Using eye movements to predict performance on reading comprehension tests

Diane Mézière (IDEALAB, Macquarie University, Rijksuniversiteit Groningen, University of Potsdam, Newcastle University), Lili Yu (Macquarie University), Erik Reichle (Macquarie University), Titus von der Malsburg (University of Potsdam, Massachusetts Institute of Technology), Genevieve McArthur (Macquarie University)

Reading comprehension is one of the most complex cognitive tasks that we engage in on a daily basis. Although many theories of reading comprehension exist, the essential cognitive skills that are predictive of reading comprehension remain unclear, making the design of valid measurements of reading comprehension difficult. In this study, we use eye-movements to examine the extent to which three different reading comprehension tests measure various cognitive skills.

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

The correlations between the three comprehension scores were moderate and statistically significant (0.59-0.63). Correlations between the comprehension scores and the eye-movement measures yielded a different pattern for each test. Scores from the YARC tended to be more highly correlated to early eye-movement measures, indicative of early reading processes such as lexical processing. Scores from the GORT showed similar correlation coefficients for both early and late eye-movement measures - typically associated with higher-level integration processes. Scores from the WRAT were more highly correlated to late eye-movement measures.

To further investigate the relationship between eye movements and comprehension scores, we ran a second set of analyses to test if eye movements could predict comprehension scores. Bayesian linear models were used to evaluate the efficacy of all combinations of our eye movement measures. Leave-one-out cross-validation (Vehtari, Gelman & Gabry, 2017) was then used to compare these models and identify the 'best' model to predict comprehension. Results from these analyses also yielded test variance. For the YARC, the best model included both early and late eye-movement measures. For the GORT, early measures appeared as the best predictors, closely followed by total reading time. For the WRAT, the best set of predictors did not include any fixation time measures but rather skipping and regression rates. Models run with the average comprehension score across the three tests indicated reading speed (number of words read per minute) and late measures as the best predictors of comprehension. In call cases, eye movements explained substantial amounts of variance over and above reading speed alone. Full models for the comprehension tests explained an average of 39% of the variance in comprehension scores (YARC: 29%; GORT: 42%; WRAT: 46%).

The results from these analyses are in line with previous studies showing that reading comprehension tests do not measure the same cognitive skills to the same extent (Keenan, Betjemann & Olson, 2008; Keenan & Meenan, 2014). Results from both sets of analyses shed light on the complexity of the relationship between eye movements and reading comprehension – eye movements can predict comprehending scores, however, the best predictors and their predictive ability are modulated by the task demands. These results have important practical implications for the use of reading comprehension tests in research 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. • Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. Journal of learning disabilities, 47(2), 125-135.