Supplementary Tables

Supplementary Table S1. Exploratory comparisons of HAST-PL task and total scores by sex (N = 173).

Score	Female (n = 73)	Male (n = 100)	Mean difference [95% CI]	p
Word Comprehension	3.79 (0.69)	3.66 (0.70)	0.13 [−0.08, 0.34]	0.209
Sentence Comprehension	3.03 (1.03)	3.04 (1.07)	−0.01 [−0.33, 0.31]	0.938
Repetition	2.95 (1.20)	3.10 (1.08)	−0.15 [−0.50, 0.20]	0.384
Naming	2.78 (1.40)	2.74 (1.51)	0.04 [−0.40, 0.48]	0.855
Word Fluency	2.62 (1.11)	2.59 (1.08)	0.03 [−0.31, 0.36]	0.876
Total score	15.16 (4.06)	15.13 (4.26)	0.03 [−1.23, 1.30]	0.957

Note. Values are mean (SD). Mean differences are female minus male. Welch's two-sample t-tests were exploratory; p values are unadjusted.

Supplementary Table S2. Ordinal one-factor CFA restricted to participants with aphasia (n = 130).

Task	Std. loading	SE	z	p	95% CI
Word Comprehension	0.517	0.102	5.062	<0.001	[0.317, 0.717]
Sentence Comprehension	0.744	0.065	11.443	<0.001	[0.617, 0.872]
Repetition	0.609	0.068	8.951	<0.001	[0.475, 0.742]
Naming	0.817	0.050	16.200	<0.001	[0.718, 0.916]
Word Fluency	0.816	0.050	16.368	<0.001	[0.718, 0.913]

Note. WLSMV estimation. Scaled $\chi^2(5) = 9.259$, $p = 0.099$; CFI = 0.988; TLI = 0.975; RMSEA = 0.081, 90% CI [0.000, 0.162]; SRMR = 0.059. All loadings were significant. T-size CFI: estimate 0.929, poor–fair 0.729, fair–close 0.824; T-size RMSEA: estimate 0.162, poor–fair 0.162, fair–close 0.138. The RMSEA estimate lies near the poor–fair boundary at the reported precision. Observed-scale reliability: $\alpha = 0.767$, $\omega = 0.799$.

Supplementary Table S3. Operating characteristics across integer-score thresholds corresponding to the ROC coordinates (N = 173).

Positive if score	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Youden J
≤5	3.08	100.00	100.00	25.44	0.0308
≤6	5.38	100.00	100.00	25.90	0.0538
≤7	7.69	100.00	100.00	26.38	0.0769
≤8	10.77	100.00	100.00	27.04	0.1077
≤9	19.23	100.00	100.00	29.05	0.1923
≤10	26.92	100.00	100.00	31.16	0.2692
≤11	28.46	100.00	100.00	31.62	0.2846
≤12	35.38	100.00	100.00	33.86	0.3538
≤13	40.77	100.00	100.00	35.83	0.4077
≤14	45.38	100.00	100.00	37.72	0.4538
≤15	55.38	97.67	98.63	42.00	0.5306
≤16	66.15	97.67	98.85	48.84	0.6383
≤17	77.69	86.05	94.39	56.06	0.6374
≤18	86.92	62.79	87.60	61.36	0.4971
≤19	96.15	32.56	81.17	73.68	0.2871

Note. Each row represents a binary rule: scores at or below the stated integer are screen-positive and higher scores are screen-negative. The sample comprised 130 participants with aphasia and 43 without aphasia. Values were calculated from the HAST-PL classification counts. Extreme all-positive/all-negative rules are omitted. The maximum observed Youden index occurred at ≤16.

Supplementary Table S4. Exploratory item-level comparisons between the aphasia and comparison groups (N = 173).

Item / subtask	Aphasia (n = 130)	Comparison (n = 43)	p
Word Comprehension Item 1	121 (93.1%)	43 (100.0%)	0.114
Word Comprehension Item 2	109 (83.8%)	42 (97.7%)	0.017
Word Comprehension Item 3	125 (96.2%)	43 (100.0%)	0.334
Word Comprehension Item 4	122 (93.8%)	38 (88.4%)	0.314
Sentence Comprehension Item 1	123 (94.6%)	41 (95.3%)	>0.999
Sentence Comprehension Item 2	89 (68.5%)	38 (88.4%)	0.010
Sentence Comprehension Item 3	89 (68.5%)	41 (95.3%)	<0.001
Sentence Comprehension Item 4	63 (48.5%)	41 (95.3%)	<0.001
Repetition Item 1	115 (88.5%)	43 (100.0%)	0.024
Repetition Item 2	97 (74.6%)	42 (97.7%)	<0.001
Repetition Item 3	106 (81.5%)	43 (100.0%)	<0.001
Repetition Item 4	45 (34.6%)	34 (79.1%)	<0.001
Naming Item 1	92 (70.8%)	42 (97.7%)	<0.001
Naming Item 2	87 (66.9%)	43 (100.0%)	<0.001
Naming Item 3	56 (43.1%)	38 (88.4%)	<0.001
Naming Item 4	79 (60.8%)	40 (93.0%)	<0.001
Word Fluency: fruits	4.9 (3.4)	9.5 (3.0)	<0.001
Word Fluency: letter K	3.3 (3.1)	7.8 (3.1)	<0.001

Note. For Tasks 1–4, values are the number (%) of correct responses and p values are from Fisher's exact test. For the two raw Word Fluency subtasks, values are mean (SD) and p values are from Welch's two-sample t-test. All p values are unadjusted.

Supplementary Table S5. Exploratory Spearman correlations with age and days since stroke (N = 173).

Score	Age ϱ	p	Days since stroke, ϱ	p
Word Comprehension	−0.082	0.286	−0.108	0.158
Sentence Comprehension	−0.164	0.032	0.010	0.900
Repetition	−0.163	0.032	0.055	0.475
Naming	−0.145	0.057	0.062	0.415
Word Fluency	−0.159	0.037	−0.014	0.854
Total score	−0.213	0.005	0.016	0.836

Note. Two-sided, unadjusted p values. Associations were computed in the pooled sample of 130 participants with aphasia and 43 without aphasia using the verified time-since-stroke data.

Supplementary Table S6. Bootstrap internal validation and cut-off selection stability (N = 173; 20,000 resamples).

A. Apparent performance and bootstrap adjustment.

Measure	Apparent value	Mean difference	Adjusted value
Full-score AUC	0.882826	0.000215	0.882612
Maximum Youden index	0.638283	0.024181	0.614102

Note. For Youden's index, the mean difference is estimated selection optimism; the adjusted value pertains to the threshold-selection procedure. For fixed-score AUC, the expected bootstrap-minus-original difference is zero, and the small observed difference reflects Monte Carlo variation. The bootstrap percentile 95% CI for AUC was [0.832111, 0.927907]. No confidence interval was calculated for the corrected Youden index. Monte Carlo standard errors of the mean differences were 0.000174 for AUC and 0.000368 for Youden's index; these quantify numerical simulation precision, not uncertainty in population performance.

B. Frequency of selecting each screen-positive rule.

Screen-positive rule	Selections	Percentage
≤16 points	10,253	51.265%
≤17 points	9,590	47.950%
≤18 points	157	0.785%
Other thresholds	0	0.000%

Note. The cut-off maximizing Youden's index was reselected in every resample across c = 0, …, 20. The lowest maximizing cut-off was the prespecified tie-break; no tied maxima occurred. Combined, ≤16 and ≤17 were selected in 99.215% of resamples. Frequencies are not probabilities that a cut-off is clinically optimal.

Supplementary Table S7. Likelihood ratios for the two principal HAST-PL screening rules (N = 173).

Screen-positive rule	LR+ (95% CI)	LR− (95% CI)
≤16 points	28.45 [4.08, 198.14]	0.347 [0.271, 0.443]
≤17 points	5.57 [2.64, 11.76]	0.259 [0.184, 0.365]

Note. Approximate log-scale 95% confidence intervals treat the rules as fixed and are not adjusted for threshold selection. No continuity correction was applied. At ≤16, classification counts were TP = 86, FN = 44, FP = 1, and TN = 42; at ≤17, TP = 101, FN = 29, FP = 6, and TN = 37. The interval for LR+ at ≤16 is very wide because of the single false positive. LR intervals are not bootstrap percentile intervals: 7,403 resamples (37.015%) had no false positives at ≤16 and therefore produced an infinite LR+.

Supplementary Table S8. Distribution of participants by centre and HAST-PL examiner in the analytical sample (N = 173).

Centre	Participants, n	Examiner code and number assessed
Department of Neurology, University Clinical Hospital No. 1, Szczecin	54	E1: 54
Department of Neurology, University Clinical Hospital, Poznań	40	E2: 40
Wielkopolska Rehabilitation Centre, Provincial Hospital, Poznań	35	E3: 35
Department of Neurology with a Stroke Unit, Prof. K. Gibiński University Clinical Center, Katowice	44	E4: 25 E5: 19

Note. Counts refer to participants retained in the final analytical sample. Examiner codes identify individual speech-language pathologists.