Measuring Cognitive Effort with Pupillary Activity and Fixational Eye Movements When Reading

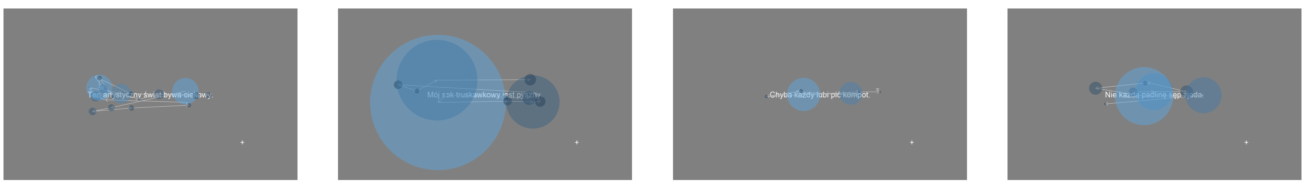
Longitudinal Comparison of Children With and Without Primary Music Education

Agata Cybulska
arodziewicz@swps.edu.pl
Institute of Psychology
SWPS University of Social Science & Humanities
Warsaw, Poland

Krzysztof Krejtz
kkrejtz@swps.edu.pl
Institute of Psychology
SWPS University of Social Science & Humanities
Warsaw, Poland

Andrew T. Duchowski
duchowski@clemson.edu
School of Computing
Clemson University
Clemson, SC, USA

Izabela Krejtz
ikrejtz@swps.edu.pl
Institute of Psychology
SWPS University of Social Science & Humanities
Warsaw, Poland



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

Figure 1: Example scanpaths (circles indicate fixations of relative duration, larger radii are longer) when reading sentences with different embedded keywords.

ABSTRACT

This article 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 2nd and 3rd grade, 107 children (8-9 y.o.) from music and general primary school were asked to read 40 sentences with keywords differing in length and frequency while their eye movements were recorded. Sentences with low frequency or long keywords received more attention than sentences with high frequent or short keywords. The word frequency and length effects were more pronounced in younger children. At the 2nd grade, music children dwelt less on sentences with short frequent keywords than on sentences with long frequent keywords. As expected LHIPA decreased over sentences with low frequency short keywords suggesting more cognitive effort at earlier stages of reading ability. This finding shows the utility of LHIPA as a measure of cognitive effort in education.

ETRA '22, June 8–11, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9252-5/22/06…$15.00
https://doi.org/10.1145/3517031.3529636

CCS CONCEPTS

• **Applied computing → Psychology**.

KEYWORDS

Low/High Index of Pupillary Activity, cognitive effort, music education, reading

ACM Reference Format:
Agata Cybulska, Krzysztof Krejtz, Andrew T. Duchowski, and Izabela Krejtz. 2022. Measuring Cognitive Effort with Pupillary Activity and Fixational Eye Movements When Reading: Longitudinal Comparison of Children With and Without Primary Music Education. In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22), June 8–11, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3517031.3529636

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 between 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) [Duchowski et al. 2020a], which we use to measure cognitive effort when reading among primary school children with and without music education. The cognitive effort is referred in the present paper as the amount of cognitive resources an individual allocates to task performance e.g., reading [de Jong 2010]. We make an attempt towards discriminating between word recognition and semantic processing with pupillary activity along with *classical* gaze (e.g., eye fixations, average fixation duration, and dwell time).

We also analyze sentence reading by primary school children with and without systematic music education. There is growing evidence that music training improves children's cognitive functioning [Jaschke et al. 2018], yet the evidence from eye tracking research is scarce. For example, Schellenberg [2004] demonstrated that a 3-year music training (vs. drama training and passive control group) in 6 y.o. children resulted in significant increase of their intelligence. Similarly, a 7-month musical training (singing and playing on instrument) of 3-4 y.o. children improved their intelligence measured with Wechsler IQ test compared to a control group [Gromko and Poorman 1998]. Another study of three-year musical training in children showed a positive relation to children's academic achievements in English and mathematics measured at higher grades [Holochwost et al. 2017].

In the present study, we examine whether developing ability to sight-read music notation facilitates sentence reading. We compare children from music and non-music schools at the 2nd and 3rd year of primary education to gauge how cognitively demanding is sentence reading for them. We present our hypotheses following a brief literature review.

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, 2005]). 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].

Music reading involves analogous note-to-sound decoding to be able to perform the corresponding sound. Both text and music reading involve decoding visual stimuli into sound components, which are more related to left hemisphere activation [Brown et al. 2006]. Brown et al. [2006] observed, in positron emission tomography, when amateur musicians vocally improvised melodic or linguistic phrases, activations in similar functional brain areas (e.g., the primary motor cortex, Broca's area, primary and secondary auditory cortices). These results suggest a significant overlap in brain mechanisms when vocalizing text and music. Further, Ahken et al. [2012] demonstrated similarities of pianists' eye movement behavior when processing syntactic incongruities in text and music reading. Incongruities in both domains (text and music) triggered more and longer fixations resulting in longer reading time suggesting that music and language processing share neural network related to syntactic processing.

During music education (e.g., playing an instrument) children learn to sight-read music notations. Sight-reading is a complex transcription task [Sloboda 1977]. With growing expertise musicians identify and process groups of notes as meaningful "chunks" [Polanka 1995]. Reading sentences, as well as music notation, is determined by hierarchical organization of the structural and visual features of the reading material [Cara and Gomez Vera 2016]. Additionally, musical and text reading use horizontal peripheral vision [Goolsby 1994]. Therefore one may assume parallels between music and language processing, since both processes require knowledge of abstract structure to guide horizontal visual processing of verbal or musical text Sloboda [1977]. Furthermore, spacing of notes can be congruent to spacing between words, therefore sentence structure can be comparable to a musical phrase.

In fact, Cara and Gomez Vera [2016] observed similarities between music and text reading in music students. Also, Li et al. [2019] demonstrated evidence that music reading experience may enhance letter identification benefiting from the fact that both text reading (in left-to-right scripts) and music reading involve left-to-right sequential symbol processing.

2.1 Reading Eye Movement Development

Most studies on eye movements in reading show that key eye movement characteristics systematically decline with age: mean fixation duration and mean number of fixations [McConkie et al. 1991; Taylor 1965]. This was classically shown early on by Buswell [1922] who compared eye movements in reading of children from different grades, starting at 1 grade (6-7 y.o.) up to the 6th grade (11-12 y.o.) along with adults. The variability in eye movement measures reflects ongoing information processing (eye-mind assumption [Just and Carpenter 1980]) and can be treated as a proxy for real-time cognitive effort in reading.

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]. Leinenger and Rayner [2017] claimed that beginners encode less information than skilled readers per fixation due to a shorter reading span [Häikiö et al. 2009] and lower familiarity with words. 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].

2.2 Pupillary Activity

Observations of pupillary activity have a long history in studying cognitive effort evoked by perceptual and mental tasks (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. Pupillary dilation in a response to a stimulus is an effect of a neural inhibitory mechanism on the parasympathetic oculomotor complex or Edinger–Westphal nucleus by LC-NE [Krejtz et al. 2020; Wilhelm et al. 1999]. It is generally considered that pupil diameter provides a "very effective index of the momentary load on a subject as they perform a mental task", and reflects the amount of material under active processing [Kahneman and Beatty 1966]. An ample of studies showed systematic pupil activity in reaction to cognitive effort [Krejtz et al. 2018; Piquado et al. 2010] and arousal [Bradley et al. 2008]. Several studies also demonstrated higher pupillary activity to low-frequency words than high-frequency words [Ledoux et al. 2016; Schmidtke 2014]. This lower reactivity of pupil to high-frequency words supports advantage of familiarity for regularly encountered words [Schmidtke 2018].

Recently, Duchowski et al. [2020b] derived a novel measure of pupil activity during information processing, the Low/High Index of Pupillary Activity (LHIPA), a ratio of low to high frequency oscillation of pupil diameter. This 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 $x(t) = \sum_{j,k=-\infty}^{\infty} c_{j,k}\psi_{j,k}(t),\ j,k \in \mathbf{Z}$, and is expected to decrease with increased cognitive effort, i.e., lower LHIPA scores suggest greater cognitive effort. LHIPA demonstrated its sensitivity in differentiating between low and high cognitively demanding tasks. It discriminated task difficulty in a series of experiments where participants performed easy and difficult mental arithmetic, an *nBack* task, eye typing [Duchowski et al. 2020b], and cognitive effort during complex decision making [Krejtz et al. 2020]. For details on implementation of LHIPA see Duchowski et al. [2020b, 2018]. To the best of our knowledge, LHIPA has never been used to capture cognitive effort of semantic processing of text while reading.

2.3 The Present Study

In the present study, we analyzed eye movement characteristics of developing reading skills: at the beginning of the second and third years of children's primary education. Children read four types of sentences, including either short or long keywords which were either of low or high frequency.

We hypothesized that reading sentences with different keywords would moderate dwell time, average fixation duration, and pupil activity. Due to a higher cognitive effort of encoding the meaning of infrequent keywords, LHIPA should decrease. As children tend to encode less information than skilled readers per fixation [Leinenger and Rayner 2017], we expected to observe longer dwell time during reading of sentences with long keywords. We also expected to observe differences in dwell time on sentences between the 2nd and 3rd year of education. Finally, since previous studies reported similarities between music reading and text reading [Cara and Gomez Vera 2016] and facilitated performance on verbal tasks after music training [Li et al. 2019], we expected differences favoring music children when compared to their peers in general primary school.

3 METHOD

To test our hypotheses, a longitudinal eye tracking study with a $2 \times 2 \times 2 \times 2$ mixed design was conducted. Children were recruited from music and general primary schools, thus the school type acting as our main between-subject factor. Three independent within-subjects variables were: keyword frequency (frequent vs. infrequent) and keyword length (long vs. short) embedded in the sentences read by children, and measurement time (2nd vs. 3rd).

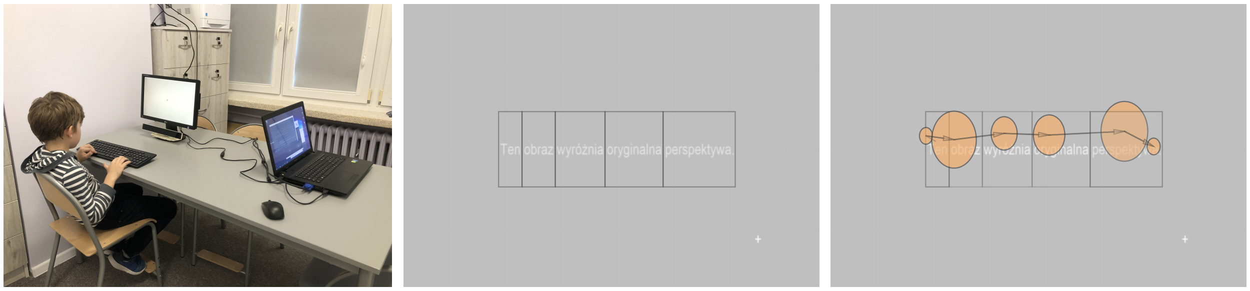
3.1 Participants

Eye tracking data were obtained from 107 children. The sample consisted of 56 children (30 girls and 26 boys) who attended music primary school and 51 children (29 girls and 22 boys) attending general primary school. These children were selected from a larger sample and matched on a fluid intelligence score (Raven's Colored Progressive Matrices [Jaworowska and Szustrowa 2003; Raven 1984]) and gender. There was no significant difference in proportion of boys to girls between children from both school types ($\chi^2(1) = 0.11$, $p = 0.73$). Participating children from music ($M = 8.28$, $SD = 0.38$) and general schools ($M = 8.36$, $SD = 0.34$) also did not differ significantly in age ($t(104.97) = 1.16$, $p = 0.25$). In the final sample children from music ($M = 31.25$, $SD = 2.53$) and general primary schools ($M = 30.37$, $SD = 3.23$) were also similar in terms of their fluid intelligence measure ($t(94.56) = 1.55$, $p = 0.12$). A 2-way mixed-design analysis of variance on intelligence score with measurement time and school type as factors showed only a main effect of measurement time ($F(1, 105) = 57.52$, $p < 0.001$, $\eta^2 = 0.092$). As expected, in the third grade children obtained higher intelligence scores ($M = 31.84$, $SD = 2.75$) than in the 2nd grade of their primary education ($M = 29.82$, $SD = 3.61$). Neither the main effect of school type ($F(1, 105) = 2.47$, $p = 0.12$) nor the interaction term between school type and measurement time ($F(1, 105) < 1$) was statistically significant.

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

3.2 Description of the Music Program

Music schools and general primary schools have the same obligatory general education curriculum taught during integrated educational 45-minute classes (11-12 units per week e.g., the development of reading, writing, basics of mathematical skills, creative problem solving). Music classes in the general education curriculum are limited to one class per week during which children learn basic musical concepts (e.g. melody, accompaniment, rhythm), reading



(a) Experimental setting (b) AOI definitions (c) fixations (circles) within AOIs

Figure 2: Experimental setting (a) Areas Of Interest (AOIs) definitions (b) Example scanpath (c).

and writing of musical notation (seven notes), and basics of singing and playing one instrument.

On top of the general education curriculum, children from music schools attend: chosen music instrument class (twice a week), rhythm class, ear training class (twice a week), and approx. once a week starting from the 2nd grade a 30-min. class with an accompanist. One of the educational outcomes in the primary music school curriculum is reading and understanding music notation and performing a selection of simple pieces of music.

3.3 Extracurricular activities

Most of the children from music and general primary schools declared that they attended extracurricular non-formal educational activities (in total 91 answered 'yes'). Children in music and general school did not differ in frequency of non-curriculum classes, $\chi^2(2) = 4.45$, $p = 0.11$. Among non-curriculum activities attended by the study participants, the following were mentioned (in frequency of appearance order): sports, foreign language, fine-art, board games and creativity, drama, instrument playing, and scouts. Two children from the general school group attended extracurricular music instrument classes (one piano and one guitar). The intensity and duration of these extracurricular activities were not recorded.

3.4 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]. 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(b)). Statistical analyses were conducted using R, the language for computational statistics [R Core Team 2020].

3.5 Procedure

Children were individually tested (Fig. 2(a)) 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(c)), then press the space bar to advance to the next sentence. Presentation of each sentence was proceeded with a $1000\,ms$ fixation cross placed at the position of the first letter in the sentence. At the end of the experimental procedure, each child received a reward for their participation. The experiment lasted approximately 20 min.

4 RESULTS

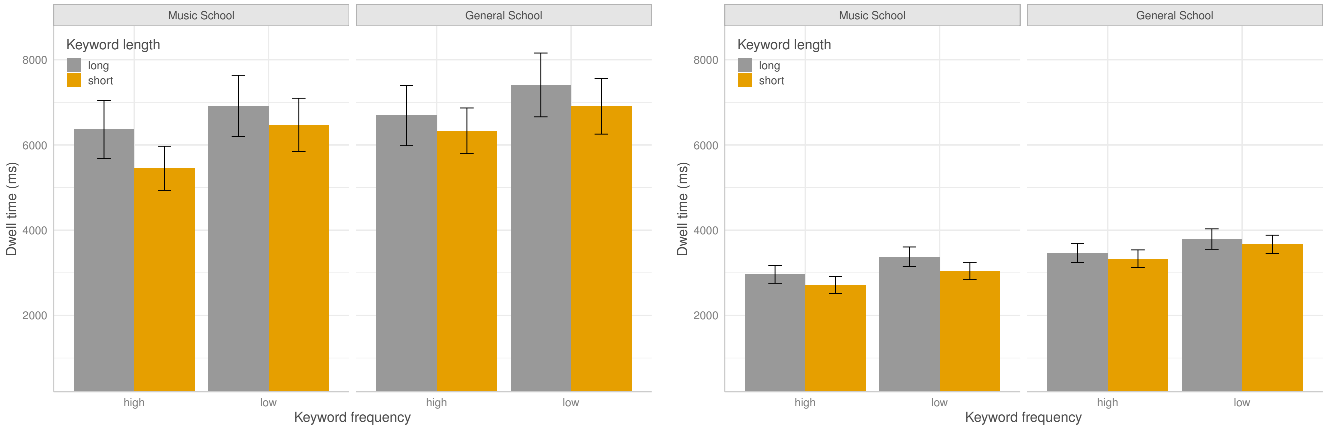
Results are analyzed in terms of dwell time, fixation duration, and the Low/High Index of Pupillary Activity (LHIPA).

4.1 Dwell Time During Sentence Reading

A mixed-design analysis of variance (2 × 2 × 2× 2) was conducted, with keyword frequency (low vs. high), length (low vs. high), time of measurement (2nd vs. 3rd school grade), and school type (music vs. general) as independent variables and dwell time during sentence reading as the dependent variable.

The analysis revealed a main effect of keyword frequency, ($F(1, 102) = 84.68$, $p < 0.001$, $\eta^2 = 0.01$) and length ($F(1, 102) = 23.05$, $p < 0.01$, $\eta^2 = 0.003$). In line with expectations, children, in general, dwell significantly longer while reading sentences containing the low frequency keyword ($M = 5.23$sec., $SE = 0.31$) than on sentences containing the high frequency keywords ($M = 4.69$sec., $SE = 0.27$). They also dwell significantly longer on sentences containing long keywords ($M = 5.15$sec., $SE = 0.32$) than on sentences with short keywords ($M = 4.77$sec., $SE = 0.26$).

Also in line with our predictions, the analysis showed a statistically significant main effect of measurement time ($F(1, 102) = 79.96$,



(a) Second grade of education

(b) Third grade of education

Figure 3: Dwell time while reading sentences with keywords of different frequency and length by children from music and general school. *Note:* bar graphs represent the estimated means and error bars represent ± 1*SE*.

$p<0.001$, $\eta^2=0.18$). At the first measurement time (2nd grade) children dwelt significantly longer on sentences ($M=6.61$sec., $SE=0.46$) than a year later ($M=3.31$sec., $SE=0.15$).

The analysis revealed also a significant interaction effect of keyword frequency and measurement time ($F(1,102)=13.53$, $p<0.001$, $\eta^2=0.001$). Pairwise comparisons with Tukey correction for multiple comparisons showed that the frequency effect was greater when children were younger (for low frequency words $M=6.97$sec., $SE=0.49$; for high frequency words $M=6.25$sec., $SE=0.43$). The difference between low and high frequency words decreased at 3rd grade (for low frequency words $M=3.49$sec., $SE=0.15$; for high frequency words $M=3.14$sec., $SE=0.14$).

Similarly, we observed a significant interaction effect between keyword length and measurement time ($F(1,102)=6.55$, $p<0.05$, $\eta^2=0.001$). Pairwise comparisons again showed that dwell time while reading sentences with long and short keywords was greater at the 2nd grade (for short words $M=6.33$sec., $SE=0.41$; for long words $M=6.89$sec., $SE=0.51$) than at the 3rd year of education (for short words $M=3.21$sec., $SE=0.14$; for long words $M=3.42$sec., $SE=0.16$).

Analysis also revealed a statistically significant 4-way interaction effect of keywords length and frequency, time of measurement and school type ($F(1,102)=4.37$, $p<0.05$, $\eta^2=0.001$), see Fig. 3. This interaction was decomposed with two separate ANOVA for measurement in the 2nd and 3rd grade. The analysis for 2nd grade showed a statistically significant interaction effect between keywords frequency, length, and school type, ($F(1,104)=4.01$, $p<0.05$, $\eta^2=0.001$), see Fig. 3a while the analysis for the measurement at 3rd grade showed that this interaction effect was not significant ($F(1,103)<1$), see Fig. 3b. Pairwise comparisons for the 2nd grade measurement indicated that all differences were significant except for the comparison between sentences with frequent long vs. short keywords in general school group. There were not significant differences between music and general school groups.

4.2 Fixation Duration When Reading Sentences

A similar (2 × 2 × 2× 2) mixed design ANOVA was conducted for average fixation duration as the dependent variable. The analysis yielded a significant main effect of keyword frequency, ($F(1,102)=10.35$, $p<0.01$, $\eta^2=0.001$). Average fixation duration was significantly longer while reading sentences with low frequency keywords ($M=177$ms, $SE=0.66$) than with high frequency keywords ($M=172$ms, $SE=0.67$).

The main effect of keyword length appeared significant, ($F(1,102)=44.59$, $p<0.001$, $\eta^2=0.005$). Short keywords in sentences evoked longer average fixation duration ($M=180$ms, $SE=0.70$) than long keywords ($M=168$ms, $SE=0.63$).

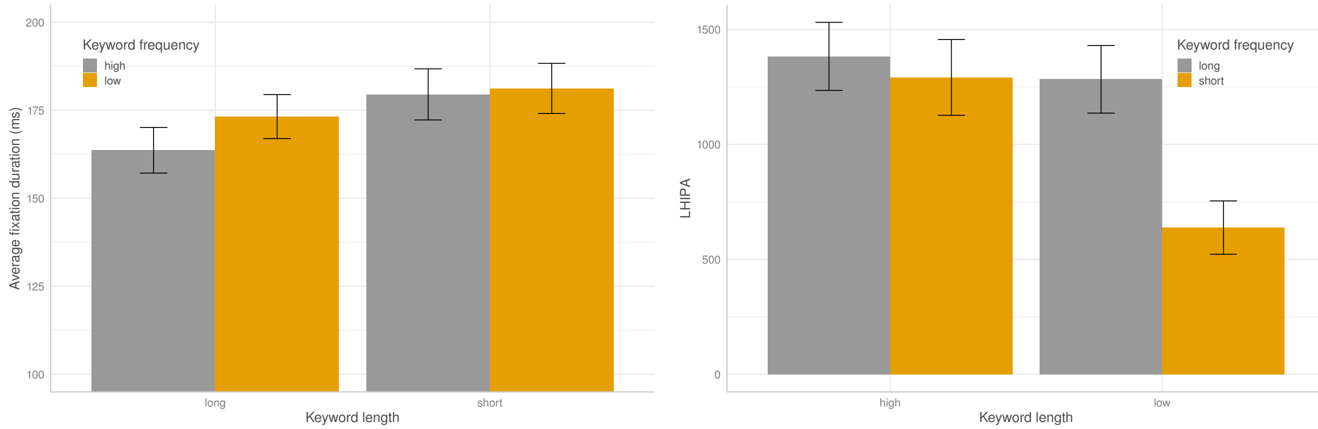
As expected, the main effect of measurement time was also statistically significant ($F(1,102)=25.89$, $p<0.001$, $\eta^2=0.074$). Second grade children had significantly longer average fixation duration while reading ($M=198$ms, $SE=0.01$) than a year later ($M=150$ms, $SE=0.005$).

The interaction effect of keywords frequency and length was also significant ($F(1,102)=4.80$, $p<0.05$, $\eta^2=0.001$), see Fig. 4a. Decomposition of this interaction with pairwise comparisons showed that difference in average fixation duration when reading sentences with high ($M=180$ms, $SE=0.73$) and low ($M=181$ms, $SE=0.71$) frequency short keywords was not statistically significant ($t(102)=0.56$, $p=0.58$). The same difference for long keywords was statistically significant ($t(102)=5.14$, $p<0.001$).

Interestingly, the main effect of school type was marginally significant, ($F(1,102)=3.13$, $p=0.08$, $\eta^2=0.019$). Music school children tend to show shorter average fixation durations while reading sentences ($M=163$ms, $SE=0.92$) than children without music education ($M=186$ms, $SE=0.95$).

4.3 Low/High Index of Pupillary Activity

For evaluating the Low/High Index of Pupillary Activity (LHIPA) score dependence on keyword length and frequency as well school type, we conducted a (2 × 2 × 2) mixed design ANOVA. Because



(a) Fixation duration (b) LHIPA

Figure 4: LHIPA and average fixation duration while reading sentences with keywords of different frequency and length by primary school children. *Note:* bar graphs represent the estimated means and error bars represent ± 1*SE*.

evaluating cognitive effort assumes that the experimental task was properly executed by participants, we decided to run this analysis only on those sentences to which participants answered correctly.

The results of these analyses showed a statistically significant main effect of keyword frequency ($F(1, 104) = 19.46$, $p < 0.001$, $\eta^2 = 0.016$). As expected, low frequency keywords evoked lower LHIPA scores on average ($M = 0.94$, $SE = 0.10$) than high frequency keywords ($M = 1.27$, $SE = 0.12$).

We also obtained a significant main effect of keyword length ($F(1, 104) = 11.87$, $p < 0.001$, $\eta^2 = 0.014$). Here inclusion of short keyword in the sentence evoked lower LHIPA scores ($M = 0.95$, $SE = 0.11$) than long keywords ($M = 1.26$, $SE = 0.13$).

Interpretation of these main effects should be made with caution since both factors, keyword length and frequency, were involved in statistically significant interaction effect ($F(1, 104) = 14.70$, $p < 0.001$, $\eta^2 = 0.014$), see Fig. 4b. Decomposition of this interaction effect with pairwise comparisons showed that the difference in LHIPA scores between frequent ($M = 1.27$, $SE = 0.13$) and infrequent ($M = 1.25$, $SE = 0.14$) long keywords was not significant ($t(104) < 1$). However, the difference for short words was statistically significant ($t(104) = 5.41$, $p < 0.001$), and in the expected direction (for frequent short words: $M = 1.28$, $SE = 0.15$ and for infrequent words: $M = 0.62$, $SE = 0.10$).

5 DISCUSSION

The main goal of the present study was to verify the sensitivity of the Low/High Index of Pupillary Activity (LHIPA) in capturing cognitive effort during reading in developing readers. In general, the study results supported our predictions.

We demonstrated that LHIPA decreased when reading sentences with infrequent, short keywords compared to reading 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 duration. In contrast, reading for decoding semantic meaning is akin to scrutinizing [Lohmeyer and Meboldt 2015] and requires substantial cognitive effort, measurable by pupillary activity, e.g., LHIPA.

Results of the present study are in line with effects of word frequency and length on dwell time and average fixation duration during reading (c.f., Tiffin-Richards and Schroeder [2015]). Sentences with infrequent and long keywords received more attention than those with short and frequent keywords (c.f., Inhoff and Rayner [1986a]; Rayner and Duffy [1986]). Shorter dwell time when reading short and frequent words suggests lower cognitive effort (c.f., Hyönä and Olson [1995]; Just and Carpenter [1980]; Kliegl et al. [2004]). Results showed that the effects are more pronounced for younger children.

Furthermore, corroborating previous studies (e.g., see McConkie et al. [1991]), we report significantly longer reading dwell times and average fixation duration at the 2nd grade than in the 3rd grade. This suggests that, at the 2nd year of primary education, the reading process is not yet fully proficient and is more cognitively demanding due to a shorter reading span [Häikiö et al. 2009] and lower familiarity with words [Schmidtke 2018]. Results also support Leinenger and Rayner's [2017] claim that children encode less information per fixation than skilled readers. According to Coltheart [2000], the semantic meaning of a word requires word familiarity and cognitive resources. At the 2nd grade, children with music education dwelt significantly less on sentences with short, high frequency words compared to long, high frequency words, this difference was not observed in general school children. Music school children tended to also have shorter average fixation duration during reading. These findings suggest that experience in music reading may facilitate sentence reading (c.f., Li et al. [2019]).

6 CONCLUSION

We claim that cognitive effort during semantic sentence decoding is captured by the Low/High Index of Pupillary Activity. Eye tracking measures, among them LHIPA, can help objectively diagnose reading difficulties at early stages of education. Future studies should test real-time monitoring of cognitive effort to encourage learners to exert more cognitive effort for effective learning, when needed [Paas et al. 2005; Sweller 2010]. Finally, future studies should also test the sensitivity and specificity of eye movement metrics in relation to the reading process.

ACKNOWLEDGMENTS

This work was financially supported in part by the Polish National Science Center, PRELUDIUM 11 grant (2016/21/N/HS6/02845) to Agata Cybulska and by a U.S. National Science Foundation grant (IIS-1748380) to Andrew Duchowski.

Open access of this article was financed by the Ministry of Science and Higher Education in Poland under the 2019-2022 program "Regional Initiative of Excellence", project number 012 / RID / 2018/19.

REFERENCES

Stephanie Ahken, Sylvie Hebert, and Ramesh Balasubramaniam. 2012. Eye Movement Patterns During the Processing of Musical and Linguistic Syntactic Incongruities. *Psychomusicology: Music, Mind and Brain* 22 (06 2012), 18–25. https://doi.org/10.1037/a0026751

Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91, 2 (1982), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Hazel I. Blythe and Holly S.S.L. Joseph. 2011. Children's eye movements during reading. In *The Oxford Handbook of Eye Movements.*, Liversedge Simon, Gilchrist Iain, and Everling Stefan (Eds.). Oxford University Press, Oxford, New York, Chapter 36, 644–662.

Hazel I. Blythe, Ascensión Pagán, and Megan Dodd. 2015. Beyond Decoding: Phonological Processing during Silent Reading in Beginning Readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41, 4 (2015), 1244–1252. https://doi.org/10.1037/xlm0000080

James R. Booth, Charles A Perfetti, and Brian MacWhinney. 1999. Quick, automatic, and general activation of orthographic and phonological representations in young readers. *Developmental psychology* 35, 1 (1999), 3–19. https://doi.org/10.1037/0012-1649.35.1.3

Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 4 (2008), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Steven Brown, Michael J. Martinez, and Lawrence M. Parsons. 2006. Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience* 23 (2006), 2791–803. https://doi.org/10.1111/j.1460-9568.2006.04785.x

Guy Thomas Buswell. 1922. *Fundamental reading habits: a study of their development.* University of Chicago, Chicago, Il, USA.

Michel Cara and Gabriela Gomez Vera. 2016. Silent reading of music and texts; eye movements and integrative reading mechanisms. *Journal of Eye Movement Research* 9 (2016). https://doi.org/10.16910/jemr.9.7.2

Max Coltheart. 2000. Commentary on Section 3 - Dual Routes from Print to Speech and Dual Routes from Print to Meaning: Some Theoretical Issues. In *Reading as a Perceptual Process*, Alan Kennedy, Ralph Radach, Dieter Heller, and Joël Pynte (Eds.). North-Holland, Oxford, 475–490. https://doi.org/10.1016/B978-008043642-5/50022-X

Max Coltheart. 2005. Modeling Reading: The Dual-Route Approach. In *The Science of Reading: A Handbook*, Margaret J. Snowling and Charles Hulme (Eds.). John Wiley & Sons, Ltd, Malden, MA, USA & Oxford, UK & Carlton, Victoria, Australia, Chapter 1, 6–23. https://doi.org/10.1002/9780470757642.ch1

Ton de Jong. 2010. Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science* 38, 2 (March 2010), 105–134. https://doi.org/10.1007/s11251-009-9110-0

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

Andrew T. Duchowski, Krzysztof Krejtz, Nina A. Gehrer, Tanya Bafna, and Per Bækgaard. 2020a. The Low/High Index of Pupillary Activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376394

Andrew T. Duchowski, Krzysztof Krejtz, Nina A. Gehrer, Tanya Bafna, and Per Bækgaard. 2020b. The Low/High Index of Pupillary Activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376394

Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. *The Index of Pupillary Activity: Measuring Cognitive Load Vis-à-Vis Task Difficulty with Pupil Oscillation.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173856

Thomas W. Goolsby. 1994. Profiles of Processing: Eye Movements during Sightreading. *Music Perception* 12, 1 (1994), 97–123. https://doi.org/10.2307/40285757

Joyce Eastlund Gromko and Allison Smith Poorman. 1998. Effect on of Music Training Preschoolers ' Performance Task. *Journal of Research in Music Education* 46, 2 (1998), 173–181. https://doi.org/10.2307/3345621

Tuomo Häikiö, Raymond Bertram, Jukka Hyönä, and Pekka Niemi. 2009. Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology* 102, 2 (2009), 167–181. https://doi.org/10.1016/j.jecp.2008.04.002

Eckhard H. Hess and James M. Polt. 1964. Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science* 143, 3611 (1964), 1190–1192.

Steven J Holochwost, Cathi B Propper, and Kelly R Fisher. 2017. Music Education, Academic Achievement, and Executive Functions. *Psychology of Aesthetics, Creativity, and the Arts* 11 (2017), 147–166. https://doi.org/10.1037/aca0000112

Jukka Hyönä and Richard K. Olson. 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, 6 (1995), 1430–1440. https://doi.org/10.1037/0278-7393.21.6.1430

Jukka Hyönä, Richard K. Olson, John Defries, David Fulker, Bruce Pennington, and Shelley Smith. 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 (1995), 1430–1440. https://doi.org/10.1037/0278-7393.21.6.1430

Albrecht Inhoff and Keith Rayner. 1986a. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics* 40 (11 1986), 431–439. https://doi.org/10.3758/BF03208203

Albrecht Werner Inhoff and Keith Rayner. 1986b. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics* 40, 6 (Nov. 1986), 431–439. https://doi.org/10.3758/BF03208203

Debra Jared, Jane Ashby, Stephen J. Agauas, and Betty Ann Levy. 2016. Phonological activation of word meanings in grade 5 readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42, 4 (2016), 524–541.

Artur C. Jaschke, Henkjan Honing, and Erik J. A. Scherder. 2018. Longitudinal Analysis of Music Education on Executive Functions in Primary School Children. *Frontiers in Neuroscience* 12 (2018), 103. https://doi.org/10.3389/fnins.2018.00103

Aleksandra Jaworowska and Teresa Szustrowa. 2003. *Test Matryc Ravena w wersji kolorowej. Formy: klasyczna i równoległa.* Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego, Warszawa, Poland.

Holly S.S.L. Joseph, Kate Nation, and Simon P Liversedge. 2013. Using Eye Movements to Investigate Word Frequency Effects in Children's Sentence Reading. *School Psychology Review* 42, 2 (2013), 207–222. https://doi.org/10.1080/02796015.2013.12087485

Marcel Adam Just and Patricia A Carpenter. 1980. A theory of reading: from eye fixations to comprehension. *Psychological Review* 87, 4 (1980), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Daniel Kahneman and Jackson Beatty. 1966. Pupil Diameter and Load on Memory. *Science* 154, 3756 (1966), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Reinhold Kliegl, Ellen Grabner, Martin Rolfs, and Ralf Engbert. 2004. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology* 16 (2004), 262–284. https://doi.org/10.1080/09541440340000213

Krzysztof Krejtz, Andrew T Duchowski, Anna Niedzielska, Cezary Biele, and Izabela Krejtz. 2018. Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS One* 13, 9 (2018), e0203629. https://doi.org/10.1371/journal.pone.0203629

Krzysztof Krejtz, Justyna Żurawska, Andrew Duchowski, and Szymon Wichary. 2020. Pupillary and microsaccadic responses to cognitive effort and emotional arousal during complex decision making. *Journal of Eye Movement Research* 13, 5 (2020), 15 pages. https://doi.org/10.16910/jemr.13.5.2

Kerry Ledoux, Emily Coderre, Laura Bosley, Esteban Buz, Ishanti Gangopadhyay, and Barry Gordon. 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 (2016), 285–305. https://doi.org/10.3758/s13428-015-0571-6

Mallorie Leinenger and Keith Rayner. 2017. What we know about skilled, beginning, and older readers from monitoring their eye movements: Implications for teaching reading. In *Reading Comprehension in Educational Settings*, José A. León and Inmaculada Escudero (Eds.). John Benjamins Publishing Company, Amsterdam, Philadelphia, USA, 1–27. https://doi.org/10.1075/swll.16.01lei

Sara T. K. Li, Susana T. L. Chung, and Janet H. Hsiao. 2019. Music-reading expertise modulates the visual span for English letters but not Chinese characters. *Journal of Vision* 19, 4 (04 2019), 10–10. https://doi.org/10.1167/19.4.10

Quentin Lohmeyer and Mirko Meboldt. 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)*. Design Society, Milan, Italy, 359–368.

George W. McConkie, David Zola, John Grimes, Paul W. Kerr, Nancy R. Bryant, and Phillip M. Wolff. 1991. Children's eye movements during reading. In *Vision and Visual Dyslexia*, John F. Stein (Ed.). Macmillan Press, London, 251–262.

Fred Paas, Juhani Tuovinen, Jeroen J. G. Van Merrienboer, and Aubteen Darabi. 2005. A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development* 53 (09 2005), 25–34. https://doi.org/10.1007/BF02504795

Jonathan W. Peirce. 2007. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods* 162, 1 (2007), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Tepring Piquado, Derek Isaacowitz, and Arthur Wingfield. 2010. Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* 47, 3 (2010), 560–569. https://doi.org/10.1111/j.1469-8986.2009.00947.x

Mark Polanka. 1995. Research Note: Factors Affecting Eye Movements During the Reading of Short Melodies. *Psychology of Music* 23, 2 (1995), 177–183. https://doi.org/10.1177/0305735695232005

Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk. 2012. *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warszawa. 331 pages.

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

John C. Raven. 1984. *Manual for the Coloured Progressive Matrices (Revised)*. Windsor, UK: NFER-Nelson.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (1998), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Keith Rayner. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62, 8 (2009), 1457–1506. https://doi.org/10.1080/17470210902816461

Keith Rayner and Susan A. Duffy. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14, 3 (1986), 191–201. https://doi.org/10.3758/BF03197692

Keith Rayner and Gary E. Raney. 1996. Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin and Review* 3, 2 (1996), 245–248. https://doi.org/10.3758/BF03212426

Monika Riegel, Małgorzata Wierzba, Marek Wypych, Łukasz Żurawski, Katarzyna Jednoróg, Anna Grabowska, and Artur Marchewka. 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 (2015), 1222–1236. https://doi.org/10.3758/s13428-014-0552-1

E. Glenn Schellenberg. 2004. Music lessons enhance IQ. *Psychological Science* 15, 8 (2004), 511–514. https://doi.org/10.1111/j.0956-7976.2004.00711.x

Jens Schmidtke. 2014. Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology* 5 (2014), 1–16. https://doi.org/10.3389/fpsyg.2014.00137

Jens Schmidtke. 2018. Pupillometry in linguistic research - an introduction and review for second language researchers. *Studies in Second Language Acquisition* 40 (09 2018), 529–549. https://doi.org/10.1017/S0272263117000195

John A. Sloboda. 1977. Phrase units as determinants of visual processing in music reading. *British Journal of Psychology* 68 (1977), 117–124. https://doi.org/10.1111/j.2044-8295.1977.tb01566.x

John Sweller. 2010. Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review* 22, 2 (June 2010), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Stanford E Taylor. 1965. Eye Movements in Reading: Facts and Fallacies. *American Educational Research Journal* 2, 4 (Nov. 1965), 187–202. https://doi.org/10.2307/1161646

Simon P. Tiffin-Richards and Sascha Schroeder. 2015. Word length and frequency effects on children's eye movements during silent reading. *Vision Research* 113 (2015), 33–43. https://doi.org/10.1016/j.visres.2015.05.008

Helmut Wilhelm, Barbara Wilhelm, and Holger Lüdtke. 1999. Pupillography: Principles and applications in basic and clinical research. In *Pupillography: Principles, methods and applications*, J. Kuhlmann and M. Böttcher (Eds.). Zuckschwerdt Verlag, Müñchen, Germany, 1–11.