Validity of and Agreement Between Self- and Surrogate-Reported Communicative Functioning in Persons with Aphasia

While there have been many helpful advances over the past 40 years, barriers to effective measurement of functional communication skills in adults with aphasia remain. First, the ability range effectively measured by current assessments frequently falls below the ability level of many community-dwelling stroke survivors (Frattali, 1992). Second, the burden of assessment associated with most functional communication assessments is high, limiting their use in the current healthcare environment (Worrall, 2001).

These limitations may be addressed through the creation of an item bank (Thissen, Reeve, Bjorner, & Chang, 2007), a set of test items that responds to a unidimensional construct existing on an ordered continuum. The items are calibrated to a common measurement scale, typically using item response theory. The calibration process assigns a difficulty value to each item and links them to a common scale, permitting individual ability estimates derived from different subsets of items to be directly compared. This allows adaptive testing, which can provide precise score estimates with minimal response burden.

This paper reports interim findings in the development of a new self- and surrogate-reported instrument for measuring communicative functioning in persons with aphasia: The Aphasia Communication Outcome Measure (ACOM). In examining the validity of the instrument, we asked the following questions: (1)Do surrogate ratings of item difficulty for the communication activities comprising the scale agree with item difficulty ratings by persons with aphasia (PWA), (2)Do surrogate ratings of person ability agree with PWA ratings of person ability, (3)Do constructs underlying more directly observable behaviors (e.g., writing) yield better agreement than those underlying less observable behaviors (e.g., comprehension), and (4)What is the concurrent validity of the ACOM with a performance-based measure of communicative ability?

METHOD

Participants were 133 PWAs and 133 surrogate respondents. PWAs met the following inclusion criteria: diagnosis of aphasia ≥1 MPO; community dwelling; self-reported normal pre-morbid speech-language function; pre-morbid literacy with English as a first language; negative self-reported history of progressive neurological disease, psychopathology, and substance abuse; ≥0.6 delayed/immediate ratio on ABCD Story Retell (Bayles & Tomoeda, 1993); ≤5 self-reported depressive symptoms on the GDRS-15 (Sheikh & Yesavage, 1986); and BDAE severity rating ≥1. Surrogate respondents met similar criteria, except for diagnosis of aphasia, and reported weekly or more-frequent contact with their respective PWA both prior to and after aphasia onset. Demographic and clinical characteristics of the sample are summarized in Table 1.

The ACOM item pool is comprised of 177 items describing various communication activities. Participants were asked to rate on a 4-point scale how effectively the PWA performs each activity. “Effectively” was defined as “accomplishing what you want to, without help, and without too much time or effort.” Sample items are presented in Table 2. Responses were collected separately from PWAs and surrogates with interviewer-assist. The protocol also included administration of the PICA (Porch, 2001) and a motor speech examination if indicated.

ANALYSIS and RESULTS

Three item subsets were selected based on a priori considerations of content and approximation of unidimensionality tested by factor analysis: verbal expression (n = 44 items), comprehension (n = 25), and writing (n = 11). For each item set, a separate Rasch Rating Scale model (Andrich, 1978) was estimated for the PWA and surrogate data, and items demonstrating poor fit according to information-weighted (infit) and outlier-sensitive (outfit) mean-square (MSQ) fit statistics were iteratively excluded. Retained items (30 verbal expression, 21 comprehension, 11 writing) obtained MSQ values between 0.7 and 1.4 (Wright & Linacre, 1994). Two exceptions were made for writing items that demonstrated slight overfit to the model for PWAs ("write a simple to-do list," infit MSQ=0.64, outfit MSQ=0.58) and surrogates ("write a shopping list," infit MSQ=0.62, outfit MSQ=0.56). There was complete overlap in the 5 items demonstrating the most extreme underfit in the PWA and surrogate samples (Table 3). The reliability of the resulting scales was 0.95, .90, and .80 for the verbal expression, comprehension, and writing scales, respectively.

To address the first research question, correlations between the Rasch item difficulty estimates obtained from the PWA and surrogate samples were computed. The correlations were high in each case: 0.91 for verbal expression, 0.95 for comprehension, and 0.99 for writing. Tests of differential item functioning indicated that two comprehension items ("understand warning signs" and "recognize when people do not understand you") obtained difficulty estimates that were reliably different between PWAs and surrogates. These items were excluded.

Next, the PWA and surrogate data were combined to obtain item difficulty estimates from the full sample. Using these estimates, ability scores in each of the three domains based on PWA and surrogate ratings were obtained for each PWA. To address the second research question, we performed two sets of analyses. First, we conducted paired-sample t-tests to determine whether, as a group, surrogate respondents gave higher or lower ability ratings than PWAs gave themselves. For each of the three domains, the mean difference was <0.03 logits, and the tests were all non-significant (ts<0.25, ps>0.8). Second, to evaluate the reliability of PWA-surrogate differences at the individual level, we compared the scores for each pair using the standard errors provided by the Rasch model to evaluate significance. For the verbal expression, comprehension, and writing scales, 46%, 38%, and 7%, respectively, of the comparisons were significant at $p<0.05$. Next, to address the third research question, we computed correlations between the PWA and surrogate scores for each domain. The correlation for the writing domain (0.80) was significantly higher than the correlation for the verbal domain (0.65, z=2.61, p=0.005), which was non-significantly higher than the correlation for the comprehension domain (0.54, z=1.38, p=0.08).

To address the final research question, we computed correlation coefficients (Table 4) between the ACOM ability scores and PICA modality scores derived from Rasch Rating Scale models of the relevant subtests (verbal: I, IV, IX, XII; comprehension: V-VII, X; and writing: A-D). In every case, the ACOM ability score correlated most strongly with the corresponding PICA modality score. In one case (surrogate ACOM writing score), the difference between the two highest correlations failed to reach significance.

DISCUSSION

These results support the validity of self-reports of communicative functioning by PWAs, and of ACOM measures of self- and surrogate-reported verbal expression, comprehension, and writing ability. The finding that PWA and surrogate ability estimates correlated most highly for the most directly observable domain is consistent with prior comparisons of patient and proxy reports of

physical and psychological functioning (Duncan et al., 2002; Hilari, Owen, & Farrelly, 2007; Williams et al., 2006). The agreement between PWAs and surrogates on item difficulty ratings and the lack of significant differences in person ability estimates at the group level suggest that self- and surrogate-reported communicative functioning may be validly measured using the same items and measurement scales. However, the moderate correlations between PWA and surrogate ability estimates and the finding of a significant proportion of reliable differences in individual cases suggest that they are not interchangeable at the individual level. Data collection is ongoing in the initial field trial of the ACOM items, and future investigations will include more detailed examinations of construct dimensionality. The goal of this project is to develop a calibrated item bank that can serve as a platform for psychometrically rigorous and socially valid computer-adaptive and short-form measurement of self- and surrogate-reported communicative functioning in aphasia.

Table 1. Demographic and clinical characteristics of the study sample, n = 133 persons with aphasia, n =133 surrogates.

	Persons with Aphasia	Surrogates
Age in Years, mean (sd)	60 (14)	60 (13)
Gender, % male	63%	26%
Race		
African American	9%	
Caucasian	80%	
Hispanic	10%	
Native American	1%	
Mixed	1%	
Education		
Primary/Middle School	0%	8%
High School	24%	22%
Some College	34%	35%
College Graduate	20%	18%
Post-Graduate Degree	14%	17%
Missing	9%	0%
Surrogate Relationship with PWA		
Spouse/Partner		58%
Parent		15%
Child		10%
Sibling		10%
Other Relative		1%
Friend		6%
Frequency of Contact with PWA		
Daily		85%
Weekly		15%
Length of Relationship in Years, mean (sd)		38.5 (15)
Marital Status		
Currently Married or Cohabitating	64%	
Divorced or Separated	22%	
Widowed	2%	
Never Married	11%	
Months Post-Onset of Aphasia, median (min-max)	49 (1-507)	
PICA Overall score, median (min-max)	12.37 (7.47-14.21)	
BDAE Severity Rating		
0	0%	
1	21%	
2	12%	
3	22%	
4	28%	

5	7%
Missing	10%
% of PWA with concomitant Motor Speech Disorder	45%

Table 2. Sample item content, item difficulty estimates, and standard deviations for Aphasia Communication Outcome Measure scales of Verbal Expression, Comprehension, and Writing. The items in each domain are scaled separately to have a mean difficulty of 0 and a unit size of 1 logit.

ACOM Domain	Item Difficulty (standard error)	How effectively do you (does your partner)… Item Content
Verbal Exp	0.89 (0.14)	Talk with a group of people?
	0.45 (0.14)	Start a conversation with other people?
SD = 0.54	0.01 (0.14)	Make small talk with neighbors?
	-0.76 (0.14)	Tell people how you feel?
	-1.37 (0.15)	Talk to your closest family member or friend?
Comprehension	1.42 (0.13)	Understand conversation in a noisy place?
	0.95 (0.13)	Understand medical insurance information?
SD = 0.89	-0.09 (0.14)	Read product labels?
	-1.76 (0.16)	Recognize the names of common objects?
Writing	2.30 (0.29)	Take notes during meetings/classes?
	-0.27 (0.24)	Communicate by email?
SD = 1.84	-0.71 (0.20)	Fill out simple forms?
	-3.01 (0.19)	Write your name?

Response Scale	Score	Category Label
	3	Completely
	2	Mostly
	1	Somewhat
	0	Not Very
	0	Do not do, because of communication disorder
	NA/missing	Do not do, for some other reason

Table 3. Item content and information-weighted (infit) and outlier-sensitive (outfit) mean-square values for the five ACOM candidate items demonstrating the poorest fit to the Rasch measurement model in the person with aphasia (PWA) and surrogate (SUR) data.

ACOM Domain	Item Content	Infit MSQ		Outfit MSQ	
		PWA	SUR	PWA	SUR
Verbal Exp	Tell people why you can't talk well	1.67	1.47	1.76	1.49
	Get your point across when you are upset or angry	1.47	2.03	1.48	2.33
	Make appointments on the phone	1.32	1.78	1.22	1.65
Comprehension	Read a book for pleasure	1.55	1.68	1.50	1.46
	Look up a number in a telephone book	1.41	1.34	1.38	1.35

Table 4. Correlations between Aphasia Communication Outcome Measure (ACOM) scores for persons with aphasia (PWA) and surrogates (SUR) in each domain, and between ACOM scores and Rasch-modeled Porch Index of Communicative Ability modality scores. All correlations are significant at $p < 0.001$. The prediction for the highest correlation between each ACOM domain score and PICA modality score is bolded. The rightmost column gives the estimate and lower bound of the 1-sided 90% confidence interval for the difference between the predicted highest correlation and the next-highest observed correlation. Significant differences are marked with an asterisk(*).

ACOM Domain		PICA Verbal	PICA Comprehension	PICA Writing	Difference Between Predicted Highest and Next-Highest Correlation (LB of 90% CI)
Verbal Expression	PWA	**0.66**	0.53	0.48	0.13 (0.07)*
	SUR	**0.71**	0.63	0.55	0.08 (0.04)*
Comprehension	PWA	0.42	**0.52**	0.35	0.10 (0.02)*
	SUR	0.46	**0.57**	0.33	0.11 (0.04)*
Writing	PWA	0.50	0.54	**0.62**	0.08 (0.01)*
	SUR	0.53	0.58	**0.62**	0.04 (-0.03)

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Bayles, K. A. & Tomoeda, C. K. (1993). *Arizona Battery for Communication Disorders of Dementia*. Tucson, AZ: Canyonlands Publishing, Inc.

Duncan, P. W., Lai, S. M., Tyler, D., Perera, S., Reker, D. M., & Studenski, S. (2002). Evaluation of proxy responses to the Stroke Impact Scale. *Stroke, 33,* 2593-2599.

Frattali, C. M. (1992). Functional assessment of communication: Merging public policy with clinical views. *Aphasiology, 6,* 63-83.

Hilari, K., Owen, S., & Farrelly, S. J. (2007). Proxy and self-report agreement on the Stroke and Aphasia Quality of Life Scale-39. *Journal of Neurology, Neurosurgery, and Psychiatry, 78,* 1072-1075.

Porch, B. (2001). *Porch index of communicative ability*. Albuquerque, NM: PICA Programs.

Sheikh, J. I. & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS) Recent Evidence and Development of a Shorter Version. In T.L.Brink (Ed.), *Clinical Gerontology: A Guide to Assessment and Intervention* (pp. 165-173). New York: Hawthorn Press.

Thissen, D., Reeve, B. B., Bjorner, J. B., & Chang, C. H. (2007). Methodological issues for building item banks and computerized adaptive scales. *Qual.Life Res., 16 Suppl 1,* 109-119.

Williams, L. S., Bakas, T., Brizendine, E., Plue, L., Tu, W., Hendrie, H. et al. (2006). How valid are family proxy assessments of stroke patients' health-related quality of life? *Stroke, 37,* 2081-2085.

Worrall, L. (2001). A survey of outcome measures used by Australian speech pathologists. *Pacific Journal of Speech, Language, and Hearing, 6,* 149-162.

Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions [On-line]. Retrieved 11-30-2009 from http://rasch.org/rmt//rmt83.htm