



Background

When assessing people for aphasia, clinicians have many instruments from which to choose. The majority of the manuals for these tools are not peer-reviewed, calling into question the trustworthiness of psychometric properties reported. Efforts have been made to describe the psychometric concurrence of different tools, but description of the quantity of peer-reviewed information available has yet to be published (Skenes et al., 1985). When choosing an assessment, varying clinical situations demand different psychometric profiles. In the case of a long-term rehab patient, for example, high intra-rater and test-retest reliability are paramount whereas acute-stage assessment requires high validity and sensitivity to ensure an accurate diagnosis.

Objective

The purpose of this systematic review is to summarize the amount of peer-reviewed quantitative information about the psychometric properties of assessments for aphasia.

Methods

- Search terms: aphasia, diagnostic, evaluat*, assess*, test, tool, instrument, scale, battery, schedule, reliability, validity, psychometrics
- Included: diagnostic or descriptive studies in which quantitative psychometric properties were established
- Excluded: screenings, assessments for apraxia of speech, and studies using participants with primary progressive aphasia; psychometric evaluations of single items from assessments and non-binary comparisons; tests and articles with original language of publication other than English
- 740 articles obtained by search--study exclusion task completed for all articles obtained; resolved differences by consensus, resulting in 84 articles
- Full-text review and appraisal completed using consensus for differences, resulting in 14 articles for review
- Data extraction performed by the authors simultaneously; only measures of reliability and validity were extracted

Article Summary and Appraisal

Author	Year	Test	Psychometric Properties Evaluated/ Reported	a	b	c	d	e
Bruce	2010	CAT	predictive validity	3	2*	3	3	3
Del Toro	2011	S-BNT	concurrent validity, person reliability, item reliability	3	2*	3	3	3
Huff	1986	BNT	construct validity, internal consistency	3	3*	3	3	3
Ross	2004	PICA, WAB	sensitivity, specificity	3	3*	2	3	3
Flanagan	1997	ACTS, BNT, RCBA	test-retest reliability	3	1	3	3	3
Gallaher	1979	TTT	test-retest reliability, internal consistency	2	1*	3	3	3
Howard	2010	CAT	concurrent validity	3	2*	3	2	3
Miller	2000	EAAT	test-retest reliability, inter-rater reliability, concurrent validity, internal consistency	3	3	3	2	3
Nicholas	1989	BNT	inter-rater reliability	3	3*	3	3	3
Paci	2015	TTT	inter-rater reliability, intra-rater reliability	2	3*	2	3	3
Park	2000	RTT	test-retest reliability, inter-rater reliability, intra-rater reliability	3	2*	2	2	3
Ryan	1998	LNNB	construct validity	3	1*	2	3	3
Shewan	1990	WAB	test-retest reliability, inter-rater reliability, intra-rater reliability, construct validity	3	3*	3	2	3
Walker	2012	PNT	test-retest reliability, concurrent validity, internal consistency	3	3*	3	2	3

Key

- | | |
|---|--|
| a: Administration and scoring standardization | 3: Good quality method |
| b: Is the sample representative? | 2: Adequate or inconsistent quality method |
| c: Is the sample large enough? | 1: Poor quality method or not reported |
| d: Are there conflicts of interest? | *: geographically restricted sample |
| e: Are the statistical measures used appropriate? | |

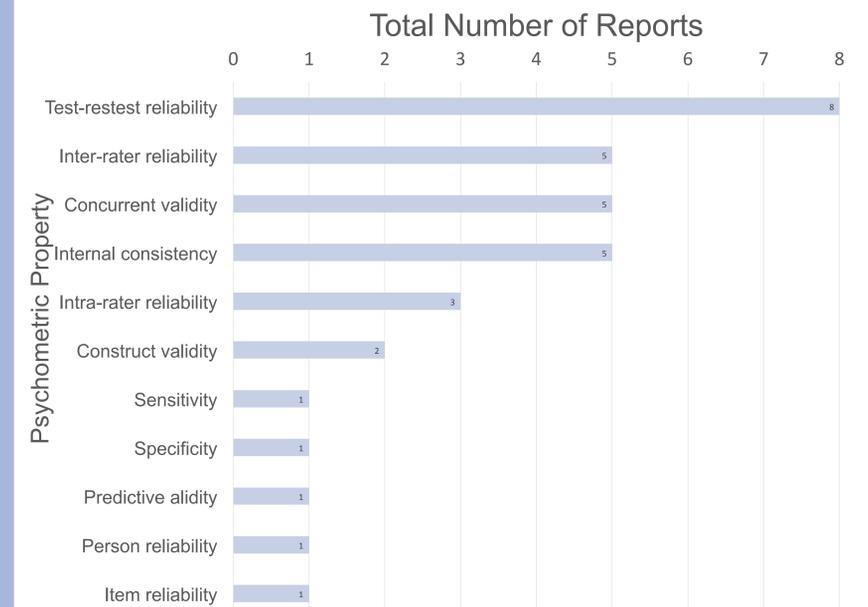
Results

Study appraisal resulted in overall ratings ranging from lesser to good quality. No psychometric properties were reported by more than one study for any test. Test-retest reliability was the most often reported measure (8/12 assessments), followed by inter-rater reliability (5/12), concurrent validity (5/12), and internal consistency (5/12). 12/14 studies used geographically restricted samples, a common limitation of diagnostic studies. On five studies, an author was common to both the article and the assessment.

Discussion

While unsurprising, the lack of peer-reviewed, quantitative information available regarding the psychometric properties of assessments for aphasia is problematic. Many of these assessments are performed at multiple stages of post-stroke recovery, requiring high temporal reliability and, in the acute stage, good validity. While this information can be obtained from manuals, the methodology may be questionable since manuals are not peer reviewed. Additionally, comparison of psychometric properties through manuals is not financially possible for speech-language pathologists in most settings. Further evaluation of established, often used aphasia assessments is needed to enable clinicians to choose the most psychometrically appropriate tools for each situation.

Summary of Reported Properties



Disclosures/Acknowledgements

This systematic review was completed as a project for SPHS 701 Introduction to Research Methods, under the guidance of Drs. Linda Watson and Jessica Steinbrenner. Neither author has any conflict of interests, financial nor intellectual. Please email michael_smith@med.unc.edu or matthew_suderman@med.unc.edu for references or questions.