Which reading measures are the strongest predictors of reading comprehension?

Jan Brasser[*1], **Vera Demberg**[2], **Lena A. Jäger**[1], **Maja Stegenwallner-Schütz**[3]

[1]Department of Computational Linguistics, University of Zurich
[2]Department of Computer Science, Saarland University
[3]Institute of Special Education, University of Koblenz

*contact: jan.brasser@uzh.ch

Reading measures derived from eye movements are commonly categorized into first-pass, second-pass, and global measures, reflecting initial word encounters, rereading, or aggregated scanpath characteristics across the entire text. Previous studies have demonstrated that eye movements reflect how well readers comprehend text (Shubi *et al.*, 2026; Mézière *et al.*, 2024), but mostly focused either on individual measures or on undifferentiated sets of measures, without necessarily considering the category of measures. Moreover, comprehension questions differ in the representations they probe, ranging from the surface text base to the fully enriched situation model. In this study, we determine which combination of first-pass, second-pass, and global measures best predicts accuracy across question types.

We analyze the eye movements of 71 native German-speaking adults reading expository and prose texts, and model the readers' comprehension accuracy using different combinations of reading measures as predictors, as part of the MultiplEYE project. Comprehension questions vary with respect to the level of text comprehension they target, spanning local (inquiring explicit information), bridging (requiring inferences), and global question types (requiring reflecting and assessing information). Measures were calculated exclusively from eye movements on annotated critical spans containing the information relevant to each question. Using Bayesian hierarchical models, we tested all combinations comprising exactly one measure from each category (four first-pass, two second-pass, and five global) to predict comprehension accuracy, and evaluated model performance (five-fold cross validation). Data were split into an exploratory set ($N = 50$) and an independent validation set ($N = 21$) to test the replicability of the identified effects.

The top-performing model contains average first-pass gaze duration (first-pass measure), regression rate (second-pass measure), and wrap-up effects (global measure). Specifically, shorter first-pass gaze durations predicted higher accuracy when responding to local and global questions, while higher regression rates were associated with higher accuracy on bridging questions. Despite wrap-up effects appearing in all top-performing models, wrap-up effects were not associated with comprehension accuracy. Validation on an independent sample failed to replicate the observed effects, likely due to limited sample size; ongoing data collection will bring the validation sample to up to 50 readers.

Increased first-pass gaze duration is a well-established index of processing difficulty, specifically regarding lexical access or syntactic integration. Although processing difficulty can in principle compromise the representation of text information at all levels, this study only confirms this for explicit information derived from the text-base and full situation models. Similarly, regression rates reflect difficulty in sentence processing leading readers to reread the text. Since regression rates are associated with improved comprehension performance, they signal successful rereading attempts. The role of wrap-up effects remains unclear and the failure to replicate in the validation models limits current generalizability of the findings. Taken together, the systematic combination of reading measures allows us to distinguish whether certain measures reflect mere processing difficulty or more mindful rereading strategies. This study points to the potential of modelling comprehension accuracy in relation to different types of questions, but the reliability of these associations remains to be established.

References

Mézière D. C., Yu L., McArthur G., Reichle E. D., von der Malsburg T. (2024). Scanpath regularity as an index of reading comprehension. *Scientific Studies of Reading 28*(1), 79–100. doi:10.1080/10888438.2023.2232063.

Shubi O., Reich D. R., Klein K. G., Angel Y., Prasse P., Jäger L. A., Berzak Y. (2026). Eyebench: Predictive modeling from eye movements in reading. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.