

Received:  
5 January 2018  
Revised:  
11 May 2018  
Accepted:  
15 June 2018

Cite as: Per Erik Eriksson,  
Thorbjörn Swenberg,  
Xiaoyun Zhao,  
Yvonne Eriksson. How gaze  
time on screen impacts the  
efficacy of visual instructions.  
*Heliyon* 4 (2018) e00660.  
doi: [10.1016/j.heliyon.2018.e00660](https://doi.org/10.1016/j.heliyon.2018.e00660)



# How gaze time on screen impacts the efficacy of visual instructions

Per Erik Eriksson<sup>a,b,\*</sup>, Thorbjörn Swenberg<sup>a</sup>, Xiaoyun Zhao<sup>a</sup>, Yvonne Eriksson<sup>b</sup>

<sup>a</sup> *Dalarna University, Sweden*

<sup>b</sup> *Mälardalen University, Sweden*

\* Corresponding author.

E-mail address: [pek@du.se](mailto:pek@du.se) (P.E. Eriksson).

## Abstract

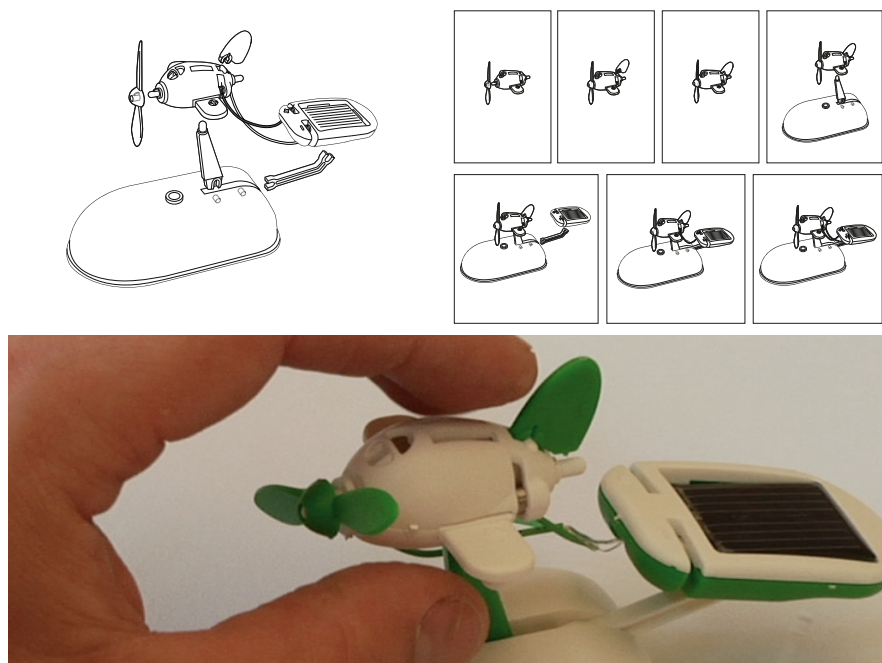
This article explores whether GTS (gaze time on screen) can be useful as an engagement measure in the screen mediated learning context. Research that exemplifies ways of measuring engagement in the on-line education context usually does not address engagement metrics and engagement evaluation methods that are unique to the diverse contemporary instructional media landscape. Nevertheless, unambiguous construct definitions of engagement and standardized engagement evaluation methods are needed to leverage instructional media's efficacy. By analyzing the results from a mixed methods eye-tracking study of fifty-seven participants evaluating their visual and assembly performance levels in relation to three visual, procedural instructions that are versions of the same procedural instruction, we found that the mean GTS-values in each group were rather similar. However, the original GTS-values outputted from the ET-computer were not entirely correct and needed to be manually checked and cross validated. Thus, GTS appears not to be a reliable, universally applicable automatic engagement measure in screen-based instructional efforts. Still, we could establish that the overall performance of learners was somewhat negatively impacted by lower than mean GTS-scores, when checking the performance levels of the entire group ( $N = 57$ ). When checking the stimuli groups individually ( $N = 17, 20, 20$ ), the structural diagram group's assembly time durations were positively influenced by higher than mean GTS-scores.

Keywords: Psychology, Education, Information science

## 1. Introduction

The mixed-methods eye-tracking study presented in this article explores how 57 students' off-line and online eye-movement behavior impact on their ability to comprehend and successfully use diagram and video assembly instructions (see Fig. 1). This exploration is based on basic statistical data analyses that all include gaze time on screen (GTS) eye-tracking (ET) data from three stimuli-groups' screening sessions (N = 17, 20, 20) in addition to data from observational video recordings of the stimuli-groups' assembly sessions.

In the screen-based instructional milieu, students and learners frequently encounter either static media – pictures, drawings and diagrams – or transient media such as various types of animations or videos. As noted by Clark and Mayer in *e-Learning and the Science of Instruction* (2016), static instructional visuals and instructional videos condition the learner's engagement differently and are therefore associated with different gain scores (pp. 224–230). Clark and Mayer distinguish between two types of engagement: psychological and behavioral. Behavioral engagement is “any overt action by the learner” during a learning activity (p. 223). Similar to



**Fig. 1.** Figure showing the instructions featured in the conducted study: The Structural Diagram (top left), the Action Diagram (top right), and a still from sequence no. 7 of the Live Action Video. Here, the original aspect ratio of the video is not preserved.

how Boucheix et al. (2013) and Fredricks et al. (2004, p.771) define behavioral engagement, this article is centered on one particular physical overt action, namely eye-movement behavior. Attending to relevant visual information with one's eyes is necessary for learning, and eye-tracking (ET) studies can provide deeper insights into visual attention and learning processes of students when they interact with different types of "diagrams" and "videos" (Boucheix; Lowe, 2010; De Koning et al., 2010; Kriz and Hegarty, 2007; Lohmeyer and Meboldt, 2015; Matthiesen et al., 2013; Ozcelik et al., 2009, 2010; Ruckpaul et al., 2015; Wang and Antonenko, 2017).

The question of whether a particular instructional design is engaging or not may be answered by analyzing assumed links between visual behavior factors and performance factors. Existing eye-tracking research that feature many different kinds of instructional visuals and various populations of learners shows that when visual behavior can be defined as optimal, engagement can be defined as successful, and learning can be expected to improve. For example, the studies by Boucheix and Lowe (2010), De Koning et al. (2010), Kriz and Hegarty (2007), Ozcelik et al. (2009, 2010) and Scheiter and Eitel (2015), all indicate that high levels of visual attention, that are assumed to relate to an increase in cognitive processing, result in superior performance outcomes. Thus, the accepted hypothesis among eye-tracking scholars that do research on visual instructions is that low gaze time scores impact on performance negatively. It is this expectation and pre-understanding that informs the statistical data analyses based on performance measures presented in this article. Here, it must be noted from the outset that "performance" refers to either students' GTS-scores (see next section on the GTS-measure) or assembly performance. Assembly performance is learners' ability to quickly, and, more importantly, accurately assemble an object, in this case a solar powered toy (see Fig. 1).

However, the concept of physical engagement is not as straightforward as it first might seem and there are a few inconclusive studies and divergent findings with regard to the relations between visual attention and learning outcomes. In brief, the crux of the matter is that the tacit assumption that attention is linked to foveal gaze direction is not always correct (Duchowski, 2007, p 12). For instance, Ozcelik et al. (2010), in their diagram-based multimedia study including forty undergraduate students, showed that shorter search times, but not overall fixation times, for signaled design elements were related to better transfer performance. Similarly, Boucheix and Lowe (2010) analyzed comprehension scores, and concluded that continuous cueing (a spreading color cue) primarily supported learners in the initial stages of processing the animations, not so much in later stages. Wang and Antonenko (2017), in their ET-study, including 37 undergraduate students, on the learning effects related to the visual presence of an instructor in videos on mathematics, conclude that no significant effects were found on learning transfer scores, but that enhanced videos attracted viewers' attention and that this led to better recall

of information as well as higher overall satisfaction. Other research-based studies establish no link at all between what is considered key eye-tracking parameters and learning, such as the in studies by [De Koning et al. \(2010\)](#), [Jarodzka et al. \(2013\)](#), [Kriz and Hegarty \(2007\)](#) and [van Marlen et al. \(2016\)](#). As noted by [Scheiter and Eitel \(2015\)](#), such divergent results, indicating that an increase in visual attention that is followed by an increase in cognitive processing does not result in better learning, call into question the absoluteness of the eye-mind assumption by [Just and Carpenter \(1980\)](#). Hence, it is not for certain that on target gaze patterns or/and longer gaze times always correlate with positive learning outcomes, although the prevailing assumption is that they should. Moreover, these divergent results cast doubt on whether it is at all feasible to assess learners' behavioral engagement via the means of eye-tracking.

Yet, considering the rapid and continuous expansion of online instructional efforts in higher education and training settings, the capacity to assess learners' engagement when interacting with visual instructions displayed on screens is unprecedentedly important. According to Henrie, Halverson and Graham, unambiguous construct definitions of engagement and standardized engagement evaluation methods are needed to leverage instructional media's efficacy in the contemporary digital learning setting. If not, digital instructional practices that leverage greater engagement cannot be satisfactorily identified ([2015](#)). Using gaze time on screen (GTS) as an ET-based engagement-measure could therefore be useful as part of unobtrusive evaluation methods suitable for the burgeoning digital, visual, educational setting.

### 1.1. The GTS-measure

One possible way of assessing learners' behavioral engagement would be to employ the GTS (gaze time on screen) measure. This, we speculate, may be technologically achieved by the employment of eye-tracking capable cameras in e-learning environments. GTS measures the time an eye-tracker can track the corneal reflections of a person's eyes. Thus, it is a global ET-data measure, that may be broken down into other AoI-specific (Areas of Interests) measures. GTS-scores then provide an on-line to off-line ratio, since GTS is total on-line time as percentage of task time. Our definition of off-line behavior (i.e. disengagement) may include, for instance, blinking, closing the eyes, looking up, or assuming postures that involve not looking directly at the screen. However, it is not within the scope of this article to establish what this off-line behavior really consists of. On-line time is made up primarily of fixations. Fixations normally constitute about 90% of an ET-recording's gaze samples. However, in the present article, we also include saccades as part of on-line time, since one of the stimuli featured in this article consists of moving images.

In ET-based research efforts, type of measurements and how variables are being operationalized require careful consideration ([Holmqvist et al., 2011](#); [Duchowski,](#)

2007). In such ET-research, GTS is normally used as an ET-data quality indicator measure. Low GTS-scores indicate “bad data”. Post experiments, properly set GTS-levels are considered to warrant complete sets of ET-data that facilitate non-distorted ET-data analyses, irrespective of type of stimuli (Hvelplund, 2011, p. 104; Sjørup, 2013, p. 105). This is deemed important (by researchers), since ET-systems are extremely sensitive to inferior system calibration, abnormally oscillating eye-movements, as well as other external error-inducing factors (Duchowski, 2007, p.178; Holmqvist et al., 2011, p. 29). Hence, GTS-scores seldom reach 100%, unless task time is extremely short (and participants do not blink). In an ET-study on cognitive load and continuity editing in documentary filmmaking by Swenberg and Eriksson (2017), the GTS-threshold was set to 2 SD under the mean value (93.5%) = 82,4%. In other words, in that study the eye-movement data sets that were associated with a lower than 82,4% GTS-score were considered substandard.

However, which eye-tracking data sets that should be considered “bad”, and which ones that should not, is not self-evident. Standardized guidelines regarding how exactly to establish relevant GTS-levels remain unspecified. Depending on the type of data streams available, a high GTS-score is likely to be around 95%, and an acceptable score around 80% (Hvelplund, 2011, pp. 103–108). In ET-research on reading and translation, considerably lower levels are regarded as “acceptable” (Sjørup, 2013). Thus, in this article, we accept that what is ‘acceptable’, or not, in terms of GTS-scores, is somewhat arbitrary. Here, it may be noted that among the participants who generated the data for the analysis of this study, the average GTS-score was 89,3%. Therefore, in this article, we do not consider certain GTS-scores, for instance, as being “high” or “low”, or falling within a range of acceptability. Instead, we simply establish GTS-means.

Task specificities and designerly issues can be factors that influence GTS-means, at least to the extent that reading text does not require the same levels of behavioral engagement as decoding instructional pictures/images. Rosenfield et al., in their study on reading from digital screens that included 16 visually-normal subjects, suggest that rapid blinking, or very little blinking, is due to the “cognitive demand” of the task (2015). However, according to cognitive load theorists, it is doubtful whether behavioral disengagement/engagement has anything to do with task difficulty (Sweller, Ayres & Kalyuga, 2011).

Another possibility is that the GTS-measure may capture visual behavior that reflects the narrative functionality of the stimuli that is used, what comics scholar McCloud discusses in terms of “closure” (1993) and what art and cognition scholar Stafford refers to as onlookers’ “binding” (2007). Perhaps periods of refocusing and “resting” ones eyes in between the depicted procedural steps might be captured by using GTS as an engagement measure. In other words, it is possible that closure, what here might be labeled as disengagement, actually equals something akin to cognitive

focus. Such possible cognitive focus would, then, be very obvious in the context of procedural instructions since all such visuals are narratives that aim to explicitly show step by step procedures (Daniel and Tversky, 2012). Still, the prevailing notion among eye-tracking scholars that employ the GTS-measure as an ET-data quality measure is that gaze time could be a result of a wide range of different external factors, not stimuli specific factors, such as an individual's fatigue, motivation, emotional state, prior knowledge, ability level or, we speculate, "mind wandering tendencies" (Loh et al., 2016). This quality is essentially what makes it valid an ET-data quality indicator measure. Likewise, in this article we propose it is this quality that would make it valid an engagement measure. Its validity depends upon its assumed applicability across media platforms.

In brief, then, in the study presented in this article, we predict that the students' GTS-scores are likely not to reflect the complexity of the learning materials used, their designs, and, consequently, most likely do not correlate with the students' assembly performance scores. To the contrary, we think, that it is more likely that, for example, ability levels or/and visual literacy capacities correlate with performance scores, rather than on-target gaze patterns (however, this remains to be verified). See Eriksson et al. (2014) on the issue of the relations between "assembly performance" and "visual literacy capacities".

## 1.2. Static and transient instructions

The conducted study features both *static* and *transient* examples of visual, procedural instructions. Transient, visual, instructions are basically videos, i.e. visual instructions that actually move and that are time-based. In the educational psychology discourse the term "animation" and the more retro sounding and ambiguous term "multimedia" are more frequently used instead of "video". In static representations, movement and the passing of time are only implied. Examples of such visuals are pictures, diagrams or/and drawings (some drawings are diagrams). The instructions in this article feature the same object to be assembled (a solar powered toy), while representing unambiguous, visual instructional archetypes that, in turn, represent commonly used instructional media types in online learning efforts. These are two diagrams (line drawings), and one live action video (see Fig. 1). First and foremost, these stimuli represent two different representation modes, the static, and the transient representation mode.

Cognitive psychologist Barbara Tversky suggests that the fleeting nature of animations is challenging for learners (Tversky et al., 2002), and that transient stimuli leave too little room for purposeful acts of interpretation (Tversky, 2011). This highlights the assumed advantage of diagrams in screen-based learning efforts, suggesting that they promote psychological engagement to a higher degree than so-called, "animations", and that, on the whole, psychological engagement surpasses

behavioral engagement with regards to learning outcomes (Clark and Mayer, 2016). Yet, quite naturally, in comparison to static visuals, transient media offers better representations of temporal aspects, for instance, how movements play out over time. Höffler and Leutner (2007) call this the *procedural-motor advantage* of animated presentations. In certain instructional contexts, this aspect is considered a success factor in terms of learners' cognitive performance (Hooijdonk and Kraemer, 2008). There are several research-based instructional media assessments informed by Cognitive Load Theory (Sweller, 1988, 2010), and Cognitive Theory of Multimedia Learning (Mayer, 2005) that aim to circumscribe the temporal aspect affordance of videos (Ayres and Paas, 2007; Boucheix and Forestier, 2017; Cojean and Jamet, 2017; Castro-Alonso, 2015; Sweller, Ayres & Kalyuga, 2011; Wong et al., 2012; Ibrahim, 2012; Ibrahim et al., 2014; Marcus et al., 2013; Merkt et al., 2011; Watson et al., 2010). Apart from the *transient effect*, these research-based studies also address a few other moderating variables that seem to mitigate the temporal affordance of transient instructional media. Videos are often inherently multimodal, and are more likely to suffer from low "semiotic clarity" (Figl et al., 2010) than static media that tends to be based on simple percepts. Moreover, videos – short abstract animations aside – tend to be information overloaded. The scanning and decoding of detail rich videos may therefore become an arduous, physical task that requires great concentration and focus. However, in theory at least, optimal viewing strategies, such as quick on-target fixations, can counter adverse effects.

As briefly discussed in the previous section on the GTS-measure, static and transient instructions' differences can also be discussed in terms of mental animation efforts, or levels of "closure". This is the process whereby humans fill in the gaps between images, and transform them conceptually into a unified idea (McCloud, 1993, pp. 60–93; Cohn, 2013). The action diagram instructional type differs from the structural diagram instructional media type in that it requires a very high degree of closure, more specifically, the moment-to-moment and action-to-action closure categories (McCloud, 1993, p.70). The structural diagram involves the aspect-to-aspect closure category, the "wandering eye" being its hallmark (p.72). In comparison, the process of closure, when it comes to most videos, would be imperceptible, continuous, and largely involuntary (McCloud, 1993, p.68). Thus, in spite of the diagrams belonging to the same instructional archetype, in some ways, the video in the presented study has more things in common with the structural diagram, than the structural diagram has with the action diagram, since both the structural diagram and the video appear more "informationally complete" (Watson et al., 2010, p.91), and thus require relatively low levels of closure. It is a possibility that "closure" conditions learners' visual behavior when they interact with static and transient procedural instructions and that "closure" manifests itself as off-line visual behavior of different degrees depending on what type of procedural instruction that triggers it.



### 1.3. The live action video instructional format

The present study includes one particular kind of transient media, i.e. a Live Action Video (LAV). Here, the LAV-format and its specific designerly aspects and narrative inclinations deserve further consideration since LAVs are associated with specific instructional affordances that differ from other transient media. Live Action Video is what a video camera records when a videographer pushes the record button and records activity, capturing the world as we know it, live in front of the lens, thereof the term “live action”. LAVs are thus photographic in nature. Hence, LAV does not require copious and costly post-production activities that aim to conform indirect perception to direct perception (Anderson and Anderson, 2005), and emotional design aspects that are conducive to learning (Mayer and Estrella, 2014) come at no extra cost. Undoubtedly, this explains why there are hundreds of millions of how-to LAVs on YouTube. With regard to live action cinematography, the father of direct perception theory, James J. Gibson thus once claimed “Moviemakers are closer to life than picturemakers” (1979, p.293).

This kind of “realism” is discussed by J.C. Castro-Alonso et al. (2015), in their study on transient affordances in procedural, instructional live action videos, in terms of the activation of embodied cognitive systems and the human movement effect. Such activation is less affected by working memory limitations, when primary information is used to assist in the acquisition of cultural knowledge (Paas and Sweller, 2012). This means that LAVs are a perfect medial vehicle for exploiting humans’ cognitive architecture, since biologically primary knowledge is integral to LAVs, as they often feature real people who make actual real movements, which, in turn, exhibit naturally occurring cues and signals. This partly explains the LAV’s popular appeal in traditional educational settings, in that it is the most cost-effective way of creating realistic video content, what Chih-Ming Chen and Chung-Hsin Wu call “video lecture styles” (2015). This is also what J.C. Castro-Alonso et al. (2015), and Wong et al. (2012) use as material for their instructional so-called “animations”. However, in comparison to (computer generated-) animations, LAVs are not burdened by unfamiliarity and the uncanny valley effect (Lowe and Boucheix, 2011). Still, the quality of cinematic realism is easily compromised by surface aspects, such as, for instance, poor resolution (Eriksson and Eriksson, 2015), and interface information overloads. Therefore, transient realism may come at a cost, due to the fact that it may cause cognitive load to exceed working memory limits (Wong et al., 2012). Nevertheless, as successful filmmakers have long since realized, a clear-cut narrative subordinates presentation aspects to the story and makes the viewing experience effortless. This is what J. D. Anderson describes as the process whereby the “questionable status” of images is supported by the story it conveys, thereby placing the irrational actions portrayed in a rational narrative context



(2005). This is to say that it is likely that an unambiguous narrative probably lessens the need for compensatory top-down processing.

#### 1.4. Research question and objective

In this article, the analysis is driven by one primary research question: *What do GTS-scores indicate about learners' performance in a learning situation that involves static and transient procedural, screen-based, visual instructions?* Given the theoretical perspectives and empirical evidence presented, the current study was designed to explore the potential use of GTS as an engagement measure within the context of visual, procedural instructions of three kinds: one structural diagram, one action diagram and one live action video that is based on the diagrams (i.e. they are information equivalent). The objective is to, first, establish the visual instructions' respective efficacy, i.e. their associated performance scores. Second, to check whether the learners' GTS-scores correlate with their performance scores. Third, to check if there are statistically significant differences between the means of three stimuli groups with regards to build time, build error and GTS-scores (using ANOVA). Fourth, to check if there are differences between the "low" and "high" GTS performance classes, regarding build time and build error in each stimuli group (using ANOVA).

#### 1.5. Contribution

Research that exemplifies ways of measuring engagement in the instructional screen-based context is rare and usually does not address engagement metrics and engagement evaluation methods that are unique to the diverse contemporary instructional media landscape. The exploration of this potential use in this article complements existing literature and could reduce the confounding roles that quality factors, medial affordances, and the diversity that the role of screen-based visual instructions plays in influencing not only engagement, but also learning outcomes. This is the rationale for exploring GTS as an engagement measure in this article. More generally, this novel approach offers two benefits. Firstly, this approach advances our understanding of the relation between learners' closure levels and engagement levels, and, in fact, closure as enacted by students in an instructional learning situation equals disengagement captured by GTS. Analyzing GTS-means in conjunction with assembly performance scores in the context of visual instructional media that is more or less narratively inclined makes this explicit. Secondly, this approach furthers our understanding of the limitations of standardized ways of measuring engagement and quality assessment methods that pertain to technology mediated learning, such as in MOOCs, what some may consider the "gamification" of instructional learning.

## 2. Method

### 2.1. Participants

There were 57 participants in total in the study, 25 male, and 32 female. The average age of the participants was 26 years, all with normal, or corrected to normal, vision. All participants were students recruited from Dalarna University, and the group consisted mainly of BA-students, with a few MA-students. About half of the participants consisted of a mixed group of Engineering/Technology/Economics students, while the rest were media students (TV/Film/Graphic Design/Commercials Production/Music Production). The participants were randomly approached (on campus) and assigned to three stimuli groups: a structural diagram group ( $N = 17$ ), an action diagram group ( $N = 20$ ), and a video group ( $N = 20$ ). All were considered novices with regard to the assembly task. This assumption was made due to the novel nature of the solar powered toy to be assembled and not formally assessed. None of the students were at the time enrolled in the researchers'/teachers' classes and some were approached by a third party when this was considered ethically correct. They all received a movie theater gift certificate (15€ value) for their participation regardless of whether their data was used or not. All participants gave informed consent (one pre experiment and one post experiment). Since 7 participants who were intended to be part of the study generated extremely inferior calibration, or did not sign the release forms (post experiment), their data sets were discarded. This explains the uneven numbers of participants in the groups. The participants were not aware of the purpose of the study when viewing, but were informed afterwards in writing, and they confirmed the use of their generated data by written consent. Data on the participants' educational background and mother tongue was also captured, in order to be able to (later) discriminate between possible background factors influencing the viewing data.

### 2.2. Instruments

The stimuli were run with SMI Experiment Centre 3.4.119 software. This is a table mounted eye-tracker (not a mobile one). No stimulus was deemed potentially harmful. We used a 9-point calibration with 4-point validation to ensure good data, even at the screen edges. This was considered critical, since one of the stimuli consists of moving images in which the AoIs (Areas of Interests) move around, and sometimes end up close to the edges of the video frames. Any participant who generated extremely inferior calibration was not included in this article. The eye movements were recorded with a SMI RED250 stationary eye-tracker, sampling eye data at 120 Hz, with iViewX 2.8.26. The video material was generated at 25 frames per second and screened on a computer screen, Dell P2211, driven by a NVIDIA GeForce GT440 video card, in 1680x1050 px resolution and mp4 codec (SMI default). Light emittance was measured to 90 cd/m<sup>2</sup> (at brightness level 65%, and contrast level

75%) for a 255-255-255 white screen. The viewing position was 60–80 cms from the screen.

### 2.3. Materials

Three kinds of visual stimuli were designed in the study: a structural diagram showing how the individual parts of a solar powered toy should be assembled; a single page action diagram showing how the toy should be assembled step-by-step; and, lastly, a live action video based on the diagrams (see Fig. 1). The visual stimuli were inspired by the technical documentation included in the box containing the solar powered toy, courtesy of *6 In 1 Educational Solar Kit*. All three stimuli were designed on the *information equivalence* premise to avoid incomparable content as recommended by Tversky et al. (2002) and Ganier and de Vries (2016). First the structural diagram was designed, this is the diagram that the most resemble the *6 In 1 Educational Solar Kit* original instruction. The diagrams were designed and produced by Peter Johansson, an expert in Graphic Design and informative illustrations at a Swedish university, with a distinct engineering and design profile. The LAV was produced by Per Erik Eriksson (the corresponding author of this article), a former professional videographer and TV-producer. The diagrams were made from a user's point of view ("POV") perspective. Presumably, this facilitates an error-free assembly process, since it enables the assembler to more easily relate his/her assembly process to the depicted object's different parts, and the progression towards an object as a completed unified whole. The black and white diagrams were made in a two-vantage point perspective, and two thicknesses of lines are used, where the thicker line is used to give the object a distinct shape and volume (Richards et al., 2007). In comparison with the LAV (that has color), there were no indications of human beings, such as hands in the drawings. Neither the diagrams, nor the video, employ text and audio. Few modalities allow fairly simple distinctions and comparisons to be made between the groups. The video was recorded on an HVX201 Panasonic HD-camcorder, 1080i resolution at 25 frames per second (fps). The video was 2 mins. and 17 secs. long. Its sequences basically adhere to the steps in the action diagram, and were designed to mimic the diagrams in their overall simplicity, their POV-perspective and framing, i.e. close-ups. It manifests techniques and design choices considered best practice within the field of instructional LAVs, and exhibits soft, high-key lighting, sharp focus, correct exposure, consistent framing and angles. See stills of stimuli in Fig. 1.

### 2.4. Procedure

First, the participants were asked to settle down for a moment, in order for all those participating to "assume a similar state of mind" (Holmqvist et al., 2011, p.115). Then, all participants were informed that they were to look at a visual assembly

instruction on a computer screen, and afterwards they were expected to assemble the object depicted in the instruction. The participants viewed one of the visual instructions (1–3) each, one participant at a time, comfortably seated in front of the screen and the speakers. This is what we consider study-time. The lab setting was configured with a wall-screen in between the participant and the researcher, in order to avoid disturbance to the participant when running the experiment. After the eye-tracking sessions, the participants were asked to assemble the toy. They were allowed to (re-) study the diagram while attempting to assemble the toy. This was considered a more ecologically valid situation. They were given the following instructions: “The goal is to assemble this item and make it complete as per the instructions you just watched. If you feel that you cannot assemble the object, you are free to discontinue at any point. If you like, you are allowed to consult the instruction on the laptop computer in front of you during the build.” According to Boucheix and Forestier (2017), providing learners with instructions such as these will facilitate more ecologically valid learning assessments. The observed build was recorded on the Panasonic HVX 201 and on the Apple’s iSight, on the laptop computer used during the build. After the build, all participants were debriefed, in order for the researchers to identify possible reasons for uneasiness, discontent and/or disengagement. Apart from a few comments that concerned the power cords to the solar panel that were difficult to attach (possibly an aspect of *task difficulty*), no relevant data was obtained from the debriefing protocols and is therefore not included in the data analysis of this article. The viewing session and the assembly session, including settle down time and debriefing, normally lasted about 30 minutes for each participant.

The study/research project was checked by the authors for ethical aspects, according to the local Bill-of-Self-Audit (Dalarna University Research Ethics Committee, 2008), and passed all stipulated criteria. Neither procedure nor stimuli was unethical in regard to Dalarna University research standards.

## 2.5. Pilot study

A pilot study (Eriksson et al., 2014) was conducted in order to assure key measurements’ validity and reliability as well as to check the design of study’s robustness. One indirect result of the pilot study was that the action diagram design was considered to be of overall low quality. In the pilot study, the action diagram group’s performance was very poor and the learners commented on the action diagram’s overall poor design that was of the multiple (and separate) pages kind. Consequently, the action diagram ended up being redesigned and made to fit into one page only. This one page design is the one employed in the current study.

In the pilot-study, *number of reviews-data* during assembly time was collected, i.e. how many times the participants revisited the stimulus displayed on the laptop computer in front of them with their eyes during assembly. However, in the pilot study we could not delineate any patterns or and establish any correlations with regards to

*number of reviews* and learners' performance. Hence this measure is not part of the analysis in the current study. It is also important to note that the pilot-study in question features other ET-measures that when combined make up GTS. The pilot-study also represents a slightly different research focus in comparison to the current study since it thematically revolved around the learning implications of learners' detailed versus focused viewing behavior (cf. [Holsanova, 2001](#)).

## 2.6. Data analysis – Measures from eye-tracking data

Gaze time on screen (GTS) was calculated from the participants' data sets by analyzing their individual eye-tracking timelines. GTS-scores were calculated by dividing the sum of time for saccades and fixations by total stimuli time, i.e. time looking at the visual instruction on the screen. Stimuli time, or study time, is total time watching the visual instruction in question ( $N = 3$ ), from the onset of the screening of the stimuli in question, until the end of the screening of it, or when the participant finished looking at it. In other words, stimuli time/study-time does not include the time the learners spent or did not spend with the stimuli during assembly. This optional, second review time during assembly appeared to vary greatly (just as it did in the pilot-study). Some learners chose not to consult the instructions at all during assembly. However, the duration or/and impact of this optional review during assembly was neither measured nor analyzed.

## 2.7. Data analysis – Calculating correct GTS-scores

In order to ensure that the GTS-scores were valid and reliable in psychometric terms we needed to make sure that the GTS-scores reflected human behavior associated with the phenomenon that is being investigated (learners' visual engagement). In other words, we needed to make sure that the loss of data – which lowers the GTS-scores – during study time reflected *relevant* offline behavior not other kinds of offline behavior. For example, it turned out that some students/learners stopped looking at the stimuli before the screening had ended. Some got up and left to assemble the toy before study time was supposed to stop. This time the computer counted as (GTS-) offline time, which is incorrect. Some learners' ET-timelines also included time that the ET-computer had mislabeled as offline time that, when scrutinized, was associated with the learner looking away from the screen talking about irrelevant matters (for example asking the researcher questions about what he/she should do next). Consequently, in some cases, we manually subtracted what the ET-computer considered offline time from the GTS-scores, i.e. making task time/study-time shorter. In one case we did the reverse, we replaced what the computer considered offline time with online time, since it was obvious that the participant was looking at the stimuli during that so called offline time. In this case this had to do with an odd reflection on the participant's eyeglasses that resulted

in that the ET-computer mislabeled online time for offline time. Still, subtraction of offline time was more common than the addition of online time among the participants' data sets. The average subtraction (all participants) was 1,07 seconds. If we consider the participants that were associated with adjusted data sets (not all participants' data sets were adjusted), the average adjustment was 3,32 seconds.

The tool in SMI Experiment Centre 3.4.119 software that was used to calculate the GTS-sets ("GTS score") was *Event Statistics*. In *Event Statistics* we used the function *Stimulus Statistics* in order to extract the data of the measures *tracking ratio* and *duration*. In order to establish correct GTS-scores we verified *tracking ratio*-data and *duration*-data with the ET-computer's video recordings (w. both audio and video in most cases) of the participants, their *Line graphs* (i.e. the participants' eye-movement timelines) and their *Bee Swarm* patterns (i.e. gaze hit patterns).

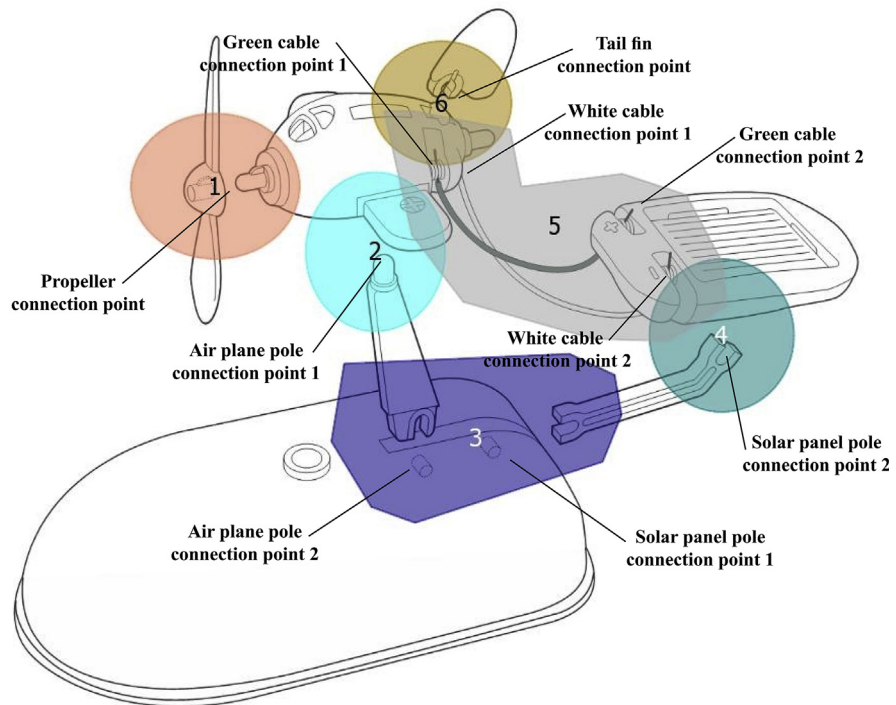
## 2.8. Data analysis – Measures from observational video

A global assembly, build error count was administered. Only errors that were attributable to the visual instructions were included. Since the instructions feature an object to be assembled, all errors relate to the object's assembly connection points. There are 10 connection points in total:

1. White cable (connection point 1)
2. White cable (connection point 2)
3. Green cable (connection point 1)
4. Green cable (connection point 2)
5. Propeller
6. Tail fin
7. Solar panel pole (connection point 1)
8. Solar panel pole (connection point 2)
9. Air plane pole (connection point 1)
10. Air plane pole (connection point 2)

Thus, dropping a cord or a part on the floor would not be considered an effect of the instructional information, whereas putting a part in the wrong place, failing to connect something properly, could be an effect of the instructional, visual information. Build errors later corrected during the assembly process were still considered build errors. For the location of the connection points, see the Structural diagram's AoIs (Fig. 2).

Overall *build time* in seconds was derived from the observational video footage. Build time was considered as the time between the start of the build process and the end of the build process, i.e. when the participant decided the object was as complete as possible.



**Fig. 2.** Figure showing the 6 AoI's of the Structural diagram and the locations of the assembly connection points. Colored areas in chart are AoIs (1–6).

## 2.9. Statistical analysis method

The statistical analysis method of this article is based on basic comparisons between the means of the three stimuli groups' respective build error, build time and GTS-scores, what we refer to as the students' "performance". In addition we employ a basic correlation analysis based on the same measures/data as well as ANOVA-tests. The null-hypotheses of the ANOVA are the following:

H0: The means of Build time in the three groups are equal

H0: The means of Build error in the three groups are equal

H0: The means of GTS score in the three groups are equal

## 3. Results

### 3.1. GTS effects on performance

According to the ANOVA there was no significant effect of build error at the  $p < .05$  level for the three stimuli groups [ $F(2, 54) = 1.233, p = 0.299$ ]. There was no significant effect of build time at the  $p < .05$  level for the three stimuli groups [ $F(2, 54) = 2.344, p = 0.106$ ]. There was no significant effect of build error at the  $p < .05$  level for the three stimuli groups [ $F(2, 54) = 1.306, p = 0.279$ ].



If we differentiate each stimuli group into two sub groups according to the GTS performance, which is, high GTS and low GTS (higher or lower than mean GTS), there was no significant effect of build error or build time at the  $p < .05$  level for each stimuli groups according to the ANOVA.

We then perform a correlation analysis checking the correlation between build error/build – time and GTS in the group as a whole ( $N = 57$ ) and the different groups, respectively. See [Table 1](#).

If we do not consider the possible influence from the specific stimuli, and check the correlation between GTS, build error and build time for the whole 57 observations. The correlation between build error and build time is small, while the correlation between GTS and build error, GTS and build time are negative, and that they are statistically significant ( $p < 0.05$ ).

When we look at the groups separately, weak correlations are found in general, except that, in Action diagram, there is negative correlation between build error and build time. In Structural diagram, strong negative correlation was found statistically significant ( $p < 0.05$ ) between GTS and build time.

Remember, correlation refers to the extent to which the examined two variables have a linear relationship with each other, not causation, as other variables may be affecting the relationship between the two variables of interest.

### 3.2. Learning performance and GTS-scores

[Table 2](#) shows the mean, standard deviation and the standard error of the mean of build error, build time and GTS-scores in the different stimuli groups. The Action diagram group has the highest mean build error, while the Video group has the

**Table 1.** Correlation coefficients between GTS, build error and build time without differentiating stimuli groups and differentiating stimuli groups.

		Build error	Build time	GTS
All	Build error	1	0.1	-0.29*
	Build time		1	-0.31*
	GTS			1
Structural diagram	Build error	1	0.42	-0.36
	Build time		1	-0.73*
	GTS			1
Action diagram	Build error	1	-0.28	-0.22
	Build time		1	-0.08
	GTS			1
Video	Build error	1	0.19	-0.32
	Build time		1	-0.28
	GTS			1

Note:  $p < 0.05$  is noted by \*.

lowest. With regard to build time, the Action diagram group again has the highest mean value, while the Structural group has the lowest build time, but with largest spread. The standard deviations for all groups in build error and build time are fairly large, which means in each stimuli group there are a number of participants who performed toward one extreme or the other. On the other hand, this indicates that the assignment of the participants to each stimuli group is fairly balanced. Overall, the Structural diagram group and the Video group outperformed the Action diagram-group, both in build error and build time. As for the GTS scores, all three stimuli groups have a mean value that is approximately equal. The standard deviations for all groups in GTS score are small and very similar. This indicates that the data sets that were consulted in order to calculate the GTS-scores were rather homogenous across the groups.

#### 4. Discussion

Research into the efficacy of different kinds of visual instructions has been extensive, and commonly focuses on signaling designs within a particular instructional genre, or the affordances of transient media versus static media. This article also belongs to this strain of research, but differs in that it categorizes visual instructions that are narratives as representing distinct degrees of closure requirements, and evaluates them by relating them to users' GTS-scores, in addition to conventional performance measures. The basic finding, here, is that visual attention disengagement, on the whole, does not appear to be a very detrimental behavior when trying to learn from procedural, visual, screen-based, instructions. This seems to be consistent with the diagram-based, assessment study of [Ozcelik et al. \(2010\)](#) that found that longer fixation times do not always lead to higher performance. Moreover, this basic finding appears to validate ET-scholar Duchowski's claim that attention is not necessarily linked to foveal gaze direction ([Duchowski, 2007](#), p 12) and the common claim by psychologists that seeing is a mental affair. All in all, the aforementioned results

**Table 2.** Mean and standard deviation of build error, build time and GTS in respective stimuli groups.

Measure		Stimuli group		
		Structural diagram (N = 17)	Action diagram (N = 20)	Video (N = 20)
Build error	Mean	2.06	2.85	1.7
	SD	2.045	2.833	2.08
Build time (s)	Mean	222.353	298.05	239.45
	SD	125.807	112.053	102.455
GTS score	Mean	0.876	0.896	0.920
	SD	0.08	0.095	0.074

shed some important light on previous inconclusive and seemingly inconsistent results in the field of learning and instructions, with regards to the fact that increased visual attention does not always lead to learners' increased performance and understanding, although the basic assumption is that it should.

#### 4.1. The GTS Measure's validity

We designed three types of visual, procedural instructions, depicting an assembly instruction and tested their respective efficacies on a group of learners ( $N = 57$ ). The video-group and the structural diagram-group outperformed the action diagram-group. We then established the learners' GTS-scores and the groups' respective GTS-means. Following this, we analyzed the learners' assembly-performance in relation to their GTS-scores. We predicted that the distribution of GTS-scores was to be relatively balanced among the stimuli groups ( $N = 3$ ). The similar GTS-means and the results from the ANOVA show that there is no significant difference among these three groups. This finding suggests that GTS-scores are not a direct consequence of type of stimuli, and do not reflect the diversity and/or complexity of the learning materials used, i.e. their respective element interactivity, but are due to some external factors. This is fortunate since GTS as an engagement measure may only prove useful (i.e. *valid*) if it circumvents some of the evaluation related constraints caused by the diversity and complexity of screen-based media interfaces.

However, it turns out that it is rather difficult to calculate correct GTS-scores. This radically lessens its potential usefulness as an automatized engagement measure in the screen-mediated instructional context. It also remains a challenge to decide what GTS-scores should be considered low and what scores should be considered high in the visual, screen-based instructional setting. In this study, the video group has the highest average GTS-scores. The video group also has the highest performance levels with regards to the build error measure. However, it is by no means for certain that the video group's performance has anything to do with GTS-scores.

#### 4.2. The live action video advantage

In the case of the video group, it is likely that this group's superior performance has to do with the *procedural motor advantage* (Hooijdonk and Kraemer, 2008; Höffler and Leutner, 2007) rather than focused eye movement behavior. We suggest that the video's efficacy is due to the activation learners' mirror neurons, what J.C. Castro-Alonso et al. (2015) discuss in terms of the activation of embodied cognitive systems and the human movement effect. This finding is consistent with the idea that instructional support tools that feature real people who make actual real movements (in this case hand movements), which, in turn, exhibit naturally occurring cues and signals, have the capacity to free up learners' cognitive recourses (cf. Paas and Sweller,

2012). In addition, it is likely that that such indexical video content facilitate more or less effortless visual decoding (cf. Kaiser et al., 2012). With regards to this, it seems likely that the LAV-medium in of itself trigger focused behavior, i.e. that recorded human movement captures attention (Franconeri and Simmons, 2005). Considering the sheer amount of live action video-based how-to videos (there are more than half a billion of how-to videos on YouTube), this may be considered an important aspect of video mediated instructional efforts in general. Yet, the effect of viewers' on-target visual attention on learning and performance is probably relatively small.

More generally, the relatively high performance scores of the video-group implies that *live action instructional videos* may leverage mediated, procedural learning efforts, and should therefore not be clumped together, as is often the case, with all sorts of other kinds of "videos" commonly regarded as on-line learning fads but which mostly hinder learning. Hence, the findings in question ascertain that well-known, laboratory-based, results that pertain to the video medium's ability to leverage learners' procedural understanding, especially ones that study the display of hand-movements, generalize to what can be considered more ecological, learning settings, i.e. settings that allow for both pre-screening and simultaneous screening/performing.

### 4.3. The structural diagram advantage

The participants ( $N = 57$ ) in the stimuli groups that have lower than the mean GTS-scores would not show that much worse performance levels than the participants with higher than the mean GTS-scores, except for the Structural group in build time. In this article, the importance of brief build times may be questioned since the visual instructions that are investigated in this article primarily aim to show how to assemble an object correctly, not quickly. Still, this intriguing result may shed some light on how the efficacies of instructional visual designs that require little "closure" (McCloud, 1993) are leveraged by high GTS-scores. However, the question remains, why this manifests itself in brief build times and not few build errors? Moreover, it is a little odd that the other kind of visual instruction that is similar in that it also requires low levels of closure – the live action video – is not associated with the same kind of potentially GTS-driven influence.

### 4.4. The action diagram disadvantage

Specifically, the findings of this article seem to indicate that suboptimal designerly efforts are unlikely to be compensated for by increased screen online time by learners. This is to suggest that the suboptimal performance levels associated with the action diagram group are not related to GTS-scores, but that the top GTS-scorers have similar overall low performance levels, in comparison with the low GTS scorers, due to the action diagram's questionable efficacy or quality. This

highlights the necessity of designerly ways to manage element interactivity and is to infer that a subpar instructional design is equally bad for all types of visual behavior. This is to infer that it is entirely possible that one key moderating variable with regard to the action diagram group's performance, is, in fact, efficacy, or quality of instructional design. In any case, we suggest it is unlikely that these poor performance levels are due to low GTS-scores, since the action diagram type, just as with comics, naturally invite closure-related behavior (cf. McCloud, 1993). However, to what degree certain visual behavior decoding styles/cognitive styles impact this, remains to be tested (cf. Holsanova, 2001; Höffler et al., 2017). The main instructional implication of this is primarily a cautionary one, and admittedly not new: if design principles and purposeful instructional strategies in the digital education settings are disregarded, dutiful and motivated students' efforts will be wasted. In other words, the low performance scores of the action diagram group put into question the assumed efficacy of the action diagram format, a presentation format often championed by scholars of instructions (Daniel and Tversky, 2012), professional designers of visual instructions, and several well-known electronics, building supply and home furnishings corporations.

#### 4.5. Limitations and future directions

Visual design elements, however subtle they may be, vie for our “undivided attention” (Stafford, 2016). However, it is not in the scope of this article to elaborate on various possible moderating factors that can be linked to specific, and often subtle, design elements. This brings us to the one major caveat with regards to using GTS as an engagement measure, namely, its crudeness. Therefore, further repeated experiments to collect more data for examination would be preferable in validating the current results, and in identifying possible external and hidden factors, such as, for example, learners' ability levels. Visual strategizing might here also be a factor that, in effect, conditions online/offline time, for instance, global eye movement behavior common in familiarization phases may influence GTS-scores. Similarly, it is also a possibility that certain eye-movement behavior that results in more or less distinct offline/online patterns, might be attributable to prior-knowledge levels (Taub et al., 2014), and/or cognitive styles (Höffler et al., 2017). However, once again, answers to questions such as these could only be established in studies that allow for a more detailed analysis with a much greater number of data streams than the relatively limited number of data streams available for analysis in this article. Studies of greater scope that include other informants than solely students might also be able to confirm or reject the results presented in this study, and make them more or less generalizable to the general public. More generally, we expect that future studies that involve other instructional media types than live action videos and diagrams will establish whether the findings of this study are of limited relevance, or have far-reaching implications for the screen-based, visual,

instructional field as a whole. In summary, then, the value GTS as an engagement measure possibly could add within the technology mediated learning milieu needs to be explored in more detail.

#### 4.6. Conclusion

In spite of the GTS measure's validity (it does not appear to be stimuli contingent), and the appealing ease of use aspects, in a technological sense, we can conclude that there are many factors that impact on learning performance in the screen-based visual instructional situation. Offline and online ratios are therefore unlikely to provide perfectly reliable silver-bullet data that may say something of great value with respect to learners' engagement across different media platforms. Just as with gaze trackers in fancy automobiles, data from such devices probably says very little about how engaging it is to drive a car, any car (although, admittedly, this remains to be tested, see [Kapitaniak, B. et al., 2015](#) for a review). In other words, GTS-scores appear to be a media independent engagement measure, and, in this context, this is a redeeming aspect, but is not reliable, it says very little about learners' engagement.

There is a reciprocal and intertwined relation between instructional design and visual attention. This relationship is complex. Weighing the practicality of GTS as an engagement measure against its potential drawbacks, including its privacy implications, we conclude that GTS when used as an engagement measure captures some aspects of this complexity. Concretely, GTS-informed assessment methods may provide a straightforward means for delineating engagement levels in instructional situations that feature structural diagrams, if, and when, learners' quick assembly durations are considered to be a success factor. Thereby, the exploration of GTS as engagement measure in this article potentially increases the capacity of educators to help students and leverage desired learning outcomes, in certain computer mediated instructional activities.

#### Declarations

##### Author contribution statement

Per Erik Eriksson: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Thorbjörn Swenberg: Contributed reagents, materials, analysis tools or data; Performed the experiments; Analyzed and interpreted the data.

Xiaoyun Zhao: Analyzed and interpreted the data.

Yvonne Eriksson: Conceived and designed the experiments; Analyzed and interpreted the data.

## Funding statement

The research project Video as Design Nexus was sponsored by Dalarna University, the European Regional Development Fund, and the Municipality of Falun, Sweden.

## Competing interest statement

The authors declare no conflict of interest.

## Additional information

No additional information is available for this paper.

## References

- Anderson, J.D., Anderson, B.F. (Eds.), 2005. Moving image theory: Ecological considerations. Chapter: Preliminary Considerations', pp. 1–6.
- Ayres, P., Paas, F., 2007. Making instructional animations more effective: a cognitive load approach. *Appl. Cognit. Psychol.* 21 (6), 695–700.
- Boucheix, J., Forestier, C., 2017. Reducing the transience effect of animations does not (always) lead to better performance in children learning a complex hand procedure. *Comput. Hum. Behav.* 69, 358–370.
- Boucheix, J., Lowe, R.K., 2010. An eye-tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learn. InStruct.* 20 (2), 123–135.
- Boucheix, J., Lowe, R.K., Putri, D.K., Groff, J., 2013. Cueing animations: dynamic signaling aids information extraction and comprehension. *Learn. InStruct.* 25, 71–84.
- Castro-Alonso, J., Ayres, P., Paas, F., 2015. Animations showing lego manipulative tasks: three potential moderators of effectiveness. *Comput. Educ.* 85, 1–13.
- Chen, C., Wu, C., 2015. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Comput. Educ.* 80, 108–121.
- Cohn, N., 2013. Visual narrative structure. *Cognit. Sci.* 37 (3), 413–452.
- Cojean, S., Jamet, E., 2017. Facilitating information-seeking activity in instructional videos: the combined effects of micro- and macroscaffolding. *Comput. Hum. Behav.* 74, 294–302.



- Clark, R.C., Mayer, R.E., 2016. *E-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, fourth ed. Wiley, Hoboken, New Jersey.
- Daniel, M., Tversky, B., 2012. How to put things together. *Cognit. Process.* 13 (4), 303–319.
- De Koning, B.B., Tabbers, H.K., Rikers, R.M.J.P., Paas, F., 2010. Attention guidance in learning from a complex animation: seeing is understanding? *Learn. InStruct.* 20 (2), 111–122.
- Duchowski, A., 2007. *Eye Tracking Methodology: Theory and Practice*.
- Eriksson, P.E., Eriksson, Y., 2015. Syncretistic images: iPhone fiction filmmaking and its cognitive ramifications. *Digit. Creativ.*
- Eriksson, P.E., Eriksson, Y., Swenberg, T., Johansson, P., 2014. Media instructions and visual behavior: an eye-tracking study investigating visual literacy capacities and assembly efficiency. In: Paper Presented at the HBiD, 2014, Conference.
- Figl, K., Derntl, M., Rodriguez, M.C., Botturi, L., 2010. Cognitive effectiveness of visual instructional design languages. *J. Vis. Lang. Comput.* 21 (6), 359–373.
- Franconeri, S.L., Simons, D.J., 2005. The dynamic events that capture visual attention: a reply to Abrams and Christ (2005). *Percept. Psychophys.* 67 (6), 962–966.
- Fredricks, J.A., Blumenfeld, P.C., Paris, A.H., 2004. School engagement: potential of the concept, state of the evidence. *Rev. Educ. Res.* 74 (1), 59–109.
- Ganier, F., de Vries, P., 2016. Are instructions in video format always better than photographs when learning manual techniques? the case of learning how to do sutures. *Learn. InStruct.* 44, 87–96.
- Gibson, J.J., 1979; 2015; 2014. *The Ecological Approach to Visual Perception: Classic Edition*, Classicition. ed. Psychology Press, Hoboken.
- Henrie, C.R., Halverson, L.R., Graham, C.R., 2015. Measuring student engagement in technology-mediated learning: a review. *Comput. Educ.* 90, 36–53.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, Oxford.
- Holsanova, J., 2001. *Picture Viewing and Picture Description: Two Windows on the Mind*. PhD Thesis, Cognitive Studies 83. Lund University.

- Van Hooijdonk, C., Kraemer, E., 2008. Information modalities for procedural instructions : the influence of text, pictures, and film clips on learning and executing RSI exercises. *IEEE Trans. Prof. Commun.* 51 (1), 50–62.
- Hvelplund, K.T., 2011. Allocation of Cognitive Resources in Translation: an Eye-tracking and Key-logging Study. PhD Dissertation. Copenhagen Business School. PhD Series 10.2011.
- Höffler, T.N., Leutner, D., 2007. Instructional animation versus static pictures: a meta-analysis. *Learn. InStruct.* 17 (6), 722–738.
- Höffler, T.N., Koć-Januchta, M., Leutner, D., 2017. More evidence for three types of cognitive style: validating the Object-Spatial imagery and verbal questionnaire using eye tracking when learning with texts and pictures. *Appl. Cognit. Psychol.* 31 (1), 109–115.
- Ibrahim, M., 2012. Implications of designing instructional video using cognitive theory of multimedia learning. *Crit. Quest. Educ.* 3 (2), 83.
- Ibrahim, M., Callaway, R., Bell, D., 2014. Optimizing instructional video for pre-service teachers in an online technology integration course. *Am. J. Dist. Educ.* 28 (3), 160–169.
- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., Gerjets, P., 2013. Learning to see: guiding students' attention via a model's eye movements fosters learning. *Learn. InStruct.* 25, 62–70.
- Just, M.A., Carpenter, P.A., 1980. A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87 (4), 329.
- Kaiser, M.D., Shiffrar, M., Pelphrey, K.A., 2012. Socially tuned: brain responses differentiating human and animal motion. *Soc. Neurosci.* 7 (3), 301–310.
- Kapitaniak, B., Walczak, M., Kosobudzki, M., Józwiak, Z., Bortkiewicz, A., 2015. Application of eye-tracking in drivers testing: a review of research. *Int. J. Occup. Med. Environ. Health* 28 (6), 941–954.
- Kriz, S., Hegarty, M., 2007. Top-down and bottom-up influences on learning from animations. *Int. J. Hum. Comput. Stud.* 65 (11), 911–930.
- Loh, K., Tan, B., Lim, S., 2016. Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies. *Comput. Hum. Behav.* 63, 943–947.
- Lohmeyer, Q., Meboldt, M., 2015. How we understand engineering drawings: an eye-tracking study investigating skimming and scrutinizing sequences. In:

Proceedings of the International Conference on Engineering Design, ICED, 2(80-02), pp. 359–368.

Lowe, R., Boucheix, J., 2011. Cueing complex animations: does direction of attention foster learning processes? *Learn. InStruct.* 21 (5), 650–663.

Marcus, N., Cleary, B., Wong, A., Ayres, P., 2013. Should hand actions be observed when learning hand motor skills from instructional animations? *Comput. Hum. Behav.* 29 (6), 2172–2178.

Matthiesen, S., Meboldt, M., Ruckpaul, A., Mussnug, M., 2013. Eye tracking, a method for engineering design research on engineers' behavior while analyzing technical systems. In: *Proceedings of the International Conference on Engineering Design, ICED*, vol. 7, pp. 277–286.

Mayer, R.E., 2005. *The Cambridge Handbook of Multimedia Learning*. University of Cambridge, New York; Cambridge, U.K.

Mayer, R.E., Estrella, G., 2014. Benefits of emotional design in multimedia instruction. *Learn. InStruct.* 33, 12–18.

McCloud, S., 1993. *Understanding Comics. The Invisible Art*. Harper Perennial, New York.

Merkt, M., Weigand, S., Heier, A., Schwan, S., 2011. Learning with videos vs. learning with print: the role of interactive features. *Learn. InStruct.* 21 (6), 687–704.

Ozcelik, E., Arslan-Ari, I., Cagiltay, K., 2010. Why does signaling enhance multimedia learning? evidence from eye movements. *Comput. Hum. Behav.* 26 (1), 110–117.

Ozcelik, E., Karakus, T., Kursun, E., Cagiltay, K., 2009. An eye-tracking study of how color coding affects multimedia learning. *Comput. Educ.* 53 (2), 445–453.

Paas, F., Sweller, J., 2012. An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educ. Psychol. Rev.* 24 (1), 27–45.

Rosenfield, M., Jahan, S., Nunez, K., Chan, K., 2015. Cognitive demand, digital screens and blink rate. *Comput. Hum. Behav.* 51, 403–406.

Richards, C.J., Bussard, N.D., Newman, R., 2007. Weighing-up line weights: the value of differing line thicknesses in technical illustrations. *Inf. Des. J.* 15 (2), 171–181.

Ruckpaul, A., Kriltz, A., Matthiesen, S., 2015. Differences in analysis and interpretation of technical systems by expert and novice engineering designers. In: *Paper*

Presented at the International Conference on Human Behavior in Design 14–17 October 2014, Ascona, Switzerland, 2(80-02), pp. 339–348.

Scheiter, K., Eitel, A., 2015. Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learn. InStruct.* 36, 11–26.

Sjørup, A.C., 2013. Cognitive Effort in Metaphor Translation: an Eye-tracking and Key-logging Study. PhD Dissertation. Copenhagen Business School. PhD Series 18-2013.

Sweller, J., 1988. Cognitive load during problem solving: effects on learning. *Cognit. Sci.* 12 (2), 257–285.

Sweller, J., 2010. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* 22 (2), 123–138.

Sweller, J., Ayres, P., Kalyuga, S., 2011. *Cognitive Load Theory*. Springer, New York.

Swenberg, T., Eriksson, P.E., 2017. Effects of continuity or discontinuity in actual film editing. *Empir. Stud. Arts.*

Stafford, B.M., 2016. Seizing attention: devices and desires. *Art Hist.* 39 (2), 422–427.

Taub, M., Azevedo, R., Bouchet, F., Khosravifar, B., 2014. Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Comput. Hum. Behav.* 39, 356–367.

Tversky, B., 2011. Visualizing thought. *Top. Cognit. Sci.* 3 (3), 499–535.

Tversky, B., Morrison, J., Betrancourt, M., 2002. Animation: can it facilitate? *Int. J. Hum. Comput. Stud.* 57 (4), 247–262.

van Marlen, T., van Wermeskerken, M., Jarodzka, H., van Gog, T., 2016. Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Comput. Human Behav.* 65, 448–459.

Wang, J., Antonenko, P., 2017. Instructor presence in instructional video: effects on visual attention, recall, and perceived learning. *Comput. Hum. Behav.* 71, 79–89.

Watson, G., Butterfield, J., Curran, R., Craig, C., 2010. Do dynamic work instructions provide an advantage over static instructions in a small scale assembly task? *Learn. InStruct.* 20 (1), 84–93.

Wong, A., Leahy, W., Marcus, N., Sweller, J., 2012. Cognitive load theory, the transient information effect and e-learning. *Learn. InStruct.* 22 (6), 449–457.