The test-retest reliability of the Western Aphasia Battery-Revised

By

Brandall Bond
B.A., Truman State University, 2017

Submitted to the graduate degree program in Speech-Language-Hearing: Sciences and Disorders and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Susan T. Jackson

Lindsey Heidrick

Julie Gatts

Ana Paula Mumy

Date Defended: 10 May 2019

The thesis committee for Brandall Bond certifies that this is the approved version of the following thesis:

The test-retest reliability of the Western Aphasia Battery-Revised

_____

Co-Chair: Susan T. Jackson

_____

Co-Chair: Ana Paula Mumy

Date Approved: 10 May 2019

Abstract

The purpose of the study was to examine the test-retest reliability of the Western Aphasia Battery-Revised (WAB-R). Ten people with stroke-induced aphasia were administered the entire WAB-R twice. Correlation coefficients were > .80 for 11/13 WAB-R sections/subtests. Paired t-tests revealed no significant difference between the means at Time 1 and Time 2 for any of the 13 WAB-R sections/subtests. Effect size was less than small for 11/13 WAB-R sections/subtests and small for two subtests. For 8/13 sections/subtests, one WAB-R standard error of measurement (SEM) represented less than 5% of the total number of points possible; thus, 68% of the time, persons with aphasia would be expected to score within 5% of the total number of points possible on repeated testing. For 5/13 sections/subtests, one WAB-R SEM represented more than 5% but less than 10% of the total number of points possible.

Acknowledgments

I would like to thank the subjects who participated in the study, and their families. I like to also thank the research assistants, Abigail Enneking and Hannah Haworth. Lastly, I would like to thank the members of my thesis committee, Julie Gatts, Dr. Lindsey Heidrick, and Dr. Susan T. Jackson.

Table of Contents

List of Tables

Chapter 1: Introduction

Standardized aphasia tests serve an important role in both research and clinical practice by providing an empirical component to evaluation. The primary metrics of psychometric quality for standardized assessment tools, including those specific to aphasia evaluation, are reliability and validity. Validity refers to the property of an instrument to measure its target. Reliability refers to the instrument's ability to consistently reproduce a result. More specifically, the several components of reliability (e.g., inter-rater reliability, test-retest reliability, internal consistency, etc.) concern the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials.

"Test-retest reliability is a statistical technique used to estimate components of measurement error by repeating the measurement process on the same participants, under conditions as similar as possible, and comparing the observations" (Lavrakas, 2008, pg. 888). The data derived from a test-retest study offer information about changes (or a lack thereof) in outcomes that can be attributed to genuine change (or genuine stability) over time. A test that is dependable and consistently demonstrates good reproducibility allows researchers and clinicians to make more informed decisions. Good test-retest reliability allows examiners to deduce that the differences observed are a product of real change and not an artifact of the instrument's instability (Aldridge, Dovey, & Wade, 2017). Thus, the purpose of reliability studies is to determine the extent to which repeated measurements yield the same results, when the variables (e.g., time, rater, environment, etc.) remain the same (Aldridge et al.; Carmines & Zeller, 1979).

There are different methods of analyzing reliability and more specifically, score stability. Even so, researchers can misinterpret the results of certain statistical measures (Aldridge et al., 2017). This often takes place as a result of the following common misconception: high

correlation between measurements equates to agreement between those measures (Aldridge et al.). It is common to see test-retest reliability assessed using bivariate correlation and tests of mean difference (e.g., paired *t*-tests) (Aldridge et al.). Here, it is important to make the distinction between tests that measure relative standing, and those that measure true stability. Neither of the aforementioned analyses are able to appropriately quantify the equality/similarity of repeated scores (agreement-based reliability) (Aldridge et al.). The parametric correlation coefficients frequently presented in reliability studies (e.g., Pearson's product-moment correlations), use a one-to-one coefficient to quantify consistency in a variable- to measure how one variable increases/decreases in relation to the increasing/decreasing of another variable (Aldridge et al.). However, this form of analysis is biased against the magnitude of the variable being compared, meaning the calculation is done without consideration of the magnitude. For test-retest reliability, this means that differences in value that are a result of intra- and inter-rater inconsistency are not detected if the differences are consistent in a sample. In clinical practice this would mean that if two patients achieved individual scores of 5/10 and 1/10 on an aphasia subtest, and two weeks later (without any change in their conditions) those same two patients achieve scores of 8/10 and 4/10, respectively, then a parametric correlation coefficient would yield a high level of statistical significance and suggest reproducibility. However, the raw data in the example show that there is some inconsistency undetected by the analysis (be it intra-rater reliability or test-retest reliability). Tests of mean difference are also subject to bias and secondary misinterpretation. Examining only the group average on a performance test can generate bias against individual score change. For example, on the first testing date, Subject A scores 100/100 on the exam and Subject B scores 50/100 ($\overline{x} = 50$), then on the second testing date, Subject A scores 50/100 and subject B scores 100/100 ($\overline{x} = 50$). Because the average of the

two scores is the same, a test of mean difference would show high congruency. A researcher might misinterpret the lack of difference between the two averages as score stability. However, neither bivariate correlation nor tests of mean difference are measures of true score stability. Therefore, these tests do not yield the information most appropriate for determining test-retest reliability.

To obtain a representative measurement of score stability, researchers have employed the standard error of measurement (SEM) as a method of analysis coupled with a percentage change score (Beckerman et al., 2001; Flanagan & Jackson, 1997). The SEM generates an estimated range of future performance scores. A single true score cannot be produced, due to the random error present in all standardized aphasia tests. A reliable test is one that minimizes random error. Researchers are able to deduce test-retest reliability between test scores by figuring the variability in measurements of the same participant; a small SEM range would suggest that an unchanged subject will score similarly on all administrations of the same test. SEM is calculated using the following equation:

$$\text{SEM} = SD\sqrt{1-r}$$

where *SD* is the standard deviation of the scores from testing date 1 and *r* is the correlation between the testing date 1 and 2 scores using Spearman's rho as the measure of correlation. The SEM allows researchers to predict future performance scores based on an observed score. The chances that a predicted performance score will not differ from the obtained performance score by more than ±1 SEM is 68 in 100. For example, Subject A scores a 10/20 on an exam whose SEM is 3. If Subject A were to be administered the exam 100 times, we would expect that in 68 of those instances, the subject would score within a range of ±1 SEM (in this example: 7/20 to 13/20). The chances that a predicted performance score will not differ from an obtained

performance score by more than ±2 SEM is 95 in 100. Using the same subject, test, and number of trials from the above example, we would expect that in 95 of those instances, the subject would score within a range of ±2 SEM (in this example: 4/20 to 16/20). Typically, the smaller the SEM, the greater the test-retest reliability (Nicholas & Brookshire). A "good" SEM value depends on the magnitude of the scores that can be expected. "For example, a SEM of 10 would be acceptable if the scores ranged from 100 to 500, but a SEM of 10 would not be acceptable if the range of scores was 0 to 20" (Nicholas & Brookshire, pg. 405). However, SEM ranges alone do not provide a straightforward measure of test-retest reliability, due to the varying significance of the actual numerical value (Nicholas & Brookshire). For example, a difference of one point on a 60-point test can contribute to the distinction between two clinical diagnoses. Thus, without knowledge of the overall context /numerical value of the test, the range established using the SEM lacks purport. Historically, a percentage change score (PC) has been used in conjunction with the SEM to determine if the range generated by the SEM is significant or would make any clinical distinction (Flanagan & Jackson, 1997; Nicholas & Brookshire). PC is calculated using the following equation:

$$PC = (SEM/Number\ of\ Items) \times 100$$

"'where PC represents the percentage change from the [number of test items] accounted for by a change in score of 1 SEM between sessions' (Nicholas & Brookshire, p. 342), and Number of Items represents the number of items in a particular test" (Flanagan & Jackson, pg. 38). The PC makes clear the value of the SEM as it pertains to the total number of test items. For example, the SEM calculated for the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983), which has 60 items, was 1.02 and the PC calculated was 1.70 (Flanagan & Jackson). Therefore, we can expect that a subject who scores 50/60 on the BNT will score within a range

of 48.98 to 51.02 (±1 SEM), in 68% of all instances. The BNT's SEM indicates that a subject will score similarly on future administrations of the test (provided both the test and the subject are unchanged). The percent change analysis reveals that one SEM represents 1.70% of the total number of items within the BNT, suggesting that changes attributed to random error would likely not influence the overall impressions of test performance. The combination of the SEM analysis and PC score yield a value of test-retest reliability not offered by bivariate correlation and tests of mean difference.

Chapter 2: Review of the Literature

For the Western Aphasia Battery-Revised (WAB-R) (Kertesz, 2007), neither the SEM and PC, nor any other statistical analysis was performed to examine test-retest reliability. Several years after the original Western Aphasia Battery (WAB) (Kertesz, 1982) was published, the updated version was released (Kertesz, 2007). The revised assessment includes the following changes: the introduction of new subtests (Supplemental Reading and Supplemental Writing), changes made in directions to examinees in the Spontaneous Speech section, new verbal and visual stimuli in the Object Naming subtest, new verbal stimuli and rearrangement of item order in the Repetition subtest, a new verbal stimulus in the Apraxia subtest, and added task instructions to the Clock Drawing portion of the Drawing subtest (Kertesz, 2007). Along with the omissions, additions, and revisions listed above, a new analysis of language scores, the "Language Quotient", was added. This additional component to scoring was made to allow the examiner to characterize language function with consideration of reading and writing abilities, using a quantitative score. While the WAB (Kertesz, 1982) has undergone extensive standardization testing, the psychometric soundness of the revised version remains relatively unknown. More specifically, the ability of the WAB-R (Kertesz, 2007) to detect the same deficits in the same person following a secondary administration (provided there is no change in the disability), is comparatively unidentified, when considering the original version.

The Western Aphasia Battery (WAB) (Kertesz, 1982) is a standardized assessment used typically to assess cognitive-linguistic skills in adults with acquired neurological disorders (e.g., post-stroke aphasia). The WAB (Kertesz, 1982) is comprised of various subtests (i.e., word fluency, object naming, etc.) that target areas of cognitive-linguistic communication commonly associated with aphasia. To measure the instrument's reproducibility, 22 adults with chronic

aphasia who were one or more years post-onset were tested and retested using the WAB (Kertesz, 1979; Kertesz, 1982). From here on, this study will be referred to as Study 1. No information about the time between testing dates is made clear by the author, nor is this information present in the manual. The researcher did, however, report that no significant medical changes took place between the two testing times (Kertesz, 1979). Kertesz (1979) reported a Pearson's *r* correlation coefficient of 0.992 and used the statistical significance ($p < 0.01$) as evidence of a lack of significant medical change and high test-retest reliability. Kertesz (1979) goes on to offer additional support in the form of a low mean difference (0.9) between test scores and retest scores. It was not stated which scores were used to calculate the Pearson's *r* correlation coefficient or the mean difference (e.g., Aphasia Quotient, Cortical Quotient, specific subtest scores, etc.). In addition to the lack of performance data, the quality and significance of this evidence is further called into question because no demographic data (e.g., age, type/severity of aphasia, etc.) is available for this initial sample population. Consequently, clinicians administering the WAB (Kertesz, 1982) in practice are unable to validate the test-retest reliability using information found in this study.

The WAB (Kertesz, 1982) underwent a second examination of test-retest reliability which from here on will be referred to as Study 2. A second group of participants (N=35) with (stable) chronic aphasia were administered the assessment on two different testing dates (Kertesz, 1979). "The mean post onset time of the first test was 2.05 years and the second, 3.91 years" (Kertesz 1979, pg. 70). A paired t-test also was conducted to compare mean scores of test performance to retest performance on all subtests. The mean Aphasia Quotient and Cortical Quotient scores (achieved on the test and the retest) were also compared. There was no significant difference in performance from Time 1 to Time 2 for most subtests; however,

statistically significant differences were observed on the Word Fluency subtest, Object Naming subtest, and the overall naming total. “Slightly significant” differences were observed on the Reading and Writing sections (Kertesz, 1979). Neither of the analyses used in Study 2 were able to capture a measurement of score stability because of the measures used to assess the data. Further, the differences that were observed, though statistically significant, were discussed without detail; therefore, no conclusions can be made about the lack of statistical significance observed in some subtests. The test-retest reliability data points of Study 2 are presented in Table 1.

Section	Means		Pearson’s *r*	
	1st Test	2nd Test	*N*-pairs	(Test 1-2)
Information Content	5.00	5.20	35	0.95
Fluency	5.45	5.28	35	0.93
Yes/No Questions	47.57	49.62	35	0.75
Auditory Word Recognition	41.11	40.51	35	0.85
Sequential Commands	51.08	49.20	35	0.90
Comprehension Total	7.05	6.97	35	0.88
Repetition	5.57	5.35	35	0.97
Object Naming	30.02	33.25	35	0.94
Word Fluency	5.11	6.28	35	0.89
Sentence Completion	6.17	5.94	35	0.88
Responsive Speech	5.48	5.60	35	0.96
Naming Total	4.67	5.12	35	0.96
Aphasia Quotient	55.36	55.48	35	0.97
Reading	63.30	63.30	30	0.92
Writing	53.37	57.22	24	0.95
Praxis	78.29	78.11	17	0.45
Drawing	16.64	17.20	25	0.79
Block Design	7.20	7.26	15	0.89
Calculation	17.58	18.00	24	0.81
Raven’s Coloured Progressive Matrices	21.61	22.19	26	0.89

Table 1. Test-retest reliability (Kertesz, 1979)

The WAB's test-retest reliability was again examined in a study that will be referred to as Study 3 (Shewan & Kertesz, 1980). A sample group of participants (N=38) with (stable) chronic aphasia were administered the assessment on two different testing dates (Shewan & Kertesz). The time between tests varied from six months to six years, six months (Shewan & Kertesz). The considerable length of time between testing dates calls into question the overall stability of the participants related to or notwithstanding the language disorder. Further, Shewan and Kertesz stated that the same examiner did not always administer the test on both occasions during the study. Shewan and Kertesz acknowledged their lack of control for 'Examiner' as a variable reporting that, "the interrater differences may have affected the results of temporal stability". However, the researchers reported no significant differences in performance on the majority of the WAB (Kertesz, 1982) subtests, suggesting reproducibility of results (Shewan & Kertesz). A Pearson's product-moment correlation coefficient was used to support the above conclusion. All correlations were greater than 0.880 and were significant at the 0.001 level, with the exception of the Praxis subtest (Pearson's r = 0.581; significance= 0.006) (Shewan & Kertesz). No further detail was given about the lack of reproducibility observed within the Praxis subtest. Moreover, it is apparent from the data reported in Study 3 that all participants were not included in the analysis. Table 2 presents the test-retest reliability data points of Study 3, and also demonstrates that some participants were excluded from the analysis of the Reading, Writing, Praxis, and Construction subtests as well as the analysis of the Cortical Quotient.

WAB Variable	Number of subjects	Pearson's *r*	Significance Level
Information Content	38	0.947	0.001
Fluency	38	0.941	0.001
Comprehension Total	38	0.881	0.001
Repetition	38	0.970	0.001
Naming Total	38	0.923	0.001
Aphasia Quotient	38	0.968	0.001
Reading	32	0.927	0.001
Writing	25	0.956	0.001
Praxis	18	0.581	0.006
Construction	14	0.974	0.001
Cortical Quotient	9	0.895	0.001

Table 2. Correlation coefficients for test-retest reliability for WAB variables (Shewan & Kertesz, 1980)

A novel examination of the test-retest reliability of the WAB-R is necessary given that the information and data provided in the manual are not representative nor applicable to the current assessment, due to the additions, omissions, and revisions made to the exam.

Chapter 3: Methods

Participants

Fourteen persons with stroke-induced aphasia were consented, and 10 of these individuals completed the study. The participants were recruited from previous studies conducted by Susan Jackson, from the American Stroke Foundation, hospitals, outpatient clinics, and aphasia treatment groups in the Kansas-Missouri bi-state area. All participants were monolingual English speakers who were literate prior to their stroke (per self-report) with adequate hearing and vision. Adequate hearing was defined as the ability to follow at least 4/5 one-step directions presented auditorily (see Appendix A). Adequate vision was defined as at least 4/5 accuracy on a simple and brief picture-matching task. The vision screening procedure included five picture sets, each containing four black and white line drawings. Vision screening accuracy was determined by the participants' ability to match a target stimulus to its replica in the stimulus set. The stimuli within each set (and the target stimulus) are listed in Appendix B. Participants who passed the screening using hearing or vision aids were required to wear/use the respective aids during all interviews and testing. Participants were excluded from the study if they presented with a history of additional neurological disease (e.g., traumatic brain injury). Participants also were excluded if they did not score within the aphasic range on the WAB-R (Kertesz, 2007), which is defined as an Aphasia Quotient of 93.8 or above. Four individuals were excluded from the study. Three individuals were excluded from the study based on their mild presentation of aphasia, and one person was excluded because he had a traumatic brain injury in addition to a stroke.

All participants included in the study were at least six months post-onset of a left-hemisphere stroke. Participants ranged in age from 36 to 76 years old (M = 64, SD = 13).

Months post onset of stroke varied from 8 months to 234 months (M = 65, SD =68). Years of education ranged from 12 years to 21 years (M = 16, SD =3). Each participant's race/ethnicity, gender, pre-morbid handedness, and occupation also were recorded. Table 3 presents the demographic information in greater detail. Additional demographic information collected from each participant included a current list of medications. All demographic data were collected via self-report using an aphasia questionnaire created for this study (see Appendix C). Each participant's medications are outlined in Appendix D.

Subject #	Age	Months Post Onset	Years of Education	Race/ Ethnicity	Gender	Handedness	Treatment (Group or Individual)	Occupation
1	67	112	13	B/AA	M	R	None	Stocker
2	36	35	18	W/C	F	R	Group	Retired
3	65	19	12	W/C	M	L	Group and Individual	Cable Installation
4	71	48	18	W/C	M	R	Group and Individual	Teacher
5	69	234	21	W/C	M	R	Group	Retired
6	72	24	21	W/C	M	R	Group	Judge
7	76	23	12	W/C	M	R	Group and Individual	Oil Digger
8	69	8	18	W/C	M	R	Individual	Retired
9	75	100	16	W/C	M	R	Group and Individual	Teacher
10	45	48	19	W/C	F	R	Group and Individual	Physical Therapist

Table 3. Demographic data. M = Male, B/AA = Black/ African-American, W/C = White/ Caucasian, R = Right, L = left

Stimuli

Test items from Part 1 and Part 2 of the Western Aphasia Battery-Revised (WAB-R) (Kertesz, 2007) constituted the stimuli; the test assessed speech content, fluency, auditory comprehension, repetition, naming, reading, writing, drawing, calculation, block design, and

apraxia. The Raven's Coloured Progressive Matrices (Raven, 1976) also were included as part of Part 2 of the WAB-R.

Procedure

Researchers first obtained approval for conducting the study from the University of Kansas Medical Center Institutional Review Board. Informed consent was obtained before testing began. Consent to participate in the study was acknowledged by signing a written consent form (see Appendix E). All participants signed their own consent. A surrogate consent form was available for participants who were unable to provide their own written consent due to language comprehension difficulty (see Appendix F), with an accompanying assent form (see Appendix G), but these additional consent forms were not needed. After providing consent, participating participants were assigned a number code to replace identifying information and preserve confidentiality (e.g., WAB-1, WAB-2, etc.). Each participant was then administered a questionnaire (see Appendix C) to collect demographic information and identify clinical characteristics incompatible with the study. Participants were then screened for adequate hearing and vision, as described above. Next, all sections and subtests of the Western Aphasia Battery-Revised (WAB-R) (Kertesz, 2007) were administered. The WAB-R was presented to all participants in the same order, with the exception of Participant 5 who was not administered the Block Design subtest, and Participant 10 who elected to discontinue the administration of the Apraxia and the Constructional, Visuospatial, and Calculation portions of the WAB-R. The examiners adhered to the administration and scoring procedures stated in the test manual. After a 2-week interval, participants were briefly interviewed about changes in medication (including dosage) and health status during the interval period, and then underwent a second administration of all sections and subtests of the WAB-R. No clinically relevant changes in health status took

place during the interval period, with the exception of Participant 3 who discontinued one of his medications. Testing and retesting took place in the participants' homes, at the Schiefelbusch Speech and Hearing Clinic, and in the Department of Hearing and Speech at the University of Kansas Medical Center. Participants were tested by one of three graduate student clinicians (BB, HH, AE). BB tested Participants 1, 3, 4, 7, W-1 (withdrawn), W-2 (withdrawn), W-3 (withdrawn), W-4 (withdrawn), 9, and 10. AE tested Participant 6, and HH tested Participants 2, 5, and 8. All participants were assessed by their same examiner at Time 1 and Time 2. All participants were tested in the same environment at Time 2 as Time 1.

Chapter 4: Results

Individual Data

The individual data are presented in Appendix H.

Data Analyses

Some of the data analyses were completed using the IBM SPSS Statistics software package (Pearson Product-Moment correlation, paired *t*-test). Other analyses were done by hand (SEM and Percent Change score). The effect size was calculated using an online calculator for Cohen's *d* (https://www.socscistatistics.com/effectsize/default3.aspx). The data consisted of the total number of points achieved on each section or subtest of the WAB-R (Kertesz, 2007). The WAB-R (Kertesz) typically divides the raw score by 10 or 20, to calculate quotient scores (Aphasia Quotient, Language Quotient, and Cortical Quotient). However, to preserve information about changes in performance, the current study used the raw scores instead of the adjusted scores generated by the WAB-R (Kertesz). Each analysis was performed on the data from all 10 participants with the exception of the Constructional, Visuospatial, and Calculation section and the Cortical Quotient, which included eight participants in the analysis because Participant 5 was not administered the Block Design subtest and Participant 10 elected to discontinue the Constructional, Visuospatial, and Calculation portions of the test. Analysis of the Apraxia subtest data included nine participants, as Participant 10 elected to discontinue this subtest as well.

Pearson Product-Moment Correlation

The purpose of performing the Product-Moment Correlational analysis was to compare the data from the current study to prior results of WAB (Kertesz, 1982) score stability, which primarily used correlational analyses to determine reproducibility. Table 4 displays the

correlation coefficients for the relation between the participants' WAB-R performance at Time 1 and Time 2.

WAB-R Sections/ Subtests	Pearson's *r* (Time 1-Time 2)	*p*-value
Aphasia Quotient	.973	.0001
Language Quotient	.940	.0001
Cortical Quotient	.915	.001
Spontaneous Speech	.866	.001
Auditory Verbal Comprehension	.151	.678
Repetition	.980	.0001
Naming	.993	.0001
Reading	.910	.0001
Writing	.897	.0001
Apraxia	.631	.069
Constructional, Visuospatial, and Calculation	.925	.001
Supplemental Reading	.842	.002
Supplemental Writing	.874	.001

Table 4. WAB-R Pearson's Product-Moment coefficients and p-values

The overall α was chosen as .05. The overall α was then divided by the number of analyses (13), which resulted in an α level of .0038 for each individual correlational analysis. Results indicated a positive, significant correlation between the test date (T1) and retest date (T2) for scores achieved on all subtests and sections except the Auditory Verbal Comprehension section and the Apraxia subtest.

Paired *t*-test and Effect Size

Although the paired *t*-test is not a valid measure of score stability, this statistical analysis was performed to compare the data from the current study to the results of previous studies examining the test-retest reliability of the WAB (Kertesz, 1982). Table 5 displays the results of a paired *t*-test analysis comparing the mean performance on the WAB-R at Time 1 and Time 2. A

significant difference in mean performance was defined as any value of *p* that was less than α (α = .0038; overall α = .05/13 [number of t-tests]).

WAB-R Sections/ Subtests	Means (*SD*)		*p*-value (2-tailed)	Cohen's *d*
	1st Test	2nd Test		
Aphasia Quotient	71.02 (*16.16*)	70.49 (*17.39*)	.689	0.032
Language Quotient	69.07 (*14.63*)	68.43 (*15.18*)	.705	0.043
Cortical Quotient	73.42 (*11.58*)	74.17 (*11.73*)	.674	0.064
Spontaneous Speech	14.50 *(3.06)*	14.00 (*3.55*)	.397	0.151
Auditory Verbal Comprehension	173.00 (*18.69*)	174.90 (*14.68*)	.791	0.113
Repetition	61.90 (*24.35*)	61.30 (*26.28*)	.735	0.024
Naming	61.70 (*30.34*)	63.70 (*30.64*)	.125	0.066
Reading	72.40 (*20.80*)	69.70 (*20.02*)	.352	0.132
Writing	55.50 (*23.94*)	54.55 (*21.14*)	.783	0.042
Apraxia	55.20 (*4.17*)	56.22 (*3.41*)	.384	0.268*
Constructional, Visuospatial and Calculation	69.37 (*15.74*)	71.81 (*14.85*)	.288	0.159
Supplemental Reading	7.4 (*4.81*)	8.00 (*5.29*)	.526	.119
Supplemental Writing	2.00 *(2.66*)	3.50 (*3.71*)	.034	.465*

Table 5. WAB-R paired t-test and effect size. * = Small effect size

There were no statistically significant differences in the mean scores from Time 1 to Time 2. Calculation of effect size was undertaken to examine at the magnitude of mean differences. Effect size is less susceptible to the fluctuations that arise due to a small number of participants, as in the current study. Effect sizes are presented in Table 5. The effect size was calculated using the following equation:

$$d = \frac{M_1 - M_2}{\frac{\sqrt{SD_1^2 + SD_2^2}}{2}}$$

where M_1 is the mean at Time 1, M_2 is the mean at Time 2, SD_1 is the standard deviation at Time 1 and SD_2 is the standard deviation at Time 2. Results were designated as having less than a small effect ($d < 0.2$), a small effect ($d = 0.2$-0.49), a medium effect ($d = 0.5$-0.79), or a large effect ($d \geq 0.8$), as recommended by Cohen (2013). There was a small effect size for the Apraxia subtest and the Supplemental Writing subtest. The remaining sections/subtests showed less than a small effect size.

Standard Error of Measurement and Percent Change Score

As previously mentioned, the SEM provides limited information about test-retest reliability when examined in isolation; thus, the Percent Change score was calculated to supplement the SEM by adding information about the number of points on each WAB-R (Kertesz, 2007) section or subtest. Standard error of measurement was calculated using the following formula:

$$\text{SEM} = SD\sqrt{1 - r}$$

The outcome of the analysis was a number used to establish a range of future performance scores based on the observed score (±1 SEM = 68% confidence interval; ±2 SEM = 95% confidence interval). If a standardized assessment is stable, an examiner can expect a small range of future performance scores in an unchanged individual, otherwise stated as a confidence interval with minimal variability. In this way, standard error of measurement quantifies the amount of random error variability in the test, allowing the examiner to predict the impact of random error on the examinee's true performance score. Table 6 displays the SEMs of the WAB-R (Kertesz, 2007) sections and subtests. The Percent Change scores (PCs) also are presented in Table 6. While it has been convention to calculate the SEM using the total number of items in a test, the distribution of items in the WAB-R varies widely. Some sections have less than 10 items

(Spontaneous Speech) and some sections have greater than 50 (Auditory Verbal Comprehension). Further, points are distributed unevenly across sections, and weighted differently within subtests. Consequently, a Percent Change score calculated using number of items within a WAB-R section/subtest would not be as meaningful as a score based on the number of points. Thus, the Percent Change score was calculated using the following equation:

$$PC = (SEM/Number\ of\ Points\ Possible) \times 100$$

The Percent Change score takes into account the total number of points possible when interpreting the SEM. For example, an SEM of 2 points indicates that 68% of the time, a person will score ±2 points of his/her initial score on repeated testing. The PC indicates that 2 points is 2% of the overall number of points possible if 100 points are possible, and 2 points is 20% of the overall number of points possible if 10 points are possible.

WAB-R Sections/ Subtests	Standard Error of Measurement	Percent Change score
Aphasia Quotient (Max= 100)	2.65	2.65
Language Quotient (Max= 100)	3.58	3.58
Cortical Quotient (Max= 100)	3.37	3.37
Spontaneous Speech (Max= 20)	1.12	5.60*
Auditory Verbal Comprehension (Max= 200)	17.22	8.61*
Repetition (Max= 100)	3.44	3.44
Naming (Max= 100)	2.53	2.53
Reading (Max= 100)	6.24	6.24*
Writing (Max= 100)	7.68	7.68*
Apraxia (Max= 60)	2.53	4.21
Constructional-Visuospatial, Calculation (Max= 100)	4.31	4.31
Supplemental Reading (Max= 20)	1.91	9.56*
Supplemental Writing (Max= 20)	0.94	4.73

Table 6. WAB-R Standard Error of Measurement and Percent Change score. * = greater than 5

The PCs that were calculated ranged from 2.65 to 9.56. This means that one WAB-R SEM represented between 2.65% and 9.56% of the total number of points possible on various

sections and subtests of the WAB-R (Kertesz, 2007). Examples will be provided to further explain the interpretation. The first example uses the Aphasia Quotient Percent Change score. The Aphasia Quotient PC was 2.65; thus, 1 SEM represents 2.65% of the total number of points possible for an Aphasia Quotient. Restated, 68% of the time, an examinee will score ±2.65 % of the total number of points possible for the Aphasia Quotient on repeated testing. The second example uses the Supplemental Reading subtest Percent Change score. The Supplemental Reading subtest PC was 9.56; therefore, 1 SEM represents 9.56% of the total number of points possible for the Supplemental Reading subtest. Accordingly, 68% of the time, an examinee will score ±9.56% of the total number of points possible for the Supplemental Reading subtest on repeated testing. The majority of the WAB-R (Kertesz, 2007) sections and subtests had PCs that were less than 5. This indicates that examinees would be expected to score within ±5% of the total number of points possible on repeated testing of the WAB-R sections/subtests, 68% of the time. Five of the subtests/sections had PCs greater than 5 but less than 10 (Spontaneous Speech, Auditory Verbal Comprehension, Reading, Writing, and Supplemental Reading).

Chapter 5: Discussion

This study reports the score stability of the WAB-R using 10 participants with chronic aphasia who were tested an average of 65 months post-onset of stroke and retested two weeks later. The discussion will focus on an interpretation of the results, as well as a comparison between the current results and the findings of the original WAB reliability studies. Possible explanations will be offered for the differences between the current study and past WAB reliability studies.

The present study employed a Pearson Product-Moment Correlation analysis and a paired *t*-test analysis as measures of statistical significance. Statistical significance refers to any considerable differences observed, according to the standard logic and procedures of inferential statistics (Bothe & Richardson, 2011). Previous studies (Kertesz, 1979) also have implemented these analyses to assess the reliability of the WAB. However, as mentioned previously, neither correlational analyses nor t-tests are appropriate measures of score stability. That being said, the correlation coefficients ranged from .151 to .993 for the 13 sections/subtests of the WAB-R, with only two sections/subtests falling below .80 (Auditory Verbal Comprehension section and Apraxia subtest). It is difficult to compare the results of the correlational analyses from the current study to the results of the correlational analyses from two previous studies. Kertesz (1979), Kertesz and Shewan (1980), and the current study have seven sections/subtests in common for the correlational analyses. Kertesz and Shewan (1980) have one additional WAB/WAB-R section in common with the correlational analyses of the current study (Cortical Quotient). Thus, not all of the current correlation coefficients can be compared to correlation coefficients from previous studies. Four of the coefficients from the previous studies were based on different denominators from those used in the current study [Auditory Verbal

Comprehension: 200 (current study) vs. 10 (Kertesz), Repetition: 100 (current study) vs. 10 (Kertesz), Naming total: 100 (current study) vs. 10 (Kertesz), and Apraxia/ Praxis: 60 (current study) vs. 100 (Kertesz)]. The current study used the raw data scores for the correlational analyses, whereas the previous reliability study (Kertesz, 1979) divided the raw score by 10 or 20 before performing the correlational analyses. It was not possible to determine the denominator of the scores used in the Kertesz and Shewan (1980) correlational analyses. A comparison of the data is presented in Table 7.

WAB/ WAB-R Section/Subtest	Current Study	Kertesz (1979)	Kertesz & Shewan (1980)
Aphasia Quotient	.973	.97	.968
Cortical Quotient	.915	-	.895
Auditory Verbal Comprehension/ Comprehension Total*	.151	.88	.881
Repetition*	.980	.97	.970
Naming/ Naming Total*	.993	.96	.923
Reading	.910	.92	.927
Writing	.897	.95	.956
Apraxia/ Praxis*	.631	.451	.581

Table 7. Coefficient comparison. * = different denominators across studies

Given the challenges of comparing the correlation coefficients from the current study with correlations from previous studies, three conclusions can be reached: 1) correlation coefficients are predominately above .85 across all three studies, 2) the Praxis/Apraxia correlation coefficient was the lowest or second lowest across all three studies (.631, .451, and .58), and 3) the Auditory Verbal Comprehension correlation coefficient from the current study was drastically lower (.151) than the correlation coefficient from the other two studies (.88 and .881). One subject in the current study improved significantly on one subtest (Sequential

Commands) within the Auditory Verbal Comprehension section, and perhaps the combination of this outlier and a small sample size conspired to result in a very low correlation coefficient.

Using the t-test as the test of statistical significance, the current study found no significant difference between the mean scores at Time 1 and Time 2 for any of the 13 WAB-R sections/subtests examined. One of the previous studies examining the test-retest reliability of the WAB (Kertesz, 1979, Study 1) did not implement a t-test to assess the differences in WAB scores at Time 1 and Time 2, but the mean difference between test and re-test scores was reported (.9). This number is difficult to interpret given we do not know which WAB scores were used to calculate the mean difference (i.e., Aphasia Quotient, Cortical Quotient, specific subtest scores). The second of two previous studies examining the test-retest reliability of the WAB (Kertesz, 1979, Study 2) did use t-tests in the analysis of test-retest reliability. The current study and the Kertesz (1979) Study 2 have seven sections/subtests in common for the t-tests. See Table 8 for the p-values associated with the current study and the description of whether there was a statistically significant difference in the Kertesz (1979) Study 2. Kertesz (1979) Study 2 did not provide information about the p-value used to determine whether there was a statistically significant difference between the mean scores at Time 1 and Time 2. Further, no definition was given for the term "slightly significant."

WAB/ WAB-R sections/subtests	Current Study	Kertesz (1979) Study 2
Aphasia Quotient	$p > .05$	NS
Auditory Verbal Comprehension/ Comprehension Total*	$p > .05$	NS
Repetition*	$p > .05$	NS
Naming/ Naming Total*	$p > .05$	"Significant"
Reading	$p > .05$	"Slightly Significant"

Writing	p > .05	"Slightly Significant"
Apraxia/ Praxis*	p > .05	NS

Table 8. Comparison of t-test findings. NS = not significant, * = different denominators across studies

Practical significance refers to the magnitude of any differences observed (Bothe & Richardson, 2011) - in this instance, how much change was observed between Time 1 and Time 2 in WAB-R outcomes. Practical significance is commonly measured using effect size. This study adhered to that convention. Two of the 13 WAB-R sections/subtests showed a small effect size (Supplemental Writing and Apraxia), and the remaining 11 WAB-R sections/subtests showed less than a small effect size, suggesting fair-good score stability.

Clinical significance refers to change that is meaningful/valuable to the patient and clinician in the context of service provision (Bothe & Richardson, 2011). In the present study, standard error of measurement (SEM) and Percent Change score were used to measure clinically significant changes in scores obtained on initial testing and re-testing of the WAB-R. No studies examining score stability of the WAB (Kertesz, 1982) have used the SEM or Percent Change score to determine test-retest reliability. Percent Change score is preferred over SEM as a measure of score stability because it accounts for the magnitude of the scores on which the SEM is based. The Percent Change score "represents the percentage change from the [the number of points] accounted for by a change in score of 1 SEM between sessions" (Nicholas & Brookshire, 1993, p. 342). In the current study, 1 SEM represented between 2.65% and 9.56% of the total number of points possible on the various sections and subtests of the WAB-R. For 8/13 WAB-R sections/subtests, 1 SEM represented less than 5% of the total number of points possible, which means that 68% of the time, the score of people with aphasia changes by less than ±5% of the total number of points possible, on repeated testing. Currently, there is no consensus on what dictates an "acceptable" PC score. Several factors might go into the determination of an

acceptable PC score, including whether a certain magnitude of change on repeated testing would result in a different aphasia classification, would result in a different severity of aphasia, or would indicate a change in the type of intervention chosen. For 5/13 subtests/sections of the WAB-R, 1 SEM represented greater than 5% but less than 10% of the total number of points possible: Spontaneous Speech, Auditory Verbal Comprehension, Reading, Writing, and Supplemental Reading.

Higher PC scores might be explained by: 1) the distribution of points within subtests, 2) a small number of items within a subtest, or 3) the clinical judgement of new clinicians. Several WAB-R subtests and sections have an uneven distribution of points per item, namely the Comprehension of Sentences subtest in the Reading section and the Sequential Commands subtest of the Auditory Verbal Comprehension section. These subtests allocate small point values to some items (i.e., 2/ 80 points) and large point values to other items (i.e., 20/80 points). This distribution of points is not conducive to score reproducibility because performance in the target language domain is influenced by a single heavily weighted item or a few heavily weighted items. Otherwise stated, the examinee need only answer one or two questions on the subtest incorrectly for the WAB-R score to plummet. Conversely, if the examinee responds correctly to these few items, the WAB-R score soars. As seen in the results of the current study, score instability can lead to changes in the level of aphasia severity and/or the classification of aphasia type. A small number of items within a subtest also has the potential to affect the test-retest reliability of a measure. A small number of items on a measure leaves the entire test susceptible to change based on a single item, limiting score stability. Consider a 5-point test, where each point amounts to 20% of the overall score. A 2-point change in score would have a drastic effect because each point is worth 20% of the overall score. On the other hand, a 2-point

change on a 20-point test would have less of an effect because each point is worth 5% of the overall score.

Lastly, some of the WAB-R sections/subtests with higher PC scores had more subjective scoring. Making clinical judgements can be especially challenging for new clinicians who have limited experience making these real-time judgements and limited clinical exposure to deficit presentations of the language domain being assessed. Thus, it is likely that inexperienced clinicians will demonstrate less consistency in making judgements of performance.

Implications

The absence of WAB-R test-retest reliability data prompted the current study. The purpose of this investigation was to determine the extent to which the WAB-R is able to accurately reproduce results on two separate occasions. Results of the study suggest that many of the components of the test yield score stability. However, clinicians should be aware of the PC score results that suggested less score stability in the following sections/subtests: Spontaneous Speech, Auditory Verbal Comprehension, Reading, Writing, and Supplemental Reading. Instability in the Spontaneous Speech scores or the Auditory Verbal Comprehension scores could result in changes in the WAB-R aphasia classification, which is one of the main purposes of an impairment level aphasia battery. Instability in the Spontaneous Speech scores or the Auditory Verbal Comprehension scores also could result in changes in the severity of aphasia determined by the WAB-R. In fact, four of the participants in the current study were assigned a different aphasia severity at Time 1 and Time 2, and two participants differed in their aphasia type at Time 1 and Time 2.

Limitations

This study was limited by its small sample size. When sample size is small there may not be enough power to detect a statistically significant difference had one existed. Further, the small sample was limited in its diversity of race, ethnicity, and gender. One person in the study was African-American, and two people in the study were women. These ratios do not accurately represent the broader population of individuals with aphasia. Because the individuals in the present study are not representative of the population, the generalizability of the results is limited. The majority of the participants in the current study had mild to moderate aphasia. One of the participants had severe aphasia at Time 1 and moderate aphasia at Time 2, and another participant had severe aphasia at both time points. None of the participants had very severe aphasia. Thus, the results of this study are not generalizable to people with severe aphasia or very severe aphasia. The study also was limited by access to testing materials. Materials were limited to a single copy of the WAB-R, restricting the availability for participant testing. The experience level of the examiners was a limitation. Each of the graduate students were relatively new clinicians whose emerging clinical judgement presented challenges in the scoring of subjective portions of the WAB-R (i.e., Spontaneous Speech and Writing).

Future Studies

Ideally, future studies would include a larger sample size with increased racial, ethnic, and gender diversity as well as varying severities of aphasia. In addition to participants with mild and moderate aphasia, future studies should include people with severe and very severe aphasia. Future studies also might incorporate the use of video recording technology as a way of preserving performance data. The WAB-R ascribes a categorization of aphasia type to the examinee based on the pattern of performance on Part 1 (i.e., Spontaneous Speech, Auditory

Verbal Comprehension, Repetition, and Naming). It was observed in the present study that some participants were promoted/demoted to a different aphasia type between T1 and T2. Future studies are needed to investigate the degree to which scores on testing and re-testing indicate a different diagnosis (e.g., aphasic vs. non-aphasic) and/or a different type of aphasia. Future studies also might survey speech-language pathologists to determine what degree of change in WAB-R scores would suggest a change to the treatment approach.

References

Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing test-retest reliability of psychological measures: Persistent methodological problems. *European Psychologist*, *22*(4), 207-218.

Beckerman, H., Roebroeck, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. M. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research*, *10*(7), 571-578.

Bothe, A. K., & Richardson, J. D. (2011). Statistical, practical, clinical, and personal significance: Definitions and applications in speech-language pathology. *American Journal of Speech-Language Pathology, 20* (3), 233-242.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment* (Vol. 17). Thousand Oaks, CA: Sage publications.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York: Routledge.

Flanagan, J. L., & Jackson, S. T. (1997). Test-retest reliability of three aphasia tests: Performance of non-brain-damaged older adults. *Journal of Communication Disorders*, *30*(1), 33-43.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test*. Austin, TX: Pro-ed.

Kertesz, A. (1979). *Aphasia and associated disorders: Taxonomy, localization, and recovery*. New York, NY: Holt Rinehart & Winston.

Kertesz, A. (1982). The Western Aphasia Battery. New York, NY: Grune & Stratton.

Kertesz, A. (2007). The Western Aphasia Battery- Revised. San Antonio, TX: PsychCorp.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage publications.

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.

Raven, J. C. (1976). *Coloured progressive matrices*. London: Lewis.

Shewan, C. M., & Kertesz, A. (1980). Reliability and validity characteristics of the Western Aphasia Battery (WAB). *Journal of Speech and Hearing Disorders*, *45*(3), 308-324.

Appendix A. (Hearing Screening-Auditory Commands)

Aided Hearing (e.g., personal amplifier, hearing aids, etc.)	Yes	No
Present each item verbally		
1. Clap your hands	0	1
2. Touch your head	0	1
3. Point to your nose	0	1
4. Stick out your tongue	0	1
5. Look at the ceiling	0	1

Score:_____/5

Appendix B: (Vision Screening-Picture Matching)

Aided Vision (e.g., glasses, bifocals, contacts, etc.) Yes No

Explain to the participant that they will be shown a single picture and asked to match it to an identical picture amongst a set of four picture. Present target stimulus first. Then, present stimulus set.

1. Stimulus Set (example) 0 1
- Spoon
- Basketball
- Elephant (target)
- Woman

2. Stimulus Set #1 0 1
- Gas
- Cake (target)
- Towel
- Chair

3. Stimulus Set #2 0 1
- Umbrella (target)
- Wrench
- Refrigerator
- Cat

4. Stimulus Set #3 0 1
- Flowers
- Dog (target)
- Book
- Candy

5. Stimulus Set #4 0 1
- Car
- Broom
- Chicken
- Saw (target)

Appendix B: (Vision Screening-Picture Matching)-continued

6. Stimulus Set #5 • Shoes • Butterfly • Carrots (target) • Whistle	0	1

Score:_____/5

Appendix C: (Initial Aphasia Questionnaire)

Participant #: _____

DOB: _____ /_____ /_____

Gender: Male Female

Consent Type: Consent Surrogate Assent

Race/Ethnicity: White/ Caucasian Black/ African-American Native-American

Hispanic/ Latin-American Asian/ Pacific Islander Other

Hx of other neurologic disease (e.g., head injury, Parkinson's, history of alcohol abuse):

Multiple Strokes: Yes No

Date of stroke: _____ Time post onset: _____yrs _____mos

Location of stroke: Left Right

Level of Education:

_____

Current Speech, Language, Cognitive Services: Yes (describe below) No

_____

_____

Current Medications:

_____

Passed Hearing Screen: Yes No Score:_____/5

Passed Vision Screen: Yes No Score:_____/5

Testing Date 1: _____ /_____ /_____

Testing Date 2: _____ /_____ /_____ Time from T1 to T2:_____

Changes in Medication: Yes No

Other:

_____

_____

_____

Appendix D. (List of Medications)

Participant 1: Calcium
Cholecalciferol
Clopidogrel
Simvastatin
Vitamin D

Participant 2: Apremilast
Aspirin
Duloxetine
Folic Acid
Hydroxychloroquine
Levothyroxine
Methylphenidate
Zolpidem

Participant 3: Clopidogrel
Alprazolam
Atorvastatin
Dextroamphetamine/Levoamphetamine
Duloxetine
Hydrocodone
Lamotrigine
Lisinopril
Meloxicam
Metformin
Pantoprazole
Trazadone

Participant 4: Aspirin
Atorvastatin
Baclofen
Carvedilol
Clopidogrel
Keppra
Lexapro
Losartan

Participant 5: Aspirin
Atorvastatin
Fluticasone
Losartan
Pantoprazole
Psyllium

Participant 6: Atorvastatin
Cholecalciferol
Clopidogrel
Cobalamin
Januvia
Lisinopril
Memantine
Metformin
Rivaroxaban
Sertraline

Participant 7: Alprazolam
Antidepressants (unspecified)
Sertraline

Participant 8: Unavailable

Participant 9: Unavailable

Participant 10: Aspirin
Atorvastatin
Cobalamin
Ferrous sulfate
Fluoxetine

Appendix E: (Consent Form)

RESEARCH CONSENT FORM

Test Re-test Reliability of the Western Aphasia Battery - Revised

Susan T Jackson
(913) 588-5937

You are being invited to join a research study being done at the University of Kansas Medical Center (KUMC) by Susan Jackson. Being in this study is optional. You can decide not to participate or stop at any time. Regardless of your decision, you will still get the same care from your health care team.

Why is this study being done?
The purpose of this study is to find out whether people with aphasia perform similarly on the Western Aphasia Battery-Revised when they are given the test on two different occasions.

What am I being asked to do?
This study involves two visits to the Hearing and Speech department at the KU Medical Center, or we can test you in your home or in a public space (library, place of worship, senior center) if you would prefer. Your participation in the study will have two parts:

Part 1
This part will take about 2 hours. We will begin the study by asking you questions about you and about your current health (for example: address, date of birth, date of stroke, handedness, medications). We will test your vision and hearing. You will be given some aphasia tests that will include naming objects, repeating words and sentences, answering questions, following some directions, reading, writing, drawing some pictures, acting out how to do certain things, and choosing the best option to complete a visual design.

Part 2
This part will take about 1.5. hours. Part 2 will take place 2-3 weeks after Part 1. You will do the same aphasia tests that you did in Part 1 (naming objects, repeating words and sentences, answering questions, following some directions, reading, writing, drawing some pictures, acting out how to do certain things, and choosing the best option to complete a visual design).

Are there risks or discomforts to consider?
There are no known risks associated with this study. You may get tired or frustrated, but you may take a break at any time or stop the session and continue again on another day.

How will confidentiality and privacy be protected?
There is a small risk of loss of confidentiality when personal information is used for research. Your information will only be used by study team members and approved researchers. When we write up our results or make presentations, we will not use any names.

Appendix E: (Consent Form)-continued

We will follow the HIPAA laws about privacy. Study records will include your health information and information we collected about you during the research. We will keep your study information indefinitely. The study information will be kept separately from your name and other personal identifiers. Study information will be shared with members of the research team. It might also be seen by people who monitor research if there was an audit.

We will do our best to protect the confidentiality of your information. If study information is shared outside KUMC, it will have your name and other direct identifiers removed. It is possible that information shared outside KUMC might be released by others and no longer protected by HIPAA laws. Removing direct identifiers will lessen this risk.

If you want to cancel your permission to use your health information, please write to Susan Jackson. The mailing address is: Department of Hearing and Speech, Mailstop 3039, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160. If you cancel permission to use your health information, we will not gather any new information about you; however, we may use and share information that was gathered before we received your cancellation.

Consent

Please talk to the research team if you have any questions about joining the study. If you have questions about the rights of research participants, you may contact the KUMC Institutional Review Board at (913) 588-1240 or IRBhelp@kumc.edu.

If you agree to join, please sign and date below. You will receive a signed copy of this form.

Printed name: _____

Signature: _____ Date _____

Appendix F: (Surrogate Consent Form)

RESEARCH CONSENT FORM-SURROGATE

Test Re-test Reliability of the Western Aphasia Battery - Revised

Susan T Jackson
(913) 588-5937

As a relative or other individual who is making decisions on behalf of a person with aphasia, you are being asked to approve his or her participation in a research study being done at the University of Kansas Medical Center (KUMC) by Susan Jackson. Being in this study is optional. The potential participant (the person for whom you are making decisions) does not have to participate in this research study. You can decide that you want the potential participant to stop being in the study at any time. Regardless of your decision, the potential participant will still get the same care from his or her health care team.

Why is this study being done?
The purpose of this study is to find out whether people with aphasia perform similarly on the Western Aphasia Battery-Revised when they are given the test on two different occasions.

What is the potential participant being asked to do?
This study involves two visits to the Hearing and Speech department at the KU Medical Center, or we can test the potential participant in his or her home or in a public space (library, place of worship, senior center) if you would prefer. Participation in the study will have two parts:

Part 1
This part will take about 2 hours. We will begin the study by asking the potential participant questions about him/herself and about his or her current health (for example: address, date of birth, date of stroke, handedness, medications). If the potential participant is not able to answer these questions, we will ask you to answer them. We will test the potential participant's vision and hearing. The potential participant will be given some aphasia tests that will include naming objects, repeating words and sentences, answering questions, following some directions, reading, writing, drawing some pictures, acting out how to do certain things, and choosing the best option to complete a visual design.

Part 2
This part will take about 1.5. hours. Part 2 will take place 2-3 weeks after Part 1. The potential participant will do the same aphasia tests that he or she did in Part 1 (naming objects, repeating words and sentences, answering questions, following some directions, reading, writing, drawing some pictures, acting out how to do certain things, and choosing the best option to complete a visual design).

Are there risks or discomforts to consider?
There are no known risks associated with this study. The potential participant may get tired or frustrated, but he or she may take a break at any time or stop the session and continue again on another day.

How will confidentiality and privacy be protected?
There is a small risk of loss of confidentiality when personal information is used for research. The potential participant's information will only be used by study team members and approved researchers. When we write up our results or make presentations, we will not use any names.

We will follow the HIPAA laws about privacy. Study records will include the potential participant's health information and information we collected about the potential participant during the research. We will keep the potential study participant's study information indefinitely. The study information will be kept separately from the study participant's name and other personal identifiers. Study information will be shared with members of the research team. It might also be seen by people who monitor research if there was an audit.

We will do our best to protect the confidentiality of the potential participant's information. If study information is shared outside KUMC, it will have the study participant's name and other direct identifiers removed. It is possible that information shared outside KUMC might be released by others and no longer protected by HIPAA laws. Removing direct identifiers will lessen this risk.

If you want to cancel your permission to use the potential participant's health information, please write to Susan Jackson. The mailing address is: Department of Hearing and Speech, Mailstop 3039, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160. If you cancel permission to use the potential participant's health information, we will not gather any new information about the potential participant; however, we may use and share information that was gathered before we received your cancellation.

Consent
Please talk to the research team if you have any questions about the potential participant joining the study. If you have questions about the rights of research participants, you may contact the KUMC Institutional Review Board at (913) 588-1240 or IRBhelp@kumc.edu.

If you agree to that you would like the potential participant to join, please sign and date below. You will receive a signed copy of this form.

Printed name: _____

Signature: _____ Date _____

Name of Potential Participant: _____

Appendix G: (Assent Form)

ASSENT FORM
Test Re-test Reliability of the Western Aphasia Battery - Revised

Susan T Jackson

(913) 588-5937

I have aphasia. I am being asked to be in a research project.

I will be asked to name objects, answer some questions, follow some directions, read some words and sentences, and do some other tasks with words.

If I sign my name to the line, it means that I want to be part of the research. I know that I do not have to do it and that I can stop being in the research at any time even if I signed. If I want to stop, all I have to do is tell a family member, one of the people who take care of me, or the researcher from the University of Kansas.

_____
Signature of Subject

Appendix H. (Individual Data)

Subject #	Age	Months post onset	Years of Education	Handedness
1	67	112	13	Right
2	36	35	18	Right
3	65	19	12	Left
4	71	48	18	Right
5	69	234	21	Right
6	72	24	21	Right
7	76	23	12	Right
8	69	8	18	Right
9	75	100	16	Right
10	45	48	19	Right
Subject #	Race/Ethnicity	Gender	Aphasia Type-Time 1	Aphasia Type-Time 2
1	Black/African American	Male	Anomic	Anomic
2	White/Caucasian	Female	Anomic	Anomic
3	White/Caucasian	Male	Broca's	Conduction
4	White/Caucasian	Male	Broca's	Broca's
5	White/Caucasian	Male	Anomic	Anomic
6	White/Caucasian	Male	Anomic	Anomic
7	White/Caucasian	Male	Anomic	Conduction
8	White/Caucasian	Male	Conduction	Conduction
9	White/Caucasian	Male	Conduction	Conduction
10	White/Caucasian	Female	Broca's	Broca's
Subject #	Aphasia Quotient-1	Aphasia Quotient-2	Language Quotient- 1	Language Quotient-2
1	87.00	87.70	76.95	84.40
2	87.40	87.80	75.80	75.10
3	73.70	77.90	75.60	75.50
4	46.10	51.30	45.60	54.30
5	82.20	82.40	87.00	87.90
6	87.50	89.00	87.20	84.00
7	76.70	70.60	63.90	59.70
8	49.50	43.40	58.90	55.90
9	65.80	65.90	71.70	64.80
10	54.30	48.90	48.10	42.75

Subject #	Cortical Quotient-1	Cortical Quotient-2	Spontaneous Speech-1	Spontaneous Speech-2
1	82.25	84.40	18	17
2	83.10	82.35	17	18
3	76.75	83.85	12	15
4	53.55	61.98	9	8
5	Missing data	Missing data	15	15
6	88.20	87.52	17	17
7	68.70	65.40	16	13
8	62.45	57.38	13	13
9	72.42	70.53	17	16
10	62.83	Missing data	11	8
Subject #	Auditory Verbal Comprehension- 1	Auditory Verbal Comprehension-2	Repetition-1	Repetititon-2
1	162	177	90	94
2	198	186	86	84
3	191	177	67	66
4	145	187	40	50
5	192	190	73	74
6	181	188	88	90
7	173	168	70	68
8	181	142	21	15
9	158	165	34	32
10	149	169	50	40
Subject #	Naming- 1	Naming- 2	Reading-1	Reading-2
1	84	86	95	90
2	82	82	56	56
3	86	85	82	67
4	28	33	51	64
5	92	93	100	98
6	89	91	90	87
7	67	71	69	66
8	6	1	72	80
9	46	55	75	60
10	37	40	34	29

Subject #	Writing- 1	Writing- 2	Apraxia-1	Apraxia-2
1	63.5	68.5	55.0	54.0
2	55.0	53.5	60.0	60.0
3	64.0	71.5	56.0	56.0
4	25.5	32.5	54.0	59.0
5	83.0	88.0	60.0	60.0
6	82.0	63.5	57.0	58.0
7	15.5	14.0	54.0	56.0
8	53.5	55.5	54.0	50.0
9	79.5	58.0	46.0	53.0
10	33.5	40.5	56.0	Missing data
Subject #	Constructional, Visuospatial & Calculation-1	Constructional, Visuospatial & Calculation -2	Supplemental Reading-1	Supplemental Reading-2
1	55.0	68.5	9	14
2	84.0	82.0	5	4
3	64.5	62.0	10	11
4	66.00	75.00	6	9
5	Missing data	Missing data	9	11
6	87.0	89.0	15	16
7	42.5	44.0	8	8
8	71.0	67.0	0	0
9	85.0	87.0	12	6
10	88.0	Missing data	0	1
Subject #	Supplemental Writing-1	Supplemental Writing-2		
1	5	8		
2	1	1		
3	2	1		
4	0	4		
5	5	7		
6	7	10		
7	0	0		
8	0	4		
9	0	0		
10	0	0		