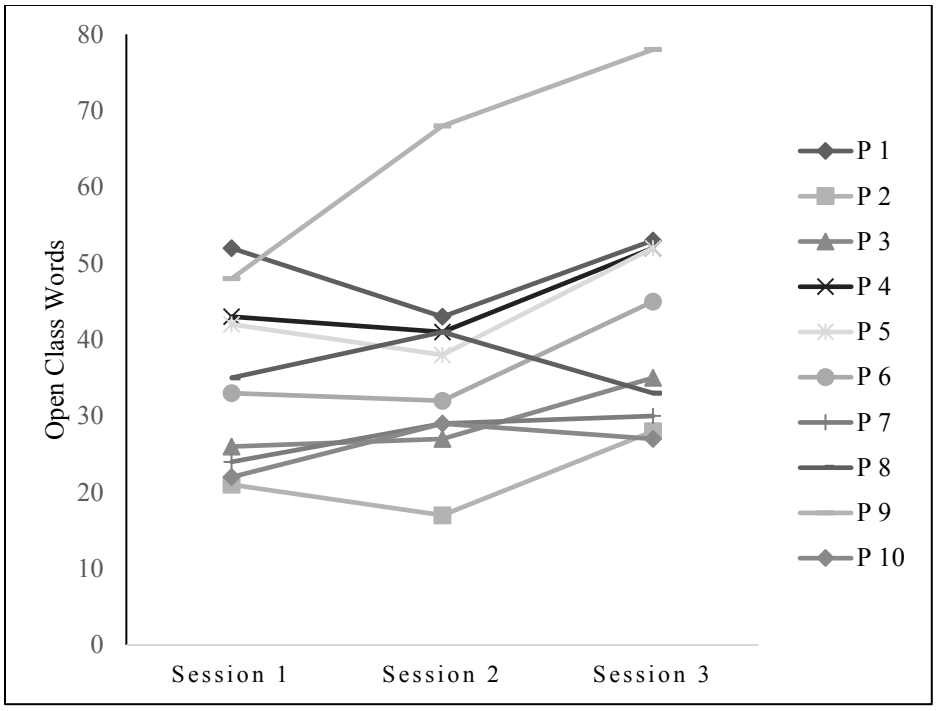

Figure 17: Number of open class words per session for the low-CIU group.

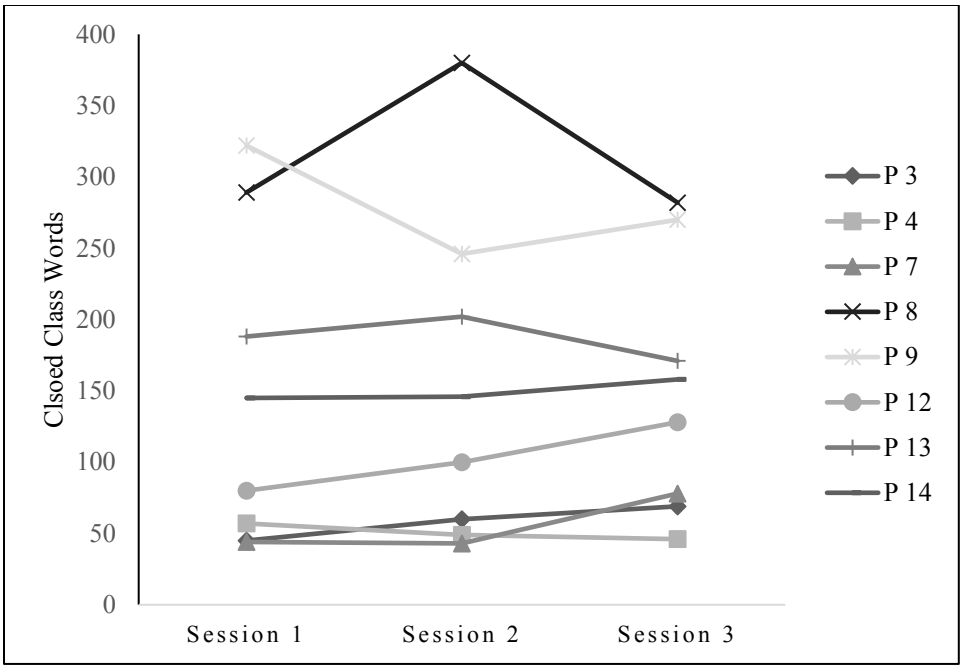

Figure 18: Number of closed class words per session for the high-CIU group.

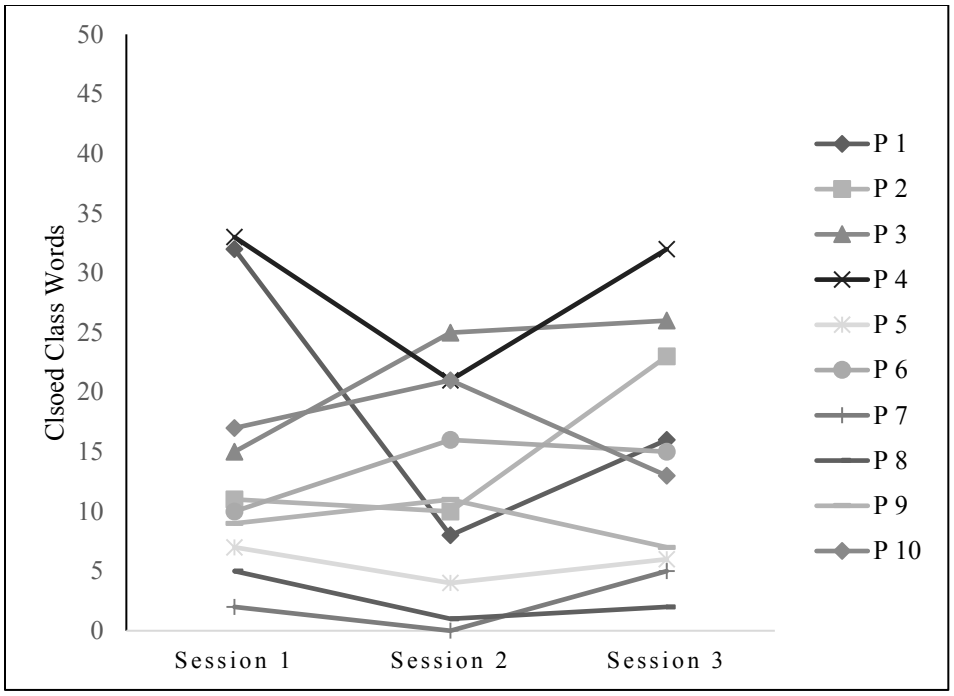

Figure 19: Number of closed class words for the low-CIU group.

DISCUSSION

The purpose of this study was to examine the test-retest stability of various aspects of language production in PWA using the Nicholas and Brookshire (1993) language elicitation system. The results of the group analyses suggest that all measures except lexical diversity may be sufficiently stable for group research (Fitzpatrick et. al, 1998). In contrast, the results of individual analyses of the 18 participants revealed lack of stability in performance across the three sessions for the majority of the measures for the majority of the participants.

Group Stability of Performance

Like Boyle (2014), group analyses for number of CIUs revealed strong correlations across sessions, suggesting that this measure derived from this elicitation task may be appropriate for the purposes of group research. The correlations found in the current investigation for the group of 18 PWA (*r*-values = 0.93 to 0.98) were greater than those found by Boyle (2014; *r*-values = 0.85 to 0.88).

The difference in CIU correlation values between the current study and that of Boyle (2014) could be related to CIU productivity. The CIUs per session averages for Boyle (2014) were 39, 40, and 40 CIUs, for session 1, session 2, and session 3, respectively. In comparison, the average CIUs per session for this investigation was 167, 166, and 172 CIUs per session, with a greater range of performance across participants.

When examining the correlation coefficients for the low-CIU group only, CIU scores per session similar to Boyle's (2014). The session-to-session correlation values for these ten low-CIU participants ranged from $r = 0.60 - 0.72$ while the high-CIU group of eight participants had correlation values ranging from $r = 0.89 - 0.98$. These results may suggest that CIU production at one sampling time may be somewhat more predictive of a subsequent sampling time when CIU production is greater. However, the higher correlations associated with the high-CIU group may also reflect a greater range of performance across participants. Aphasia severity and word-retrieval severity is likely associated with CIU production and stability but was not addressed in the current study. Cameron et al. (2010) found that participants with *less* severe aphasia had a greater dispersion (range) of scores over 5 sessions, which could appear to be contrary to the findings from this investigation. However, Cameron et al. did not examine stability of performance over sessions relative to severity; dispersion of scores does not directly translate to stability coefficients.

The results of the group analyses for number of words also revealed strong correlations between sessions (r –values = 0.85 - 0.92). Like the number of CIU correlations, the correlation values found in this investigation were higher than those reported by Boyle (2014; r-values = 0.74 - 0.84).

The results of the group analyses revealed strong correlations across sessions for percent CIUs (r-values = 0.73 to 0.95). This is lower than reported by Brookshire and Nicholas (1994; $r = 0.94$) but similar to data reported by Boyle (2014; $r = 0.61 - 0.95$). For percent CIUs, the correlation between session 2 and session 3 found by Boyle (2014; $r = 0.95$) and this investigation ($r = 0.95$) indicate this session 2 to session 3 change in

percent CIUs is the strongest across sessions for all participants. Brookshire and Nicholas (1994) only reported correlations between session 1 and session 2, so it is not possible to compare correlations yielded between session 2 and session 3. This higher correlation between session 2 and session 3 could be due to participant familiarity with the language task expectations on the second and third administration that could have potentially created an increased efficiency of language production.

Correlations for numbers of nouns, verbs, open class words, and closed class words were strong for the session to session comparisons. This was expected, given that strong associations were found for total CIU across sessions and the preceding counts were derived from the CIU counts. However, a similarly strong association could not be assumed for specific grammatical form classes. These findings extend the existing research on the test-retest stability of discourse using the Nicholas and Brookshire (1993) discourse elicitation procedures. Researchers expect strong temporal reliability when gauging group treatment outcomes of production of number of nouns or verbs with similar participants with aphasia.

The correlation coefficient for lexical diversity suggests poor stability to moderate stability across sessions (r values = 0.22 to 0.51). Boyle (2014) reported lexical diversity reliability coefficients that exceeded at 0.7 for all session comparisons. One likely cause of this difference across studies is the different methods used to measure lexical diversity. The current investigation used a moving average of type-token ratio whereas Boyle (2014) used a type-token ratio (vocabulary size divided by length). Covington and McFall (2010) recommend the moving average of type-token ratio as being more appropriate for measuring lexical diversity; a moving average of type-token

ratio accounts for changes in text length, whereas type-token ratio does not. Additionally, the participants had a narrow range of lexical diversity ratios for each testing time and across testing times (range = 0.6 to 0.9). Since *r*-correlation measures the linear relationship of two variables, the participants may not have had sufficient variability for the *r*-correlation to measure the data points appropriately.

Individual Stability of Performance

Upon examination of individual stability in performance, no participant achieved stability across all measures for all session comparisons, although Participant 15 achieved stability across all sessions for all measures except number of open class words and number of closed class words. The measure for which the greatest number of participants demonstrated stable performance was percent CIUs; 8 of 18 participants achieved stability for all three testing times. The measures for which the lowest number of participants demonstrated stable performance was number of CIUs, only Participant 15 achieved stability for all three testing times.

The findings concerning instability in percent CIU production for individual participants is consistent with reports by Boyle (2014) and Cameron et al. (2010) who also reported variation that exceeded Brookshire and Nicholas' (1994) original report. Specifically, Brookshire and Nicholas reported that none of their 20 participants exceeded a 10% CIU difference across sessions. In the current investigation, 8 of the 18 participants demonstrated at least a 10% difference across the sessions, with performance being classified as "unstable" for 10 of the 18 when considering the SEM.

Possible Sources of Session-to-Session Variability in Discourse

Premeasurement training of participants may have influenced outcomes relative to performance stability and that may account for Brookshire and Nicholas' (1994) differing findings. Specifically, Brookshire and Nicholas (1994) gave each participant practice and training trials on 2 procedural discourse stimuli to ensure that the participant understood the task. The feedback consisted of the examiner telling the participant whether or not the response was satisfactory in length and content. If the examiner considered the response unsatisfactory, the participant was given suggestions about other ideas to include in their response. These practice trials were not included in the methods of Cameron et. al (2010), Boyle (2014), or in the current investigation; premeasurement practice is not efficient for clinical use (Nicholas & Brookshire, 1993). Based on the current findings concerning individual performance, such training may be advisable despite the time-cost.

Another factor that may have influenced outcomes relative to performance stability is differing time between sessions across investigations and number of samples obtained. In the present investigation, language samples were obtained at consistent time intervals (initial sample, 1 week after initial sample, and 4 weeks after initial sample). Boyle (2014) elicited language samples between 2 and 7 days apart. Both Brookshire and Nicholas (1994) and Cameron et al. (2010) elicited language samples between two and ten days apart.

In the present investigation, SEM was used to judge stability because it is objective and statistically appropriate. Reliability coefficients have been used by other investigations to gauge individual stability (e.g., Boyle, 2014); however, SEM may be a more appropriate measure because it allows for statements about individual stability

(Harvill, 1991). This investigation is the first to use SEM to judge stability of discourse in individuals PWA. Cameron et. al (2010) and Brookshire and Nicholas (1994) used descriptions of percent change to judge test-retest stability, which may be clinically relevant. However, this method is more subjective than using a measure such as SEM.

SEM has its own limitations as well in the context of this investigation. Although the participants were split up into high CIU and low CIU groups, SEM may disguise or enhance score differences in participants who scored on the extremes of the two groups (Harvill, 1991). Given the lack of stability of performance for individual participants using the SEM metric, determination of individual performance variability through repeated pretreatment measurement may be necessary to ascertain changes that may be associated with intervention.

Another source of individual variability could stem from severity of aphasia, severity of word-retrieval deficits, time postonset of aphasia, and aphasia type. It was beyond the scope of this investigation to examine these factors in relation to group or individual scores; however, these participant characteristics have been examined previously in relation to test-retest stability. Boyle (2014) found a moderate positive correlation with aphasia severity and number of words. Since aphasia severity impacts quantity and quality of discourse in PWA, it should be considered as a possible influencing factor in this investigation. Boyle (2014) also found that months postonset also had a moderate correlation with number of words, CIUs, and percent CIUs; age was poorly correlated with number of words, number of CIUs, and percent CIUs.

It has been suggested that language variability is a hallmark characteristic of language in PWA (Kolk, 2007). If variability is a defining feature of aphasia, it may

difficult to reliably measure true performance or change in performance within individuals with a relatively unconstrained method such as the procedure developed by Nicholas and Brookshire (1993). Repeated measurement to establish the extent of variability within an individual may be necessary. However, group performance appears to be sufficiently stable across sampling times to allow measurement of change at the group level.

APPENDIX A

SAMPLE TRANSCRIPT

A [CATINTREEPICTURE] the [det] girl [n].
A the [det] kitty [n].
A tricycle [n].
A the [det] dog [n].
A {the} the [det] man [n] was [ccv] [S].
A {the police man it's the police man it's fire house}
A the [det] two [adj] fire [adj] {man}.
A the [det] bird [n] [29wd] [15ciu].

A [DOGATECAKE] the [det] dog [n].
A the [det] lady [n] was [ccv] [S].
A{the} [pro] it 's [ccv] {happy} birthday [n] [S] {thing it's}.
A the [det] little [adj] boy [n] was [ccv] mad [adj] [S] {the let's see little boy}.
A {and the the wife no it's something it's}.
A the [det] girl [n] {and the girl happy birthday thing} [42wd] [15ciu].

A [SUNDAY] Sunday [n] {it's} football [n] {yeah} downstairs [n] {the} [7wd] [3ciu].

A [DISHES] {it's let's see mom'll do this one I'll do the drier but it's not a drier} [21wd] [0ciu].

A [FIGHTSEQUENCE] {the} they [pro] 're [ccv] married [adv] [S].
A the [det] guy [n] 's [ccv] {like} [S].
A the [det] [n] lady 's [ccv] [S].
A {the the lady was it's}.
A the [det] man [n] was [ccv] {sad} very [adj] sad [adj] [S].
A {the it's} the [det] crash [n] {it's} [30wd] [13ciu].

= total words: 129
= total CIU: 46

APPENDIX B

CONVENTIONS FOR CODING LANGUAGE TRANSCRIPTS

Correct information units (CIUs). The number of CIUs are bracketed at the end of each stimulus.

Word count. The number of narrative words are bracketed at the end of each stimulus.

Number of verbs. All CIU words that are main verbs are marked with [v]. Verb phrases are counted as one verb.

Number of nouns. All CIU words that are nouns are marked with [n].

Well-formed sentences. Sentence scoring was adapted from the conventions of Saffran et al. (1989). Well-formed sentences are marked with [S] at the end of each sentence. Only CIU words are considered for analyzing sentence structure. Utterances scored as well-formed sentences must conform to one of the following structural types:

Noun + Main Verb. These structures need not be semantically coherent; violations of strict subcategorization and selection restrictions should be ignored. For example, the following are noun + main verb sentences: “Cinderella rode the house,” and “Cinderella put”

Noun + Copula + Adjective. The following are examples of this sentence sub-

type: “Cinderella is beautiful,” but not “Cinderella beautiful” *Noun + Copula + Prepositional Phrase*. The prepositional phrase must contain minimally Preposition + Noun Phrase, in addition to the Noun + Verb requirement described above, in order for the utterance to qualify as a sentence. The following is an example of this sub-type: “Cinderella is at the prince,” but not “Cinderella is with.”

Embedded Subjects do not count as separate sentences. Syntactically well-formed but semantically anomalous sentences are scored as well-formed. Omission of obligatory arguments renders a sentence well-formed, but violation of selectional restrictions does not.

Open class words. Open class words are calculated for all CIU words in SALT using the Explore function. The following types of words are counted as open class words: nouns, verbs, adjectives, adverbs with the following exceptions and inclusions:

Numbers. Words that are numbers are considered open class words with the exception of the use of *one* as a pronoun (e.g., “He was the one”).

Adjectives. Words that are adjectives are marked with [adj] and adverbs are marked with [adv].

Closed class words. Closed class words are calculated for all CIU words in SALT using the Explore function. The following types of CIU words are counted as closed class words: determiners, conjunctions, pronouns, auxiliary verbs, and prepositions.

Determiners. Words that are determiners are marked [det]; conjunctions are marked [conj]; pronouns are marked [pro]; auxiliary verbs are marked as [aux]; prepositions are marked as [prep].

Non-CIU words. Non-CIU words are enclosed in curly brackets and not considered in any language analyses.

REFERENCES

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

Armstrong, E., & Ferguson, A. (2010). Language, meaning, context, and functional communication. *Aphasiology, 24*(4), 480-496.

Armstrong, E., Ciccone, N., Godecke, E., & Kok, B. (2011). Monologues and dialogues in aphasia: Some initial comparisons. *Aphasiology*, *25*(11), 1347-1371.

Berko-Gleason, J., Goodglass, H., Obler, L., Green, E., Hyde, M., & Weintraub, S. (1980). Narrative strategies of aphasics and normal-speaking subjects. *Journal of Speech and Hearing Research*, *23*, 370-382.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*, *13*(3), 236-249.

Boyle, M. (2014). Test-retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research*, *57*(3), 966-978.

Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research*, *37*(2), 399-407.

Cameron, R. M., Wambaugh, J. L., & Mauszycki, S. C. (2010). Individual variability on discourse measures over repeated sampling times in persons with aphasia. *Aphasiology, 24*(6-8), 671-684.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology*, *26*(10), 1250-1277.

Covington, M. A. (2007). MATTR user manual (CASPR research report 2007–05). Atheus, GA: Institute for Artificial Intelligence, The University of Georgia.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, *17*, 94–100.

Covington, A., & McFall, J. MATTR. Computer software. Http://ai1.ai.uga.edu/caspr/ Vers. 2.0.3018.28419. *University of Georgia Research Foundation*. Web.

Disimoni, F. G., Keith, R. L., & Darley, F. L. (1980). Prediction of PICA overall score by short versions of the test. *Journal of Speech and Hearing Research*, *23*(3), 511-516.

Doyle, P. J., McNeil, M. R., Spencer, K. A., Jackson Goda, A., Cottrell, K., & Lutig, A. P. (1998). The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. *Aphasiology*, *12*(7/8), 561-574.

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25*(11), 1414-1430.

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech-Language Pathology*, *22*(2), S397-S408.

Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*, *2*(14), i-74.

Goodglass, H., & Kaplan, E. (1983). *The Boston Diagnostic Aphasia Examination*. Boston, MA: Lea & Febiger.

Hinkley, J. J. (2007). *Narrative-based practice in speech-language pathology.* San Diego, CA: Plural Publishing.

Kaplan, E., Goodglass, H. & Weintraub, S. (1983). *The Boston Naming Test*. Boston, MA: Lea & Febiger.

Kertesz, A. (2007) *Western Aphasia Battery – Revised*. San Antonio, TX: Psychological Corporation.

Miller, J., & Iglesias, A. (2012). *Systematic Analysis of Language Transcripts (SALT), Clinical Version 2012* [Computer Software]. Middleton, WI: SALT Software, LLC.

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 and Hearing Research*, *36*, 338-350.

Olness, G. S. (2006). Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology*, *20*(2-4), 175-187.

Papathanasiou, I., P. Coppens, & Potagas, C., (2013) Aphasia and related neurogenic communication disorders. Burlington, VT: Jones and Bartlett Learning.

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

Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18*(12), 1075-1091.

Raymer, A.M., & Gonzalez Rothi, L.J. G. (2008). Impairments of word comprehension and production. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic disorders* (5th ed., pp. 607-631). Baltimore, MD: LippincottWilliams & Wilkins.

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.

Ulatowska, H.K., Allard, L., & Bond Chapman, S. (1990). Narrative and procedural discourse in aphasia. In Y. Joanette & H.H. Brownell (Eds.), *Discourse ability and brain damage: Theoretical and empirical perspectives* (pp. 180-198). New York: Springer-Verlag.

Ulatowska, H. K., Reyes, B. A., Santos, T. O., & Worle, C. (2011). Stroke narratives in aphasia: The role of reported speech. *Aphasiology, 25*(1), 93-105.

Wambaugh, J. L., Wright, S., Nessler, C., & Mauszycki, S. C. (2014). Combined aphasia and apraxia of speech treatment (CAAST): Effects of a novel therapy. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2191-2207.

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

Yorkston, K.M., & Beukelman, D.R. (1980). An analysis of connected speech: Samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, *45*, 27-36.