

# Scoring the Quick Aphasia Battery: Training, definitions, fidelity



Katarina L. Haley<sup>1</sup>, Adam Jacks<sup>1</sup>, Marcia Rodriguez<sup>1</sup>, Lorelei Johnson<sup>2</sup>

<sup>1</sup>University of North Carolina, Chapel Hill NC, <sup>2</sup>Atrium Health, Charlotte NC



## Background

- The Quick Aphasia Battery (QAB)<sup>1</sup> is a multidimensional and reliable assessment tool with the desirable quality of an approximate administration time of 15 minutes.
- Scores inform both severity estimation and qualitative language profiling. Subtests use a graded scoring system.
- We recently adopted the QAB for a large study on speech production after left hemisphere stroke. To maximize assessment fidelity, we standardized administration procedures and trained our team on what we thought was the most subjective component of the test—the Connected Speech rating.

### Connected Speech Rating (trained).

10 features: length/complexity, speech rate, agrammatism, paragrammatism, anomia, empty speech, semantic paraphasia, phonemic paraphasia/neologisms, self-corrections, and overall communication impairment.

Severe (0)	Marked (1)	Moderate (2)	Mild (3)	Normal (4)
Evident in most or all utterances	Pervasive but not ubiquitous	Frequent but not pervasive	Detectable but infrequent	Not present or within the normal range

### Five Language Subtests (not trained):

Word comprehension, Sentence comprehension  
Picture naming, Repetition, Reading aloud

Unrelated foil or no response within 6 sec.	h partial comprehension (1)	Idiosyncratic response (2)	Correct but delayed >3 sec, self-corrected, or repeated (3)	Correct (4)
Unrelated response or no response within 6 sec. (0)	Some relation to target (1)	At least ½ of phonemes/ words are correct (2)	Correct but delayed >3 sec, self-corrected, or repeated (3)	Correct (4)

## Purpose

To evaluate interrater agreement for the Quick Aphasia Battery following manual review and training on the connected speech ratings.

## Methods

- Two experienced clinicians were trained to score the QAB alongside the rest of our research team. One had comprehensive clinical research experience; the other had significant research experience.
- Training progressed from reviewing and clarifying manual instructions<sup>1</sup> to administering the test to healthy adults and subsequently to volunteers portraying different types of aphasia.
- Scoring definitions were reviewed and expanded. The research team constructed supplementary manual notes and checklists and devised a slightly refined Connected Speech rating rubric (partially illustrated below). This work was followed by practice ratings on video recorded assessment sessions.

Severe (0)	Marked (1)	Moderate (2)	Mild (3)	Normal (4)
Evident in <b>all or almost all</b> utterances	Pervasive but not ubiquitous. <b>In most of the utterances but not in all or almost all of the utterances.</b>	Frequent but not pervasive. <b>In a lot of utterances, but not in most utterances.</b>	Detectable but infrequent. <b>Occasional utterances; barely noticeable.</b>	Not present or within the normal range

- To estimate interobserver agreement, the two clinicians independently reviewed video samples from 14 chronic stroke patients with aphasia for whom scores and media were shared with the research community via Aphasia Bank<sup>3</sup>.

## Results

- Interrater **agreement was strong for the trained subtest** (\*Connected Speech) and satisfactory, albeit somewhat lower, for most of the subtests that were not trained.

Point-to-point inter-observer agreement (%):

QAB subtest	Perfect agreement	Agreement within one scale level	Agreement within two scale levels
Connected Speech*	69	98	100
Word Comprehension	96	99	99
Sentence Comprehension	87	93	95
Picture Naming	74	87	95
Repetition	76	93	100
Reading aloud	69	90	96

## Discussion

- The trained connected speech ratings showed satisfactory rating fidelity (somewhat stronger than the reliability data provided with the original validation and reliability study<sup>1</sup>, where agreement was 94% within one scale level and 52% for a perfect match).
- Because scoring for the other subtests appeared straightforward, we assumed that observer agreement would be strong. This was not the case. Some issues were resolved by consulting the manual<sup>1</sup>. Other reasons that were consequential for scoring reliability:
  - Unclear how to time 3 and 6 seconds
  - Unclear what is “a complete attempt” (e.g. pe-pale-pe-pel-palie-palcil) -> what to score?
  - Difficult to differentiate fragment+complete response vs self-correction
  - Unclear what constitutes an “apraxic error”
- We refined operational definitions, added observation checklists, and are currently implementing video vignettes for rater training and calibration purposes.
- Our teams will calibrate monthly and monitor interobserver agreement throughout the study
- There is no reason to believe that the QAB is unique in generating disagreement among scorers. It is our experience that researchers and clinicians report that scoring manuals do not answer many of their questions, leaving them to develop their own definitions.
- Less than 7% of treatment studies provide information about scoring or assessment training<sup>2</sup>. It seems wise to radically increase that percentage and consider sharing training resources across clinics and laboratories.
- 15 minutes administration time is accompanied by significant time for training and scoring. Researchers and clinicians should allocate appropriate time to ensure valid and reliable test results.

## References

- Wilson SM, Eriksson DK, Schneck SM, Lucanie JM. A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. PLoS ONE. 2018 Feb 9;13(2):e0192773.
- Richardson JD, Hudspeth Dalton SG, Shafer J, Patterson J. Assessment fidelity in aphasia research. Am J Speech Lang Pathol. 2016 Dec 1;25(4S):S788–97.
- MacWhinney, B., Fromm, D., Forbes, M. & Holland, A. (2011). AphasiaBank: Methods for studying discourse. Aphasiology, 25,1286-1307.