AJSLP

Research Article

Assessing the Amount of Spontaneous Real-World Spoken Language in Aphasia: Validation of Two Methods

Michelle M. Haddad,[a] Edward Taub,[a] Gitendra Uswatte,[a,b] Margaret L. Johnson,[c] Victor W. Mark,[a,d,e] Ameen Barghi,[f] Ezekiel Byrom,[a] Xiaohua Zhou,[e] and Christina M. Rodriguez[a]

Purpose: The purpose of this article is to present the results of a study evaluating the psychometric properties of 2 new measures that exclusively assess the amount of real-world spoken language in patients with aphasia.
Method: Forty individuals with aphasia were evaluated on several measures of spoken language in real-world settings. The Verbal Activity Log (VAL; Johnson et al., 2014) has participants, aided by caregivers, indicate current amount and quality of real-world spoken language compared with before stroke. In addition, digital voice recorders objectively measured the amount of real-world spoken language. The Communicative Effectiveness Index (Lomas et al., 1989), a previously validated measure of functional communication, was used as a comparison measure. Nineteen participants received follow-up assessment ≥ 3 weeks later.
Results: Validity was supported by Pearson correlations between spoken language recordings and the VAL, $r(38) = .70$, $p < .001$. Likewise, correlation with the Communicative Effectiveness Index was strong, $r(38) = .73$, $p < .001$. Test–retest reliability for both VAL and audio recording was high, with intraclass correlations ≥ .96 and .90, respectively.
Conclusions: These results present preliminary evidence for the reliability and validity of the VAL and spoken language recording for assessment of the amount of real-world spoken language in aphasia. As a simple patient-reported outcome, the VAL may assist diverse therapies for aphasia.

Improving spoken language participation in real-life situations is an important goal of many aphasia rehabilitation approaches (Granger, 1998). However, in many studies, there is no focus on assessing this specific aspect of language per se (as opposed to other nonspoken language communicative modalities; Eadie et al., 2006; World Health Organization, 2001). Previous studies have found evidence that impairments in spoken language limit participation in activities of daily living (Baylor et al., 2013). Improvement of real-world spontaneous spoken language is the primary goal of Constraint-Induced Aphasia Therapy II (CIAT II), a novel intervention that has shown promise in a pilot study (Johnson et al., 2014). Assessing spontaneous spoken language in life situations, therefore, is an important indicator of therapeutic success for CIAT II, as well as other therapies targeted toward improving real-world spoken language participation. However, most current assessment in this area either focuses on spoken language performance in the clinic or laboratory, following prompting from an examiner, or evaluates general communication efficacy during everyday activities, without emphasis on spoken language characteristics themselves. This is of concern because the amount of spoken language when prompted and trained by a therapist in a treatment setting may be considerably greater and of better quality than the amount and quality of spoken language actually exhibited in real-life situations where language is either not prompted or the surroundings are stressful for the individual (Croteau & Le Dorze, 2006; Croteau, Vychytil, Larfeuil, & Le Dorze, 2004). Therefore, standard language assessments used in the laboratory or clinic may overestimate a patient's actual spontaneous spoken language use in the life situation. In addition, examiner-prompted assessments do not take into account patient report, and thus do

[a]Department of Psychology, University of Alabama at Birmingham
[b]Department of Physical Therapy, University of Alabama at Birmingham
[c]Department of Communication Science and Disorders, Samford University, Birmingham, AL
[d]Department of Neurology, University of Alabama at Birmingham
[e]Department of Physical Medicine and Rehabilitation, University of Alabama at Birmingham
[f]Department of Medicine, Harvard Medical School, Boston, MA
Correspondence to Edward Taub: etaub@uab.edu

Editor: Krista Wilkinson
Associate Editor: Daniel Kempler
Received August 5, 2015
Revision received May 20, 2016
Accepted September 22, 2016
https://doi.org/10.1044/2016_AJSLP-15-0115

Disclosure: The authors have declared that no competing interests existed at the time of publication.

not utilize the unique perspective of the person who has aphasia (Brown, 2010).

There are several measures that evaluate functional communication in daily living activities (Worral, McCooey, Davidson, Larkins, & Hickson, 2002). However, each measure has characteristics that make it desirable to develop a new instrument to accomplish this purpose. The Communicative Effectiveness Index (CETI) is an inventory of communicative behavior, including both spoken language and nonspoken language elements, during commonly encountered daily activities; it is completed by the caregiver rather than the patient (Lomas et al., 1989). The American Speech-Language-Hearing Association Functional Assessment of Communication Skills (ASHA-FACS; Frattali, Thompson, Holland, Wohl, & Ferketic, 1995) and the Communication Activities of Daily Living Scale, 2nd Edition (Holland, Frattali, & Fromm, 1998) assess communication disability in social communication and daily living skills (Worral et al., 2002). However, these tests use clinician observation in the treatment setting or in the life situation, rather than patient report, thereby reducing their real-world utility for many purposes because the presence of a clinician is likely to serve as a cue, based on treatment setting experience, to improve amount or quality of spoken language performance beyond what might be exhibited spontaneously. The Communicative Participation Item Bank (CPIB) is a patient-reported outcome designed to assess impairment of participation in spoken language activities owing to a variety of disorders (Baylor et al., 2013). However, completion of the CPIB requires reading and writing by the patient, and because difficulties with those skills are common concomitants of spoken language deficit in poststroke aphasia, its accuracy may be questionable in more impaired aphasia patients. In addition, the CPIB items are phrased in terms of interference of a spoken language–related condition with activities of daily living, rather than in terms of the spoken language characteristics themselves, which may be of primary concern for a particular investigation. Moreover, the CPIB instructs the patient to consider all aspects of a spoken language–related situation, including other health and environmental conditions. Therefore, it may not be specific enough for a direct assessment of spontaneous spoken language in real-life situations per se. Finally, it was developed for and validated with individuals with conditions in which dysarthria is a major problem, and not on individuals with poststroke aphasia alone.

For these reasons, our laboratory developed a new patient-centered measure called the Verbal Activity Log (VAL) and an objective measure using a spoken language recorder. This was undertaken in conjunction with the development of CIAT II because its main objective is to improve spontaneous spoken language in real-life situations, and a test instrument was not available to specifically quantify this aspect of communication (Johnson et al., 2014). The VAL may also be of value in quantifying the effect of other treatments for which improving spontaneous real-world spoken language is a major concern, such as surgical interventions (Wilson et al., 2015) and cranial stimulation (Shah, Szaflarski, Allendorfer, & Hamilton, 2013). The VAL is a structured interview that compares the current amount and quality of spoken language in a variety of behaviorally distinct real-life situations with these characteristics prior to disease onset. Communication by means of gesturing, nonverbal vocalization, or writing, does not affect the test score. The items on the VAL were chosen through a combination of literature review, expert judgment, and patient report of common challenges experienced by persons with aphasia. The second measure studied here, the audio recording measure, was originally developed for assessing the validity of the Amount scale of the VAL. However, it has the advantage of being unequivocally objective, which would be a marked advantage for certain experimental purposes. However, the speed and ease of administering the VAL would make it preferable for use in a clinical setting. The current study is a preliminary evaluation of the psychometric properties of the VAL as well as the objective measure of amount of spoken language from audio recording in real-life situations in a sample of individuals with poststroke aphasia. Development and psychometric evaluation of the VAL follow classical test theory (Allen & Yen, 1979).

Methods

Participants

Participants were recruited through the Speech Language Pathology programs at the University of Alabama at Birmingham and the University of Montevallo, Montevallo, Alabama. They were included in the study if they met the following criteria: (a) nonfluent or fluent aphasia with ability to comprehend and comply with instructions during the screening interview (mean Western Aphasia Battery Revised [WAB-R; Shewan & Kertesz, 1980] aphasia quotient was 72.1, range 11–100; for WAB-R subtest scores see the lower half of Table 1), (b) mild to severe aphasia secondary to stroke as reported by the referring therapist or medical record review, (c) availability of a caregiver for in-laboratory procedures, (d) English spoken as a first language, and (e) ability to comprehend and comply with instructions during the screening interview and other preliminary interactions. Individuals who had taken part in previous communication studies by this laboratory were excluded. In addition, individuals who were less than 6 months poststroke were not given follow-up testing to avoid confounding evaluation of test–retest reliability with the effects of spontaneous recovery. In total, 47 participants were recruited for the study, 40 completed the initial testing, and 19 received repeat testing three or more weeks after the initial session. Figure 1 summarizes the sample characteristics.

Individuals who retested were significantly more chronic than those who did not. However, this difference is expected given that individuals less than 6 months poststroke were excluded from retesting. There were no other significant differences between participants who completed retesting and those who did not. Table 1 summarizes the participants' characteristics.

Table 1. Demographic, stroke-related, and spoken language–related characteristics of all participants, including those who returned for retest and those who did not.

Characteristic	All (*N* = 40)	Test–retest (*n* = 19)	Not retested (*n* = 21)
Demographic			
Age (*M* ± *SD*)	55.9 ± 16.3	58.7 ± 12.4	53.3 ± 19.1
Female, no. (%)	21 (52.5)	9 (47.4)	12 (57.1)
European American, no. (%)	31 (77.5)	14 (73.7)	17 (81.0)
Stroke related			
Primarily left insult, no. (%)	33 (82.5)	14 (73.7)	19 (90.5)
Chronicity, years (*M* ± *SD*)	4.7 ± 5.0	7.34± 5.8	2.1 ± 2.2
Communication related			
CETI score (*M* ± *SD*)	57.3 ± 17.9	59.3 ± 14.3	55.4 ± 20.8
WAB-R aphasia quotient (*M* ± *SD*)	72.1 ± 26.7	71.5 ±31.8	72.6 ± 22.4
Spontaneous speech	14.2 ± 5.8	14.7 ± 6.5	13.8 ± 5.2
Auditory verbal comprehension	8.3 ± 2	8.1 ± 2.3	8.4 ± 1.7
Repetition	6.9 ± 3.3	6.9 ± 3.8	6.9 ± 2.8
Naming and word finding	7.1 ± 3.1	7.0 ± 3.5	7.2 ± 2.7
VAL Amount scale (*M* ± *SD*)	2.3 ± 1.1	2.3 ± 1.2	2.3 ± 1.1
Spoken language %	4.9 ± 5	4.9 ± 5.4	4.8 ± 4.8

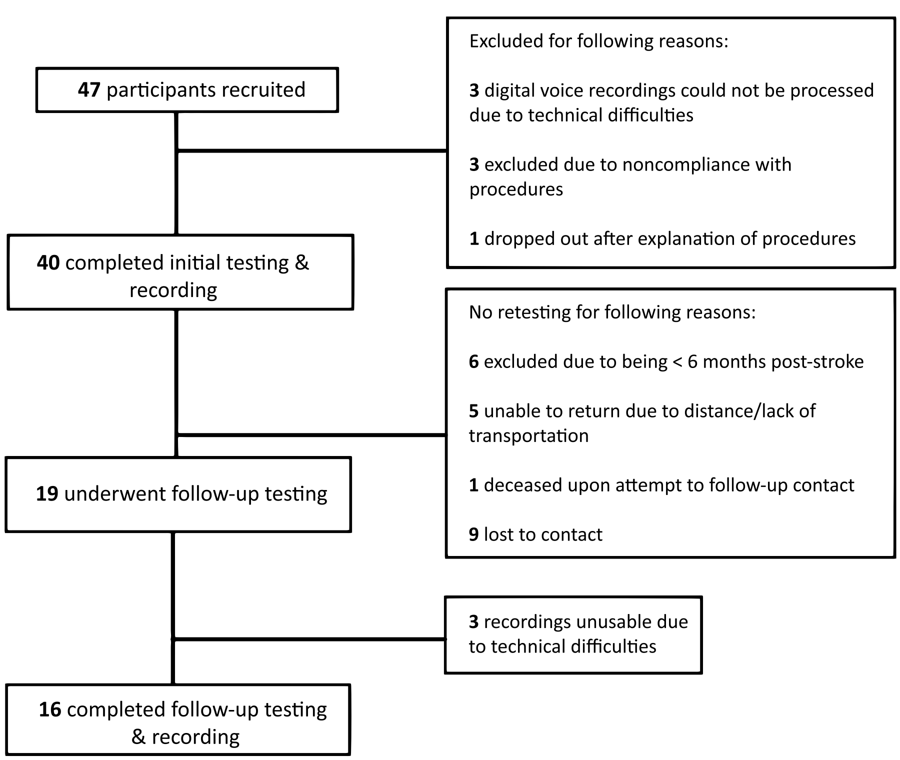
Measures and Apparatus

VAL

This structured interview is designed to assess spontaneous spoken language participation in real-life situations. The items on the VAL were developed by two senior speech-language pathologists and two psychologists with expertise in both rehabilitation and development of outcome measures for behaviorally based interventions. Content was chosen on the basis of literature review and patient report of common challenges experienced by persons with aphasia. Current instruments were extensively reviewed, and the VAL was designed to address gaps in assessment measures currently available to the field. The VAL was initially piloted with several patients with aphasia prior to the current, formal validation study.

The VAL is administered to the patient and caregiver jointly. During test administration, the patient serves as the primary source of response with the caregiver verifying the report. The protocol in the event of a disagreement is to

Figure 1. Study enrollment and completion.



have the examiner resolve the disagreement after asking the participant to demonstrate the item, although in the present study agreement was achieved by discussion between participant and caregiver in each case. Patients were asked to rate the amount and quality of their spoken language during 12 commonly encountered daily activities (see Appendix A). Two of the original items have been eliminated following factor analysis (see Results). The rating scales ask the participant to compare amount and quality of spoken language during the previous week to spoken language before illness onset. Ratings are made on a Likert scale from 0 to 5 that expresses current spoken language as a percentage relative to prestroke (e.g., a rating of 3 on the Amount scale means that participants were using half as much spoken language after stroke as they did before stroke.) Each whole-number score is paired with a brief description (see Appendix A). Ratings in half steps are permitted. The rating scale and anchoring definitions are presented in 24-point font in front of the patient, who may respond either verbally or nonverbally by pointing, depending on ability and comfort level. Each response is verified and probed by scripted questions asked by the examiner to improve accuracy and confirm the response. In addition, the examiner may ask the participant to demonstrate an item if further rating clarification is needed. Patients are asked first to rate their amount of spoken language for each item; then the Quality scale is administered. Some of the items are about activities that vary mainly in context (e.g., spoken language during a meal, or spoken language to order in a restaurant) or are separate activities within the same context (e.g., answering the phone, making a phone call). However, these activities are quite different from a behavioral point of view for a person with aphasia. They pose different challenges to whether such a person will speak or not in real-life situations.

CETI

This instrument is completed by a caregiver and measures verbal and nonverbal functional communication in 16 activities of daily living (Lomas et al., 1989). It is a reliable and valid measure of functional communication (Bakheit, Carrington, Griffiths, & Searle, 2005). It should be noted that the CETI includes items addressing nonverbal communicative behaviors and thus assesses a related but different construct than the VAL.

WAB-R

This language assessment measures both receptive and expressive language produced by a patient on request. It was used in this study to characterize participants' language deficits. The WAB-R is an established measure of language performance made on request in the treatment setting (Shewan & Kertesz, 1980).

Real-World Recording

Olympus VN-8100PC Digital Voice Recorders (Olympus Corporation of the Americas, Center Valley, PA) were used to sample real-world spoken language. The recorder was worn by the participant in a small case either inside a chest pocket or pinned to the shirt in the same area. Participants were instructed to wear the recorder for a minimum of 8 waking hours while going about their daily activities, usually on the day of testing and the following day. The maximum recording time was 23 hours, which was the recorder's capacity. Participants were instructed not to selectively activate or deactivate the recorder unless necessary for privacy or safety (e.g., when in the bathroom). Recordings were processed using Sony SoundForge 9.0 (Sony Creative Software, Inc., Middleton, WI).

Procedures

Participants and caregivers attended each testing session together. All participants provided informed consent, in some cases with the aid of the caregiver. At the initial testing session, the participant and caregiver were administered the VAL as soon as consent was obtained. Next, the caregiver was asked to fill out the CETI while the participant completed the WAB-R with the examiner. The participant and caregiver were then shown how to use the spoken language recorder and sent home to complete the real-world portion of the study for the remaining waking hours of that day and the following day if needed to reach at least 8 hr of recording time. Participants who underwent retesting completed the VAL and the digital voice recording a second time, 21 days or more after the initial testing session (Mdn = 56 days, interquartile range = 28–315 days). All procedures were approved by the University of Alabama at Birmingham Institutional Review Board.

Data Reduction

Prior to processing each recording sample, the first 15 min after device activation were removed because this encompassed the period when each participant was leaving the laboratory and the therapists were still present; they could therefore serve as learned cues to spoken language optimization. The remaining sample was divided into 15-min segments from which 8 segments were randomly selected to calculate the participant's real-world spoken language time. This was accomplished by assigning a randomly generated number to each segment and selecting the eight segments with the lowest numbers.

Data Processing

The targeted content for sound editing in this study was spoken language, defined as meaningful, intelligible, and complete words produced by the participant. Utterances that did not fit this definition were excluded from the SoundForge edited sample. More detailed descriptions and examples of possible content are provided in Table 2. The final product was a recording file that contained only the participant's spoken language. The duration of this recording divided by 120 min (i.e., the eight 15-min segments randomly selected for analysis) was defined as the spoken language percentage and used as our objective measure of amount of spoken language in the life situation.

Real-world sound samples were processed and analyzed by two independent raters, who first practiced editing samples from AphasiaBank (MacWhinney, Fromm, Forbes,

Table 2. Recorded sample editing criteria.

Category	Criteria
1 (Keep)	• Complete, intelligible words, allowing for dysarthria or phonemic paraphasias, e.g., "rinking" for "thinking" or "shindow" for "window." • Repeated words that are complete and intelligible. • Fillers that are words (e.g., counting, or repeated phrases such as "for sure"). • Exclamations that are words (e.g., okay, whoops). • Words produced while laughing (but not laughter alone). • First and last consonants that are part of a complete word, even if there is some delay in production. • Sounds after any pause < .5 s will be counted as part of the preceding word.
2 (Delete)	• Any background noises or other voices in the sample. • Any sound produced by the participant that is not a word (e.g., sound effects, biological noises). • Fillers that are not words (e.g., um, uh). • Singing. • Exclamations that are not words (e.g., uh-oh, um, huh, mmhmm, grunting). • Laughter that does not include words. • Parts of words (e.g., "Thursd–" or "Decem…"). • Repeated word-initial consonants that are not part of the completed word (e.g., "p-p-p-cup"). • Silence of any duration immediately preceding and immediately following any nonlanguage sound. • Any pause ≥ .5 s.

& Holland, 2011), an online research database. The practice samples consisted of patients with aphasia responding to open-ended prompts posed by an examiner. The intraclass correlation was high (0.99) during both rater training and study sample processing, indicating that excellent interrater reliability had been achieved (Lexell & Downham, 2005; see Appendix B for training procedures).

Statistical Analysis

To evaluate the psychometric properties of the VAL, an exploratory factor analysis was performed to identify its factor structure and to determine whether items needed to be dropped or grouped as separate scales to derive one or more unidimensional scales. Factors were identified using a criterion of retaining factors with eigenvalues above 1, and an item was retained with factor loadings above .6 on its respective factor but not more than .3 on another factor (cf. Matsunaga, 2010). On the basis of retained items, internal consistency was estimated. A confirmatory factor analysis with retained items using maximum likelihood estimation was performed with Amos 21 to confirm the factor structure and provide fit indices; Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values above .90 are indicative of good fit as are root-mean-square error of approximation (RMSEA) values below .08 (Byrne, 2001). Stability of the VAL was evaluated by paired *t* tests of Test 1 and 2 scores; a significant change would be interpreted as a lack of test stability. Test–retest reliability was evaluated by Type 2 intraclass correlations (ICCs; Shrout & Fleiss, 1979) between Tests 1 and 2. ICCs were characterized as poor (≤ .40), fair to good (.40–.75), or excellent (≥ .75; Lexell & Downham, 2005). Validity coefficients were characterized as weak, (r = .10–.29), moderate (r = .30–.49), and strong ($r \geq 0.50$; Cohen, 1988). Convergent validity was evaluated by Pearson correlation of VAL scores with objective spoken language percentage from the recordings. Concurrent validity of the VAL and spoken language recording for measuring amount of speech in real-life situations was evaluated by comparing scores from them to scores on the CETI (Lomas et al., 1989), using Pearson correlations. Because the correlation between responses on the Amount and Quality scales was very high, $r(38) > .95$, indicating that information from the scales is redundant, we report analysis of only the Amount scale here. Statistics for the Quality scale may be found in Appendix C.

Results

Item Analysis

Exploratory factor analysis of the 12-item VAL identified two factors with eigenvalues above 1, wherein two items ("Did you introduce yourself to others?" and "Did you verbally order for yourself in a restaurant?") did not load onto the primary dimension at the .6 criterion despite different orthogonal rotation efforts; those two items also demonstrated the lowest item-total correlations (.54 and .37, respectively). Because it was not possible to identify a meaningful construct that would be represented uniquely by these two items, the decision was made to drop these two items from the test. The exploratory factor analysis without these two items supported a single dimension with an eigenvalue above 1, explaining 65.66% of the variance, with factor loadings for items at .72 and higher. Internal consistency of this reduced measure was α = .94, with all item-total correlations ranging from .81 to .84. In addition, the confirmatory factor analysis with the remaining 10 VAL items indicated acceptable fit of a unidimensional model (CFI = .98, TLI = .97, RMSEA = .07). Table 3 summarizes the item characteristics.

Test–Retest Reliability and Validity

Test–retest reliability and stability were supported for the VAL Amount scale and the spoken language recording measure (see Table 4). Convergent validity for measuring

Table 3. Amount scale item characteristics.

Item number	Item-total correlation	*M*	*SD*	Skewness	Endorsement frequency					
					0	1	2	3	4	5
1	.71	3.2	1.1	1.30	4	11	7	13	2	2
2	.70	3	1.9	−1.10	13	9	2	7	5	4
3	.79	3.1	1.7	−0.39	10	5	9	5	8	3
4	.75	2	1.6	0.00	19	3	7	6	3	2
5	.72	2.8	1.1	−1.30	4	3	7	16	7	3
6	.78	2.2	2.3	0.23	6	5	4	10	8	7
7	.83	3.8	1.3	−0.54	3	1	5	11	14	6
8	.74	2.6	1.3	0.17	3	2	10	7	9	9
9	.78	3.3	1.7	−0.24	7	3	11	9	5	5
10	.71	3.2	1.1	1.30	5	7	3	15	8	2

amount of speech in real-life situations was supported for the VAL and spoken language recording measure. The correlation between the VAL Amount scale scores and spoken language percentage from the digital recordings was strong, $r(38) = .70$, $p < .001$. Concurrent validity of the VAL and CETI was also supported; the correlation between the VAL Amount scale and caregiver report scores on the CETI was strong, $r(38) = .73$, $p < .001$, whereas spoken language percentage was moderately correlated with CETI scores, $r(38) = .48$, $p < .01$. Descriptive statistics for these measures are presented in Table 1. The psychometric properties of the VAL Quality scale are reported in Appendix C.

Discussion

The purpose of the Amount scale of the VAL and the audio recording measure is to evaluate the amount of spontaneous real-world spoken language. In particular, they are designed specifically to measure spoken language in isolation from other compensatory forms of communication. This may be the primary objective of a therapeutic intervention, which is the case for CIAT II. When comparing the potential utility of the two instruments, the objective nature of the audio recording measure would be of primary value in experimental situations. However, use of the measure is time-consuming and complex. Moreover, because the present study shows that the VAL is strongly correlated with the objective audio recording measure, there would seem to be little loss in accuracy in using it rather than the audio recording measure and much to be gained for clinical purposes in the greater ease and reduced time to use it. Moreover, the VAL has the additional advantage of pinpointing the behavioral situations that are the greatest problem for spoken language use by patients with nonfluent aphasia and that would be of greatest value for attention in treatment.

The wording of each VAL item and of the general instructions is specific to spoken language (e.g., specifying delivery of a phone message verbally). This makes the VAL distinct from established tools for measuring communication outside the treatment setting, such as the CETI, ASHA-FACS, and CPIB. It does not make the VAL superior to these instruments, but it does make it different for the measurement of amount of spoken language per se, and of potential value for accomplishing the objective for which it was explicitly designed. The CETI and the ASHA-FACS include explicit items that are about nonverbal communication modalities, as well as items that could be interpreted as referring to either spoken language or other modalities. The CPIB, although designed as a spoken language–focused measure, has items that could be interpreted otherwise (e.g., "Taking a phone message" which could be done in writing or by spoken language). Moreover, the CPIB general directions instruct the participant to consider "all issues that may affect how you communicate…including…other health conditions, or features of the environment" (Baylor et al., 2013, p. 1205). This might be interpreted as including compensatory forms of communication, such as gesture, in accomplishing the tasks being assessed, and therefore overestimate the person's actual spoken language use. Prior research has indicated that including multiple communicative modalities in a single instrument may make scores difficult to interpret (Baylor et al., 2013).

Table 4. Test–retest data.

Measure	Test 1 (*M* ± *SD*)	Test 2 (*M* ± *SD*)	Change (*M* ± *SD*)	Stability (*t*)[a]	Reliability (ICC)[b]
VAL Amount mean ($n = 19$)	2.4 ± 1.3	2.6 ± 1.1	+0.2 ± 0.6	1.90, *ns*	.96
Spoken language (%), (recording) mean ($n = 16$)	4.3 ± 5.3	4.0 ± 4.9	−0.3 ± 2.1	0.47, *ns*	.90

[a]Absence of significant change from Test 1 to Test 2 indicates good stability. [b]ICCs are characterized as poor (≤ .40), fair to good (.40–.75), or excellent (≥ .75).

Another feature that makes the VAL distinct is its joint administration to patient and caregiver. This is different from the CETI and the ASHA-FACS, which are completed by observers. The CPIB is a patient-reported outcome only. Joint administration capitalizes on the strengths of both assessment types by allowing the patient to self-report current spoken language characteristics, whereas the caregiver's verification may help improve accuracy of the assessment. The examiner may further ask the patient to demonstrate an item to clarify ratings. The patient is also provided with a large-print version of the VAL rating scale that is read aloud by the examiner. Responses may be given verbally or by pointing to a number on the scale. The combined verbal and visual content of test administration makes the VAL adaptable for patients with varying levels of spoken language difficulties, and also eliminates the need for reading comprehension, which is required by the CPIB.

The items on the VAL were chosen to encompass a variety of commonly encountered situations, which are behaviorally distinct from one another and in which persons with aphasia might show a differential tendency to speak or not speak. Many of the items on the VAL are similar to those on existing measures of functional communication, although they are different in important respects, because the VAL serves a different purpose than these previously established tools. Table 5 provides a comparison of items on the VAL to those on the CETI, CPIB, and ASHA-FACS. The VAL, CETI, and CPIB all include items about communicating with familiar people and other items about communicating with strangers. All four measures have items about talking on the phone, but the VAL differentiates between answering the phone and simply communicating on the phone in any context. This distinction is behaviorally important because persons with aphasia may participate in the latter behavior but not the former. Regarding phone use, both the VAL and the CPIB include items about taking a message. However, the VAL specifies delivering the message verbally, whereas the CPIB only asks about *taking* the message, without specifying modality or delivery to another person. Only the VAL and the CPIB include items that specifically address speaking in the community, which is also likely to be more challenging than speaking within the familiar home environment.The VAL also includes one item that is not reflected in any of the three established measures (i.e., use of full sentences), which represents higher-level spoken language function.

Despite the broad range of tasks evaluated by the VAL, factor analysis suggests that the 10 retained items measure a single underlying construct. The retest reliability coefficient was strong for the VAL Amount scale, and the Amount scale was strongly correlated with caregiver report scores on the CETI as well as real-world spoken language percentage. By comparison, the CETI, an established measure of real-world communication, produced only a moderate correlation with spoken language percentage obtained from recordings.

The VAL Amount scale validity coefficient compares favorably with self-report measures of motor performance when validated against objective metrics. For example, a recent meta-analysis comparing motor self-report measures to objective activity monitoring found median validity coefficients of .30–.39 (Helmerhorst, Brage, Warren, Besson, & Ekelund, 2012). The VAL scale validity coefficient, $r = .70$, is comparable to the best-performing motor questionnaires examined by the meta-analysis. To our knowledge, no other spoken language participation measure has been validated by comparison to objective real-world recording. This study did not attempt to validate the Quality of Speech scale against an objective measure. However, the data presented in Appendix C suggest that the other psychometric properties of the scale were strong.

As noted, the VAL was administered jointly to the participant and caregiver in the present study. However, this instrument could be administered to persons without a caregiver. Participant-only administration is the standard mode for the Motor Activity Log (Taub et. al., 1993) on which the VAL is modeled. The accuracy of responses with this type of administration for the Motor Activity Log is indicated by the strong correlation of this real-world motor measure with accelerometry data obtained for the same time period being evaluated (Uswatte et. al., 2005, 2006). The VAL originated in the context of research in which CIAT II was developed. During CIAT II, a caregiver was present in all phases of the work. This was taken advantage of to increase the accuracy of the participant's responses. The same mode of administration was retained for the present study so that the psychometric data obtained would apply to the CIAT II work. It would be of value to examine the psychometric properties of the VAL when administered to the participant only in future work with adults with aphasia similar to those studied here, (i.e., with nonfluent aphasia or fluent aphasia without severe impairment in comprehension).

Study Limitations

Quantifying the amount of spoken language as carried out in the present study would not be meaningful in persons with Wernicke's aphasia in terms of communication effectiveness, and therefore this measure in its present form would not be suitable for use with such patients. Because this was not an intervention study, these data cannot provide information on the VAL's sensitivity to change. There was no objective assessment of spoken language quality here. The primary focus of analysis was on spoken language amount. Because this preliminary study supported overall reliability and validity of the VAL for measuring amount of speech in real-life situations, more detailed analyses would be appropriate in future studies with larger sample sizes. Validation studies of the Quality of Speech scale would also be appropriate. In addition, the suitability of the VAL for persons with more than mild impairment in auditory comprehension and memory is unknown because such individuals were absent from this study. Future studies might investigate this question, along with the utility of pictorial supports for test administration for these individuals.

Table 5. Comparison of VAL items with those on existing measures of functional communication.

VAL *"How much/How well…"*	CETI *"Indicate how well the participant performs the task"*	CPIB *"Does your condition interfere with…"*	ASHA-FACS *Clinician or caregiver rates on Communicative Independence and Qualitative Communication scales.*
Did you talk while eating a meal with others you did know?	Having a one-to-one conversation with you.	Talking with people you know.	—
Did you talk to an employee in an office, store, or public place?	Participating in a conversation with strangers.	Communicating while you are out in your community.	—
Did you answer the phone?	—	—	—
Did you take a message from someone on the phone and deliver it to someone verbally?	—	Taking a phone message.	—
Did you tell a story or relay an event to someone?	Getting involved in group conversations that are about him or her.	Talking with family or friends about something you are planning to do with them.	—
Did you use sentences when talking?	—	—	—
Did you start a conversation with your caregiver or others in your family?	Having a spontaneous conversation.	Starting a conversation with someone you know.	Initiates communication with other people.
Did you talk about any of your needs?	Communicating physical problems such as aches and pains.	—	Makes needs or wants known.
Did you make a phone call and talk, or take over a phone call and talk?	—	Making a phone call to get information.	Exchanges information on the phone.
Did you talk with others while in a group?	Being part of a conversation when it is fast and there are a number of people involved.	Communicating with a small group of people.	Participates in group conversation.

Note. Em Dashes indicate that there was no item in comparison instrument equivalent to indicated item in the VAL.

Conclusions

The VAL was designed to fill a gap in current assessment tools, specifically the need to evaluate and target amount of real-world spoken language. Although the VAL was initially designed in parallel with the CIAT II intervention and was validated here in poststroke patients with aphasia, it may be appropriate for broader clinical and research application. The VAL would be of use to anyone wishing to specifically study amount of spoken language use in isolation from other communication modalities. Moreover, in clinical settings, the VAL may be preferable to the recorder measure of speech introduced in this article because it is both less burdensome and provides information that may be of value in focusing treatment efforts.

Acknowledgment

This research article was supported in part by American Heart Association Grant 12PRE1137002. We thank Lyn Turkstra for a critical reading of the article.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks-Cole Publishing Company.

Bakheit, A. M., Carrington, S., Griffiths, S., & Searle, K. (2005). High scores on the Western Aphasia Battery correlate with good functional communication skills (as measured with the Communicative Effectiveness Index) in aphasic stroke patients. *Disability and Rehabilitation, 27*(6), 287–291. https://doi.org/10.1080/09638280400009006

Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research, 56*(4), 1190–1208. https://doi.org/10.1044/1092-4388(2012/12-0140)

Brown, M. (2010). Participation: The insider's perspective. *Archives of Physical Medicine and Rehabilitation, 91*(9, Suppl.), S34–S37. https://doi.org/10.1016/j.apmr.2009.11.030

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

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

Croteau, C., & Le Dorze, G. (2006). Overprotection, "speaking for," and conversational participation: A study of couples with aphasia. *Aphasiology, 20,* 327–336. https://doi.org/10.1080/02687030344000616

Croteau, C., Vychytil, A., Larfeuil, C., & Le Dorze, G. (2004). "Speaking for" behaviours in spouses of people with aphasia: A descriptive study of six couples in an interview situation. *Aphasiology, 18,* 291–312. https://doi.org/10.1080/02687030500475051

Eadie, T. L., Yorkston, K. M., Klasner, E. R., Dudgeon, B. J., Deitz, J. C., Baylor, C. R., . . . Amtmann, D. (2006). Measuring communicative participation: A review of self-report instruments in speech-language pathology. *American Journal of Speech-Language Pathology, 15*(4), 307–320. https://doi.org/10.1044/1058-0360(2006/030)

Frattali, C. M., Thompson, C. M., Holland, A. L., Wohl, C. B., & Ferketic, M. M. (1995). The FACS of life ASHA facs—A functional outcome measure for adults. *Asha, 37*(4), 40–46.

Granger, C. V. (1998). The emerging science of functional assessment: Our tool for outcomes analysis. *Archives of Physical Medicine and Rehabilitation, 79,* 235–240. https://doi.org/10.1016/S0003-9993(98)90000-4

Helmerhorst, H. J., Brage, S., Warren, J., Besson, H., & Ekelund, U. (2012). A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *International Journal of Behavioral Nutrition and Physical Activity, 9*(103), 1–55. https://doi.org/10.1186/1479-5868-9-103

Holland, A., Frattali, C., & Fromm, D. (1998). *Communication activities of daily living*. Austin, TX: Pro Ed.

Johnson, M. L., Taub, E., Harper, L. H., Wade, J. T., Bowman, M. H., Bishop-McKay, S., . . . Uswatte, G. (2014). An enhanced protocol for constraint-induced aphasia therapy II: A case series. *American Journal of Speech-Language Pathology, 23,* 60–72. https://doi.org/10.1044/1058-0360(2013/12-0168)

Lexell, J. E., & Downham, D. Y. (2005). How to assess the reliability of measurements in rehabilitation. *American Journal of Physical Medicine and Rehabilitation, 84,* 719–723. https://doi.org/10.1097/01.phm.0000176452.17771.20

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The Communicative Effectiveness Index: Development and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders, 54*(1), 113–124. https://doi.org/10.1044/jshd.5401.113

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

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don't, and how-to's. *International Journal of Psychological Research, 3*(1), 97–110.

Shah, P. P., Szaflarski, J. P., Allendorfer, J., & Hamilton, R. H. (2013). Induction of neuroplasticity and recovery in post-stroke aphasia by non-invasive brain stimulation. *Frontiers of Human Neuroscience, 7,* 888. https://doi.org/10.3389/fnhum.2013.00888

Shewan, C. M., & Kertesz, A. (1980). Reliability and validity characteristics of the Western Aphasia Battery (WAB). *Journal of Speech and Hearing Disorders, 45,* 308–324. https://doi.org/10.1044/jshd.4503.308

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Taub, E., Miller, N. E., Novack, T. A., Cook, E. W., III, Fleming, W. C., Nepomuceno, C. S., . . . Crago, J. E. (1993). Technique to improve chronic motor deficit after stroke. *Archives of Physical Medicine Rehabililation, 74,* 347–354.

Uswatte, G., Taub, E., Morris, D., Light, K., & Thompson, P. A. (2006). The Motor Activity Log-28: Assessing daily use of the hemiparetic arm after stroke. *Neurology, 67,* 1189–1194.

Uswatte, G., Taub, E., Morris, D., Vignolo, M., & McCulloch, K. (2005). Reliability and validity of the upper-extremity Motor Activity Log-14 for measuring real-world arm use. *Stroke, 36,* 2493–2496.

Wilson, S. M., Lam, D., Babiak, M. C., Perry, D. W., Shih, T., Hess, C. P., . . . Chang, E. F. (2015). Transient aphasias after left hemisphere resective surgery. *Journal of Neurosurgery, 123,* 581–593. https://doi.org/10.3171/2015.4.JNS141962

World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. Geneva, Switzerland: Author.

Worral, L., McCooey, R., Davidson, B., Larkins, B., & Hickson, L. (2002). The validity of functional assessments of communication and the Activity/Participation components of the ICIDH-2: Do they reflect what really happens in real-life? *Journal of Communication Disorders, 35,* 107–137. https://doi.org/10.1016/S0021-9924(02)00060-6

Appendix A

Verbal Activity Log (VAL)

Patients and their caregivers were jointly asked to rate the amount and quality of a patient's spoken language in 12 different situations compared with before stroke. The questions are listed below and are followed by the two rating scales. The manual for administering the instrument may be obtained by contacting the corresponding author.

1. Did you talk while eating a meal with others you did know?
2. Did you talk to an employee in an office, store, or public place (e.g., post office, market)?
3. Did you answer the phone?
4. Did you take a message from someone on the phone and deliver it to someone verbally?
5. Did you tell a story or relay an event to someone?
6. Did you use sentences when talking?
7. Did you start a conversation with your caregiver or others in your family?
8. Did you talk about any of your needs?
9. Did you make a phone call and talk, or take over a phone call and talk?
10. Did you talk with others while in a group?

Note. Items eliminated after factor analysis:
Did you introduce yourself to others?
Did you verbally order for yourself in a restaurant?

Verbal Activity Log (VAL) Scales

Amount Scale
0 – My speech was not used at all for that activity. **(0%)**
.5
1 – Occasionally used speech but only very rarely. **(very rarely used – 10%)**
1.5
2 – Sometimes used speech, but rarely. **(rarely used – 25%)**
2.5
3 – Used speech about half as much as before the stroke. **(half the time – 50%)**
3.5
4 – Used speech almost as much as before the stroke and did not rely on my caregiver very often. **(frequently – 75%)**
4.5
5 – Used speech often as before the stroke. **(100%)**

How Well Scale
0 – My speech was not used at all for that activity. **(0%)**
.5
1 – My speech was used for that activity but was not helpful. **(very poor – 10%)**
1.5
2 – My speech was of some use during that activity but I required help and my speech was slow. **(poor –25%)**
2.5
3 – My speech was used for that activity but it was slow and I only used 2-4 words. **(fair – 50%)**
3.5
4 – My speech was used for that activity and was almost normal but not quite as fast or as accurate as before the stroke. **(almost as good as before stroke – 75%)**
4.5
5 – My speech for that activity was as good as before the stroke. **(100%)**

Appendix B

Rater Training

Real-world spoken language samples were processed and analyzed by two independent raters following the protocol detailed in the Data Reduction section. Raters were first trained on the basic functions of the Sony SoundForge program. Then they practiced editing samples from AphasiaBank, an online interdisciplinary research database. The recordings featured individuals with aphasia completing a variety of spoken language tasks in a laboratory environment.

The trainer compared the length and characteristics of the raters' edited practice samples to troubleshoot any differences in protocol interpretation. At the end of the training period, the two raters independently edited 10 randomly selected 15-min samples from AphasiaBank. The intraclass correlation of the ratings of the two raters was 0.99. The training period was approximately 80 hr.

Appendix C

VAL Quality Scale

Test development and item analysis

Item number	Item-total correlation	*M*	*SD*	Skewness	Endorsement frequency 0	1	2	3	4	5
1	.77	2.3	1.3	−.128	4	8	8	9	10	1
2	.70	1.9	1.6	.185	12	6	5	7	8	1
3	.76	2.3	1.7	−.146	10	4	4	7	12	3
4	.77	1.6	1.8	.522	20	1	3	4	6	3
5	.80	2.4	1.4	−.125	4	5	10	10	8	3
6	.86	2.8	1.6	.435	6	2	6	9	8	9
7	.83	2.9	1.4	−.459	3	3	9	7	11	7
8	.83	3.0	1.5	−.413	3	3	8	6	11	9
9	.85	2.5	1.6	−.280	7	5	5	9	10	4
10	.84	2.5	1.4	−.356	4	4	6	15	8	3

Factor analysis was conducted on the VAL Quality scale using scores from the 10 items that were retained after factor analysis of the VAL Amount scale. Exploratory factor analysis of these Quality scale item scores identified a single factor, which accounted for 71.13% of the variance, with factor loadings from .76 to .89 and $\alpha = .95$. A confirmatory factor analysis also affirmed the unidimensional structure of this scale (CFI = .98, TLI = .96, RMSEA = .09).

Test–retest reliability and validity

Test–retest reliability was supported for the VAL Quality scale, ICC = .98, as was stability, $t(18) = -1.4$, *ns*. Convergent validity for the VAL Quality scale was also supported; the correlation with spoken language percentage from the digital recordings was strong, $r(38) = .60$, $p < .001$. Concurrent validity of the VAL Quality scale was also supported by a strong correlation with the CETI, $r(38) = .73$, $p < .001$.