

Fixation duration and the learning process: an eye tracking study with subtitled videos

Shivsevak Negi
Indian Institute of Technology
Bombay, Mumbai, India

Ritayan Mitra
Indian Institute of Technology
Bombay, Mumbai, India

Learning is a complex phenomenon and education researchers are increasingly focussing on processes that go into it. Eye tracking has become an important tool in such research. In this paper, we focus on one of the most commonly used metrics in eye tracking, namely, fixation duration. Fixation duration has been used to study cognition and attention. However, fixation duration distributions are characteristically non-normal and heavily skewed to the right. Therefore, the use of a single average value, such as the mean fixation duration, to predict cognition and/or attention could be problematic. This is especially true in studies of complex constructs, such as learning, which are governed by both cognitive and affective processes. We collected eye tracking data from 51 students watching a 12 min long educational video with and without subtitles. The learning gain after watching the video was calculated with pre- and post-test scores. Several multiple linear regression models revealed a) fixation duration can explain a substantial fraction of variation in the pre-post data, which indicates its usefulness in the study of learning processes; b) the arithmetic mean of fixation durations, which is the most commonly reported eye tracking metric, may not be the optimal choice; and c) a phenomenological model of fixation durations where the number of fixations over different temporal ranges are used as inputs seemed to perform the best. The results and their implications for learning process research are discussed.

Keywords: Learning process; Eye tracking; Fixation duration distribution; Multiple linear regression; Subtitled educational video.

Introduction

Research in education investigates learning through three distinct prisms. The first prism deals with what goes into learning, which includes the learning environment, learning content, and teacher and learner characteristics. The second prism involves the learning outcomes that are often based on summative features, such as grade point averages (GPA) or test scores. The third prism concerns itself


with the learning processes that link the first and the third prisms (Harteis, Kok, & Jarodzka, 2018). This area of research investigates the cognitive and affective mechanisms, strategies, and mental states that learners employ and how those might interact to produce the complex construct of learning. There is comparatively less research on learning processes but it has found a new lease of life with advanced computer hardware and psychophysiological sensor technologies such as eye tracking (Hegarty, Mayer, & Monk, 1995; Kardan & Conati, 2012), facial emotion recognition (Bahreini, Nadolski, & Westera, 2016), EEG (Berka et al., 2007) and GSR (McNeal, Spry, Mitra, & Tipton, 2014). These technologies have the ability to track neurophysiological and behavioral parameters at an unprecedented temporal resolution, which can shed light into the minute by minute breakdown of the underlying processes in learning, such as cognition, affect, metacognition

Received April 12, 2020; Published August 16, 2020.

Citation: Negi, S., & Mitra, R. (2020). Fixation duration and the learning process: an eye tracking study with subtitled videos. *Journal of Eye Movement Research*, 13(6):1.

Digital Object Identifier: 10.16910/jemr.13.6.1

ISSN: 1995-8692

This article is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). 

and motivation, or what is often referred to as the CAMM processes (Azevedo, Taub, & Mudrick, 2017). What was earlier possible only through post-hoc surveys, interviews, human observations, and think-aloud protocols can now be achieved easily through non-intrusive computer technologies and that too at an unprecedented resolution.

Eye tracking has been particularly useful in the study of learning processes. For example, Liu (2014) used total fixation duration, a commonly used eye tracking metric, to study learners' reading processes when working with concept maps. The author found shorter total fixation duration near task-relevant areas when concept maps were used as a study aid implying easier comprehension of content. The scan-paths of fixations revealed concept map reading strategies. These results help us understand not only if concept mapping was useful as a study aid, which can be equally obvious from pre-post test scores in a between-subjects study design, but they also tell us how and why we observed differences in learning gain. Similarly, Mitra, McNeal, and Bondell (2017) studied graphical problem solving using eye tracking data. They delineated problem-relevant graphical elements and used fixations to reveal strategic differences between high and low performers. As pupil diameters are correlated with cognitive load the authors used this proxy relationship to differentiate between high and low performers based on their cognitive load demands.

Role of average fixation duration in learning process research

Just and Carpenters' influential eye-mind hypothesis proposed that there is no significant latency between what the eyes fixate on and what the mind processes (Just & Carpenter, 1980). This set the stage for modern-day eye tracking research in the early 1980s. Ever since, average fixation duration has occupied a key space in eye tracking research (Rayner, 1998; Rayner, Chace, Slattery, & Ashby, 2006; Underwood, Hubbard, & Wilkinson, 1990). As much as this early body of work was situated in reading research the usefulness of fixation duration soon started percolating to other areas such as visual search, auditory language processing, mathematics, numerical proficiency, problem-solving, multiple representations, etc. (Rayner, 1998). Between 1990 and 2000 approximately 1300 scientific articles had fixation duration in the title. That count has multiplied more than ten times between 2010 and the publication date of this article.

Rayner (1998) in a comprehensive review of eye tracking research of the past twenty years, had suggested that eye movement data held great promise in information processing tasks but significant differences exist between task categories and it may not be prudent to generalize findings across categories. This caveat is easily verified through a meta-analytic study of expertise development in visual comprehension, which found high heterogeneity in effect sizes (Gegenfurtner, Lehtinen, & Säljö, 2011). Average fixation durations showed no clear pattern (indicating opposing directionality between studies) for either task complexity or visualization characteristics. In particular, the sum of evidence revealed that although experts indeed differed from novices in terms of having shorter fixation durations, they did not differ significantly from intermediates. More importantly and quite counter-intuitively, the intermediates seemed to have longer fixation durations than the novices.

The idea that average fixation duration should have some bearing on the processing demands of a task is almost intuitive. More complex tasks should require more processing time and hence it is expected that subjects may need to look longer to process all the information provided by the stimuli. Likewise, with growing expertise this processing demand relaxes because the expert can pull schemata relevant to the task from long-term memory, thereby reducing processing demands (Mitra, McNeal, & Bondell, 2017). A much less appreciated aspect of average fixation durations is whether the underlying construct of average fixation duration is stable across contexts. For example, if the task is a short duration visual search where participants need to quickly locate a target among distractors then the underlying construct that mediates average fixation duration could be attention, alertness, focus, concentration or some similar construct (Byrne, Anderson, Douglass, & Matessa, 1999). However, if the task changes to attending an online lecture for a long duration (several minutes to an hour) and the goal is comprehension (as opposed to search) then fixations could represent a broader range of underlying constructs that they are responsive to. If we simply use average fixation duration to indicate learning then it would be a mistake because fixation durations are not likely to be influenced by some 'super-construct' of learning. Instead, we argue, fixation durations could be influenced by latent affective processes that lead to learning. For example, a learner could be distracted, bored, confused or engaged, at different times during the lecture, which would influence his learning and should, hypothetically, leave their imprint

on fixation durations as well. Given this complexity and temporal longevity of learning, the use of an average value for fixation durations as a proxy of learning would be a simplistic approximation, and possibly erroneous, in the context of educational research.

Classification of fixation types

The distribution of fixation durations is quite predictable, famously non-normal and characteristically skewed to the left with typical median durations of 200-250 ms, mean durations of 300-350 ms, and an extended right tail of long and very long fixations (Rayner, 1998; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky et al., 2019) (see Fig. 1). In eye movement research fixations have been traditionally classified into ambient and focal types. Ambient visual fixations are short in duration and do not permit for conscious identification of visual objects. On the contrary, focal fixations are comparatively long in duration and reflect the conscious perception of objects. In spatial processing research, they seem to reflect either involuntary spatial processing (ambient) or conscious perception of visual objects (focal). Although the general definition and role of ambient and focal fixations are agreed upon their classification is not as straight-forward.

Velichkovsky et al. (2002) originally characterized fixations as ambient or focal based on their durations and the amplitude of the succeeding saccade. Unema et al. (2005) further suggested that larger saccade amplitudes and shorter fixation durations during the initial image viewing period represented ambient processing and that smaller saccade amplitudes and longer fixation durations during the later viewing period represented focal processing. Providing further support for such classification between ambient and focal visual processing, Helmert et al. (2005) found that viewers were more likely to remember objects when fixation duration was long but the subsequent saccade amplitude was short, compared to when fixation duration was short and the subsequent saccade amplitude was long. Krejtz et al. (2012b) used an ambient/focal attention coefficient, defined as the relation between the current fixation duration and the subsequent saccade amplitude to

classify fixations. Krejtz et al. (2016) formally defined the ambient/focal attention coefficient proposed by Krejtz et al. (2012b) as K , which acts as a dynamic indicator of fluctuations between ambient and focal visual processing modes and permits statistical comparison between individuals and groups. They suggested negative values of coefficient K represent ambient processing and positive values represent focal processing. The interpretation of the null value of K remained ambiguous; it may mean long (short) fixations followed by long (short) saccades. Although both seem to be rare, a better understanding of the underlying process is desirable yet currently unavailable. In another study, Holmqvist et al. (2011) discussed a method of classification of focused (focal) versus overview (ambient) visual behavior. This technique was based more on transitions between areas of interest (AOI) rather than on the relation of fixation durations and subsequent saccade amplitudes, and therefore, the findings are not easily generalizable to studies that did not look at transitional analysis of eye movement data. In the aforementioned studies, it seems that classification was primarily based on saccade amplitude with relatively less reliance on exact fixation durations. Follet et al. (2011) even claimed that classification of visual fixations can be achieved using saccade amplitude alone and discounted the role of fixation duration in such classification (Table 1).

A different approach in ambient-focal classification is based only on fixation durations without considering saccade characteristics at all (Velichkovsky, 1999). Such classification suggests short fixations (50-150 ms) to be of ambient type while long fixations around 300-500 ms are considered to be focal (Velichkovsky et al., 2019). These two populations of fixations, ambient and focal, with small differences in the actual ranges used by researchers (see Table 1), have often been reported in eye-movement research. It may be noted, however, the fixation durations for the corresponding categories are different across the two classification schemes that we discussed. For example, a common range in the fixation only classification scheme happens to be around 300-500 ms, which is greater than

Table 1. Summary of literature review based on ambient and focal fixation classification.

Classification Scheme	Context	Ambient	Focal	Reference	Comments
Fixation duration and saccade amplitude	Static and dynamic scene perception	90 - 260 ms with Saccade amplitude >5°. Actual range <180 ms	Above 260 -280 ms Saccade amplitude < 5°. Actual range >180 ms	Helmert et al., 2005	Only 20 - 600 ms range was analyzed.
	Visual perceptual processing- Dynamic road traffic scenario	90-300 ms (large scale ambient fixation), long-range saccade amplitude	Above 280-300 ms with short-range saccades amplitude, Attentive processing	Velichkovsky et al., 2002	Fixations around 200 ms were categorized as ambient and around 400 ms as focal. Fixations <90 ms discarded.
	Visual perceptual processing - Complex picture	Relatively long-range saccades, >5°, Short fixations	Relatively short-range saccades, <5°, Long fixations	Pannasch et al., 2011	
	Scene viewing-initial and late viewing phase	<180 ms followed by saccade amplitudes larger than average subsequent saccade	>180 ms followed by saccade amplitudes smaller than average subsequent saccade	Unema et al., 2005	Only less than 600 ms range was analyzed.
	Scene viewing-different age groups	<180 ms followed by large saccades (> 5°)	>180 ms followed by short saccades (< 5°)	Helo et al., 2014	Fixations less than 90 ms were discarded.
	Gaze-controlled interface	100 - 250 ms z-transformed saccade amplitude <0	Individual mean fixation duration. (270 - 380 ms)	Velichkovsky et al., 2014	
	Developmental changes in focal-ambient visual processing	<180 ms followed by large saccades (> 5°)	>180 ms followed by short saccades (< 5°)	Krishna et al., 2017	Only 20 - 400 ms range was analyzed.
	Free scene viewing	Saccade amplitude (Mean = 2.5°)	Saccade amplitude (Mean = 11°)	Follet et al., 2011	Only saccade amplitude was used.
Fixation duration	eSports- Pro and immature players	50 - 150 ms	300 - 500 ms	Velichkovsky et al., 2019	
	Information processing in risky decisions	<250 ms	>500 ms	Glockner et al., 2011	Also reported 251-500 ms fixations but those were not used to test hypothesis.
	Language processing	Mean = 225 ms	Mean = 400 ms, more effortful processing	Rayner, 1998	Focal range implies more effortful processing.
	Information visualization	<200 ms, involuntary processing	300 - 500 ms	Bylinskii et al., 2015	Focal processing is referred to as deeper exploration of visual elements. Study also stated that fixations < 300 ms are not encoded in memory.
	Memory performance, levels of performance	120 - 250 ms Shallow perceptual processing	250 - 450 ms Semantic encoding Also mentioned 300 -450 ms range	Velichkovsky, 1999	

the values of focal fixations suggested in the classification based on fixation and saccade amplitude (Table 1).

The fixation durations in an eye tracking dataset vary from as small as 50 ms to several seconds (over 2000 ms) (Buswall, 1935; Velichkovsky et al., 2019). The aforementioned ambient/focal classification of fixations does not adequately represent this wide range. Indeed, as indicated in Table 1, several authors have not used the full range of fixation durations and have restricted their analyses to a few hundreds of milliseconds. Such decisions are possibly justified based on the type of tasks that have accompanied such studies as explained in the Introduction. However, in case of learning tasks, restricting our analysis to such a narrow range of fixation durations could be fraught with errors as very long fixations could indicate negative academic affects such as confusion, boredom, and frustration. Some empirical results suggest very long fixation durations to be indicative of negative affect such as those related to confusion (Ehmke & Wilson, 2007; Goldberg & Helfman, 2010; Salminen et al., 2018) or cognitive load (Goh et al., 2013), although a potential confound with interest cannot be ruled out (Granka, Feusner, & Lorigo, 2008).

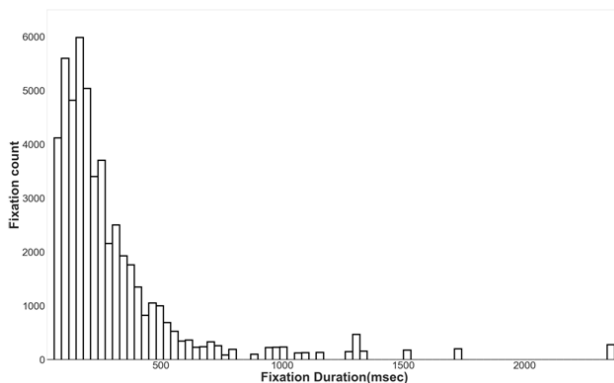


Figure 1. Representative frequency distribution of raw fixation durations (n = 51238) from a single participant.

Research Questions

In this study we look at fixation duration distributions and their usefulness in the study of learning processes. As outlined in the Introduction and Table 1, our understanding of fixation duration is largely from studies outside the context of learning and can, therefore, pose a challenge. Hence, evaluation of the role of fixation duration in learning process research is quite timely and useful. Fixation

durations are typically right-skewed with a long tail and previous workers have outlined the heterogeneity of fixation vis-a-vis the type of cognitive processing that they represent. In view of such evidence, we think it is apt to re-evaluate the role of average fixation duration in learning process research. The use of a mean fixation duration may be unjustified both from a geometric perspective as well as a process perspective. A skewed distribution is unlikely to be well-represented by a single mean value of the fixation durations and studies that have used this common metric (Ozcelik, Karakus, Kursun, & Cagiltay, 2009; Hsiao-Ching She e al., 2009; Asselen et al., 2010; Meng-Jung Tsaia et al., 2011; Kuei-Pin Chien et al., 2015; Kimario Nizetha Daniel & Eiji Kamioka, 2017) could have extracted useful insights had they incorporated fixation distribution characteristics in their analysis. This concern is purely geometric in nature and is solely dependent on the shape of the distribution regardless of the cause behind such a distribution. More importantly, in a context as complex and lengthy as learning, a single mean may be a conceptually unsound choice as well because of the underlying affective controls of learning and how those affective states influence fixation durations. Specifically, for the purpose of this study, we have the following research questions.

RQ1: Is there empirical evidence to justify the use of average fixation duration for research on learning processes?

RQ2: Is average fixation duration a *necessary* and *sufficient* measure to study learning processes?

RQ3: Is there an alternative model for the representation of learning processes with fixation durations?

Methods

Participants

A total of 51 male participants (16-18 years old with no vision problem) from a polytechnic institute in India voluntarily participated in the study. The exclusively male population was not because of selection bias as enrolled students were exclusively male as well. The goal of polytechnic institutes in India, as elsewhere, is to make students employable for entry-level industry jobs by imparting industrial training using an apprenticeship model of learning. This institute in particular is geared toward supplying the lowest rung of employment such as for jobs as

an electrician, automobile mechanic, plumber, refrigerator technician, etc. As a result, a large majority of the student population are from socio-economically backward strata of the society with an average annual income of less than 4600 USD. The participants had very poor English comprehension skills and everyone spoke the same native language, Marathi. Though the medium of instruction in the institute was English (L2) the rudimentary English language skill, both verbal and written, necessitated any conversations with the students to be conducted primarily in the Marathi language. The poor financial condition compels them to leave mainstream education such as 4 years engineering/medical/commerce degree programs, and join such short term diploma courses (one to two years in duration) that impart job-ready industrial skills. The minimum qualification of enrolment was passing of the 10th grade examination.

The study was approved by the local institutional ethics committee (No.IITB-IEC/2019/012) and all participants provided informed written consent.

Materials

The materials used in this study included a demographic data sheet, educational video, a pre-test, and a post-test. The demographic questions pertained to age, enrolled courses, last class attended, family income, etc. In consultation with the course instructor, a 12 minute long educational video introducing the basic concept and architecture of the World Wide Web and the internet was chosen. The voice-over was in Indian English and the video lecture contained presentation slides with textual and graphical information. The VLC player was used to embed either English or Marathi subtitles. A maximum of two lines were used to display subtitles on the screen. Before eye tracking data collection participants' prior knowledge about the topic was assessed with the pre-test comprising 14 multiple choice questions, which were designed in consultation with the course instructor. Similarly, 14 multiple choice questions, different from the pre-test questions but of similar difficulty level, were used to formulate the post-test questionnaire. The pre- and post-test questionnaires were drafted in English and language experts translated all questions to Marathi so that the students could have both versions, English and Marathi, when answering the questions. Both the tests were untimed.

Procedure

The participants were instructed to watch the video lecture after completing the demographic survey and the pre-test questionnaire. A screen-based Tobii eye tracker (X3-120) operating at 120 Hz recorded the eye movements of the participants as they watched the video lecture. The stimulus was displayed on a 17-inch screen and participants were seated comfortably in a sufficiently illuminated room on a stable chair at a distance of approximately 680 mm from the stimulus screen. The minimum fixation duration that can be recorded by the eye tracker was 66 milliseconds. A five-point calibration procedure was used to calibrate each participant's eyes before the experiment. During the recording session, participants were not allowed to use the mouse and keyboard or pause video. Tobii Pro Lab (version 1.123) was the software used for data collection and analysis. A post-test was administered immediately after the participants completed watching the video lecture.

The video was watched with Marathi subtitles (MS) by 19 participants, with English subtitles (ES) by 16 participants and with no subtitles (NS) by 16 participants. One goal of our study was investigating the impact of subtitles within educational context hence a between-subjects study design was chosen (Kruger, 2013; Kruger, Hefer & Matthew, 2014). We will report group differences of self-reported measures, performance and eye tracking measures separately. In this article, however, we choose not to focus on group differences. All statistical analyses were conducted with Jamovi (Version 0.9.5.13; Love, 2018), an open-source software project.

Results

Learning gain

A total of 6 participants (3 from MS, 1 from ES and 2 from NS) were rejected at the onset of analysis due to very low eye gaze data collection. Learning gain was computed using Hake's formula given by $(S^{\text{post}} - S^{\text{pre}}) / (S^{\text{max}} - S^{\text{pre}})$, where S^{post} represents the post-test score, S^{pre} represents pre-test score and S^{max} represents the total score for the test (Hake, 1998). The learning gains of 2 participants in the ES group and 1 participant in the NS group were more than 2 SD ($SD=28.7$) away from the combined group (MS+ES+NS) mean ($M=0.439$) and excluded from further analysis (Fig. 2). One way ANOVA of learning gains

between the three groups was insignificant [$F(2, 39) = 0.524, p = 0.596$].

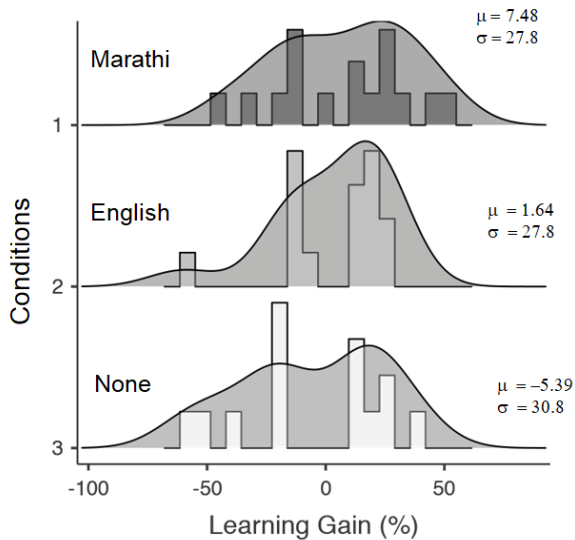


Figure 2. Histograms representing distributions of Hake's learning gain calculated for each participant under three different conditions (Marathi subtitle, English subtitle and No subtitle). The group means and standard deviations are reported.

Eye Movement Analysis

Two areas of interest (AOIs) were identified based on heat maps (Fig. 3), namely, content ($Content_{AOI}$) and subtitle ($Subtitle_{AOI}$). The content area displayed the main learning content; however, such nomenclature does not imply subtitles do not provide any learning content. The descriptive statistics of fixations recorded for the three groups are provided in Table 2. The pooled fixation durations from the MS and ES groups were used for the multiple regression analysis ($N=29$). The data from the NS group ($N=13$) were analyzed separately (see Discussion). The reason for clubbing the two groups with subtitles for analysis stems from a learning process perspective. Subtitles generate a lot of attention as evident from the heat maps of fixations (Fig. 3a, 3b). The conscious act of reading a subtitle or not can influence understanding of the content. Although the subtitle language can influence such comprehension, as suggested by the difference in fixation density in Figs. 3a and 3b ($Subtitle_{AOI}$), the process of sense-making from subtitled videos is expected to remain similar across language groups, as revealed in the similarities between Fig. 3a and 3b. The same cannot be said about videos without any subtitling (Fig. 3c). Furthermore,

from an analytical perspective, this grouping ensures parity between NS and the other two groups as the areas of interest, $Content_{AOI}$ and $Subtitle_{AOI}$, are relevant only for the MS and ES groups but not for the NS group.

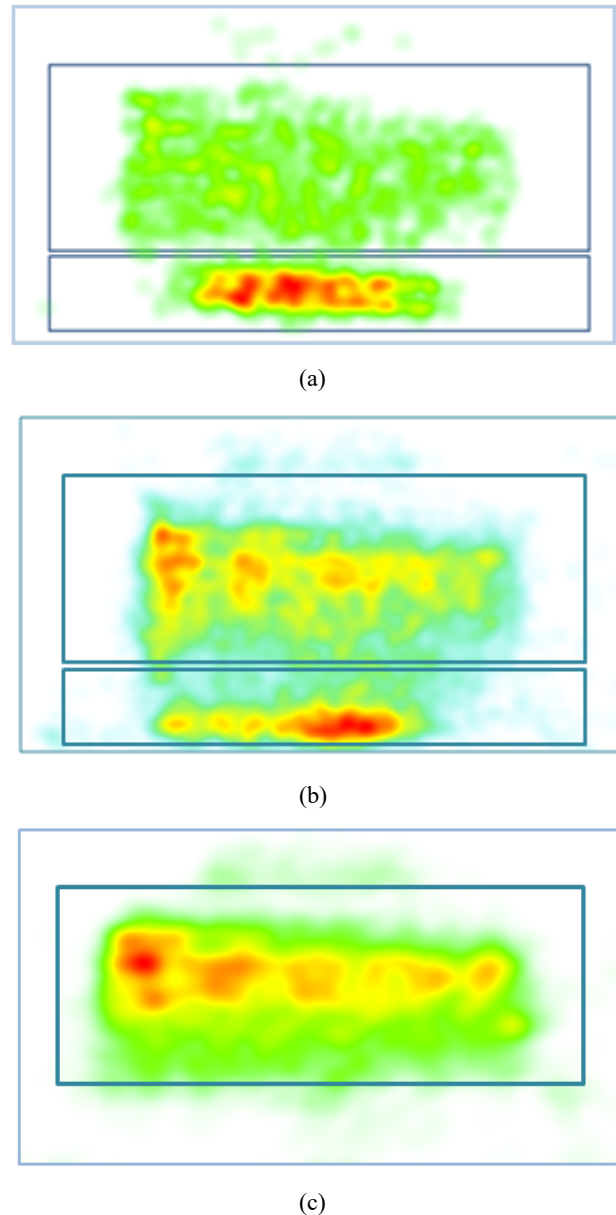


Figure 3. Heat maps for (a) MS, (b) ES, and (c) NS groups. It is evident that MS and ES groups paid attention to the subtitle area apart from attending to the content area (more red/dark shades). Red/dark shade indicates high density of fixations and green/light shade indicates low density. The two analytical sections are the $Content_{AOI}$ (upper box) and $Subtitle_{AOI}$ (lower box, when present).

Table 2. Descriptive statistics for the three groups (MS - Marathi subtitle; ES - English subtitle and NS - No subtitle) and two areas of interest (AOI) – content and subtitle. FD stands for fixation duration in milliseconds. FC stands for fixation counts and the subscripts refer to the range in milliseconds. Kurtosis and Pearson’s coefficient of skewness are also reported.

Groups	FD _{Mean}	FD _{Median}	Std Dev	Skewness	Kurtosis	FC ₍₆₆₋₁₅₀₎	FC ₍₃₀₀₋₅₀₀₎	FC _(>1000)
Content _{AOI}								
MS	330	241	395.5	0.67	61.9	101682	128551	11866
ES	321	242	345.9	0.72	92.9	112722	112915	13764
NS	274	208	288.4	0.80	73.9	132946	63267	8984
Subtitle _{AOI}								
MS	240	208	158	0.61	9.0	36065	79254	816
ES	272	183	333	0.89	26.5	29605	46392	1852

Multiple linear regression analyses were conducted to model the variability in post-test scores caused by different subsets of independent variables, namely, the pre-test scores, the mean of fixation durations, the median of fixation durations, and the three distinct non-continuous and non-overlapping fixation duration categories that are based on prior work (see Introduction and Table 1), after controlling for subtitle language. The number of fixations less than 150 ms (but greater than 66 ms), F₁₅₀, loosely correspond to the ambient fixations of prior work. Similarly, the number of fixations between 300 and 500 ms, F₃₀₀₋₅₀₀, represents focal fixations, and the number of fixations greater than 1000 ms, F₁₀₀₀, includes fixations that may be considered to be too long. The results from the multiple linear regression models are provided in Table 3. For all the models, assumptions of autocorrelation and normality of residuals were not violated and corresponding test results remained within reasonable values. Some specific cases of multicollinearity are discussed in the appropriate context below.

Subtitle

For the preliminary model we use only pre-test scores, which expectedly predicts post-test scores and explains 21% of the variation (F=4.82, p= 0.017, AIC_w=0.018, Adjusted R² = 0.21). On adding mean as a predictor the model remains comparable (F= 3.75, p= 0.024, AIC_w=0.018, Adjusted R² = 0.23). Upon adding median to the list of predictors there is a substantial increase in predictability to 34% (F=4.6, p= 0.007, AIC_w=0.08, Adjusted R² = 0.34). We do note that the obvious concern of multicollinearity is valid in this case as revealed by the high VIF (VIF=11.87). However, contrary to popular belief “the fact that some or all predictor variables are correlated among themselves does not, in general, inhibit our

ability to obtain a good fit nor does it tend to affect inferences about mean responses or predictions of new observations” (Kutner, 2013, p. 289), although it does affect the coefficients and p-values of the predictors that show multicollinearity.

Excluding both the measures of central tendency, only pre-test scores and F₁₅₀ provide a reasonably good fit with an adjusted R² value of 0.3 (F=5.0, p= 0.007, AIC_w=0.048). Adding F₃₀₀₋₅₀₀ to the model does not yield any substantially different model fit (F=3.92, p= 0.014, AIC_w=0.029, Adjusted R² = 0.295). However, adding F₁₀₀₀ clearly improves the model to a predictability of 39.5% (F=4.66, p= 0.004, AIC_w=0.217, Adjusted R² = 0.395). A more parsimonious model with marginally better adjusted R² value and lowest AIC is obtained by excluding F₁₅₀ from this model (F=5.85, p= 0.002, AIC_w=0.590, Adjusted R² = 0.409). This could indicate some collinearity issues between F₁₅₀ and F₃₀₀₋₅₀₀ although the VIF measure were reasonable in this case (F₁₅₀, Tolerance = 0.6, VIF = 1.67; F₃₀₀₋₅₀₀, Tolerance = 0.88, VIF = 1.14; Pearson’s r = 0.61).

Content

The model for pre-test remains the same for both categories. However, the model does not appreciably improve upon adding mean (F=3.92, p= 0.02, AIC_w=0.253, Adjusted R² = 0.24) and median (F=2.93, p= 0.042, AIC_w=0.093, Adjusted R² = 0.216) although the model with pre-test and mean performs the best out of these three models. Without the central tendency measures and adding only F₁₅₀ to pre-test scores makes for a worse fit (F=3.13, p= 0.043, AIC_w=0.093, Adjusted R² = 0.19). Adding F₃₀₀₋₅₀₀ does not improve the model fit (F=2.41, p= 0.077, AIC_w=0.056, Adjusted R² = 0.17). Adding F₁₀₀₀ improves the model substantially with a predictability of 27.8% (F=3.16, p= 0.026, AIC_w=0.253, Adjusted R² = 0.278).

Table 3. Results of multiple linear regression models. Independent variables (IVs) are listed in the far left column. Each model has all the IVs from the one above except in the last model of Subtitle_AOI where all the IVs are listed. Significance levels < 0.01 , < 0.05 and < 0.1 are marked as **, *, and /, respectively.

Subtitle_AOI														
Model	Std. Coefficients (95% CI)						Overall Model Test					AIC		Adjusted R ²
	Subtitle language	Pre-test	Mean	Median	F	df1	df2	p	AIC	Δ AIC	AICw			
Pre-test	-0.23(-0.58/0.13)	0.43*(0.073/0.78)			4.82	2	26	0.017	113	7	0.018	0.21		
	-0.23(-0.58/0.19)	0.37*(0.004/0.73)	-0.21(-0.56/0.15)		3.75	3	25	0.024	113	7	0.018	0.23		
	-0.005(-0.39/0.38)	0.41*(0.074/0.75)	-1.34*(-2.4/-0.27)	1.2*(0.12/2.3)	4.6	4	24	0.007	110	4	0.080	0.34		
	F_150			F_300-500			F_1000							
150	-0.19(-0.52/0.15)	0.49*(0.15/0.78)	0.33 \mp (-0.0023/0.65)		5	3	25	0.007	111	5	0.048	0.3		
500	-0.13(-0.5/0.23)	0.51***(0.15/0.88)	0.21(-0.21/0.63)	0.02(-0.26/0.66)	3.92	4	24	0.014	112	6	0.029	0.295		
1000	-0.08(-0.41/0.26)	0.45*(0.11/0.80)	0.13(-0.27/0.52)	0.31(-0.13/0.75)	4.66	5	23	0.004	108	2	0.217	0.395		
Pre-test	-0.063(-0.4/0.26)	0.47***(0.133/0.80)		0.40*(0.064/0.73)	5.85	4	24	0.002	106	0	0.590	0.409		
	Content_AOI													
Model	Std. Coefficients (95% CI)						Overall Model Test					AIC		Adjusted R ²
	Subtitle language	Pre-test	Mean	Median	F	df1	df2	p	AIC	Δ AIC	AICw			
Pre-test	-0.23(-0.58/0.13)	0.43*(0.07/0.78)			4.82	2	26	0.017	113	0	0.253	0.21		
	-0.22(-0.60/0.13)	0.36 \mp (-0.01/0.72)	-0.24(-0.6/0.13)		3.92	3	25	0.02	113	0	0.253	0.24		
	-0.23(-0.60/0.13)	0.35 \mp (-0.02/0.72)	-0.4(-1.11/0.32)	0.19(-0.53/0.9)	2.93	4	24	0.042	115	2	0.093	0.216		
	F_150			F_300-500			F_1000							
150	-0.22(-0.58/0.15)	0.44*(0.07/0.82)	-0.056(-0.43/0.32)		3.13	3	25	0.043	115	2	0.093	0.19		
500	-0.2*(-0.57/0.18)	0.46*(0.08/0.85)	-0.12(-0.54/0.3)	0.14(-0.3/0.53)	2.41	4	24	0.077	116	3	0.056	0.17		
1000	-0.15(-0.5/0.2)	0.48*(0.12/0.84)	-0.33(-0.77/0.11)	0.46 \mp (-0.03/0.96)	3.16	5	23	0.026	113	0	0.253	0.278		

Discussion

RQ1: Is there empirical evidence to justify the use of fixation durations for research on learning processes?

In all the multiple linear regression models pre-test scores alone have a low yet consistent explanatory power of the variability observed in the post-test scores. This is only to be expected because pre-test scores test prior knowledge, which is an important indicator of performance in post-tests. If fixations had not played any role in the underlying learning process then the addition of such predictors should not have improved the predictive power of the models. However, most of our subsequent models that used at least one measure of fixation duration (either a central tendency measure or a fixation duration count measure) explain more of the variability in the learning gain compared to the model with only pre-test scores. This indicates that something fundamentally linked to performance in the tests was being captured by the fixation measures and learning would be the most likely candidate for that.

The model results provide further insight. The models for the subtitle section with a highest adjusted R^2 value of 0.409 clearly outperform the models of the content section with a maximum adjusted R^2 value of 0.278. In other words, fixation duration should have higher explanatory power when subtitles were included. Heat maps of fixations for the MS and ES groups also indicate the relative importance participants placed on these two sections (Fig. 3a, 3b). Indeed, if no subtitles were provided, then fixation durations should have low explanatory power and the overall model fit should deteriorate as learning can be expected to suffer without any form of subtitling.

In this context, it would be useful to check how such a model performs with the NS group data ($N=13$) where there was no subtitling. As expected, the model predictability was low for any combination of predictors, including pre-test scores only. The best model was with pre-test scores alone ($F=2.60$, $p=0.135$, Adjusted $R^2=0.118$). By adding fixation based predictors the model fit worsened. Thus, neither the pre-test score nor any of the fixation measures, or any of the combinations thereof, were able to predict the variability observed in the post-test scores with reasonable success. This is consistent with our expectation that without any subtitling this group would remain confused about the material being taught and the poor

performance could be symptomatic of either poor comprehension or random test-taking.

We also need to ask a related question; can we quantify the learning that takes place? A definitive answer to this question is outside the scope of this work but we do want to highlight the fact that the results open up interesting possibilities for using fixation duration as a *process metric*. In this operationalization of the term, a process metric is a small step ahead of a *process measure*, which we define as simply a tool or a set of tools to understand the learning process. For example, scan paths have been used as a tool to reveal the process of learning (Goh et al., 2013) and can be considered as a process measure. However, for it to be considered as a metric the variability of scan paths across tasks and user types need to be quantified. This would be a challenging task given the exploratory nature of scan paths and the inherent qualitative nature of interpretation that usually accompanies it. In our opinion, something that is easily quantifiable such as fixation duration would be more suitable as a learning process metric. These results suggest such a possibility but before we arrive at such a metric generalizability of these results need to be tested with a variety of learning media and contexts.

RQ2: Is the mean fixation duration a necessary and sufficient measure to study learning processes?

Several studies have used the arithmetic mean of fixation durations. As discussed earlier the fixation duration distribution is typically right-skewed and it has been suggested that the mean may not be capturing the fixation duration distribution adequately. In RQ2, we have proposed to test this premise by asking whether the mean is a necessary and/or a sufficient measure of fixation durations.

The best model for the subtitle section (Adjusted $R^2=0.409$) as well as the content section (Adjusted $R^2=0.278$) does not require the mean which suggests that mean is not a necessary measure. Also, we do see a marked improvement in model fit when the median is used with the mean (Adjusted $R^2=0.23$ increases to 0.34). This suggests that simply using the mean may not sufficiently capture the variability in test performance.

This finding also flags a more subtle concern in using a mean value for fixation durations and one that we have discussed earlier in the Introduction. The fact that even a combination of mean and median can fall short of explaining the variability observed in the dependent variable indicates that the problem probably rests on what those values

truly represent than the geometric nature of the distribution. If the usage of the mean of fixation durations suffered only because of the underlying non-normality and/or skewness then addition of a median would have invariably improved the model fit. However, the fact that this is not seen in the data indicates a different underlying cause. Averages of any kind, be it mean or median, assumes the values being averaged to be of one kind. If learning were a unidimensional construct and the fixation durations reflected that construct and only that construct alone, then averaging would have, at least, had some conceptual underpinning. However, several emotions regulate our learning and each of these emotions may have different fixation duration signatures. In that case, using an average measure is more than mathematically unsound; it is conceptually unsound as well. In the following section, we propose an alternative model to better reflect the fixation duration distribution within a learning context.

RQ3: Is there an alternative model for the representation of learning processes with fixation durations?

From the results discussed so far, we propose a phenomenological tripartite affective model of fixation durations. In this model, very small fixations (less than $\sim F_{150}$) represent negative valence (not focusing, being distracted, etc.), medium duration fixations ($\sim F_{300-500}$) represent positive valence (such as, focused attention) that is beneficial to learning and very long durations of fixation ($\sim F_{1000}$) represent negative valence (e.g., zoning out or confused) that is not beneficial to learning. This tripartite model is loosely in keeping with the observations from previous studies that were presented in the Introduction. We, however, refrain from pinning down the exact emotions vis-a-vis their fixation duration signatures and restrict ourselves to positive valence/beneficial to learning and negative valence/detrimental to learning classification only. Based on our empirical results if such a model is found to be valid we would argue the explanatory power of the measures of central tendency is only as much as those measures are tied to the most task-relevant fixation category. For example, the mean may have a greater predictive power in a learning scenario if and only if the mean value represents the category (ies) of fixation; namely, F_{150} , $F_{300-500}$, and F_{1000} , that was (were) most relevant to the task and/or user. Likewise, mean and median together would have greater predictive power if the two categories best represented by these values were relevant for the task and/or user.

We now use empirical evidence from this study to validate such a tripartite model. Fig. 4 demonstrates an ensemble of confidence intervals (95% CI) for the standardized coefficients of the various models outlined in Table 3. The pre-test score coefficients have 95% CIs that are greater than zero, which indicates, as expected, an unambiguous positive correlation between pre- and post-test scores.

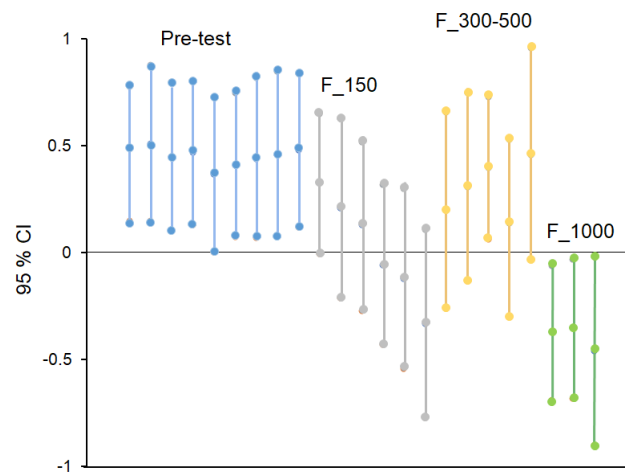


Figure 4. An ensemble of confidence intervals (95%) from the regression models listed in Table 3. The labels align with the first CI of that category. Light blue, grey, yellow and green color bars represent CIs for the pre-test, < 150 ms fixation duration, 300-500 ms fixation duration and > 1000 ms fixation duration, respectively.

All 95% CIs for F_{1000} are less than zero indicating that as supposed in our model very long fixations (lasting for more than one second) are indeed detrimental for learning. Therefore, we can be sure, at the very least we have a bipartite model of fixation. Evidence for a tripartite model, where very small fixations are also detrimental and a middle range of fixations, also referred to as focal fixations in some contexts, that is beneficial to learning, is more nuanced. As expected from the model, most 95% CIs of $F_{300-500}$ are greater than zero, which indicates these fixations are indeed positively correlated with post-test scores. Additionally, the effect sizes (standardized coefficients) of the pre-test and the $F_{300-500}$ fixations are similar indicating equal contribution to learning outcomes. This result supports our proposed model as the $F_{300-500}$ fixations seem to be contributing positively and adequately toward learning.

However, for the proposed tripartite model to work the F_{150} fixations need to be negatively correlated with the

learning outcome. The interpretation of F_150 is less straightforward as the 95% CIs traverse both negative and positive territories in Fig. 4 and also overlap with F_300-500. Therefore, it seems that the valence (positive vs. negative academic emotion) that was well differentiated at the upper end of the fixation duration range (between F_300-500 and F_1000) is less so at the lower end.

There could be several reasons for the above results. First and foremost, the tripartite model may not be accurate. However, there is a marked dip in the 95% CIs of F_150 fixation durations into negative territory, which is not the case for F_300-500. Likewise, we see the possibility of the F_300-500 coefficients to be also becoming negative although demonstrably they are mostly in positive territory. These two facts suggest a more likely alternative. The differentiation between productive and non-productive fixations at the low end of the spectrum is likely more fuzzy and context-dependent. For example, from the literature we know that very short fixations can suggest expertise; a proficient reader does not need to fixate for long durations (Gegenfurtner et al., 2011; Bertram et al., 2013). However, small duration fixations may also suggest distraction as we had originally proposed. Hence, instead of outright rejection of a tripartite phenomenological model, we believe we have evidence for a softer version of this model. In such a model, very long fixations are detrimental, which is followed by a beneficial range of fixations whose lower limit can extend to very small fixations and one which could be dependent on task/learner characteristics. Only after reaching this limit the fixations may contribute negatively toward learning. Likewise, the range of detrimental fixations could well expand into the 300-500 ms range, which again would again be sensitive to the context.

Limitations

- 1) This study only explores a specific learning context. If we attach too much significance to this study alone, we risk making the same mistakes as the studies that do not take into consideration task and domain differences. More studies with a variety of contexts are required to generalize the proposed model.
- 2) The boundaries of fixation durations were drawn loosely from literature and based on fixations alone. However, we do not make any claim for the specific

numbers to be invariable. Instead, the goal of the article was to suggest a plausible affect-based fixation duration model that can explain variation in test scores.

- 3) Finally, we lacked evidence to suggest a truly tripartite model and the boundary at the lower end (between F_150 and F_300-500) seemed to be fuzzy. However, from this one study we cannot also be sure whether the boundary at the upper end (between F_300-500 and F_1000) is as clear-cut as it seems, especially because, in some learning contexts, distraction and interest are likely to be two contrasting affect that can influence such long fixation durations.
- 4) The analysis solely employs fixation duration ranges. As outlined in the Introduction, both saccade and fixation have been used to classify fixations into ambient and focal types. Therefore, it is quite conceivable that classifying fixation with information from saccadic eye movement would improve the model results.

Conclusion

The findings revealed that fixation duration can be a useful metric to trace the learning process. Within the context of this study, central tendency measures of fixation durations could not adequately explain the variability observed in the post-test scores. A tripartite phenomenological model based on affective processes that underlie learning was proposed. In this model, short or ambient fixations (<150 ms) and very long fixations (>1000 ms), contribute negatively to learning, whereas focal fixations (300-500 ms) contribute positively. So far we have been able to validate a softer version of this model where the demarcation of useful from useless fixation at the lower end of the spectrum is fuzzy. One possible reason for the ambiguity in the classification of short fixations into either detrimental or beneficial for learning could be the observed effect of expertise on fixation durations. As experts tend to require shorter fixations than novices, the F_150 range is possibly representing expertise as well as a negative learning affect.

Ethics and Conflict of Interest

The authors declare that the contents of the article are in agreement with the ethics described in <http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html> and

that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

The authors are grateful to four anonymous reviewers for their comments, which greatly helped in improving the manuscript. This work was supported by Seed and IRITPP funds provided to Ritayan Mitra by the Indian Institute of Technology Bombay.

References

- Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies.
- Bahreini, K., Nadolski, R., & Westera, W. (2016). Towards multimodal emotion recognition in e-learning environments. *Interactive Learning Environments*, 24(3), 590–605.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 78(5), B231–B244.
- Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The effect of expertise on eye movement behaviour in medical image perception. *PloS one*, 8(6).
- Buswell, G. T. (1935). How people look at pictures: a study of the psychology and perception in art.
- Bylinskii, Z., Borkin, M. A., Kim, N. W., Pfister, H., & Oliva, A. (2015, October). Eye fixation metrics for large scale evaluation and comparison of information visualizations. In *Workshop on Eye Tracking and Visualization* (pp. 235-255). Springer, Cham.
- Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). Eye tracking the visual search of click-down menus. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 402–409. ACM.
- Chien, K. P., Tsai, C. Y., Chen, H. L., Chang, W. H., & Chen, S. (2015). Learning differences and eye fixation patterns in virtual and physical science laboratories. *Computers & Education*, 82, 191-201.
- Ehmke, C., & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but Not as We Know It-Volume 1*, 119–128. British Computer Society.
- Follet, B., Le Meur, O., & Baccino, T. (2011). New Insights into Ambient and Focal Visual Fixations using an Automatic Classification Algorithm. *I-Perception*, 2(6), 592–610. <https://doi.org/10.1068/i0414>
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552.
- Glöckner, A., & Herbold, A. K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, 24(1), 71-98.
- Goh, K. N., Chen, Y. Y., Lai, F. W., Daud, S. C., Sivaji, A., & Soo, S.-T. (2013). A comparison of usability testing methods for an e-commerce website: A case study on a Malaysia online gift shop. *2013 10th International Conference on Information Technology: New Generations*, 143–150. IEEE.
- Goldberg, J. H., & Helfman, J. I. (2010). Comparing information graphics: A critical look at eye tracking. *Proceedings of the 3rd BELIV'10 Workshop: BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, 71–78. ACM.
- Granka, L., Feusner, M., & Lorigo, L. (2008). Eye monitoring in online search. In *Passive eye monitoring* (pp. 347–372). Springer.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74.
- Harteis, C., Kok, E. M., & Jarodzka, H. (2018). The journey to proficiency: Exploring new objective methodologies to capture the process of learning and professional development.
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A

- comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18.
- Helmert, J. R., Joos, M., Pannasch, S., & Velichkovsky, B. M. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 27, No. 27).
- Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision research*, 103, 83-91.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329.
- Kardan, S., & Conati, C. (2012). Exploring gaze data for determining user learning with an interactive simulation. *International Conference on User Modeling, Adaptation, and Personalization*, 126–138. Springer.
- Kimario, N. D., & Kamioka, E. (2017). Scan Path Based Approach to Detect Learner's Concentration in Distance Learning System.
- Krejtz, I., Szarkowska, A., Krejtz, K., Walczak, A., & Duchowski, A. (2012, March). Audio description as an aural guide of children's visual attention: evidence from an eye-tracking study. In *Proceedings of the symposium on eye tracking research and applications* (pp. 99-106).
- Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning ambient/focal attention with coefficient K. *ACM Transactions on Applied Perception (TAP)*, 13(3), 1-20.
- Krishna, O., Yamasaki, T., Helo, A., Pia, R., & Aizawa, K. (2017). Developmental changes in ambient and focal visual processing strategies. *Electronic Imaging*, 2017(14), 224-229.
- Kruger, J. L., Hefer, E., & Matthew, G. (2013, August). Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts. In *Proceedings of the 2013 Conference on Eye Tracking South Africa* (pp. 62-66).
- Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, 7(5). <https://doi.org/10.16910/jemr.7.5.4>.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (Vol. 5). New York: McGraw-Hill Irwin.
- Liu, P.-L. (2014). Using eye tracking to understand learners' reading process through the concept-mapping learning strategy. *Computers & Education*, 78, 237–249.
- Love, J., Dropmann, D., & Selker, R. J. (2018). Jamovi (Version 0.9). Jamovi. *Amsterdam, Netherlands: Jamovi Project*.
- McNeal, K. S., Spry, J. M., Mitra, R., & Tipton, J. L. (2014). Measuring student engagement, knowledge, and perceptions of climate change in an introductory environmental geology course. *Journal of Geoscience Education*, 62(4), 655–667.
- Mitra, R., McNeal, K. S., & Bondell, H. D. (2017). Pupillary response to complex interdependent tasks: A cognitive-load theory perspective. *Behavior Research Methods*, 49(5), 1905–1919.
- Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education*, 53(2), 445-453.
- Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics*, 73(4), 1120-1132.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10(3), 241–255.
- Salminen, J., Jansen, B. J., An, J., Jung, S.-G., Nielsen, L., & Kwak, H. (2018). Fixation and Confusion: Investigating Eye-tracking Participants' Exposure to Information in Personas. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 110–119. ACM.

- She, H. C., & Chen, Y. Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers & Education*, 53(4), 1297-1307.
- Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, 58(1), 375-385.
- Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationships between reading skill, reading speed, and visual inspection. *Language and Speech*, 33(1), 69-81.
- Unema, P. J., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual cognition*, 12(3), 473-494.
- van Asselen, M., Sampaio, J., Pina, A., & Castelo-Branco, M. (2011). Object based implicit contextual learning: A study of eye movements. *Attention, Perception, & Psychophysics*, 73(2), 297-302.
- Velichkovsky, B. M. (1999). From levels of processing to stratification of cognition Converging evidence from three domains of research. *Stratification in cognition and consciousness*, 15, 203.
- Velichkovsky, B. M., Rotherth, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 145-156.
- Velichkovsky, B. B., Rumyantsev, M. A., & Morozov, M. A. (2014, December). New Solution to the Midas Touch Problem: Identification of Visual Commands Via Extraction of Focal Fixations. In *IHCI* (pp. 75-82).
- Velichkovsky, B. B., Khromov, N., Korotin, A., Burnaev, E., & Somov, A. (2019, September). Visual Fixations Duration as an Indicator of Skill Level in eSports. In *IFIP Conference on Human-Computer Interaction* (pp. 397-405). Springer, Cham.