EVALUATING THE IMPACT OF INTERACTIVE CINEMA IN AUDIOVISUAL LITERACY AND EDUCATION: AN EYE-TRACKING STUDY[1]

*Giorgos Dimitriadis**
geodim@enl.auth.gr

*Katerina Gouleti**
agouleti@enl.auth.gr

*Michalis Kokonis**
kokonis@enl.auth.gr

**School of English – Aristotle University, Thessaloniki*

Abstract

The present paper explores the use of interactive cinema in education from the perspective of young students' responsiveness. With the use of an eye-tracker, gaze data was collected from students aged between 10-13 years who watched a short film extract, enhanced with a number of interactive elements or "hotspots" that appeared during playback. During the experiment, the eye-tracker collected data on the behavioral patterns of

[1] This research is co-financed by Greece and the European Union (European Social Fund – ESF) through the Operational Program "Human Resources Development, Education and Lifelong Learning 2014-2020" in the context of the project "Implementing Audiovisual Media in Education: Evaluation and Application of Eye-tracking Data" (MIS 5005088).

https://doi.org/10.33919/dasc.23.6.8

student participants with regard to their willingness to access the interactive elements on screen, the time they devoted to reading them, possible optimal positions of those elements, etc., also studied in relation to demographic information concerning age group, area of living, and gender of the participants. The aim of the experiment was to assess students' responses in the context of exploring the prospects for using interactive cinema in education in order to teach elements of audiovisual literacy as well as any other cross-curricular content.

Keywords: Film, interactivity, multimedia learning, eye-tracking, audiovisual literacy.

1. Introduction

The integration of film in the teaching and learning (T-L) process is a practice with a generally positive impact. As such, combining it with new technologies of interactivity and immersion within the context of multimedia learning should be aimed at further improving and facilitating learners' experience, always in relation to the desired learning outcomes. Nevertheless, considerations arise in relation to the introduction of new tools in T-L, as well as the effective use of existing ones. One might consider the integration of films in education as being somewhere in between: on one hand, films are already being used in T-L in several ways by instructors of all levels specializing in various teaching subjects; but on the other, films do not seem to enjoy the established status other teaching aids do, such as literature for instance. Despite its dominance in entertainment, film has been struggling with securing a permanent place in T-L for quite some time now, at the same time that interactivity is gaining ground in people's everyday routine. Taking that as a starting point, the present paper addresses the role that film can indeed play in T-L: based on an experiment that tests the use of interactive narrative film in a learning activity, the project utilizes eye-tracking to monitor and assess student response, and, by extension, the viability of using this form of interactive cinema as the basis of educational resources.

There was a conscious attempt in the project both to test a form of film that matches the increasingly haptic way of interacting with visual texts today, and to support it with the level of reliability that eye-tracking data can provide. The fact that audiovisual content is delivered to viewers today to a large extent with the use of devices such as computers, smartphones, tablets or smart TVs involves accessing audiovisual texts through touch screens or mouse clicks that require interaction with some form of user interface. This condition creates new possibilities for using such texts in education, but also poses research challenges, since a better understanding of the cognitive operations involved in such an interaction is required. Eye-tracking was deemed to be the proper tool to address those challenges; because of its ability to monitor viewers' visual attention, eye-tracking is "very suited to study differences in attentional processes evoked by different types of multimedia and multi-representational learning materials" which usually involve the use of computers (Van Gog & Scheiter 2010: 95).

The form and interface of an interactive form of cinema in the project was based on the communicative and informative potential of an audiovisual text in the context of multimedia learning. It was also driven by relevant research indicating that "it is not the media that matter, but how they are used" (Fletcher & Tobias 2005: 118), and that a lot of care is required in the "design, format and configuration of the content shown" in multimedia learning (Molina et al. 2018: 45). That way, teaching materials may benefit from a form of audiovisual narrative text that is not only entertaining but also open to an array of possibilities. Such possibilities include the integration of additional, external information that is combined with the audiovisual text to significantly expand the quality and quantity of information it can provide: interactive films of this kind are enhanced with multimodal[2] connections to other audiovisual content or texts, that build on the knowledge that those films can offer. The element of interactivity also enriches the functionality and appeal of such films: the success of a properly designed interactive audiovisual text lies in its ability to invite engagement with it, which is a strongly desired feature in T-L activities as well. At the same time, interactivity respects students' learning pace by allowing them to access that extra information at will. In other words, interactive cinema that has been successfully integrated in an educational activity can render it an appealing source of information worth exploring through multimedia learning.

2. Foundations of the present study

a. Film and Multimedia Learning

The impetus behind this project has been the increasing exposure of people, especially of younger ages, to multimedia of all sorts, and the con-

[2] The terms "multimodal" and "multimedia" are strongly related to each other in T-L. In fact, Moreno & Mayer (2007) have used them interchangeably (p. 309), referring to a definition of "multimodal learning environments as learning environments that use two different modes to represent the content knowledge: verbal and non-verbal" (Paivio, 1986, as cited in Moreno & Mayer 2007: 310). In the latter mode, the authors include both static graphics such as photos, images, maps, etc., as well as dynamic graphics such as video (p. 310).

sequent need for further research on how this can be put to good use. Relevant studies make this need obvious; in Sweden, for instance, a significant increase has been observed in children's use of various types of digital media since the early 2010s, with children aged between 9-14 years spending an average of 87' per day online (Gidlöf et al. 2012: 330), an age group which includes the one that the present study focuses on. This observation directs attention to the possible ways of enabling a better interaction with the content of that media. Specifically concerning the comprehension of film within the present context, several definitions have been suggested for the concept of visual literacy, for instance the one offered by Nöth (2003) as "the ability to decode the pictorial repertoire of the media without indexical or iconic support;" or Messaris's (1994) multi-levelled understanding of visual literacy as ranging between comprehending the content of visual media, to developing skills for recognizing their aesthetic qualities (as cited in Scheiter et al. 2009: 78). Considering the lack of tools to accurately determine students' visual literacy (Scheiter et al. 2009: 78), as well as the general observation that today people's processing capabilities are under increasing strain (Lajoie & Nakamura 2005: 489), it seems very reasonable to steer research attention towards exploring the ways this added information is, or can be, handled more efficiently. Acquiring more insight into processes of visual literacy and multimedia learning would enable students to respond better to material such as film within T-L conditions.

This need for further research should be geared towards the nature of multimedia content and its connection to film in particular. Although a final definition for multimedia still seems rather elusive, film seems to be an integral part of its various versions. Collins et al. (1997) refer to this difficulty and opt for a definition that includes at least three out of a list of six audiovisual components presented on a computer, such as video, sound and text (pp. 3–4). Another early definition describes multimedia as "the combination of various digital media types such as text, images, sound and video, into an integrated multi-sensory interactive application or presentation to convey a message or information to an audience" (Neo & Neo 2001: 20). A simpler, more versatile, and presently more relevant definition is the one by Mayer (2005b) who defines multimedia as "presenting both words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video)" (p. 2). Although the term remains distinct from that of *hypermedia*, the two concepts have been associated with each other (Collins et al. 1997: 5; Dillon & Jobst 2005: 569), especially when "the interactive aspects of multimedia" come into play (Lajoie & Nakamura 2005: 490).

The incorporation of our version of interactive film in T-L, therefore, falls well within the scope of *multimedia learning* and *instruction*, that is, the process by which learners are exposed to a combination of material presented in both words and images, which is based on Mayer's *multimedia principle*: this simultaneous exposure to words and images is believed to significantly facilitate learning (Fletcher & Tobias 2005: 117–118; Mayer 2005a: 31–32, 2005b: 3, 2009: 4–5, 2014: 385). Mayer's theory is based on the assumption that "learners are limited-capacity dual encoders who actively process information in order to integrate it meaningfully with their existing knowledge" (Dillon & Jobst 2005: 570). There are some additional parameters in the present project related to multimedia learning. First, there are debatable indications that the use of video or animation is actually more attractive than still pictures as one of the types of media used in education (Fletcher & Tobias 2005: 123; Takacs & Bus 2016); and second, with regard to the form of the material, studies reveal the importance of presenting the diverse sources of information in multimedia in a contiguous manner; Mayer's principles of *temporal contiguity* and *spatial contiguity* stress the increased effectiveness for learning when words and pictures are presented simultaneously, as well as close to one another on a page or screen (Ayres & Sweller 2014: 140–143; Fletcher & Tobias 2005: 121; Mayer 2005c: 184). In fact, one of these two channels of information has also been found to be effective when it is aural, which directly relates to the perception of film texts (Ayres & Sweller 2014: 143). In the present study, the film extract that incorporated the added interactive elements is in line with these principles: all the interactive elements that were used contained text as well as, in some cases, an explanatory image; they were also placed within the film borders during playback and thus coincided with the film features that they explained.

b. Advancing technologies in education: interactivity and the role of eye-tracking

The degree to which films and other audiovisual material can be used for educational purposes is also affected by changes in available technologies. The rapidly spreading availability of streaming services and online content creates the conditions for reconsidering the use of films in T-L. The increasing commercially available speed of internet connections, technological innovations, and wide availability and affordability of computers have also gradually allowed audiovisual content to be incorporated in T-L activities (Lajoie & Nakamura 2005: 490). Technology can be beneficial for students, allowing the use of "attractive and versatile teaching electronic materials providing information in the form of text and images, moving graphic elements as well as synchronized verbal information" (Molina et al. 2018: 45). Precisely within these conditions, research is shifted towards learners' interaction with multimodal material as a way of learning through active engagement rather than simple observation (Lajoie & Nakamura 2005: 490; Renkl & Atkinson 2007: 235).

It becomes obvious that the concepts of interactivity and multimodality lie at the heart of these research considerations. The commercial shift towards interactive forms of entertainment[3] is the technological outcome of abundance and availability of content, as much as it is an opportunity for novelty in T-L. In an early account of using interactive videos in T-L, Norris et al (1990) had highlighted the need for more "reliable experimental evidence" on the use of IT in education, while stressing the positive effects of interactive video on students' learning pace, enjoyment and motivation (cited in Collins et al. 1997: 21). In the early 2000s non-interactive videos were considered "much less effective for creating contexts that students can explore and reexamine, both individually and collaboratively" (Bransford et al. 2000: 209) and studies in interactivity and learning were considered insufficient (Kettanurak et al. 2001: 542). Despite the fact that interactivity still also lacks a fixed definition (Domagk et al. 2010: 1024–1025; Moreno & Mayer 2007: 310; Renkl & Atkinson 2007: 235), it is considered today by many as the most promising form of educational technology (Domagk et al. 2010: 1024). With the changing technological landscape allowing interactive forms of multimedia to increase dramatically, and with a wealth of digital tools and film content also readily available for individual and educational use, the same need for more targeted research still persists, probably more than ever before.

The significance of using eye-tracking in a project such as this lies not only in the kind of knowledge that it can provide, but also in the fact that relevant eye-tracking research still seems to be relatively limited. In contrast to the prominent use of traditional research tools[4] to comprehend the cognitive impact of multimedia learning, methods that directly study "the cognitive and perceptual processes underlying these effects are relatively rare" (Van Gog & Scheiter 2010: 95). Eye-tracking can facilitate the study

[3] Gaming is probably the spearhead of interactive technologies, but even something as common today as video-on-demand (VoD) constantly engages viewers with forms of interactivity, enabling them to control their entertainment experience.

[4] Van Gog & Scheiter (2010) note that research on the cognitive impact of multimedia learning has generally been "based on (transfer test) performance measures, sometimes combined with measures of cognitive load and/or time-on-task" (p. 95).

of those processes by enabling researchers to obtain immediate, raw data about the visual behavior of participants. Within the wider benefits of using it for studying human-computer interfaces (Sungkur et al. 2015: 1786), especially when working with younger participants, eye-tracking "provides information not consciously controlled by the students," offering an insight into "their interests and preferences, which is more difficult to obtain using traditional techniques" (Molina et al. 2018: 45), and more specifically into the "cognitive process of learning" (Lai et al. 2013: 91). Responding to the current need for obtaining data on the way multimedia learning is cognitively processed, eye-tracking can be particularly helpful in studies that include "multimedia multi-representational learning materials" as it can "provide unique information concerning what medium or representations are visually attended to, in what order, and for how long" (Van Gog & Scheiter 2010: 95); this, in turn, can help research on multimedia learning "overcome the limitations of self-reporting measurements" (Alemdag & Cagiltay 2018: 413). Despite the fact that the need for more eye-tracking research on the combination of text and images in education has been repeatedly pointed out (Jacob & Karn 2003: 587; Schmidt-Weigand 2009: 92), relevant work on this field still remains inadequate and in need of further contributions (Alemdag & Cagiltay 2018: 415).[5] We believe that the present work addresses several of the underrepresented characteristics in relevant studies, specifically with regard to the researched school subject, the age groups of participants and the country of origin.[6]

3. Description and process of data collection

a. Target group

The participant sample selected for the project included 82 students attending the last two grades of Primary school and the first two grades of

[5] Dogusoy and Cagiltay (2009) provide an account of educational research that has been carried out with the use of eye-tracking. The studies that they present in their work include, but are not limited to, multimedia learning.

[6] Alemdag and Cagiltay (2018) provide an overview of available literature in English on eye-tracking and multimedia learning published in the period 2010-2016. First, film was not researched in any of the considered studies as it does not appear in the table presenting the various researched subjects provided by the authors. Second, the percentage of studies with Secondary education participants (Middle school in particular) amounted to 5.8%, with Primary education participants (mentioned as Elementary students) constituting 1.2%, and mixed students – without mentioning specific age groups – a mere 1.2%. Finally, the country of origin for published papers also showed relatively little diversity: 65% were from Europe, with 25 out of 39 being from Germany (pp. 417–418).

Secondary school, which, in the Greek educational system, generally include the ages between 10 and 13. This selection was based on the cognitive abilities of students in relation to both the material used in the activity and their familiarity with computers: further widening the age gap between the two educational levels (recruiting students of e.g. 12 and 16 year of age) would probably and quite predictably produce unequal results given the nature of the present experiment. More specifically, such a selection may have been more straightforward in terms of the comparison of the two levels; it might also have been more applicable internationally, given the fact that the separation between Primary and Secondary education at the age of 10-13 reflects specifically the Greek educational system. Nevertheless, considering a wider age gap such as this would be more appropriate once the purposes of the present experiment have been explored: first a basic understanding of student engagement with interactive film needs to be established, before a subsequent, more fine-grain research can place more emphasis on specific age groups. Such follow-up research could also focus on additional parameters, making use of material that is e.g. more grade-appropriate, or differentiated in other ways as well.[7] Taking all this into consideration, separating participants into groups of Primary and Secondary school students but remaining within student groups of comparable skills also reflects a more practical scope: if a specific form of interactive film such as this one is to be implemented in only one level of education, this experiment can contribute to making that choice a more informed one.

In addition to the first demographic parameter, two additional ones were considered. The second one reflects the gender of participants: a balanced ratio of male (40) to female (42) participants was maintained. Finally, a third parameter of selection considered the area of residence based on population; a simple separation into urban, semi-urban and rural areas was deemed preferable, based on the geographical distribution of population in Greece.[8] The overlapping of all three parameters (age, gender and area of residence) in different groups within the same pool of participants allows the sample to serve the study of several demographic profiles at the same time. A balanced representation of each demographic group within the same parameter was maintained during the selection of participants, and

[7] The findings presented at the end of this paper indicate such possible directions that could be subsequently explored.

[8] For the urban area, a city of >1 million inhabitants was selected; the semi-urban and rural areas were a town of approximately 20000 inhabitants and two villages of approximately 1000 inhabitants respectively, both of which are fairly common sizes in the geographical distribution of population in Greece.

the responses of students across all demographic groups were tested with the same visual material and under the exact same conditions. Considering this co-existence of different demographic characteristics in each student, the final number of participant groups for each demographic parameter from the entire sample of 82 participants is presented in Figure 1:

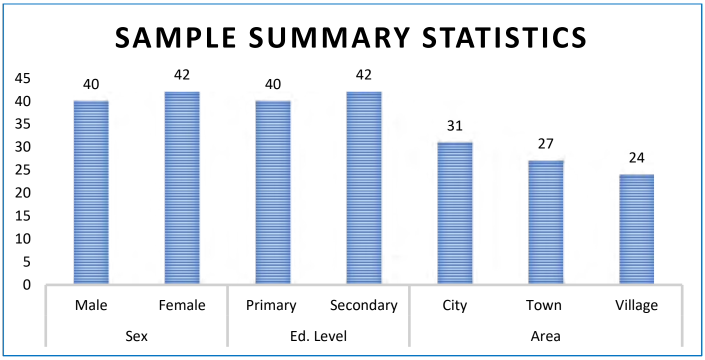

Figure 1: Breakdown of participant sample

b. Technical considerations

With regard to the interactive film that was used, an age-appropriate 4min. 24sec. film extract was selected from a mainstream narrative movie that participants had not seen before.[9] The extract featured some essential cinematic qualities, such as over-the-shoulder shots of characters in dialogue, camera movement, non-diegetic music, CGI effects, contrast between light and shadow, and location shooting. The extract comprised a long tracking shot and a dialogue scene. Interactive elements were spread equally over both these parts and their content included text and pictures. That content was kept very simple, so that participants would be able to go through each one of them in a few seconds at a normal reading rate. The entire content, i.e. both the film extract itself and the interactive elements, was in Greek so that language would not pose an additional barrier while interacting with the activity and was dissociated from any specific class material taught at school. The reasons for the latter choice were based on the fact that the study, in this particular stage, explores the students' level of engagement with interactive film in general; the feasibility of specific learning objectives within individual subjects can subsequently be ex-

[9] Before engaging with the activity, participants were asked if they had seen the specific film or had any recollection of it, and all of them replied negatively.

plored, once the conditions for optimal interactivity in a T-L environment have been better understood. In other words, relating the content of the film to a specific school subject would be irrelevant before establishing first the potential prospects of this T-L method. It would probably also create impractical complexities, as even the same or similar school subjects are taught differently in each educational level. Finally, there could even be interference with the results themselves: with material tied to a specific school subject, the preference that some students may have for that subject might affect their level of engagement with the activity, thus compromising the collected data.

With regard to the equipment, setup, and software, the conditions aimed to replicate as closely as possible a real-life T-L scenario, in which computers featuring interactive video would probably be used individually by students inside their schools. For this reason, the experiment was conducted entirely inside the participants' schools. Apart from the convenience of bringing researchers to schools rather than student participants to the researchers' lab, the "ecological validity" of a real-life environment (Duchowski 2007: 160) was deemed a desired feature, while still maintaining adequate control of the experiment conditions, for instance the importance of calibrating the eye-tracker (Bojko 2013: 178; Majaranta & Bulling 2014: 46–47). A *Tobii X2-60* portable eye-tracker was used, along with a commercially popular and industry-standard 15.6" screen laptop. The film extract was made interactive in the online platform *Wirewax*:[10] a set of 8 clickable objects, or "hotspots,"[11] were placed on the film, appearing during playback with approximately 10 to 35sec. intervals (Table 1). For each hotspot, first a neutral image appeared on screen, resembling an animated countdown clock (Fig. 2a),[12] prompting participants to click on it so that the educational content would open (Figs. 2b-2d); if participants did not click on the animated clock within 10 seconds, the hotspot would disappear permanently. The reason for this was the fact that the content of each hotspot explained the exact part of the film extract where it appeared, with

[10] See https://wirewax.app/.

[11] The term "hotspots" is used by *Wirewax* for the interactive elements that can be used with clips in the platform.

[12] Although the specific design that was used was selected from a pre-determined library within *Wirewax*, our selection considered the hotspot size, non-intrusiveness and informativity: the selected hotspots were large enough to notice but without covering too much of the film action on the screen; they were also of a non-overtly conspicuous white & grey color, and resembled a counter-clockwise index motion to signal the ten-second availability of each hotspot.

the latter functioning at the same time as a form of visual example; therefore, hotspots lingering on the screen for more time during playback would not only disconnect their content from their examples but would also start overlapping each other on screen. These timings and durations provide participants with ample time to notice and read each hotspot respectively. Table 1 breaks down the grouping of hotspots with regard to their timestamp (i.e. their time of appearance/availability inside the film extract):

Table 1: Timestamps and grouping of hotspots

Timestamp in the film extract (min:sec)	**Hotspot No.**	**Hotspot Groupings in Tobii Studio**	
00:03 – 00:13	01	Group 01	Group 09 - Tracking
00:35 – 00:45	02	Group 02	
01:15 – 01:25	03	Group 03	
01:45 – 01:55	04	Group 04	
02:30 – 02:40	05	Group 05	Group 10 - Dialogue
03:05 – 03:15	06	Group 06	
04:03 – 04:13	07 08	Group 07 Group 08	

Figure 2a: The selected design for the hotspot (magnified), featuring an imitation of clock indexes in countdown motion.

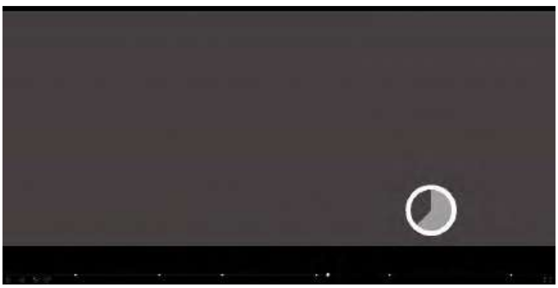

Figure 2b: A hotspot as it first appeared on the screen; the image shows its actual size relative to that of the film playback (grey area) and controls underneath it.

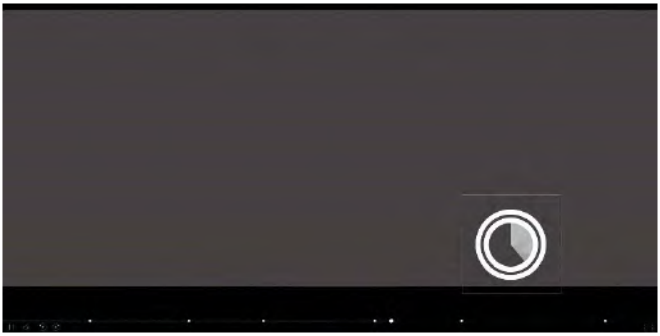

Figure 2c: The small change in the appearance of a hotspot when the mouse pointer hovered over it (i.e. before clicking on it).

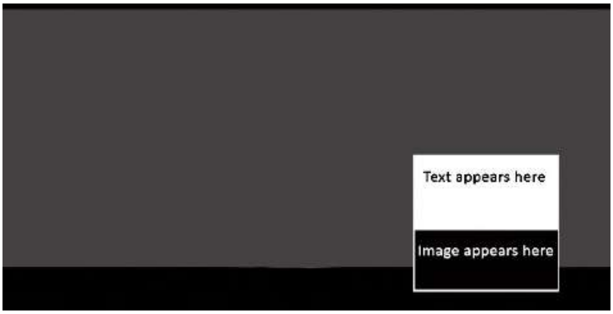

Figure 2d: The appearance of an average open hotspot content, after participants clicked on it.

The interactive video was subsequently inserted in Tobii Studio,[13] the eye-tracker software, so that AOIs (Areas of Interest) would be placed over each hotspot in order for the eye-tracker to collect gaze data only from the hotspots rather than the entire screen. In order to combine data from the same hotspots across all individual participant recordings, AOIs were combined in groups: for example, Hotspot Group 01 included the gaze data of Hotspot 01 collected from all participant recordings combined. Moreover, hotspots 01-04 appeared during the tracking shot that contained no characters or dialogue, whereas hotspots 05-08 appeared during the dialogue scene which mostly comprised over-the-shoulder shots and close-ups. Consequently, in order to consider potential differences in the participants' responses between these two parts of the video, each of these two sets of hotspots was also assigned to additional groups, Group 09 (Tracking) and Group 10 (Dialogue) respectively. Finally, hotspots 07 & 08 were set to appear and disappear simultaneously. Although they were a single item on

[13] See https://www.tobii.com/.

screen, an additional option was presented to participants after the hotspot was accessed, leading to content outside the film extract; therefore, a separate AOI was used for that option so that the eye-tracker would be able to check the responsiveness of participants to it.

4. Evaluation of data

a. Types of Data Considered

The specific data parameters studied are the following: (i) *time to first fixation*, (ii) *total visit duration*, and (iii) *time to first mouse click*. Specifically, *time to first fixation* generally indicates how quickly students responded visually to the AOIs thus also revealing their readiness to shift their attention to the hotspots.[14] The second parameter, *total visit duration*, measures the total time that participants spent looking inside the hotspots, which indicates an overall willingness of participants to maintain visual interaction with them[15] as a marker of sustained interest in the interactive parts of the project. Finally, *time to first mouse click* measures not only the swiftness but also the conscious willingness of participants to access the interactive parts of the film, and, as such, it was a key factor in evaluating the overall results in relation to the main focus of the project, which is to approach the educational viability of interactive narrative films of this form. Mouse clicks are normally expected to be more task-driven and voluntary; in other words, whereas participants' eyes may be involuntarily attracted by the sudden appearance of the hotspot as a "bottom-up" change in the salient features of the image, which can thus cause a more reflexive saccadic behavior, controlling the mouse and clicking on specific visual prompts is a much more voluntary action, thus also easier to understand as a conscious, task-related decision. Figures 3-5 present the mean *time to first fixation* – TFF (Fig. 3), *total visit duration* – TVD (Fig. 4), and *time to first mouse click* – TFMC (Fig. 5) in seconds, for all participant groups, as well as a comparison between TFF and TFMC (Fig 6). The following two sections break down the findings based on the content of the film and the demographic profile or participants respectively.

[14] This could also provide an indication of the successful (or not) design of the hotspots.

[15] *Total visit duration* differs from *total fixation duration* as it includes the total time of fixations as well as the total time of saccadic activity inside each AOI (Kim et al. 2012: 2423). Despite our temporary blindness during saccades (Gidlöf et al. 2012: 332), the total time of engaging visually with hotspots combining both fixations and saccades is more pertinent here, as the experiment presently focuses on the total time participants spent inside hotspots.

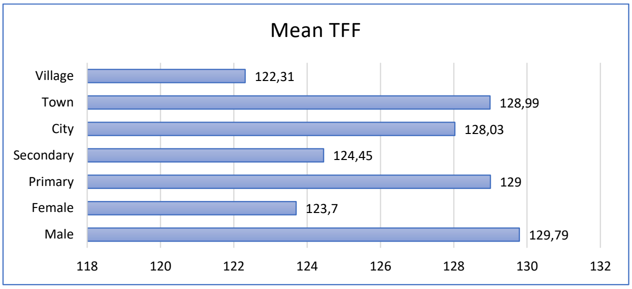

Figure 3: Mean Time to First Fixation.

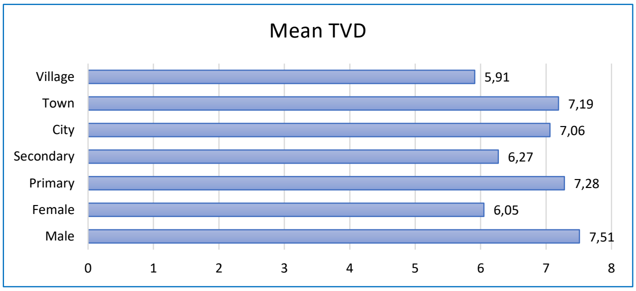

Figure 4: Mean Total Visit Duration.

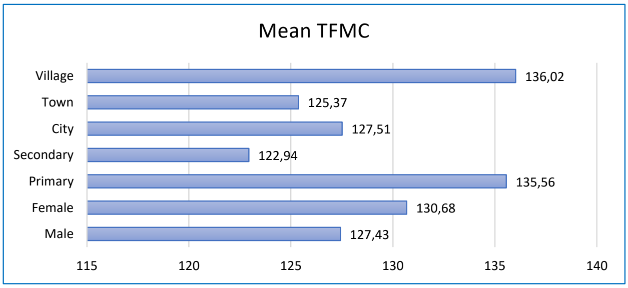

Figure 5: Mean Time to First Mouse Click.

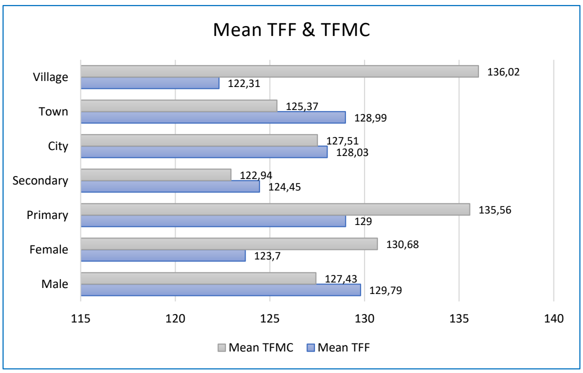

Figure 6: Comparison between mean Time to First Fixation and mean Time to First Mouse Click.

b. Findings I: Interactive Content

The first set of findings is related to the characteristics of the material used, both with regard to the features of the film extract and the way hotspots were inserted in order to enhance it. The research hypotheses related to these factors were the following:

H1. The interaction of participants with hotspots may be affected by the content featured in the underlying film.

H2. The engagement of participants may be affected by the positioning of hotspots on the film frame.

H3. An interactive movie can increase participants' willingness to delve deeper in the material taught.

The following paragraphs present the findings related to these specific hypotheses.

As described earlier, the appearance of hotspots occurred in a sequence with approx. 10 to 35sec. intervals, and identical hotspots were grouped together across all participant recordings. The fact that they appeared in a sequence obviously means that no direct comparison among them can be exported based on statistical analyses. However, comparing the time between the participants fixating on a hotspot and clicking on it can provide valuable information. In Figure 7, the curves for the average *time to first fixation* and the average *time to first mouse click* as marked in the left Y-axis are presented in conjunction with their difference (*Δ*) in the right Y-axis:

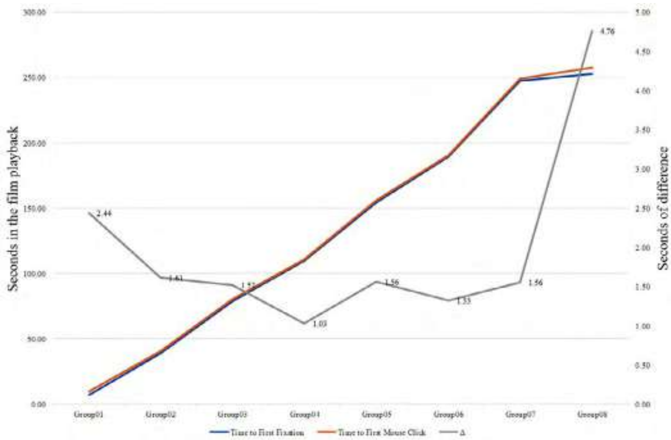

Figure 7: Comparison of the average time between Time to First Fixation and Time to First Mouse Click in the left Y-axis, with their difference (Δ) in the right Y-axis. Numbered "Groups" in the graph refer to the grouping of equivalent hotspots across recordings (see end of section 3b). Also, since the graph refers to separate hotspots, it does not include Groups 09 & 10 which combine data for sets of hotspots (See Table 1).

It can be observed that on average, this difference is decreasing until hotspot 04, and generally stabilizes between hotspots 05 and 07, implying that, on average, participants were progressively quicker in their reactions to the hotspots possibly due to their increasing learning curve. The lowest average difference of 1.03" is marked at the end of the tracking shot (hotspot 04), and the second lowest difference is reported in hotspot 06. The biggest difference is found in hotspot 08 where the gap between the average *time to first fixation* and the average *time to first mouse click* is 4.76", representing the time that participants needed for considering the additional option they were presented with after the hotspot was accessed.

Figure 7 demonstrates that there was a gradually faster tendency in the first part of the extract (hotspots 01-04, tracking shot) to use the mouse right after the first fixation. That tendency was partly restrained and stabilized when the underlying film changed form in the second part (hotspots 05-07, dialogue). It is obvious that, after the compositional features of the film extract changed, participants required some time to get accustomed to the new form. More specifically, without the speed of using the mouse regressing to the normally anticipated low levels of hotspot 01, the dialogue scene, being significantly different from the preceding tracking shot which had no visible editing, no words spoken and no specific characters participating, required participants to re-adjust the way they would perceive the new scene which contained characters conversing.

There is one additional important element that needs to be mentioned here and will provide further insight into the present research questions. Although participants experienced hotspots 07 and 08 as one, they were analyzed separately (using separate AOIs) because hotspot 08 provided participants with the option to freely access additional online material at will. This provides important information about the behavior of participants when given an option that would take them outside the material contained in the activity: selecting one of the two virtual buttons included inside hotspot 08 opened a normal internet browser outside the film extract, with information on the plot and music score of the film. It is noteworthy that only 8 out of the 82 participants clicked on this option at all; the sample therefore clearly exhibited an extremely low interest in accessing anything beyond the confines of the given material.

c. Findings II: Demographic Considerations

The other important aspect concerning the findings relates to the demographic profile of participants, specifically their level of education (primary, secondary), gender (male, female), and area of living (rural, semi-urban, urban). This particular part of the experiment extracted conclusions regarding the overall visual behavior of participant groups rather than focusing on individual hotspots. The research hypotheses associated with these parameters were the following:

H4. Primary school students are expected to have slower and, overall, less interaction with the video content compared to Secondary school students.

H5. Male and female students are expected to have similar levels of interaction with the content.

H6. The geographical factor (area of living) may have some connection to the participants' degree of interaction with the content.

Before going into the findings, it is important to stress that some difference in gaze behavior can be expected between children and adults. As a general observation, differences in the demographic profile of viewers may be among the factors that affect whether bottom-up or top-down cognitive mechanisms prevail during film watching (Dyer & Pink 2015). Specifically about age, Cohen refers to Day's (1975) observations that children differ from adults "in the speed, efficiency, systematicity, and exhaustiveness of visual scanning," as well as Mackworth and Bruner's (1970) findings that "children do not fixate the most informative areas of a picture as frequently as adults" (as cited in Cohen 2017: 273). Based on that, the level of attention of young students to the most informative areas of the screen is expected to be different compared to that of adults.

Returning to the investigation of the hypotheses themselves, the data generally support H4 (Table 2). Specifically, the data on *time to first fixation* revealed a noticeable difference between Primary and Secondary education students. Secondary education participants established their first fixation by an average of 4.55" earlier than Primary education participants. The second parameter, *total visit duration*, on the other hand, demonstrates less interaction with hotspots by Secondary education participants, at an average of 6.27" in contrast to 7.28" of Primary education participants. Given the intentional choice of using small-sized hotspots, both these visit durations are sufficient to read their content under normal circumstances, but this relatively small difference in seconds still translates to Secondary education participants devoting approximately 13.8% less time to the hotspots. Finally, *time to first mouse click* reveals the most significant difference so far: Secondary education participants interacted with hotspots by clicking on them for the first time by an average of 12.62" faster than Primary education participants. Any interaction with the film was an optional task for all participants, and both groups were almost equally represented in the 73 out of 82 participants that used the mouse: 38 out of 73 (approx. 52.05%) were from Secondary education, whereas 35 out of 73 (approx. 47.95%) were from Primary education. These observations certainly reveal a higher level of interaction by Secondary education students, which supports H4 and thus draws attention to either the extent of applicability of such material in Primary education or, at least, to the type of interactivity or material that can be used at that level of education.

Table 2: Data comparison between Primary & Secondary education participants

Grade	mean(tff_al~n)	mean(tvd_al~n)	mean(tfmc_a~n)
Primary	129.00	7.28	135.56
Secondary	124.45	6.27	122.94

→ Ed. Level = Primary

Variable		Obs	Mean	Std. Dev.	Min	Max
TFF	~n	40	128.99	9.67	97.83	145.03
TVD	~n	40	7.28	3.93	.64	15.37
TFMC	~n	35	135.56	32.72	96.6	253.96

→ Ed. Level = Secondary

Variable		Obs	Mean	Std. Dev.	Min	Max
TFF	~n	42	124.45	19.20	59.33	153.02
TVD	~n	42	6.27	3.12	.61	19.32
TFMC	~n	38	122.93	26.37	27.12	169.73

Moving on to H5, the gender of participants produced relatively mixed results (Table 3). The *time to first fixation* revealed a wide gap, as female participants fixated on hotspots by an average of 6.09" earlier than male participants. This observed speed in female students accessing the interactive content of the film can be correlated with the findings from the *total visit duration*: in a combined average of 6.78" between male (7.51") and female participants (6.05"), the latter were found to spend less time seeing interactive hotspots by 1.46". In other words, whereas female participants visually accessed interactive hotspots much faster than male participants, they spent approximately 19.3% less time on them throughout the project.

The last parameter, *time to first mouse click*, produced the same considerations as in the Primary-Secondary education set of participants, with regard to the adequate representation of both male and female participants in the sample of those who indeed used the mouse to open the interactive hotspots. In this case as well, the two groups were represented almost equally (35 female over 38 male). With regard to the way they performed, male participants interacted with the film generally earlier by average, making their earliest and latest first mouse clicks, respectively, 23.89” earlier and only 4.22” later compared to female participants.

Table 3: Data comparison between male & female participants

Gender	mean(tff_al~n)	mean(tvd_al~n)	mean(tfmc_a~n)
Male	129.79	7.51	127.43
Female	123.70	6.05	130.68

→ gender = Male

Variable		Obs	Mean	Std. Dev.	Min	Max
TFF	~n	40	129.79	9.68	97.36	142.43
TVD	~n	40	7.51	3.73	1.44	19.32
TFMC	~n	38	127.43	29.61	27.12	253.96

→ gender = Female

Variable		Obs	Mean	Std. Dev.	Min	Max
TFF	~n	42	123.69	18.97	59.33	153.02
TVD	~n	42	6.05	3.26	.61	14.45
TFMC	~n	35	130.68	30.87	51.01	249.74

Finally, H6 assumed a level of connection between the area of residence and the participants’ engagement with the material (Table 4).[16] It appears that all three parameters yielded similar results for the city and town residents, which were markedly different from those for village residents:

[16] This specific hypothesis (H6) considers possible connections between the participants’ areas of residence and their interaction with film-based multimedia learning activities, but the inherent complexity of explaining the various possible factors related to it (location, infrastructure, etc.) exceeds the purposes of the present study.

Table 4: Data comparison for participants from different areas of residence:

Area	mean(tff_al~n)	mean(tvd_al~n)	mean(tfmc_a~n)
City	128.03	7.06	127.51
Town	128.99	7.19	125.37
Village	122.31	5.91	136.02

→ area = City

Variable		Obs	Mean	Std. Dev.	Min	Max
TFF	~n	31	128.02	12.14	79.32	142.43
TVD	~n	31	7.05	3.06	1.84	12.88
TFMC	~n	29	127.51	16.75	86.86	154.86

→ area = Town

Variable		Obs	Mean	Std. Dev.	Min	Max
TFF	~n	27	128.99	14.86	84.9	153.02
TVD	~n	27	7.19	3.97	.64	19.32
TFMC	~n	25	125.36	27.61	27.12	169.73

→ area = Village

Variable		Obs	Mean	Std. Dev.	Min	Max
TFF	~n	24	122.30	19.06	59.33	141.88
TVD	~n	24	5.91	3.66	.61	14.45
TFMC	~n	19	136.01	45.68	51.01	253.96

Although the latter exhibit a significantly faster *time to first fixation*, their *total visit duration* was much smaller, and their *time to first mouse click* occurred much later than residents of both the other areas. In other words, rural participants' attention was captured by hotspots much faster, but they spent a lot less time looking at them and were much slower to interact with them. The parameter *time to first fixation* can more easily be attributed to bottom-up factors (i.e. related to the change in the stimulus caused by the sudden appearance of the hotspots), whereas *total visit duration* and *time to first mouse click* can be more directly associated with the given task, thus revealing a more conscious, top-down engagement with the material. As such, it appears that the findings support H6: the attention of participants from rural areas was more quickly attracted by the appearing stimulus only by 4.82% compared to the combined average of urban and semi-urban areas, but their haptic engagement with the interactive film was more noticeably smaller, by 7.57%.

5. Discussion

It was assumed in H1 that there are optimal conditions that facilitate engagement with the interactive material, related to the content of the underlying film extract. It was also revealed (Fig. 7) that participants gradually interacted faster and generally maintained that interaction with the hotspots as the film extract progressed (with the exception of the special case of hotspot 08); this demonstrates a desired increasing learning curve that is very encouraging with regard to the potential applicability of interactive films of this form and length in teaching scenarios, as participants generally appeared to grow increasingly comfortable with an interface and a film they had not seen before. Overall, considering the possibility that dialogue scenes in film could be a less compelling spectacle for these age groups in visual terms compared to the visually elaborate tracking shot, the dialogue scene seems to generally lend itself better to being used with interactive hotspots.

Given the above findings, the assumption in H2 of an optimal positioning of hotspots on screen could not be conclusively verified. Still, it is important to note that, unlike the way the first half of the film extract was modified for the experiment (hotspot set 01-04), in the second half (hotspot set 05-08) all hotspots were placed in the lower part of the screen, where subtitles usually appear; the fact that participants should normally have had some visual experience with subtitles, since foreign audiovisual material in Greece is commonly subtitled, could have made hotspots seem more naturally placed. Therefore, maintaining interaction with them was probably also facilitated by their position.

Hotspot 08 was specifically used in order to test H3, which assumed that participants would demonstrate willingness to learn more about the film by exiting the playback of the given extract and accessing other related material. It became clear from the results that this hypothesis cannot be verified, considering the sharp drop in the *total visit duration* in hotspot 08 (the dual option hotspot), as well as the fact that only 8 out of 82 participants clicked on any of the two options it provided. Although we believe that this particular hypothesis deserves to be further explored in a more specialized experiment, the participants' unwillingness to access content outside the film extract directs our attention to the probability that film extracts that use only local resources would be more effective and appealing to students. Finally, these findings, in conjunction with the fact that the difference (Δ) between *time to first fixation* and the average *time to first mouse click* (Fig. 7) gradually decreased and remained relatively stable between hotspots 01-

07 only to spike again in hotspot 08, supports both the assumed learning curve described earlier in H1, as well as the fact that clicks were not made randomly. This further enhances the observation that H3 cannot be verified: in a generally increasing and stable level of interest between hotspots 01-07, when hotspot 08 ceased to be interesting the time to first mouse click plummeted.

The demographic profile of the participants also offered valuable information. Indeed, correlating all three parameters reveals that Primary education participants demonstrated less readiness and haptic interaction with the interactive film, as predicted in H4. Nevertheless, the fact that Secondary education participants were quicker to fixate and click on hotspots, but spent less time on them, does not necessarily also mean better or faster learning of the material included in the hotspots; it simply reinforces the assumption that Secondary education participants are more swift when engaging with interactive material, and that they can go through the content faster, which can be related to their age difference. Given the fact that the film extract was the same for all participants, this could be an indication that the parameters of design and appearance of the hotspots (time, size, pacing of film extract, etc.) may need to vary even between groups of such relatively small age difference. After all, the higher average *total visit duration* observed in Primary education participants is an indication that they were indeed willing to interact with the material; instead, factors such as content or less familiarity with computers may have caused the difference, which nonetheless remains a thing to consider.

The gender and area of residence of participants were probably the most challenging sets of groups to compare, also delivering comparable results. First of all, despite the fact that H5 predicted similar levels of engagement between male and female participants, the findings are mixed: the attention of male participants was slower to capture initially, but both their gaze retention by hotspots and their physical response were higher than that of female participants. Based on these findings, H5 remains inconclusive albeit intriguing; it is one of the main parameters that should be addressed in a more specialized experiment, as it could relate to broader social factors of accessibility and educational or entertainment opportunities.

With regard to the findings related to H6 and the area of residence, the fact that the *village* group showed faster initial engagement but less gaze retention and slower physical response than both the *city* and *town* groups is not a straightforwardly explicable finding. Even if one assumes variations in the amount or even the type of technological stimuli available in the everyday activities of each group of participants, the rural areas (*villages*) where the study took place are in very close geographical proximity to the semi-urban ones (*town*); any assumed technological or entertainment facilities in the latter are fairly accessible to the former as well. Adding to the fact that both these areas are relatively farther than the urban area (*city*), one might expect similar findings in the *village* and *town* groups, in contrast to the *city* group, which was not the case, as the *village* group was the one that exhibited less interaction with the material in relation to the other two. Overall, regardless of any assumptions assigned to a relative geographical proximity, the marked differences in the data may point to an increased familiarity with using computers and interactive audiovisual interfaces as well as with movie watching. This, in turn, may translate to better skills in engaging with interactive technology, especially while watching a narrative movie. Therefore, the differences in the parameters of *total visit duration* and *time to first mouse click*, which can be associated with such learned skills, are much larger than those of *time to first fixation*, since the latter relates to a bottom-up response which is more commonly shared by all participants.

Following the last two sets of findings, that pertain to H5 and H6 specifically, there are two notable implications to consider in relation to the area of residence and the gender of participants. With regard to the former implication, geographical location seems to be a factor that affects the participants' willingness to interact with the material, but whether it is a matter of population or one of distance from larger towns and cities is yet to be revealed with a more focused study. About the latter implication, nevertheless, the question remains as to why similar differences are also demonstrated in the male-female difference. In both these cases, if the observations on demographic parameters are preliminary evidence of rural and female students receiving less exposure to technology and film watching at a young age, this is certainly a matter that requires further and more specialized research and, if found to be true, remedying.

6. Limitations, conclusion and future prospects

Despite the interesting insights that the present data offer, it is obvious that there is significant research ground that needs to be covered before we are able to discuss optimal ways of using interactive film in T-L. With regard to content, for instance, there are considerations on the degree of superiority of dynamic visuals compared to static ones in multimedia learning (Lewalter 2003, as cited in Lajoie & Nakamura 2005: 493); or the fact that background music or other irrelevant sounds may inhibit the learning process (Fletcher & Tobias 2005: 122), features that are natural components of film and are therefore in need of more targeted and focused study, in addition to testing a larger variety in the types of selected films. With regard to the type and form of the interactive material, care was taken to treat hotspots and their content in a way that would make them seem as naturally integrated in the film extract as possible; still, the practically unlimited options offered by hypermedia, interface design, the internet, and the digitization and integration of various streams of knowledge into one another create potentially endless combinations, the individual efficiency of which is yet to be explored. Regarding eye-tracking specifically, though invaluable in research such as this one, its limitations in the present context relate to the fact that observing a stimulus does not necessarily also mean comprehending it, the study of which requires complementary tools (Hyönä 2010: 173). Finally, additional uses of eye tracking are still being explored. For instance, learners may get performance feedback from the playback of their own or others' eye-tracking recordings; this means that eye-tracking can become itself a possible component of multimedia learning materials rather than merely a tool to design and develop such materials (Van Gog & Scheiter 2010: 98).

It is obvious that these are only a few of the possible directions that can be further explored in this field. The present study aspires to be the springboard for additional research which will eventually lead to a more solid and comprehensive integration of film into education, whether it is for purposes of audiovisual literacy specifically, or in support of any other teaching subject. The overall findings suggest that the field is quite promising, especially given the positive collaboration of the participating students. The more detailed relevant studies become, i.e. exploring the parameters addressed in the present study at a finer grain, the more informed the design of teaching material that incorporates film in T-L will be.

References

Alemdag, E., K. Cagiltay. 2018. "A Systematic Review of Eye Tracking Research on Multimedia Learning". *Computers & Education*, Vol. 125, 413–428. Available at: https://doi.org/10.1016/j.compedu.2018.06.023.

Ayres, P., J. Sweller. 2005. The Split-attention Principle in Multimedia Learning. In Mayer, R. E. (ed.). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 135–146. Available at: https://doi.org/10.1017/CBO9780511816819.009.

Bojko, A. 2013. *Eye Tracking the User Experience: A Practical Guide to Research*. Rosenfeld Media.

Bransford, J. D., A. L. Brown, R. R. Cocking (eds.). 2000. *How People Learn: Brain, Mind, Experience, and School (expanded edition)*. National Academy Press.

Cohen, K. M. 2017 [1981]. The Development of Strategies of Visual Search. In Fisher, D. F., R. A. Monty and J. W. Senders (eds.). *Psychology Library Edition: Perception, Vol. 8. Eye movements: Cognition and visual perception*. Routledge, 271–288.

Collins, J., M. Hammond and J. Wellington. 1997. *Teaching and Learning with Multimedia*. Routledge.

Dillon, A., J. Jobst. 2005. Multimedia Learning with Hypermedia. In Mayer, R. E. (ed.). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 569–588. Available at: https://doi.org/10.1017/CB O9780511816819.035.

Dogusoy, B., K. Cagiltay. 2009. An Innovative Way of Understanding Learning Processes: Eye Tracking. In Jacko, J. A. (ed.). *Lecture Notes in Computer Science: Vol. 5613. Human-Computer Interaction, Interacting in Various Application Domains*. Springer, 94–100. Available at: https://doi.org/10.1007/978-3-642-02583-9_11.

Domagk, S., R. N. Schwartz and J. L. Plass. 2010. "Interactivity in Multimedia Learning: An Integrated Model". *Computers in Human Behavior*, Vol. 26, No 5, 1024–1033. Available at: https://doi.org/10.1016/j.chb.2010.03.003.

Duchowski, A. 2007. *Eye Tracking Methodology: Theory and Practice* (2nd ed.). Springer.

Dyer, A. G., S. Pink. 2015. "Movement, Attention and Movies: The Possibilities and Limitations of Eye Tracking?". *Refractory: A Journal of Entertainment Media*, 25. Available at: https://refractory-journal.com/dyer-pink/.

Fletcher, J. D., S. Tobias. 2005. The Multimedia Principle. In Mayer, R. E. (ed.). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 117–133. Available at: https://doi.org/10.1017/CBO9780511816819.008.

Gidlöf, K., N. Holmberg and H. Sandberg. 2012. "The Use of Eye-tracking and Retrospective Interviews to Study Teenagers' Exposure to Online Advertising". *Visual Communication*, Vol. 11, No. 3, 329–345. Available at: https://doi.org/10.1177/1470357212446412.

Hyönä, J. 2010. "The Use of Eye Movements in the Study of Multimedia Learning". *Learning and Instruction*, Vol. 20, No. 2, 172–176. Available at: https://doi.org/10.1016/j.learninstruc.2009.02.013.

Jacob, R. J. K. and K. S. Karn. 2003. Commentary on Section 4 - Eye tracking in Human-computer Interaction and Usability Research: Ready to Deliver the Promises. In Hyönä, J., R. Radach and H. Deubel (eds.). *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Elsevier, 573–605.

Kettanurak, V. (Nui), K. Ramamurthy and W. D. Haseman. 2001. "User Attitude As a Mediator of Learning Performance Improvement in an Interactive Multimedia Environment: An Empirical Investigation of the Degree of Interactivity and Learning Styles". *International Journal of Human-Computer Studies*, Vol. 54, No. 4, 541–583. Available at: https://doi.org/10.1006/ijhc.2001.0457.

Kim, S.-H., Z. Dong, H. Xian, B. Upatising and J. S. Yi. 2012. "Does an Eye Tracker Tell the Truth about Visualizations?: Findings While Investigating Visualizations for Decision Making". *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, No. 12, 2421–2430. Available at: https://doi.org/10.1109/TVCG.2012.215.

Lai, M.-L., M.-J. Tsai, F.-Y. Yang, C.-Y. Hsu, T.-C. Liu, S. W.-Y. Lee, M.-H. Lee, G.-L. Chiou, J.-C. Liang and C.-C Tsai. 2013. "A Review of Using Eye-tracking Technology in Exploring Learning from 2000 to 2012". *Educational Research Review*, Vol. 10, 90–115. Available at: https://doi.org/10.1016/j.edurev.2013.10.001.

Lajoie, S. P., C. Nakamura. 2005. Multimedia Learning of Cognitive Skills. In Mayer, R. E. (ed.). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 489–504. Available at: https://doi.org/10.1017/CBO9780511816819.031.

Majaranta, P., A. Bulling. 2014. Eye Tracking and Eye-based Human-computer Interaction. In Fairclough, S. H., K. Gilleade (eds.). *Advances in Physiological Computing*. Springer-Verlag, 39–65. Available at: https://doi.org/10.1007/978-1-4471-6392-3_3.

Mayer, R. E. 2005a. Cognitive Theory of Multimedia Learning. In Mayer, R. E. (ed.). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 31–48. Available at: https://doi.org/10.1017/CBO9780511816819.004.

Mayer, R. E. 2005b. Introduction to Multimedia Learning. In Mayer, R. E. (ed.). *The Cambridge Handbook of Multimedia Learning*. Cambridge

University Press, 1–16. Available at: https://doi.org/10.1017/CBO9780511816819.002.

Mayer, R. E. 2005c. Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In Mayer, R. E. (ed.). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 183–200. Available at: https://doi.org/10.1017/CBO9780511816819.013.

Mayer, R. E. 2009. *Multimedia Learning* (2nd ed.). Cambridge University Press.

Mayer, R. E. 2014. Multimedia Instruction. In Spector, J. M., M. D. Merrill, J. Elen and M. J. Bishop (eds.). *Handbook of Research on Educational Communications and Technology* (4th ed.). Springer, 385–399. Available at: https://doi.org/10.1007/978-1-4614-3185-5_31.

Molina, A. I., Ó. Navarro, M. Ortega and M. Lacruz. 2018. "Evaluating Multimedia Learning Materials in Primary Education Using Eye Tracking". *Computer Standards & Interfaces*, Vol. 59, 45–60. Available at: https://doi.org/10.1016/j.csi.2018.02.004.

Moreno, R., R. Mayer. 2007. "Interactive Multimodal Learning Environments". *Educational Psychology Review*, Vol. 19, 309–326. Available at: https://doi.org/10.1007/s10648-007-9047-2.

Neo, M., K. Neo. 2001. "Innovative teaching: Using Multimedia in a Problem-based Learning Environment". *Educational Technology & Society*, Vol. 4, No. 4, 19–31. Available at: https://www.jstor.org/stable/jeductechsoci.4.4.19.

Renkl, A., R. K. Atkinson. 2007. "Interactive Learning Environments: Contemporary Issues and Trends. An Introduction to the Special Issue". *Educational Psychology Review*, Vol. 19, 235–238. Available at: https://doi.org/10.1007/s10648-007-9052-5.

Scheiter, K., E. Wiebe and J. Holsanova. 2009. Theoretical and Instructional Aspects of Learning with Visualizations. In Zheng, R. Z. (ed.). *Cognitive effects of multimedia learning*. Information Science Reference, 67–88.

Schmidt-Weigand, F. 2009. The Influence of Visual and Temporal Dynamics on Split Attention: Evidences from Eye Tracking. In Zheng, R. Z. (ed.). *Cognitive Effects of Multimedia Learning*. Information Science Reference, 89–107.

Sungkur, R. K., M. A. Antoaroo and A. Beeharry. 2015. "Eye Tracking System for Enhanced Learning Experiences". *Education and Information*

Technologies, Vol. 21, 1785–1806. Available at: http://dx.doi.org/10.1007/s10639-015-9418-0.

Takacs, Z. K., A. G. Bus. 2016. "Benefits of Motion in Animated Storybooks for Children's Visual Attention and Story Comprehension. An Eye-tracking Study". *Frontiers in Psychology*, Vol. 7, Article 1591. Available at: https://doi.org/10.3389/fpsyg.2016.01591.

Van Gog, T., K. Scheiter. 2010. "Eye Tracking As a Tool to Study and Enhance Multimedia Learning". *Learning and Instruction*, Vol. 20, No. 2, 95–99. Available at: https://doi.org/10.1016/j.learninstruc.2009.02.009.