Supplementary Materials

This document provides additional information about the scoring and reliability of our three reading comprehension tests: the YARC, GORT, and WRAT.

YARC

Scoring: The raw scores for comprehension, reading rate, and summary were transformed into ability scores. The ability scores for comprehension and reading rate were then converted into standard scores, with a mean of 100 and a standard deviation of 15. No standard score was calculated for the summary task. All scoring was done with the YARC online conversion tool (https://rgt.testwise.net/YARC_Aust_Pri_index.htm).

Test reliability: The test reliability of the YARC was calculated using Cronbach's alpha. The overall reliability coefficients for the passages used in this study was 0.71 for comprehension, and 0.77 for the summary, which indicates good reliability. The validity of the comprehension questions was assessed by calculating the percentage of participants who could answer the questions correctly without reading the text. For the items used in this study, the percentages of participants answering correctly ranged from 0 to 18.6%, with a mean of 2.5%. The validity of the summary questions was assessed by correlations with the comprehension scores. Because the correlations were low for all items, the (original test) authors note that the summary scores may not be entirely reliable and should be used with caution.

GORT

Scoring: The maximum score was 5 for reading rate, accuracy, and comprehension, and 10 for fluency. For reading rate and accuracy, each score (0-5) corresponds to a range of reading times and number of errors indicated in the manual, thus participants are assigned a score based on these cut-offs. For example, a participant who reads item 6 in 45 seconds and makes 3 reading errors would receive a score of 5 for reading rate, 4 for accuracy, and 9 for

fluency. The comprehension score is the sum of all questions answered correctly. These scores were then transformed into scaled scores as per the manual instructions. These scores have a mean of 10 and a standard deviation of 3. Participants who scored 8-12 were classified as 'average', 7-6 as 'below average', and 5-4 as 'poor'. The sum of the comprehension and fluency scaled scores was then calculated to obtain a participant's reading oral index scores from the manual. This index is a standard score with mean of 100 and standard deviation of 15.

Test reliability: Test reliability was calculated with Cronbach's alpha for reading rate, accuracy, comprehension, and fluency, and with Guilford's alpha for the oral reading index. The coefficient alphas range from 0.91 to 0.97, indicating good reliability. Test validity was informed by the percentage of questions that could be answered correctly without reading the passages. The majority of the questions (88%) were only answered correctly by 10% or less of the sample, indicating that most of the questions were passage-dependent.

WRAT

Test Reliability: Test reliability was measured with Cronbach's alpha. The reliability coefficients for the word reading and sentence comprehension subtests are respectively 0.92 and 0.93 for the green form, indicating good reliability.

This document contains tables with the output of the models ran with only the monolingual speakers in our sample (n =60).

Table B1

YARC: Intercepts and Estimates of the Best Ten Models and Full Model for Monolingual Speakers

Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Full Model
Intercept	93.12	92.95	93.21	93.12	93.06	93.12	93.35	93.25	92.82	92.92	93.00
Speed (wpm)	6.10	5.78	6.57	6.05	6.97	6.67			6.98	6.76	6.56
Average Fixation Duration		6.67							7.29	8.42	5.01
Saccade Length					-1.67			6.29	-1.68		0.26
Skipping	-5.22	-4.760.	-4.16	-6.02	-5.03	-5.12		-6.97	-4.61	-5.24	-5.50
First-Fixation Duration	7.88			7.79	8.87						3.75
Gaze Duration	-13.73	-12.15	-5.41	-14.35	-15.10	-6.18	-3.65	-6.83	-13.15	-17.80	-17.88
Regressions				-1.73		-1.94		-6.17		-2.98	-2.97
Go-Past Time	9.28	8.51	8.08	10.03	9.86	9.13		7.01	9.27	6.57	7.04
Total-Reading Time										7.06	6.36
R2	0.26	0.25	0.22	0.27	0.27	0.24	0.9	0.24	0.27	0.29	0.31
ELPD_LOO	-232.70	-232.71	-232.92	-233.21	-233.28	-233.32	-233.39	-233.52	-233.59	-233.63	-235.99

Notes: Models 1 to 10 are ordered based on their ELPD LOO (descending). Green = 95% credibility interval does not include 0; white: 95% credibility interval includes 0; blank: predictor not included in the model.

Table B2

GORT: Intercepts and Estimates of the Best Ten Models and Full Model for Monolingual Speakers

Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Full Model
Intercept	91.43	91.43	91.17	91.07	91.24	91.30	91.34	91.14	91.17	91.29	91.37
Speed (wpm)	-5.22										-5.22
Average Fixation Duration	-9.92		-8.99		-9.95	-9.71				-8.92	-8.07
Saccade Length	4.30	3.58	3.99	3.26	3.60	2.44	3.74	3.55	3.09	2.70	4.64
Skipping											-1.21
First-Fixation Duration	12.38		10.01		11.11	13.45	2.69	1.32	1.44	11.30	11.75
Gaze Duration						-4.72					-1.96
Regressions										1.58	-0.02
Go-Past Time		-2.46	-2.57					-3.09			-0.91
Total-Reading Time	-7.79			-2.40	-2.84		-7.93		-3.22	3.84	-5.64
R2	0.34	0.25	0.31	0.24	0.31	0.30	0.30	0.27	0.26	0.32	0.37
ELPD_LOO	-194.55	-194.71	-195.88	-194.97	-195.04	-195.06	-195.18	-195.26	-195.28	-195.33	-199.76

Notes: Models 1 to 10 are ordered based on their ELPD LOO (descending). Green = 95% credibility interval does not include 0; white: 95% credibility interval includes 0; blank: predictor not included in the model.

Table B3

WRAT: Intercepts and Estimates of the Best Ten Models and Full Model for Monolingual Speakers

Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Full Model
Intercept	107.14	107.02	107.10	107.11	107.02	107.40	107.05	107.10	107.02	107.13	107.09
Speed (wpm)	9.53	11.02	9.06	10.90	9.34	7.54	10.49	9.60	9.73	9.78	8.37
Average Fixation Duration							1.07				-3.56
Saccade Length								1.15			3.10
Skipping	-3.98	-4.49	-3.19	-4.09	-3.93	-2.47	-4.25	-4.87	-3.85	-3.95	-5.34
First-Fixation Duration		1.66			2.78						3.46
Gaze Duration			2.94	1.68					2.38		3.23
Regressions	2.74	3.36	3.92	2.95	4.46		3.15	2.67	3.16	2.73	3.32
Go-Past Time			-3.53		-3.19						-3.90
Total-Reading Time									-1.85	0.27	-0.22
R2	0.34	0.37	0.38	0.37	0.38	0.30	0.36	0.35	0.37	0.37	0.41
ELPD_LOO	-212.90	-213.30	-213.31	-213.42	-213.53	-213.67	-213.67	-213.83	-213.90	-213.94	-217.28

Notes: Models 1 to 10 are ordered based on their ELPD LOO (descending). Green = 95% credibility interval does not include 0; white: 95% credibility interval includes 0; blank: predictor not included in the model.

Table B4

Average: Intercepts and Estimates of the Best Ten Models and Full Model for Monolingual Speakers

Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Full Model
Intercept	97.46	97.45	97.47	97.45	97.46	97.47	97.46	97.47	97.47	97.48	97.44
Speed (wpm)	4.24	6.57	4.52	3.66	3.89	7.49	4.91	4.36	6.36	7.94	6.70
Average Fixation Duration											-8.75
Saccade Length		2.61		2.40	0.70				2.14		3.82
Skipping		-5.42	-0.88	-2.27		-2.34			-3.40	-3.99	-6.69
First-Fixation Duration											10.99
Gaze Duration											-5.11
Regressions								0.51			-1.79
Go-Past Time		8.76				3.28	0.84		2.90	8.64	10.37
Total-Reading Time		-5.86								-5.22	-3.82
R2	0.22	0.33	0.23	0.26	0.23	0.26	0.23	0.23	0.29	0.30	0.38
ELPD_LOO	-205.74	-206.27	-206.29	-206.41	-206.60	-206.64	206.66	-206.67	-206.71	-206.72	-209.54

Notes: Models 1 to 10 are ordered based on their ELPD LOO (descending). Green = 95% credibility interval does not include 0; white: 95% credibility interval includes 0; blank: predictor not included in the model.