Can eye movements be used to predict reading comprehension ability?

Diane Mézière (IDEALAB, Macquarie University, University of Groningen), Lili Yu (Macquarie University), Erik Reichle (Macquarie University), Titus von der Malsburg, (University of Stuttgart, MIT), Genevieve McArthur (Macquarie University)
diane.meziere@hdr.mq.edu.au

Reading is one of the most complex cognitive tasks that we engage in on a daily basis. While several theories of reading comprehension exist, which cognitive skills are the most predictive of reading comprehension remains unclear. This in turns makes the design of valid reading comprehension assessments difficult. In this study, we investigated the possibility of using eye-movements to predict reading comprehension.

Three widely-used reading comprehension tests were administered to 79 adults with no history of reading difficulties. In the *York Assessment for Reading Comprehension* (YARC; Snowling et al., 2009), participants read two long passages silently, and answered comprehension questions. In the *Gray Oral Reading Test* (GORT-5; Wiederholt & Bryant, 2012), participants read eleven short passages aloud, and answered comprehension questions. In the sentence comprehension subtest of *Wide Range Achievement Test* (WRAT-4; Wilkinson & Robertson, 2006), participants read thirty-one sentences with a missing word, and had to provide the missing word (cloze procedure). Participants' eye movements were monitored while they read the test items.

The correlations between the three comprehension scores were statistically significant but only moderate (0.59-0.63). Correlations between eye-tracking measures and comprehension scores are shown in Table 1. The correlations yielded a different pattern for each of the three tests. Scores from the YARC were more highly correlated to early eye-movement measures (e.g., gaze duration), associated with early processes such as lexical processing. Scores from the GORT tended to be equally highly correlated to both early and late eye-movement measures (e.g., total reading time), indicative of higher-level integration processes. Scores from the WRAT tended to be more highly correlated to late eye-movement measures (go-past time, total reading time).

Secondly, we used linear models to investigate the potential of eye movements to predict reading comprehension scores. We ran Bayesian linear models to evaluate the efficacy of all combinations of eight eye-movement measures and reading speed (wpm). We then used leave-one-out cross-validation (Vehtari, Gelman & Gabry, 2017) to compare these models and identify the 'best' model, and the best subset of eye-tracking measures to predict comprehension. The results from these analyses are shown in Table 2. These results also yielded different patterns between the three tests. The scores from the YARC were best predicted by reading speed, first-pass skipping rate, gaze duration, and go-past time. Scores from the GORT were best predicted by average fixation duration, forward saccade length, and first-fixation duration, closely followed by total reading time. Scores from the WRAT were best predicted by reading speed, as well as skipping and regressions rates, and interestingly the best predictors did not include any measure of fixation durations. Finally, comprehension scores averaged across the three tests were best predicted by reading speed, go-past time, and total reading time. In each case, eye movements explained substantial amounts of variance over and above reading speed alone. Models with all nine predictors explained an average of 39% of the variance in comprehension scores (YARC: 29%; GORT: 42%; WRAT: 46%).

Our results are in line with previous research suggesting that reading comprehension assessments do not all measure the same cognitive skills to the same extent (e.g., Keenan, Betjemann & Olson, 2008). Additionally, results from both correlations and linear models shed light on the complexity of the relationship between eye-movement behaviour and reading comprehension. Indeed, while eye movements can be used to predict comprehension scores, their predictive ability is modulated by the varying task demands of reading comprehension measures. These results have important practical implications for the use of reading comprehension measures in research, schools, and clinical settings, as well as theoretical implications about the relationship between eye movements and reading comprehension.

Table 1: Correlations between comprehension scores and eye movements

Measure	YARC	GORT	WRAT	Average
Global				
Speed	0.23*	0.30*	0.57*	0.48*
Av. Fix. Dur.	-0.17	-0.11	-0.28*	-0.22$^{p=0.058}$
Saccade Length	0.15	0.41*	0.26*	0.32*
First-Pass				
Skipping	-0.03	0.09	0.16	0.07
First-Fix. Dur.	-0.19	-0.07	-0.26*	-0.21$^{p=0.06}$
Gaze Dur.	-0.26*	-0.35*	-0.29*	-0.33*
Late				
Regression	-0.02	0.12	-0.09	-0.03
Go-Past	-0.09	-0.30	-0.43*	-0.34*
Total Time	-0.17	-0.36*	-0.50*	-0.41*

Note: This table shows the correlation coefficients between eye-movement measures and comprehension scores for each test. * = $p < 0.05$

Table 2: Outputs of the 'Best' and Full Models

	YARC		GORT		WRAT		Average	
Predictors	Best Model	Full Model	Best Model	Full Model	Best Model	Full Model	Best Model	Full Model
Intercept	90.61	90.58	90.96	90.93	105.75	105.74	95.60	95.61
Speed (wpm)	5.95	9.29	-5.49	-5.32	11.84	10.32	7.84	7.47
Av. Fix. Dur.		8.55	-12.94	-11.47		-3.17		-6.86
Saccade Length		-4.29	4.64	5.31		0.87		1.98
Skipping	-4.82	-3.30		-1.64	-4.66	-4.89	-3.94	-5.68
First-Fix. Dur.	7.30	1.55	15.21	14.85	1.82	4.95		8.64
Gaze Dur.	-13.15	-19.09		-2.73		0.67		-3.30
Regression		-0.05		-0.97	4.13	4.20		-0.60
Go-Past	9.34	7.66		0.14		-0.81	9.81	11.09
Total Time		6.10	-8.28	-6.24		-1.63	-6.90	-6.61

Note: This table shows the estimated coefficients of the Bayesian linear models for the three comprehension tests and the average of the three test scores. For each, the output of the "best" model according for the leave-one-out cross-validation and the output of the full model are presented. Green cells indicate the 95% credibility interval does not include zero, yellow cells indicate the 90% credibility interval does not include zero, blank cells indicate the 90% credibility interval includes zero.

References: Snowling, M.J, et al. (2009). YARC. GL Publishers • Wiederholt, J. L., & Bryant, B. R. (2012) GORT5. Pro-Ed. • Wilkinson, G. S., & Robertson, G. J., (2006). WRAT4. Pearson. • Vehtari, A., Gelman, A., & Gabry, J. (2017). Statistics and computing, 27(5), 1413-1432. • Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. Scientific Studies of Reading, 12(3), 281-300.