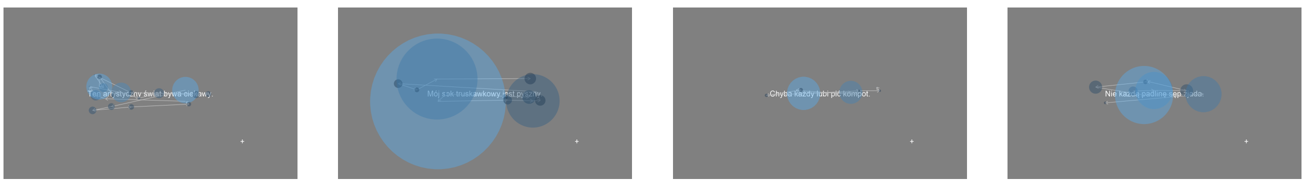
Evaluating Cognitive Effort With Pupillary Activity When Reading

Evidence from Eye Movements of Primary School Children

Agata Cybulska
Izabela Krejtz
Krzysztof Krejtz
Institute of Psychology
SWPS University of Social Science & Humanities
Warsaw, Poland

Andrew T. Duchowski
School of Computing
Clemson University
Clemson, SC, USA



(a) long and frequent (b) long and infrequent (c) short and frequent (d) short and infrequent

Figure 1: Exemplary scanpaths (circles indicate fixations of relative duration, larger radii are longer) when reading sentences with different embedded keywords: (a) long and frequent, (b) long and infrequent, (c) short and frequent, and (d) short and infrequent.

ABSTRACT

This paper evaluates the Low/High Index of Pupillary Activity (LHIPA), a measure of cognitive effort based on pupil response, in the context of reading. At the beginning of 3rd grade, 179 children (8-9 y.o.) were asked to read 40 sentences with keywords differing in length and frequency while their eye movements were recorded. In line with expectations, pupillary activity (LHIPA) was smaller for sentences with infrequent keywords. This finding shows the utility of LHIPA as a measure of cognitive effort in educational settings.

CCS CONCEPTS

• **Applied computing → Psychology**.

KEYWORDS

reading, pupillary activity, cognitive effort, word-frequency effect, word-length effect, children

ACM Reference Format:

Agata Cybulska, Izabela Krejtz, Krzysztof Krejtz, and Andrew T. Duchowski. 2021. Evaluating Cognitive Effort With Pupillary Activity When Reading: Evidence from Eye Movements of Primary School Children. In *EMICS '21: ACM CHI '21 Workshop on Eye Movements as an Interface to Cognitive State, May 14, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

Reading is a complex two-stage process of 1) decoding text symbols and 2) obtaining their semantic meaning [Blythe et al. 2015; Booth et al. 1999]. The second stage requires substantial cognitive resources from the reader since it loads executive functions (working memory, reasoning) and memory processes. Despite a large body of literature on eye tracking in reading [Rayner 1998], there is still a need for eye movement-related metrics that could differentiate these two stages of text recognition, especially in the context of early education. Such a metric could serve as a foundation for future systems supporting teachers and students in early detection of reading problems as well as in alleviating cognitive overload when learning. The present paper empirically tests a promising new metric, the Low/High Index of Pupillary Activity (LHIPA), which we use to measure cognitive effort of primary school children when reading.

2 BACKGROUND

Theories of reading development and word recognition posit that there are two pathways for decoding the meaning of words (e.g., the dual-route model of reading for meaning [Coltheart 2000]). In the first, printed words are decoded into their phonological representations, and then the word meaning is activated. In the second pathway, semantic representation is directly activated from print. The contribution of each pathway depends on word familiarity and reading skills [Jared et al. 2016].

The degree of exposure to a given word changes its decoding time and is reflected in eye movement patterns even among developing readers [Joseph et al. 2013]. More fixations on words and longer reading times may indicate difficulties in decoding information during reading [Rayner 1998]. Ample studies have demonstrated that beginners of reading are less efficient in decoding words (for review see Blythe and Joseph [2011]). Experienced readers need less time to extract visual information because of ease of encoding of the fixated word [Rayner 2009].

An extensive body of eye-tracking studies on reading has also demonstrated that words differ in the probability of being fixated due to their frequency and length. Longer words tend to be fixated more often than shorter words (the word length effect [Just and Carpenter 1980]). Frequent words receive less attention than infrequent words (the word frequency effect [Inhoff and Rayner 1986b; Rayner and Raney 1996]). Tiffin-Richards and Schroeder [2015] found that the interaction of word frequency and length on fixation duration and viewing time is more pronounced among children than adults. Moreover, several studies observed greater pupil dilation to infrequent words than frequent words due to greater retrieval effort [Ledoux et al. 2016; Schmidtke 2014].

Observations of pupillary activity have a long history in cognitive psychology (for review, see Beatty [1982]), starting with seminal works by Hess and Polt [1964] and Kahneman and Beatty [1966] who demonstrated pupil dilation with increased task difficulty. Recently, Duchowski et al. [2020] introduced the Low/High Index of Pupillary Activity (LHIPA) and demonstrated its sensitivity in differentiating between low and high cognitively demand tasks. This promising metric reflects moment-to-moment pupil diameter changes and is thought to be insensitive to effects of luminance. LHIPA is computed as the ratio of low to high frequency $x_{\psi}^{1/2\log_2(n)}(t)/x_{\psi}^{1}(2^{1/2\log_2(n)}t)$ of the wavelet coefficients $\psi_{j,k}(t)$ of the pupil diameter signal $x(t)=\sum_{j,k=-\infty}^{\infty} c_{j,k}\psi_{j,k}(t),\ j,k\in\mathbf{Z}$, and is expected to decrease with increased cognitive effort. LHIPA was shown to discriminate task difficulty in a series of experiments where participants performed easy and difficult mental arithmetic, an *nBack* task, eye typing [Duchowski et al. 2020], and cognitive effort during complex decision making [Krejtz et al. 2020]. To our best knowledge, LHIPA has never been used to capture cognitive processing of text while reading.

Reading sentences with different keywords should thus moderate number of fixations, their average duration, and pupil activity. Due to a higher cognitive effort of encoding meaning of infrequent keywords, LHIPA should decrease.

3 METHOD

To test the hypotheses an eye-tracking experiment with a 2 × 2 within-subjects design was conducted. The two independent variables were: keyword frequency (frequent vs. infrequent) and keyword length (long vs. short) embedded in the sentences read by children.

Participants. Eye tracking data were collected from 179 children of age between 8 and 9 years ($M=8.79$, $SD=0.41$). The sample was comprised of 57% girls. All children were attending 3rd grade of primary school.

Each child's parents provided written consent and each child provided verbal consent. The study was approved by the Institutional Research Board of the first author's institution (approval number 47/2016).

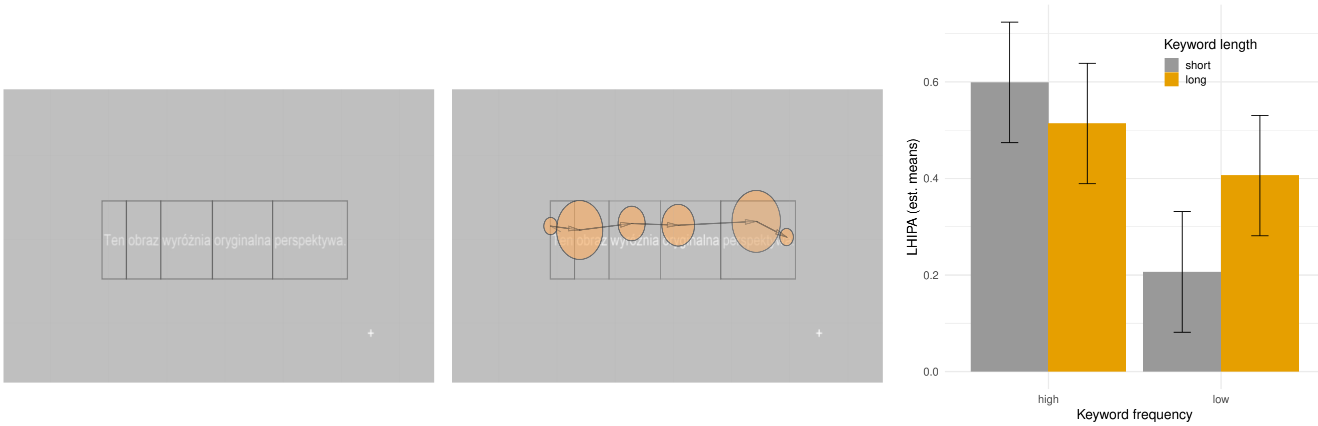
Reading Task and Eye Tracking. The reading task consisted of forty 5-word sentences (see Fig. 1). Each sentence included a keyword (mean keyword position in the sentence $M=3.38$, $SD=1.03$). The keywords, taken from the Nencki Affective Word List (NAWL) Database [Riegel et al. 2015], were either short (3 or 4 letters) or long (10 or 11 letters) [Hyönä et al. 1995] of low or high frequency. The frequency was measured as the number of occurrences per million words [Przepiórkowski et al. 2012]. To control for sentence understanding, a multiple-choice question appeared after 8 sentences. The task was created with PsychoPy 2 [Peirce 2007]. The sentences were displayed on a 22-inch LED monitor with 1920 × 1080 resolution and 75 Hz refresh rate. Participants' eye movements were recorded with a GazePoint 3 eye tracker with 60 Hz sampling rate and 0.5–1 degree visual angle accuracy. Fixation detection for eye movement analysis followed Duchowski's [2017] velocity-based saccade detection using the Savitzky-Golay derivative filter with velocity threshold set to 10°/s. Analysis of fixations used only those fixations captured within Areas Of Interest (AOIs) defined over individual sentence words (Fig. 2(a)). Statistical analyses were conducted using R, the language for computational statistics [R Core Team 2020].

Procedure. Children were individually tested in a dedicated school room. First, each child was familiarized with the eye tracker while seated in front of a monitor at a viewing distance of about 60 cm, and a standard five-point eye tracker calibration was performed (average calibration error was < 0.5° visual angle). Second, each child was asked to silently read each sentence, then look at the plus sign at the bottom right corner of the display (Fig. 2(b)), then press the spacebar to advance to the next sentence. At the end, each child received a reward for their participation. The experiment lasted approximately 20 min.

4 RESULTS

First, a series of 2 × 2 within-subjects analyses of variances was conducted on fixation count (FC), and average fixation duration (AFD), depending on keyword frequency and length (see Table 1 for detailed statistics). Interaction effects were decomposed with *post-hoc* pairwise comparisons using *Tukey HSD* correction.

As expected, the ANOVAs showed that longer keywords triggered more fixations and longer AFD than short keywords. Infrequent words yielded longer AFD, but not more fixations while reading the sentences. Interaction effects were also significant. The interaction effect on FC showed that sentences with frequent short words received significantly fewer fixations than the other types of sentences, see Table 1. The interaction for AFD suggests that average fixation duration was significantly longer for sentences including infrequent rather than frequent keywords ($\hat{\delta} = -84.90,\ t(350) = 4.99,\ p < 0.001$). There was no significant difference in AFD between sentences including short keywords of different frequency ($\hat{\delta} = -27.30,\ t(350) = 1.61,\ p = 0.11$), see Table 1. Understanding the meaning of the sentences with infrequent



(a) AOI definitions (b) fixations (circles) within AOIs (c) LHIPA for correct answers

Figure 2: Example scanpath processing: (a) Areas Of Interest (AOIs) are defined so that only (b) fixations detected within AOIs are processed further via statistical analysis, e.g., fixation count or (c) LHIPA while reading sentences with different types of keywords. *Note:* Error bars represent ±1 *SE*.

Table 1: The effects of length and frequency on eye movement metrics. Note: standard errors are in parentheses.

variable	frequency	length: long	length: short	effect	F test results
avg. fixation count	frequent	$7.69(0.13)$	$6.59(0.13)$	length	$F(1,175)=109,06,\ p<0.001,\ \eta^2=0.05$
	infrequent	$7.44(0.13)$	$6.97(0.13)$	frequency	$F(1,175)=1.05,\ p=0.31,\ \eta^2<0.001$
				interaction	$F(1,175)=16.38,\ p<0.001,\ \eta^2=0.008$
avg. fix. duration	frequent	$659ms(36.90)$	$749ms(36.90)$	length	$F(1,175)=25.18,\ p<0.001,\ \eta^2=0.004$
	infrequent	$740ms(36.90)$	$776ms(36.90)$	frequency	$F(1,175)=22.19,\ p<0.001,\ \eta^2=0.003$
				interaction	$F(1,175)=5.65,\ p<0.02,\ \eta^2<0.001$

long keywords is more cognitively demanding than reading sentences with frequent short words. These differences in cognitive demands should be reflected by the LHIPA. To test this prediction, we ran ANOVA on this pupil measure. To make sure that children decoded the meaning of words, we analyzed only sentences for which children provided accurate comprehension responses. In line with expectations, we observed a main effect of keyword frequency ($F(1,104)=7.45,\ p<0.01,\ \eta^2=0.01$). LHIPA was significantly lower when reading sentences with infrequent keywords ($M=0.31,\ SE=0.1$) than sentences with frequent keywords ($M=0.56,\ SE=0.1$), see Fig. 2(c).

5 DISCUSSION AND CONCLUSION

Results of the present study are in line with word frequency and length effects for fixation count and average fixation duration during reading, see e.g., Tiffin-Richards and Schroeder [2015]. Sentences with infrequent long keywords received more fixations than those with short and frequent keywords (c.f., Inhoff and Rayner [1986a]; Rayner and Duffy [1986]). The shortest fixation duration was observed for long words (c.f., Hyönä and Olson [1995]; Just and Carpenter [1980]; Kliegl et al. [2004]). In general, we reported relatively long fixation durations during reading by children (see also McConkie et al. [1991]) which suggests that their reading process was not yet fully proficient and more cognitively demanding due to unfamiliar words [Schmidtke 2018]. Results also support Leinenger and Rayner's [2017] claim that children encode less information than skilled readers. According to Coltheart [2000], the semantic meaning of a word requires word familiarity and cognitive resources.

We claim that cognitive effort during meaning decoding is captured by the Low/High Index of Pupillary Activity. We demonstrated that LHIPA was lower for sentences with infrequent keywords compared to sentences with frequent keywords, but only for correctly decoded sentences. This suggests that children reading without decoding the semantic meaning of sentences is akin to skimming [Lohmeyer and Meboldt 2015], which can be distinguished by eye movement characteristics such as fixation count and duration. In contrast, reading for decoding semantic meaning is akin to scrutinizing [Lohmeyer and Meboldt 2015] and requires substantial cognitive resources which can be measured with pupillary activity, e.g., LHIPA.

We conclude that LHIPA captures cognitive effort during semantic sentence decoding. Eye tracking measures, among them LHIPA, may help to objectively diagnose reading difficulties at early stages of education. Future studies should be specifically designed to test sensitivity and specificity of eye movement metrics in relation to the reading process.

ACKNOWLEDGMENTS

This work was financially supported by the National Science Center, PRELUDIUM 11 grant (2016/21/N/HS6/02845) to Agata Rodziewicz.

REFERENCES

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2):276.

Blythe, H. and Joseph, H. (2011). Children's eye movements during reading. In Simon, L., Iain, G., and Stefan, E., editors, *The Oxford Handbook of Eye Movements.*, chapter 36, pages 644–662. OUP Oxford, The address of the publisher.

Blythe, H. I., Pagán, A., and Dodd, M. (2015). Beyond decoding: Phonological processing during silent reading in beginning readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(4):1244–1252.

Booth, J. R., Perfetti, C. A., and MacWhinney, B. (1999). Quick, automatic, and general activation of orthographic and phonological representations in young readers. *Developmental psychology*, 35(1):3.

Coltheart, M. (2000). Commentary on section 3 - dual routes from print to speech and dual routes from print to meaning: Some theoretical issues. In Kennedy, A., Radach, R., Heller, D., and Pynte, J., editors, *Reading as a Perceptual Process*, pages 475–490. North-Holland, Oxford.

Duchowski, A. T. (2017). *Eye Tracking Methodology: Theory & Practice.* Springer-Verlag, Inc., London, UK, 3rd edition.

Duchowski, A. T., Krejtz, K., Gehrer, N. A., Bafna, T., and Bækgaard, P. (2020). The low/high index of pupillary activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–12, New York, NY, USA. Association for Computing Machinery.

Hess, E. H. and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611):1190–1192.

Hyönä, J. and Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: effects of word length and word frequency. *J. Exp. Psychol. Learn. Mem. Cogn.*, 21(6):1430–1440.

Hyönä, J., Olson, R. K., Defries, J., Fulker, D., Pennington, B., and Smith, S. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21:1430–1440.

Inhoff, A. and Rayner, K. (1986a). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40:431–439.

Inhoff, A. W. and Rayner, K. (1986b). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Percept. Psychophys.*, 40(6):431–439.

Jared, D., Ashby, J., Agauas, S. J., and Levy, B. A. (2016). Phonological activation of word meanings in grade 5 readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(4):524–541.

Joseph, H. S. S. L., Nation, K., and Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psych. Rev.*, 42(2):207–222.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.*, 87(4):329–354.

Kahneman, D. and Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154(3756):1583–1585.

Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology - EUR J COGN PSYCHOL*, 16:262–284.

Krejtz, K., Żurawska, J., Duchowski, A., and Wichary, S. (2020). Pupillary and microsaccadic responses to cognitive effort and emotional arousal during complex decision making. *Journal of Eye Movement Research*, 13(5).

Ledoux, K., Coderre, E., Bosley, L., Buz, E., Gangopadhyay, I., and Gordon, B. (2016). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behavior research methods*, 48(1):285–305.

Leinenger, M. and Rayner, K. (2017). What we know about skilled, beginning, and older readers from monitoring their eye movements: Implications for teaching reading. In León, J. A. and Escudero, I., editors, *Reading Comprehension in Educational Settings*, Studies in Written Language and Literacy, 16, pages 1–27. John Benjamins Publishing Company.

Lohmeyer, Q. and Meboldt, M. (2015). How We Understand Engineering Drawings: An Eye Tracking Study Investigating Skimming and Scrutinizing Sequences. In *Proceedings of the International Conference on Engineering Design*, ICED '15, pages 1–15, Milan, Italy. Design Society.

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, N. R., and Wolff, P. M. (1991). Children's eye movements during reading. *Vision and visual dyslexia*, 13:251–262.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of Neuroscience Methods*, 162(1):8–13.

Przepiórkowski, A., Bańko, M., Górski, R. L., and Lewandowska-Tomaszczyk, B. (2012). *Narodowy Korpus Języka Polskiego.* Wydawnictwo Naukowe PWN, Warszawa.

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.*, 124(3):372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8):1457–1506.

Rayner, K. and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition*, 14(3):191–201.

Rayner, K. and Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin and Review*, 3(2):245–248.

Riegel, M., Wierzba, M., Wypych, M., Żurawski, Ł., Jednoróg, K., Grabowska, A., and Marchewka, A. (2015). Nencki Affective Word List (NAWL): the cultural adaptation of the Berlin Affective Word List–Reloaded (BAWL-R) for Polish. *Behavior Research Methods*, 47(4):1222–1236.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology*, 5:1–16.

Schmidtke, J. (2018). Pupillometry in linguistic research - an introduction and review for second language researchers. *Studies in Second Language Acquisition*, 40:529–549.

Tiffin-Richards, S. P. and Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research*, 113:33–43.