

Supplemental Material S2. Two-sided z -tests for independent proportions for test set stratifications of exact match accuracy for Experiments 1–4.

Exp	Comparison	z	p
1. Baseline	Human agreement = 100% vs. human agreement < 100%	6.232	< .001
	Human confidence > median vs. human confidence \leq median	6.401	< .001
	WAB-R AQ > median vs. WAB-R AQ \leq median	4.698	< .001
	Fluent participants vs. nonfluent participants	2.691	.007
2. Controls	Human agreement = 100% vs. human agreement < 100%	8.882	< .001
	Human confidence > median vs. human confidence \leq median	9.227	< .001
	WAB-R AQ > median vs. WAB-R AQ \leq median	5.295	< .001
	Fluent participants vs. nonfluent participants	3.151	< .001
3. PWA	Human agreement = 100% vs. human agreement < 100%	12.528	< .001
	Human confidence > median vs. human confidence \leq median	12.149	< .001
	WAB-R AQ > median vs. WAB-R AQ \leq median	4.714	< .001
	Fluent participants vs. nonfluent participants	2.649	.008
4. Controls + PWA	Human agreement = 100% vs. human agreement < 100%	13.050	< .001
	Human confidence > median vs. human confidence \leq median	12.711	< .001
	WAB-R AQ > median vs. WAB-R AQ \leq median	4.299	< .001
	Fluent participants vs. nonfluent participants	2.811	.005

Note. Fluent participants are those with Wernicke, anomic, conduction, or transcortical sensory aphasia, or those considered “non-aphasic” by the WAB-R. Nonfluent participants are those with Broca, global, or transcortical motor aphasia. 46 out of 332 total sessions had unavailable WAB-R results and were excluded just from analyses involving WAB-R scores. Exp = experiment; PWA = people with aphasia; WAB-R AQ = Western Aphasia Battery–Revised Aphasia Quotient.