Computers & Education 125 (2018) 413–428

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

A systematic review of eye tracking research on multimedia learning

Ecenaz Alemdag*, Kursat Cagiltay

Middle East Technical University, Department of Computer Education and Instructional Technology, 06800, Ankara, Turkey

ARTICLE INFO

Keywords:
Multimedia/hypermedia systems
Human-computer interface
Evaluation methodologies
Teaching/learning strategies

ABSTRACT

This study provides a current systematic review of eye tracking research in the domain of multimedia learning. The particular aim of the review is to explore how cognitive processes in multimedia learning are studied with relevant variables through eye tracking technology. To this end, 52 articles, including 58 studies, were analyzed. Remarkable results are that (1) there is a burgeoning interest in the use of eye tracking technology in multimedia learning research; (2) studies were mostly conducted with college students, science materials, and the temporal and count scales of eye tracking measurements; (3) eye movement measurements provided inferences about the cognitive processes of selecting, organizing, and integrating; (4) multimedia learning principles, multimedia content, individual differences, metacognition, and emotions were the potential factors that can affect eye movement measurements; and (5) findings were available for supporting the association between cognitive processes inferred by eye tracking measurements and learning performance. Specific gaps in the literature and implications of existing findings on multimedia learning design were also determined to offer suggestions for future research and practices.

1. Introduction

Multimedia learning materials, including both visual and verbal modes of instructional messages, are commonly utilized in diverse learning settings. In face-to-face classroom environments, instructors can supplement verbal instruction with static or dynamic images owing to the advances in visualization technologies (Mayer, 2014a). In addition, with the proliferation of mobile technologies and online courses, learners can easily access multimedia materials without being limited by time and place. Despite their prevalent use, the way in which cognitive activities occur in multimedia learning has recently drawn attention (Desjarlais, 2017). Interviews, behavioral assessments, and self-reporting measurements have been mostly used to make inferences about information processing in multimedia learning (Rodrigues & Rosa, 2017). However, self-reporting measurements are not capable of capturing the temporal fluctuations in cognitive processes. There is a need for measurements that directly indicate cognitive processing during multimedia learning (Mayer, 2017).

To overcome the limitations of self-reporting measurements, eye tracking technology can be used to explore cognitive activities in multimedia learning. Using eye movement measurements obtained through this technology provides the opportunity to test assumptions about where people look during the integration of text and pictures (Tabbers, Paas, Lankford, Martens, & Van Merriënboer, 2008). Moreover, educators who gather eye tracking data can enhance the design of their multimedia materials on the basis of learners' information-processing capabilities. However, the body of knowledge that interprets eye tracking measurements in terms of

* Corresponding author.
E-mail addresses: ecenazalemdag@gmail.com (E. Alemdag), kursat@metu.edu.tr (K. Cagiltay).

https://doi.org/10.1016/j.compedu.2018.06.023
Received 21 December 2017; Received in revised form 21 June 2018; Accepted 24 June 2018
Available online 28 June 2018
0360-1315/

learners' cognitive activities is not enough (Lai et al., 2013). To fill this gap, Lai et al. (2013), in their literature review, explained the use of eye tracking technology in learning studies. The authors classified studies with respect to various areas of cognitive development examined by eye tracking technology and the connections established between eye movement measures and several learning issues. This work revealed that conceptual development, perception, and language were the most-often studied areas of cognitive development; patterns of information processing and the effects of instructional strategies were also prevalent learning themes indicated by eye movements. Although Lai et al. (2013) provided general learning topics studied by eye tracking technology, their research is inadequate in offering specific explanations of cognitive activities in the domain of multimedia learning. Thus, research that investigates information processing and the variables affecting knowledge acquisition or construction in multimedia learning is necessary (Desjarlais, 2017). In addition, analyzing studies published in different databases after 2012 is crucial to collect and synthesize the latest findings on this topic. The current study systematically reviews up-to-date literature on the use of eye tracking technology in multimedia learning in diverse databases. In particular, it aims to explain the cognitive processes that eye tracking measurements are associated with, the potential factors that can affect eye tracking measurements, and the relation between eye tracking measurements and learning performance in multimedia learning research.

2. Background

2.1. Multimedia learning

Multimedia refers to presenting both words and pictures (Mayer, 2014a). Words can be either printed or spoken text in the form of sound or audio. Pictures can be either static graphics, such as diagrams, photos, and illustrations, or dynamic graphics, such as animation and video. Multimedia learning is defined as “building mental representations from words and pictures” (Mayer, 2014a, p. 2). It can occur in diverse contexts, including multimedia presentations, e-learning, computers games, simulations, and virtual reality environments, that allow learners to process information in both verbal and pictorial form.

Research on multimedia learning hypothesizes that multimedia instructional messages designed in accordance with how the human mind works can lead to a more meaningful learning (Mayer, 2014a). Based on this hypothesis, cognitive theory of multimedia learning describes how people learn from instructional multimedia materials (Mayer, 2014a). This theory comprises three assumptions: dual-channel, limited-capacity, and active processing. First, according to Paivio's dual coding theory (1986) and Baddeley's working memory model (1999), the dual-channel assumption states that humans have separate channels with which they process visual and auditory information. Second, according to Baddeley's working memory model (1999) and Sweller's cognitive load theory (1999), the limited-capacity assumption states that the amount of information people process in each channel at one time is limited. Third, the active processing assumption assumes that humans are active agents while selecting, organizing, and integrating incoming information to construct coherent mental representations (Mayer, 2014a).

Although several basic and advanced multimedia design principles have been derived on the basis of the assumptions of cognitive theory of multimedia learning (Mayer, 2014a), there is a lack of direct and objective evidence indicating the aforementioned cognitive processes in multimedia learning (Liu & Chuang, 2011). This lack stems from the difficulty in measuring cognitive processes, including attention to multimedia elements, cognitive load, and visual search (Chuang & Liu, 2012). Eye tracking technology enables researchers and practitioners to record temporal changes in learners' visual attention to obtain evidence about multimedia information processing (Hyönä, 2010). Therefore, eye tracking technology can be a useful tool to study attention allocation during learning in detail (van Gog & Jarodzka, 2013).

2.2. Eye tracking technology in multimedia learning research

Eye tracking technology provides an online protocol allowing the study of attention processes, including “what is attended first and for how long, what is attended next and for how long, how much switching of attention is done between different components of the learning materials, what components are linked together during attentional switching” (Hyönä, 2010, p. 174). Using eye tracking technology enables researchers to make inferences about which items in the scene were interesting for the person and how the scene was perceived by that person (Duchowski, 2003).

Through eye tracking technology, two main types of measurements are obtained: fixations and saccades (van Gog & Jarodzka, 2013). Fixation describes the stable state of the eye at one point. According to the eye-mind hypothesis (Just & Carpenter, 1980), eye fixations reflect the attention process. Saccade describes quick eye movement between fixations, which shows the change in the focus of visual attention (van Gog & Jarodzka, 2013). Different specific measurements can be gathered for fixations and saccades. Lai et al. (2013) categorize them into three scales: temporal, spatial, and count. Temporal scale includes measurements that indicate time spent in specific eye movements. Total fixation duration, average fixation duration, and time to first fixation are some of the eye movement measurements in the temporal scale. Spatial scale includes measurements related to “locations, distances, directions, sequences, transactions, spatial arrangement or relationships of fixations or saccades” (Lai et al., 2013, p. 93). Examples of spatial scale measurements are saccade length and fixation sequence. Count scale includes measurements that indicate the frequency of specific eye movements. Total fixation count and inter-scanning count are two of the eye movement measurements in the count scale. In addition to fixations and saccades, pupil size and blink rate are other common eye movement measurements. With the help of these specific eye movement measurements, how learners process information in different formats can be investigated (Liu, Lai, & Chuang, 2011) as it is assumed that there is a relationship between eye movements and cognitive processes (Just & Carpenter, 1976). In particular, eye movement measurements can reveal visual attention on the items in the scene, change in the focus of visual attention

(Just & Carpenter, 1980), depth of processing information (Rayner, 1998), and difficulty in processing (Jacob & Karn, 2003).

Despite the affordances of eye-tracking technology in revealing cognitive processes, researchers in the field of education have recently started to use it (Lai et al., 2013; Yang, Chang, Chien, Chien, & Tseng, 2013). The number of the studies that investigate eye movements in multimedia learning is low (Chuang & Liu, 2012; Johnson & Mayer, 2012). Therefore, it is necessary to explore how the eye tracking technology can be utilized to examine cognitive processes in multimedia learning. This will also provide support and validation for the inferences and conclusions derived from prior multimedia learning studies. Furthermore, particular gaps in the literature and suggestions for the design of multimedia learning environments can be detected on the basis of existing eye tracking research. Therefore, this review paper presents a detailed analysis and synthesis of current eye tracking research on multimedia learning. This study is guided by the following main and sub-research questions:

1. How are eye tracking measurements used in multimedia learning research?
   - What are the cognitive processes that eye tracking measurements are associated with in multimedia learning research?
   - What are the potential factors that can affect eye tracking measurements in multimedia learning research?
   - What is the relation between eye tracking measurements and learning performance in multimedia learning research?

3. Method

3.1. Search strategy

The Web of Science, Education Resources Information Center, Education Source, and PsycINFO databases were selected to access to the related research studies in this review, because they are general reference tools most commonly used to identify primary sources by researchers (Frankel, Wallen, & Hyun, 2012). In this literature search, "multimedia" and "learning" keywords were accompanied by the following keywords using the "OR" Boolean operator: "eye move*", "eye track*", "gaze move*", and "gaze track*". To examine the highest quality and most up-to-date research, this literature search was limited to the studies published as articles in peer-reviewed journals indexed by the aforementioned databases from 2010 to 2016. This literature search was also restricted to English articles that include full text.

3.2. Study selection process

The initial literature search returned 156 articles. First, 76 duplicate articles obtained from more than one database were removed. Second, the remaining 80 articles were screened to determine whether they met the inclusion criteria (Table 1). The original research studies using eye tracking technology in multimedia learning were selected for this review. They were also required to answer at least two sub-research questions of this literature review. Seventeen articles were excluded based on their titles and abstracts because of the violation of inclusion criteria (Table 1). Among the 17 removed articles, seven were not an original research article; eight were not related to the educational research area; one was not related to the use of eye tracking technology; and one was not based on multimedia learning. Third, the full texts of the remaining 63 articles were critically evaluated to ensure that they adequately addressed this researcher's questions. As a result of the evaluation, 11 articles were eliminated; seven of these did not use eye tracking technology to examine cognitive processes; two did not investigate the usability of a multimedia system; and two studies did not use multimedia learning environments. At the end of the study selection process, 52 articles were determined to meet the inclusion criteria for this review. As more than one experiment or study were reported in some articles, 58 studies were recorded for analysis in this review.

3.3. Analysis and coding of the studies

The articles selected for this review were analyzed with respect to demographics, methodologies, and the use of eye tracking measurements, respectively. Demographics for each article were recorded, including publication year, countries where the research was conducted, and the journal in which the research was published.

Concerning methodological characteristics of each article, this literature review analyzed the learners chosen for the research, subject domains of the learning materials used, and eye tracking measurements. Learners in the studies were classified as kindergarten students, elementary school students, middle school students, high school students, college students, older adults, or mixed.

Table 1
Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Article was original and empirical research article. Article was in the area of educational research. Multimedia learning environment was used in the research. Eye-tracking technology was used in the research. Article includes findings that answer at least two sub-research questions of this review.	Article was review or commentary article. The aim of the study was to create non-learning related models based on participants' eye movements. The study investigated usability of a system.

Learning topics of the materials were divided into eight subject domains: physics, biology, chemistry, earth sciences and geography, statistics, geometry, psychology, and other. To arrange eye movement measurements, this study used Lai et al.'s (2013) classification table, which includes temporal, count, and spatial categories.

Regarding the use of eye tracking measurements in multimedia learning research, cognitive activities studied were divided into three main processes in line with the active processing assumption of cognitive theory of multimedia learning (Mayer, 2014a): selecting, organizing, and integrating. To specify the factors affecting eye tracking measurements, all sections of the articles were analyzed. To code data and determine categories, cognitive theory of multimedia learning (Mayer, 2005), cognitive-affective theory of multimedia learning (Moreno, 2006), and basic and advanced principles of multimedia learning (Mayer, 2014b) were used. Several categories, including multimedia learning principles, individual differences, metacognition, emotion, and multimedia content, emerged as data were reviewed based on these theories and principles. Finally, the relation between eye tracking measurements and learning performance was identified by synthesizing common and conflicting findings among the reviewed studies. The types of learning performance were specified as a first step to achieve our aim. These types of learning performance included recall or retention, transfer, comprehension, factual knowledge, and problem solving. Then, both significant and non-significant correlations between eye movement measures and learning performance were examined. Based on the main findings, general conclusions related to visual search efficiency, amount of attention allocated to relevant pictures, and integrative transitions were drawn.

Inter-rater reliability in coding was examined calculating Kappa coefficients. First, after determining the frameworks to be used in coding, the first author coded all of the studies. Themes or categories that were derived during coding were also noted and defined by the first author. Second, a meeting was conducted between the first author and an advanced doctoral student who was experienced in eye tracking research and volunteered to code some portion of the literature reviewed by the authors as a second rater. The first author and second rater analyzed one study together by discussing the code list. Third, the second rater reviewed 16 studies (28%) out of 58 with respect to their methodological characteristics and the use of eye tracking measurements. Fourth, the Kappa values between the two raters were calculated. According to Landis and Koch's (1977) benchmarks, substantial agreement was achieved between raters in terms of the learners chosen for the research, $k = .78$; subject domain of the learning materials, $k = .75$; the cognitive process that eye tracking measures are associated with, $k = .74$; type of learning outcome, $k = .77$; and relationship between eye tracking measures and learning performance, $k = .66$. Substantial agreement was also obtained in terms of the potential factors affecting eye tracking measures, $k = .58$. Agreement on the scale of eye tracking measurements was almost perfect, $k = .94$. Finally, the two raters discussed disagreements on some codes, especially those pertaining to potential factors affecting eye tracking measurements, and resolved them. The second rater's misunderstanding about the coding list was determined to cause a lower Kappa value. After correction of the misunderstanding, the second rater agreed with the first author on her codes. For the remaining codes on which the two raters disagreed, the first author took into account the second rater's suggestions and revised some codes. In addition, she reassessed all of the reviewed studies considering these suggestions.

4. Findings and discussion

Eye tracking studies on multimedia learning are first defined with respect to demographics and methodological characteristics in this section. Then, these studies are synthesized to explain how eye movement measurements are used. In particular, the following issues are addressed: cognitive processes that eye tracking measurements are associated with, potential factors that can affect eye tracking measurements, and relations between eye tracking measurements and learning performance in multimedia learning research.

4.1. Overview of the studies

4.1.1. Demographics of the studies

Almost half of the reviewed 52 articles were published in the last two years (Fig. 1). van Gog and Jarodzka (2013) state that research has increased on the use of eye tracking technology to examine cognitive activities in computer-based learning environments. In particular, this study reveals that there is a burgeoning interest in eye tracking research on multimedia learning. Limitations of interviews, behavioral assessments, or self-reporting measurements for deriving valid interpretations regarding learning processes (Rodrigues & Rosa, 2017) and increasing accessibility to eye-tracker devices (Desjarlais, 2017) may have motivated researchers to use this technology in recent years.

Eye tracking studies on multimedia learning were conducted by researchers in Europe (65%), North America (18%), Asia (12%), Australia (3%), and Africa (2%) (Table 2). Although the studies were distributed by different countries, Germany was the leader of this particular area of research, with a total of 25 articles. Specific scholars in Germany who conducted more than one research study in this area in seven years affected the distribution of the studies by country. Considering the limited number of eye tracking studies on multimedia learning (Chuang & Liu, 2012; Johnson & Mayer, 2012), it is critical for researchers in other countries to contribute to this nascent research area. They can replicate existing studies with the learners in their countries and new materials to discuss generalizability of findings over different cultures and provide strong empirical findings to be used in future meta-analyses (Mayer, 2017).

The titles of the journals that include more than one eye tracking study on multimedia learning are listed in Table 3. There were 17 articles that were published in 17 different journals and were categorized as other journals in Table 3. The journals of Computers in Human Behavior, Learning and Instruction, Applied Cognitive Psychology, and Computers & Education were the most common ones.

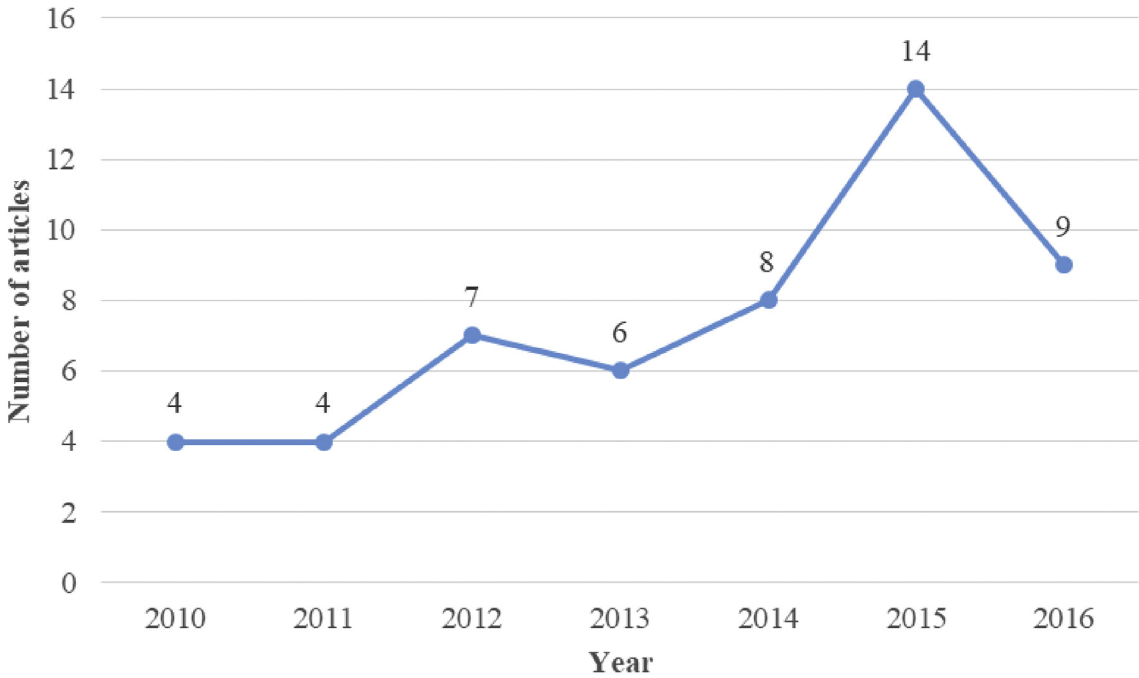

Fig. 1. The distribution of eye tracking research articles on multimedia learning by year (N = 52).

Table 2
The countries where eye tracking research articles on multimedia learning are conducted.

Continent	Country	Number of articles
Europe	Germany	25
	Italy	5
	The Netherlands	3
	Turkey	2
	United Kingdom	1
	Poland	1
	France	1
	Switzerland	1
North America	USA	10
	Canada	1
Asia	Taiwan	6
	Japan	1
Australia	Australia	2
Africa	South Africa	1

Note. Since some studies were conducted by the researchers in different countries, total number of the countries exceeds the number of articles.

Table 3
Journals in which eye tracking research articles on multimedia learning were published (N = 52).

Journal	Number of articles
Computers in Human Behavior	10
Learning and Instruction	6
Applied Cognitive Psychology	5
Computers & Education	4
Frontiers in Psychology	2
Frontline Learning Research	2
Journal of Educational Multimedia and Hypermedia	2
Reading Research Quarterly	2
Zeitschrift für Pädagogische Psychologie (German Journal of Educational Psychology)	2
Other journals that include one relevant article	17

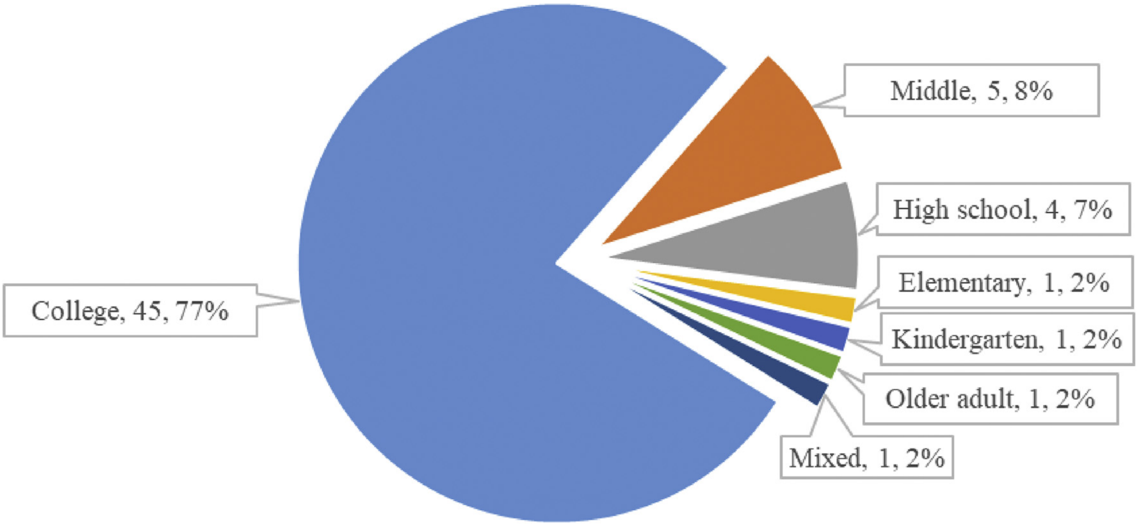

Fig. 2. Learner types chosen in the eye tracking studies on multimedia learning (N = 58).
Note. Data labels include category, number of included studies, and percentage of included studies.

4.1.2. Methodological characteristics of the studies

Of the 58 studies, 45 were conducted with college students, five with middle school students, four with high school students, one with elementary students, one with kindergarten students, one with older adults, and one with both college students and alumni (Fig. 2). College students were the main participant group on the use of eye tracking technology in multimedia learning research, which is similar to the findings of other research reviews about technology-enhanced learning (Drysdale, Graham, Spring, & Halverson, 2013; Hwang & Tsai, 2011; Wu et al., 2012). On the contrary, the amount of available research on the use of eye tracking technology in the education of older adult or younger learners is limited. Considering that multimedia learning has become quite prevalent for learners in various age groups due to increasing e-learning opportunities (Mayer, 2017), it is of great importance to ensure that all students benefit from multimedia materials equally. Therefore, researchers need to extend use of eye tracking technology to examine learning process of different types of learners. Further research can be conducted with kindergarten students, K-12 students, or older adults as the participant group to enhance the generalizability of the existing research results or provide new insights about learning of these groups.

Eye movements of the participants in the reviewed studies were collected while they were using different learning materials. As seen in Table 4, physics (n = 18) was the most common subject domain for the learning materials used in eye tracking studies, followed by biology (n = 11), earth sciences and geography (n = 6), and chemistry (n = 5). Correspondingly, physics topics, especially structure and functioning of systems, have been frequently taught in most multimedia learning studies (Eitel, Scheiter, & Schüler, 2013). Visual representations that make the phenomena more concrete and show multiple processes and relations are crucial to learning in the field of science (Cook, 2006). Therefore, science textbooks or websites include different types of visuals and text

Table 4
Subject domains of the learning materials used in the eye tracking studies on multimedia learning (N = 58).

Subject domain	Number of study	Sample topic
Physics	18	Ray optics (Lenzner, Schnotz, & Müller, 2013), the transformation of solar radiation into electricity (Skuballa, Fortunski, & Renkl, 2015), functioning of pulley system (Eitel et al., 2013), functioning of the toilet flushing system (Eitel, 2016), functioning of the suction cup of a sink plunger (Mason, Pluchino, & Tornatora, 2013a)
Biology	11	ATP syntheses (Korbach, Brünken, & Park, 2016), food chain (Mason, Tornatora, & Pluchino, 2015a), human circulatory system (Scheiter & Eitel, 2015), cell division (Stalbovs, Scheiter, & Gerjets, 2015), immunization (Park, Knörzer, Plass, & Brünken, 2015a)
Other	9	Cooking (Wang, Tsai, & Tsai, 2016), story (Takacs & Bus, 2016)
Earth Sciences & Geography	6	Truth about Dinosaurs (Yang et al., 2013), wind formation (Chuang & Liu, 2012), types of thunderstorms (Liu et al., 2011)
Chemistry	5	Atomic orbitals (Chen, Hsiao, & She, 2015), the ideal gas laws (O'Keefe, Letourneau, Homer, Schwartz, & Plass, 2014), positive charge, proton attack, electron density, and hydroxide attack (Williamson, Hegarty, Deslongchamps, Williamson, & Shultz, 2013)
Mixed	3	Geography and biology (Zhao, Schnotz, Wagner, & Gaschler, 2014)
Psychology	2	Cognitive theory of multimedia learning (Jamet, 2014)
Statistics	2	Correlation matrix (Rey, 2014)
Geometry	1	Angles (van Marlen, van Wermeskerken, Jarodzka, & van Gog, 2016)
History	1	African history (Romero-Hall, Watson, & Papelis, 2014)

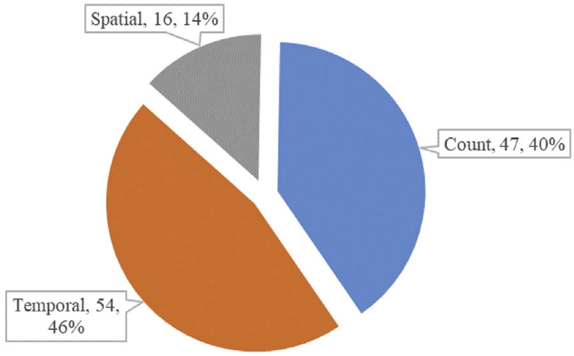

Fig. 3. Eye tracking measures collected in multimedia learning studies (N = 58).
Note. Data labels include category, number of included studies, and percentage of included studies.

that require processing for successful comprehension (Mason, Tornatora, & Pluchino, 2013b). The critical role of verbal and pictorial information in science learning may have motivated researchers to examine the visual attention of the learners studying science materials.

As a main or additional data source, the temporal, spatial, and count scales of eye movement measurements were collected in the studies. Of the 58 research studies, 54 gathered temporal scales, 47 gathered count scales, and 16 gathered spatial scales of eye tracking measurements (Fig. 3). In addition to these measurements, in one study, participants' pupil size was recorded through eye tracking technology. Parallel to the findings of Lai et al.'s (2013) study, spatial measurements were the least common measure in the reviewed studies. Spatial scale comprises fixation positions, fixation sequence, and scan path patterns (Lai et al., 2013). It requires mostly a qualitative analysis of the scan paths to obtain these measurements. Although the scan path analyses reveal how learning occurs from moment to moment (Hyönä, 2010), few studies investigate them in detail (Krejtz, Duchowski, Krejtz, Kopacz, & Chrząstowski-Wachtel, 2016). This could be due to the difficulty in qualitative analysis and synthesis of the scan paths obtained from different participants.

4.2. Use of eye tracking measurements in multimedia learning research

4.2.1. Cognitive processes in multimedia learning research

According to the cognitive theory of multimedia learning, selecting, organizing, and integrating are the three cognitive processes that build coherent mental representations (Mayer, 2014a). Eye movements in multimedia learning studies can be analyzed from this theoretical perspective (Scheiter & Eitel, 2017). Table 5 indicates the three cognitive processes in multimedia learning investigated with different eye movement measurements. As eye tracking technology is used to examine visual attention, processing of on-screen words and pictures defined as the areas of interest (AOIs) was examined in this review.

4.2.1.1. Selecting. Selecting occurs when learners attend to relevant elements in multimedia material so as to carry them into working memory (Mayer, 2014a). The selection process can be detected on the basis of learners' first fixations on multimedia elements (Park et al., 2015a) to infer visual search and the salience and relevance of multimedia elements (Scheiter & Eitel, 2017).

Table 5
Cognitive processes eye tracking measures are associated with in multimedia learning studies.

Cognitive process	Number of study	Sample eye movement measures
Selecting	24	Time to first fixation on AOI The number of times a cued item was fixated within 2 s Distance between position of cued item and current gaze position The place of first five fixations First-pass time on AOI Proportion of fixations on diagram or text
Organizing	47	Total fixation count on AOI Total fixation duration on AOI Total reading time on AOI Average fixation duration on AOI Fixation positions on AOI Pupil size on AOI
Integrating	33	Number of transitions between (corresponding) text and picture Look from (corresponding) text to picture fixation time Look from (corresponding) picture to text fixation time Sum of the saccade paths between text and picture Scan paths

In some of the reviewed studies, researchers interpreted visual search in the selection process by examining how much time had passed before the learners' first fixation on a relevant text or image (e.g., Park, Korbach, & Brünken, 2015b; Scheiter & Eitel, 2017; van Marlen et al., 2016), how far the relevant item was from the learners' current gaze position (Glaser & Schwan, 2015, Experiment 2), and whether a cued item was fixated within 2 s (e.g., Jamet, 2014). In addition, selective attentional distribution between text and images during initial processing was analyzed by using the measurements of the place of the first five fixations (Schmidt-Weigand, 2011, Experiment 1 & 2; Schmidt-Weigand, Kohnert, & Glowalla, 2010a), first pass time on the text or images (e.g., Mason et al., 2013a; Mason et al., 2013b), and the proportion of fixations on diagrams or text (e.g., Johnson & Mayer, 2012, Experiment 1, 2, & 3).

4.2.1.2. Organizing. Organizing occurs when learners make connections among words or images in order to create coherent verbal or pictorial models in working memory (Mayer, 2014a). "Organization processes are assumed to occur during fixating the processed information" (Park et al., 2015a, p. 32). The total fixation duration and the number of fixations can indicate the amount of attention allocated to a textual or pictorial AIO (Scheiter & Eitel, 2017). Table 5 lists different fixation measurements used to make inferences about the organization process in the reviewed studies.

According to Rayner (1998), longer fixation duration can indicate deeper processing. Correspondingly, there are some studies in which fixation duration was used to make inferences about the depth of processing (e.g., Glaser & Schwan, 2015; Park et al., 2015b). The number of fixations on AOI is another eye movement measurement that allows researchers to infer learners' intensity of processing (e.g., Park et al., 2015b; Scheiter & Eitel, 2015). Besides fixation measurements, successive fixation positions on AOI were examined with scan paths to determine learners' sequence of attention during organizing words or images (e.g., Krejtz et al., 2016; Liu et al., 2011; Skuballa et al., 2015).

It is important to note that fixation duration can be used to identify cognitive load as well as the depth of processing (Kruger & Doherty, 2016). In particular, longer mean fixation duration on a stimulus can indicate a greater processing difficulty (Jacob & Karn, 2003; Krejtz et al., 2016; Liu & Chuang, 2011). Some studies examined in this review (e.g., Chen et al., 2015; Krejtz et al., 2016; Schüler, Scheiter, & Gerjets, 2012; Yang et al., 2013) drew conclusions about difficulties in processing by using mean fixation duration. Therefore, interpretations of fixation duration in relation to cognitive processing are changeable (Kickmeier-Rust, Hillemann, & Albert, 2011). Another eye-based measurement of cognitive load is increase in pupil size (van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009). For example, among the reviewed studies, only Chuang and Liu (2012) used pupil size to infer cognitive load in segmented multimedia and assessment pages. However, it is necessary to acknowledge that pupil size is too sensitive for the effects of other factors such as changes in light and brightness; therefore, researchers should be careful about interpreting this measure (van Gog & Jarodzka, 2013).

4.2.1.3. Integrating. Integration occurs when learners establish connections between pictorial and verbal models and relevant prior knowledge (Mayer, 2014a). The integration process can be examined by looking at learners' transitions between text and pictures (Arndt, Schüler, & Scheiter, 2015; Hegarty & Just, 1993; Scheiter & Eitel, 2017).

Correspondingly, integration was investigated with metrics of switching between textual and pictorials elements in the reviewed studies (Table 5). Successful integration processes were analyzed by dividing text and pictures into several AOIs for ensuring that integration occurs between corresponding pieces of multimedia information (e.g., Johnson & Mayer, 2012, Experiment 1, 2, & 3; Mason, Pluchino, & Tornatora, 2015a; Mason, Tornatora, & Pluchino, 2015b). Moreover, transitions between two text segments (e.g., Mason et al., 2013b) or two visuals (e.g., O'Keefe et al., 2014; Rau, Michaelis, & Fay, 2015, Experiment 2 & 3) were also considered as integrating processes in some studies. Lastly, in addition to count and temporal eye movement measurements, scan paths were qualitatively analyzed to identify learners' integration processes and reading patterns for textual and pictorial information (e.g., Krejtz et al., 2016; Liu & Chuang, 2011; Liu et al., 2011).

Furthermore, it is crucial to report that integrative transitions were used to imply difficulty in coordinating multimedia elements. For example, Bauhoff, Huff, and Schwan (2012) interpreted a low number of transitions between two widely separated visuals as a higher working memory load. In addition, Wang et al. (2016) explained the high occurrence of switching between video, including audiovisual information and text, as information overload.

4.2.2. Potential factors that can affect eye tracking measurements in multimedia learning

The potential factors that can affect eye tracking measurements in multimedia learning were categorized into multimedia learning principles, multimedia content, individual differences, metacognition, and emotions (Fig. 4). While the most investigated factor was multimedia learning principles, the least examined factor was emotions. In some articles, effects of more than one factor were investigated.

4.2.2.1. Multimedia learning principles. Eye movement measurements can validate and support the conclusions of prior multimedia learning studies (Liu et al., 2011). Correspondingly, the majority of the reviewed studies examined the effect of multimedia learning principles on cognitive activities using eye tracking technology (Fig. 4). Among 11 principles, the signaling or cueing principle was the most investigated factor in multimedia learning studies ($n = 9$), followed by modality ($n = 7$), spatial contiguity ($n = 7$), and coherence or seductive details ($n = 5$).

The signaling principle appeared to positively guide learners' attention. The signaling or cueing principle suggests that better learning occurs in the learning materials when cues are available to guide learners' attention to the relevant information in the material or highlight the organization of the material (van Gog, 2014). The common outcome derived from signaling studies in this review is that signaled multimedia materials direct learners' visual attention, which corroborates the signaling principle (e.g., Jamet,

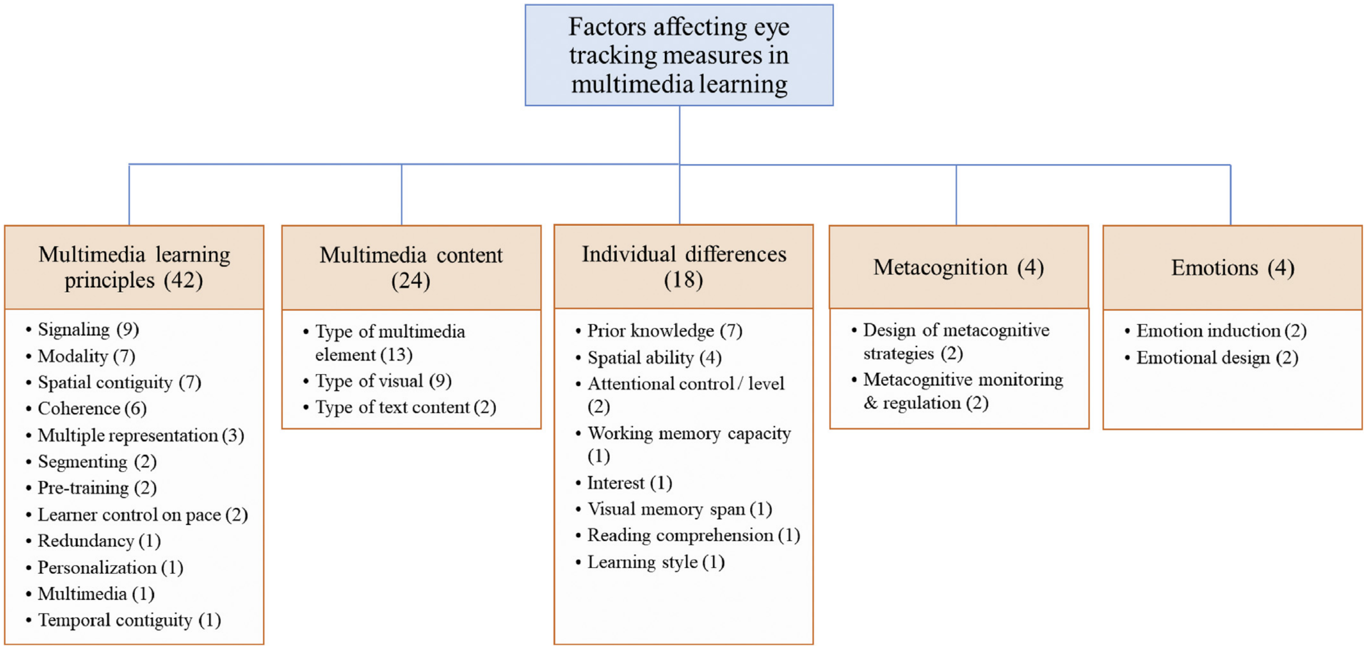

Fig. 4. Potential factors that can affect eye tracking measures in multimedia learning (number of the studies included to that theme or subtheme).
Note. Since effects of more than one factor were examined in some studies, total frequency exceeds the number of studies ($N = 58$).

2014; Glaser & Schwan, 2015, Experiment 2; Mason et al., 2013a; Mason et al., 2015a). For example, Jamet (2014) colored boxes in a diagram, and Mason et al. (2013a) used labels for pictures accompanying text. They found that learners allocated more attention to signaled pictorial elements. Eye movement modeling examples (EMMEs) displaying the eye movements of a person studying multimedia materials were a novel method for signaling in the reviewed studies. Mason et al. (2015a) used this method in their study and found that experimental group viewing EMMEs displayed more integrative transitions between text and picture than the control group.

Application of the modality principle generally increased attention to visuals in the reviewed studies. The modality principle suggests that better learning occurs in learning materials when words accompanying visuals are spoken rather than written (Mayer, 2001). The seven studies in this review (Antonietti, Colombo, & Di Nuzzo, 2015; Schmidt-Weigand, 2011, Experiment 1 & 2; Schmidt-Weigand et al., 2010a, Experiment 1 & 2; Schmidt-Weigand, Kohnert, & Glowalla, 2010b; Schüler, Scheiter, & Gerjets, 2011) compared learners' visual attention and performance in multimedia learning environments including spoken or written text. The most prevalent result was that the learners in the spoken-text group spent more time for viewing visualizations than the learners in the written text-group.

Seductive details and spatially separated multimedia elements were generally found to negatively influence attention toward pertinent multimedia elements and their integration. First, seductive details or the coherence principle suggests that better learning occurs in learning materials when irrelevant text and pictures are excluded (Mayer, 2001). The findings of the reviewed studies support the distracting effect of seductive details. Although irrelevant or mismatched pictures grabbed initial and then ignorable attention in some studies (e.g., Lenzner et al., 2013, Experiment 1; Rey, 2014; Rop, van Wermeskerken, de Nooijer, Verkoeijen, & van Gog, 2016), seductive texts or pictures were found to decrease attentional focus on relevant content and the integration process (e.g., Korbach et al., 2016; Park et al., 2015b). Second, the spatial contiguity principle suggests that better learning occurs in learning materials when words and pictures are placed near each other (Mayer, 2001). Findings supporting this principle could be observed in the reviewed studies. In particular, more attentional focus (Bayram & Mutlu Bayraktar, 2012) and integrative transitions between text and pictures (Johnson & Mayer, 2012, Experiment 1, 2, & 3), text and tables (Clinton, Morsanyi, Alibali, & Nathan, 2016), and two visuals (Bauhoff et al., 2012) were determined in materials in which multimedia elements were close to each other.

4.2.2.2. Multimedia content. Another factor affecting eye tracking measurements was multimedia content that included categories of type of multimedia element, type of visual, and type of text content.

With regard to type of multimedia element, attentional focus on text and visuals was compared in nine articles (Chuang & Liu, 2012; Johnson & Mayer, 2012; Liu & Chuang, 2011; Morrow et al., 2012; Ozcelik, Arslan-Ari, & Cagiltay, 2010; Schmidt-Weigand, 2011; Schmidt-Weigand et al., 2010b, 2010a; Schüler et al., 2011). The common finding of the studies was that learners give more attention to text than to pictures. This implies that people prefer a more text-dominant learning process in multimedia environments.

Concerning type of text content, the effects of texts with visual or spatial content were compared in two studies (Schüler et al., 2012; Schüler et al., 2011). The authors argued that processing text with spatial content and coordinating eye movements during learning can interfere with learning based on the limited capacity of the visuospatial sketchpad. However, the authors found no significant effect of type of text content on fixation time or mean fixation duration; therefore, they concluded that type of text content does not result in different levels of difficulty during processing.

Regarding the type of visual content, the effects of two- and three-dimensional visuals (Liu & Chuang, 2011), static and dynamic visuals (Chen et al., 2015; Krejtz et al., 2016; Schmidt-Weigand, 2011; Takacs & Bus, 2016; Wang et al., 2016), photos and conceptual graphics (Yang et al., 2013), and content and context visuals (Suvorov, 2015) were compared in the studies. The prevailing finding was that learners allocated more visual attention while processing dynamic visuals than while processing static visuals. For example, comparing the effect of three-dimensional static and dynamic visuals of atomic orbitals on eye movements, Chen et al. (2015) found that learners made a higher number of fixations and spent more inspection time during their learning process inspecting three-dimensional dynamic visuals. Similarly, Takacs and Bus (2016), who examined the eye movements of kindergarten children viewing a storybook with motion pictures and static illustrations, revealed that motion pictures received more visual attention than static illustrations.

4.2.2.3. Individual differences. Although individual differences are one of the multimedia learning principles, they were discussed under a new theme because there were several types of individual differences that needed to be examined comprehensively. Many of the reviewed articles controlled the effect of individual differences on learning performance, but 13 of them examined their effects on processing information as one of their research aims. Furthermore, almost all of them analyzed the effect of individual differences with other factors as listed in Fig. 4. The individual differences investigated were prior knowledge (Bauhoff et al., 2012; Korbach et al., 2016; Mason et al., 2013b; Morrow et al., 2012; Ozcelik et al., 2010; Park et al., 2015b; Yang et al., 2013), spatial ability (Chen et al., 2015; Korbach et al., 2016; Mason et al., 2013b; Park et al., 2015b), attentional control/level (Rey, 2014; Bayram & Mutlu Bayraktar, 2012), working memory capacity (Krejtz et al., 2016), interest (Wang et al., 2016), visual memory span (Bauhoff et al., 2012), reading comprehension (Mason et al., 2013b), and learning style (Cao & Nishihara, 2012).

In the reviewed studies, prior knowledge was mostly found to be the factor differentiating cognitive processing of multimedia information. For example, more knowledgeable older adults in Morrow et al.'s (2012) study spent more time to read text and view visuals in their first pass, but they allocated more visual attention to inspect relevant pictures in their second pass than less knowledgeable ones. In the study by Yang et al. (2013), students who have existing domain knowledge spent more attention on texts, keywords in the text, and some particular areas in graphics, and they performed more integrative processes for the text and pictures.

In contrast to prior knowledge, the findings were mixed with regard to the relation between spatial ability and information selection and processing. Spatial ability was a significant factor that changes visual attention distribution in Che et al.'s (2015) and Park et al.'s (2015b) studies. To illustrate, Chen et al. (2015) determined that learners with low spatial ability allocated more inspection time to the dynamic three-dimensional representations than learners with high spatial ability. On the other hand, spatial ability did not have effect in Korbach et al.'s (2016) and Mason et al.'s (2013b) studies. Lastly, to discuss the effect of other aforementioned individual differences (i.e., interest or visual memory span) on eye movements with different studies, there is a need for further empirical evidences.

4.2.2.4. Metacognition. Another factor influencing eye movements in multimedia learning is metacognition, including the subthemes of metacognitive monitoring and regulation and design of metacognitive strategies. In interactive multimedia environments, metacognitive support helps students to think, process information, and monitor their learning (Gordon, 1996). In other words, "metacognitive factors mediate learning by regulating cognitive processing and affect" (Moreno, 2006, p. 151). However, a limited number of studies has investigate metacognitive processes in multimedia learning (van Gog & Jarodzka, 2013). Similarly, among the reviewed studies, metacognition has been discussed less often in conjunction with eye tracking technology.

Four studies have indicated the effects of metacognition on visual attention (Antonietti et al., 2015; Eitel, 2016; Ruf & Ploetzner, 2014; Stalbovs et al., 2015). Regarding design of metacognitive strategies, Ruf and Ploetzner (2014) used multimedia learning environments that differ in the presence of self-monitoring questions and presentation of cognitive learning aids. They found that learning aids with dynamic presentation modes led to more transitions between content areas and support areas. In another study, Stalbovs et al. (2015) aimed to support learners' information processing during multimedia instruction with different types of implementation intentions or if-then plans. They revealed that text-picture integration and mixed implementation intentions resulted in more transitions between pictures and the corresponding text in multimedia learning. Metacognitive monitoring and regulation resulted from learners' repeated studying and testing (Eitel, 2016) and requests for viewing pictures (Antonietti et al., 2015) were also a factor influencing visual attention distribution.

4.2.2.5. Emotions. According to the cognitive-affective theory of multimedia learning, emotions are one of the factors that "mediate learning by increasing or decreasing cognitive engagement" (Moreno, 2006, p. 151. Regarding the specific role of positive emotions, there are two conflicting hypotheses: "emotions as extraneous cognitive load" and "emotions as facilitator of learning" (Um, Plass, Hayward, & Homer, 2012). Among the studies reviewed, there were only three studies that investigated the impact of emotion induction or emotional design on eye movement measurements (Knörzer, Brünken, & Park, 2016; Park et al., 2015a; Romero-Hall et al., 2014).

There were contradictory findings regarding the facilitator or suppressor effect of emotions on cognitive processes in multimedia learning. In the reviewed studies, emotions were triggered either by emotion-induction methods applied before multimedia learning, such as listening to certain music, recalling a happy/sad past event (Knörzer et al., 2016), and reading self-referencing statements (Park et al., 2015a), or by emotional design features embedded in the multimedia learning environment, such as anthropomorphisms (Park et al., 2015a) and pedagogical agents with emotional expressions (Romero-Hall et al., 2014). While Park et al. (2015a) and Romero-Hall et al. (2014) determined that participants who used learning materials including positive emotional design features allocated more visual attention to the pictures, Knörzer et al. (2016) found that participants with positive emotion induction allocated significantly less time and fewer fixations on the relevant information than participants with negative emotion induction. Therefore, more research studies on this topic are needed to contribute to the resolution of this debate.

4.2.3. Relation between eye tracking measurements and learning performance in multimedia learning research

There is a lack of evidence for the relation between eye movement measurements and students' learning outcomes (Chen et al., 2015; Lai et al., 213; Scheiter & Eitel, 2015). Of the 58 studies reviewed in this study, only about half of them (n = 29) investigated the direct relation between eye tracking measurements and learning performance. Specific instruments used to assess learning were tests for recall or retention (n = 18), transfer (n = 11), comprehension (n = 8), factual knowledge (n = 4), and problem solving (n = 2). Common findings obtained from these studies are explained in the following three themes.

4.2.3.1. Visual search efficiency was positively associated with learning performance. Visual search efficiency describes how fast the observer locates relevant visual information when corresponding words are heard or read. To prevent split attention and enhance visual search in multimedia materials, visual references can be signaled in on-screen texts or while corresponding words are being narrated (Tabbers, Martens, & van Merriënboer, 2004). Otherwise, learning is hindered due to extraneous cognitive load experienced during visual search (Tabbers, Martens, & Merriënboer, 2004). Correspondingly, eye tracking research (e.g., Ozcelik et al., 2010; Scheiter & Eitel, 2015) indicates that spending less time detecting relevant visual information in signaled multimedia materials is associated with improved learning performance.

4.2.3.2. Amount of attention allocated to relevant pictures was mostly positively related to learning performance. According to multimedia principle, "people can learn more deeply from words and pictures than from words alone" (Mayer, 2005, p. 1). To corroborate multimedia learning principles, it is necessary to examine both cognitive processes and their relation with learning performance (Mason et al., 2013b). In the reviewed studies' multimedia learning environments, visual attention to relevant pictures indicated with fixations or saccades were mostly positively associated with some learning performance scores (e.g., Eitel, 2016; Eitel et al., 2013; Korbach et al., 2016; Scheiter & Eitel, 2015; Skuballa et al., 2015). This finding highlights the importance of processing pictures, even though people prefer to rely on text in multimedia learning. In addition, this evidence supports the multimedia principle to indicate positive effect of use of both verbal and visual channels on learning. However, it is also important to note that higher visual attention to pictures does not always bring about higher learning performance. To illustrate, mean fixation duration on a picture may indicate cognitive load as well as depth of processing (Kruger & Doherty, 2016). Consequently, as mean fixation duration on a picture increases, learning performance might decrease owing to cognitive load (Ozcelik et al., 2010). Nonetheless, more research studies investigating the relation between mean fixation duration and learning performance are required in order to draw stronger inferences about cognitive load with eye movement measurements.

4.2.3.3. Integrative transitions usually correlate with learning performance. Successful integration of verbal and pictorial models is the most critical step in multimedia learning for sense-making (Mayer, 2014a), text comprehension, and learning from illustrated text (Mason et al., 2015a). Integration of two or more pictures complementing each other, facilitating comprehension of one of the representations, or helping to construct deeper understanding in a domain, is also important for better learning (Ainsworth, 2014). Correspondingly, research studies that examine the relation between eye movement measurements of integration and learning performance generally report significant correlations. Although there are some conflicting findings (e.g., Arndt et al., 2015; Clinton et al., 2016; Schmidt-Weigand et al., 2010b), transitions between relevant text and pictures (e.g., Mason et al., 2013a; Mason et al., 2013b; Mason et al., 2015a; Mason et al., 2015b; Stalbovs et al., 2015, Experiment 1) and two visuals (O'Keefe et al., 2014) were usually positively associated with some learning performance scores. It is also necessary to acknowledge that higher switches between two multimedia elements due to the difficulty of coordinating incoming information may result in worse performance. For example, transitions between text and video including audiovisual information was a negative predictor of recall due to information overload in Wang et al.'s (2016) study.

On the whole, eye movement measurements can be used for three purposes in multimedia learning research. They are investigating (a) cognitive processes of selecting, organizing, and integrating, (b) factors affecting these processes, and (c) correlation between the cognitive processes and learning performance. Fig. 5 depicts the relations among usage purposes of eye tracking measurements in multimedia learning research.

5. Conclusion

This study systematically reviews eye tracking research on multimedia learning to synthesize current literature and reveals how eye tracking technology can be used. The study indicates a burgeoning interest in the use of eye tracking technology in multimedia learning research. Research designs that include using college students as participants, science materials as learning materials, and temporal and count scale of eye movement measurements as data sources were the common methodological characteristics of these studies. Eye tracking measurements were used to make inferences about the cognitive processes of selecting, organizing, and integrating in the reviewed studies. Multimedia learning principles, multimedia content, individual differences, metacognition, and emotions were the potential factors that can influence eye movement measurements. Among these factors, metacognition and emotion received less attention from the researchers. Lastly, there were findings supporting the relation between cognitive processes inferred with eye-tracking measures and learning performance.

The rising need for quality multimedia materials in different learning environments highlights importance of eye tracking research that can assist in learning with multimedia materials. This review shows how researchers have benefited from eye tracking technology to study processing of multimedia information with relevant variables. Particular research gaps were also identified. Therefore, this study can guide future researchers and practitioners aiming to enhance multimedia learning. The study supports use of eye tracking technology in educational research areas to investigate learning processes in detail rather than merely looking at learning outcomes.

6. Suggestions for future research

Several gaps in eye tracking research on multimedia learning have been determined in this review. Regarding methodologies of the reviewed studies, the majority were conducted with college students and science learning materials. It is important to replicate existing studies with different types of learners, learning objectives, and content areas to identify boundary conditions of multimedia principles (Mayer, 2017). Hence, more research studies should be conducted with kindergarten students, K-12 students, and older adults using learning materials on the subjects different from science. In addition, eye movement measurements in spatial scale were the least frequently used measurements in eye tracking research on multimedia learning. Fixation position, fixation sequence, and scan path patterns in this scale can show spatial sequences of visual attention over time in detail. However, there are a limited number of studies that examine multimedia information processing with such measurements obtained from scan paths (Krejtz et al., 2016). Qualitative analysis of these measurements can be valuable in supplementing or supporting quantitative eye movement measurements in future research. Moreover, transition matrices can be used to summarize directions of saccades in a scene and to obtain entropy measurements that can quantify the randomness of scan path distributions over defined AOIs (Acartürk & Habel, 2012).

Concerning the potential factors affecting eye movement measurements, 11 multimedia principles were examined in the reviewed studies. There are other basic and advanced multimedia principles which need validation with eye tracking technology. Some of them are image, generative drawing, and imagination effect. Also, new techniques such as observing instructor drawing (Fiorella & Mayer, 2016) and watching first-person video modeling examples (Fiorella, van Gog, Hoogerheide, & Mayer, 2017) have recently been determined to enhance learning in multimedia environments. Authors of these studies suggest the use of eye tracking technology to

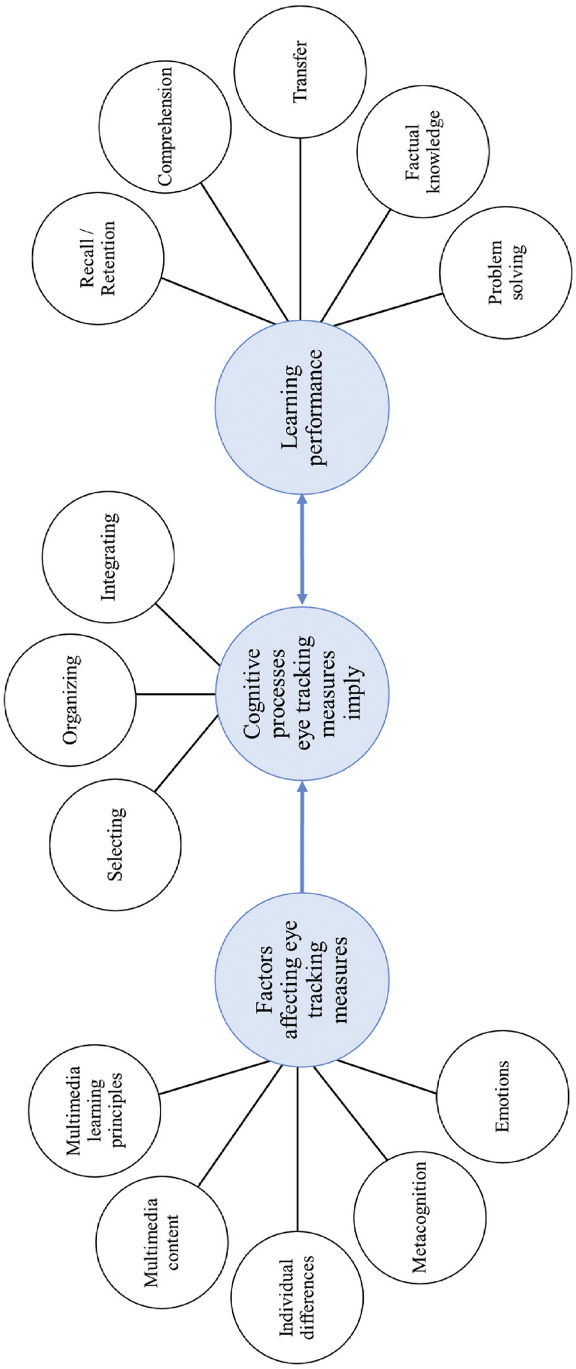

Fig. 5. Summary of how to use eye tracking measures in multimedia learning research.

explain effects of the techniques on cognitive processes. Another research gap is related to the role of individual differences in eye movements in multimedia learning due to scarce literature and conflicting findings about the impact of some differences (e.g., spatial ability, working memory capacity, and interest). Transition matrix entropy values can be used to detect individual differences in eye movement transitions between AOIs (Krejtz, Szmidt, Duchowski, & Krejtz, 2014). This review also indicates that the least examined factors influencing eye movements in multimedia learning were metacognition and emotions. Mayer (2017) also states that metacognition and motivation are two understudied areas of multimedia learning. Therefore, further investigation about these factors is needed to elucidate how metacognition and emotions affect cognitive processes in multimedia learning.

Finally, this research determined that the relation between eye tracking measurements and learning performance was investigated in only half of the reviewed studies. Therefore, the association between visual attention and learning outcomes has been less investigated. To provide more evidence about eye-mind hypothesis (Just & Carpenter, 1980) in multimedia learning research, future studies should analyze the relation between the cognitive processes of eye movement measurements and learning performance.

7. Suggestions for future practices

In this study, findings of eye-tracking research on multimedia learning were synthesized. There were common findings which can guide the design of future multimedia learning environments. First, the signaling principle was deemed positive in guiding learners' attention. In particular, some studies revealed that signaling enhanced visual search efficiency to find relevant information and improved learning performance. Hence, it is crucial to include cues such as color coding, labeling, and verbal references in the multimedia learning environments.

Second, irrelevant and spatially separated multimedia elements were generally found to have a negative effect on attentional focus on relevant information and integrative transitions. Therefore, the multimedia principles of coherence and spatial contiguity were validated with eye movement measurements. Thanks to different evidences corroborating these principles, it can be more strongly suggested that instructional designers need to pay attention to exclude seductive details and put relevant items close to each other.

Third, reviewed studies indicated that people prefer a more text-dominant learning process in the multimedia instructions consisting of on-screen text and pictures. In addition, higher visual attention to relevant pictures mostly resulted in higher learning performance in the studies. Therefore, it is critical to direct learners' attention to pictures in multimedia learning. As a design solution, verbal cues which can guide learners' attention to the pictures can be included with texts. In this way, learners can also make more integrative and corresponding transitions between text and pictures.

Fourth, reviewed studies mostly report positive correlations between integration of words and pictures and learning performance; therefore, it is important for learners to successfully integrate multimedia elements. However, coordinating visual and verbal working memory during integration requires significant effort and needs efficient use of cognitive capacity (Mayer, 2014a). Hence, instructional designers should facilitate learners' attempts at integrating multimedia elements. Some design solutions offered by reviewed studies are eye movement modeling examples illustrating integrative reading, cues in text as mentioned in the third suggestion, implementation intentions, and spatial contiguity between multimedia elements.

Finally, some eye movement measurements were associated with the cognitive processes of selecting, organizing, and integrating based on the cognitive theory of multimedia learning in this study. In addition, reviewed research supports that prior knowledge is a differentiating factor in processing multimedia information. By tracking and analyzing learners' eye movements in real time, adaptive systems can make inferences about cognitive state of the learners who have different levels of prior knowledge (van Gog & Jarodzka, 2013). Therefore, cognitive processes and related eye movement measurements identified in this review can contribute to the future design of adaptive multimedia systems. However, it is important to note that changing interpretations of some eye movement measurements and environmental conditions affecting eye movements can obstruct the development of accurate algorithms for adaptive learning environments.

References

The article included to the review.

Acartürk, C., & Habel, C. (2012). Eye tracking in multimodal comprehension of graphs. *Proceedings of the workshop on technology-enhanced diagrams research: Vol. 887*, (pp. 11–25).

Ainsworth, S. (2014). The multiple representation principle in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 464–486). New York, NY: Cambridge University Press.

*Antonietti, A., Colombo, B., & Di Nuzzo, C. (2015). Metacognition in self-regulated multimedia learning: Integrating behavioural, psychophysiological and introspective measures. *Learning, Media and Technology, 40*(2), 187–209. https://doi.org/10.1080/17439884.2014.933112.

*Arndt, J., Schüler, A., & Scheiter, K. (2015). Text–Picture integration: How delayed testing moderates recognition of pictorial information in multimedia learning. *Applied Cognitive Psychology, 29*(5), 702–712. https://doi.org/10.1002/acp.3154.

Baddeley, A. D. (1999). *Human memory.* Boston: Allyn & Bacon.

*Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance matters: Spatial contiguity effects as trade-off between gaze switches and memory load. *Applied Cognitive Psychology, 26*(6), 863–871. https://doi.org/10.1002/acp.2887.

*Bayram, S., & Mutlu Bayraktar, D. (2012). Using eye tracking to study on attention and recall in multimedia learning environments: The effects of design in learning. *World Journal on Educational Technology, 4*(2), 81–98.

*Cao, J., & Nishihara, A. (2012). Understand learning style by eye tracking in slide video learning. *Journal of Educational Multimedia and Hypermedia, 21*(4), 335–358.

*Chen, S. C., Hsiao, M. S., & She, H. C. (2015). The effects of static versus dynamic 3D representations on 10th grade students' atomic orbital mental model construction: Evidence from eye movement behaviors. *Computers in Human Behavior, 53*, 169–180. https://doi.org/10.1016/j.chb.2015.07.003.

*Chuang, H. H., & Liu, H. C. (2012). Effects of different multimedia presentations on viewers' information-processing activities measured by eye-tracking technology. *Journal of Science Education and Technology, 21*(2), 276–286. https://doi.org/10.1007/s10956-011-9316-1.

*Clinton, V., Morsanyi, K., Alibali, M. W., & Nathan, M. J. (2016). Learning about probability from text and tables: Do color coding and labeling through an interactive-user interface help? *Applied Cognitive Psychology, 30*(3), 440–453. https://doi.org/10.1002/acp.3223.
Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education, 90*(6), 1073–1091.
Desjarlais, M. (2017). The use of eye gaze to understand multimedia learning. In C. Was, F. Sansost, & B. Morris (Eds.). *Eye-tracking technology applications in educational research* (pp. 122–142). Hershey PA, USA: Information Science Reference.
Drysdale, J. S., Graham, C. R., Spring, K. J., & Halverson, L. R. (2013). An analysis of research trends in dissertations and theses studying blended learning. *The Internet and Higher Education, 17*, 90–100. https://doi.org/10.1016/j.iheduc.2012.11.003.
Duchowski, A. T. (2003). *Eye tracking methodology: Theory and practice* (2nd ed.). London: SpringerVerlag.
*Eitel, A. (2016). How repeated studying and testing affects multimedia learning: Evidence for adaptation to task demands. *Learning and Instruction, 41*, 70–84. https://doi.org/10.1016/j.learninstruc.2015.10.003.
*Eitel, A., Scheiter, K., & Schüler, A. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology, 27*(4), 451–461. https://doi.org/10.1002/acp.2922.
Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology, 108*(4), 528–546. https://doi.org/10.1037/edu0000065.
Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology, 109*(5), 653–665. https://doi.org/10.1037/edu0000161.
Frankel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York: The McGraw-Hill Companies, Inc.
*Glaser, M., & Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *Journal of Educational Psychology, 107*(4), 1006–1018. https://doi.org/10.1037/edu0000044.
van Gog, T. (2014). Signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 263–278). New York, NY: Cambridge University Press.
van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo, & V. Aleven (Eds.). *International handbook of metacognition and learning technologies* (pp. 143–156). New York: Springer.
van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior, 25*(2), 325–331. https://doi.org/10.1016/j.chb.2008.12.021.
Gordon, J. (1996). Tracks for learning: Metacognition and learning technologies. *Australian Journal of Educational Technology, 12*(1), 46–55.
Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717–742. https://doi.org/10.1006/jmla.1993.1036.
Hwang, G. J., & Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 42*(4), E65–E70. https://doi.org/10.1111/j.1467-8535.2011.01183.x.
Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*(2), 172–176. https://doi.org/10.1016/j.learninstruc.2009.02.013.
Jacob, R. J. K., & Karn, K. S. (2003). Eye-tracking in human-computer interaction and usability research: Ready to deliver the promises [Section commentary]. In J. R. Hyona, & H. D. Radach (Eds.). *The mind's eyes: Cognitive and applied aspects of eye movements*. Oxford, UK: Elsevier Science.
*Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior, 32*, 47–53. https://doi.org/10.1016/j.chb.2013.11.013.
*Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*(2), 178–191. https://doi.org/10.1037/a0026923.
Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480.
Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.
Kickmeier-Rust, M. D., Hillemann, E., & Albert, D. (2011, July). Tracking the UFO's paths: Using eye-tracking for the evaluation of serious games. *International conference on virtual and mixed reality* (pp. 315–324). Springer Berlin Heidelberg.
*Knörzer, L., Brünken, R., & Park, B. (2016). Facilitators or suppressors: Effects of experimentally induced emotions on multimedia learning. *Learning and Instruction, 44*, 97–107. https://doi.org/10.1016/j.learninstruc.2016.04.002.
*Korbach, A., Brünken, R., & Park, B. (2016). Learner characteristics and information processing in multimedia learning: A moderated mediation of the seductive details effect. *Learning and Individual Differences, 51*, 59–68. https://doi.org/10.1016/j.lindif.2016.08.030.
*Krejtz, K., Duchowski, A. T., Krejtz, I., Kopacz, A., & Chrząstowski-Wachtel, P. (2016). Gaze transitions when learning with multimedia. *Journal of Eye Movement Research, 9*(1), 1–17. https://doi.org/10.16910/jemr.9.1.5.
Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. *Proceedings of the 2014 symposium on eye-tracking research and applications*. New York, NY: ACM.
Kruger, J. L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology, 32*(6), 19–31. https://doi.org/10.14742/ajet.3084.
Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., et al. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90–115. https://doi.org/10.1016/j.edurev.2013.10.001.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
*Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science, 41*(5), 811–831. https://doi.org/10.1007/s11251-012-9256-z.
*Liu, H. C., & Chuang, H. H. (2011). An examination of cognitive processing of multimedia information based on viewers' eye movements. *Interactive Learning Environments, 19*(5), 503–517. https://doi.org/10.1080/10494820903520123.
*Liu, H. C., Lai, M. L., & Chuang, H. H. (2011). Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers' cognitive processes. *Computers in Human Behavior, 27*(6), 2410–2417. https://doi.org/10.1016/j.chb.2011.06.012.
*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. *Computers in Human Behavior, 65*, 448–459. https://doi.org/10.1016/j.chb.2016.08.041.
*Mason, L., Pluchino, P., & Tornatora, M. C. (2013a). Effects of picture labeling on science text processing and learning: Evidence from eye movements. *Reading Research Quarterly, 48*(2), 199–214. https://doi.org/10.1002/rrq.41.
*Mason, L., Pluchino, P., & Tornatora, M. C. (2015a). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology, 41*, 172–187. https://doi.org/10.1016/j.cedpsych.2015.01.004.
*Mason, L., Tornatora, M. C., & Pluchino, P. (2013b). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education, 60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011.
*Mason, L., Tornatora, M. C., & Pluchino, P. (2015b). Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: An eye-movement study. *Reading and Writing, 28*(6), 851–872. https://doi.org/10.1007/s11145-015-9552-5.
Mayer, R. E. (2001). *Multimedia learning.* Cambridge: Cambridge University Press.
Mayer, R. E. (2005). Introduction to multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.
Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.). *The Cambridge handbook of multimedia learning* (pp. 43–71). New York, NY: Cambridge University Press.
Mayer, R. E. (2014b). *The Cambridge handbook of multimedia learning.* New York, NY: Cambridge University Press.
Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning,* 1–21. https://doi.org/10.1111/jcal.12197.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149–158. https://doi.org/10.1111/j.1365-2729.2006.00170.x.
*Morrow, D., D'andrea, L., Stine-Morrow, E. A., Shake, M., Bertel, S., Chin, J., et al. (2012). Comprehension of multimedia health information among older adults with chronic illness. *Visual Communication, 11*(3), 347–362. https://doi.org/10.1177/1470357212446413.
*O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., & Plass, J. L. (2014). Learning from multiple representations: An examination of fixation patterns in a science simulation. *Computers in Human Behavior, 35*, 234–242. https://doi.org/10.1016/j.chb.2014.02.040.
*Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110–117. http://dx.doi.org/10.1016/j.chb.2009.09.001.
*Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015a). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education, 86*, 30–42. https://doi.org/10.1016/j.compedu.2015.02.016.
*Park, B., Korbach, A., & Brünken, R. (2015b). Do learner characteristics moderate the seductive-details-effect? A cognitive-load-study using eye-tracking. *Educational Technology & Society, 18*(4), 24–36.
Paivio, A. (1986). *Mental representations: A dual coding approach.* New York, NY: Oxford University Press.
*Rau, M. A., Michaelis, J. E., & Fay, N. (2015). Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers & Education, 82*, 460–485. https://doi.org/10.1016/j.compedu.2014.12.009.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.
*Rey, G. D. (2014). Seductive details and attention distraction–An eye tracker experiment. *Computers in Human Behavior, 32*, 133–144. https://doi.org/10.1016/j.chb.2013.11.017.
Rodrigues, P., & Rosa, P. (2017). Eye-tracking as a research methodology in educational context. In C. Was, F. Sansost, & B. Morris (Eds.). *Eye tracking technology applications in educational research* (pp. 1–26). Hershey PA, USA: Information Science Reference.
*Romero-Hall, E., Watson, G., & Papelis, Y. (2014). Using physiological measures to assess the effects of animated pedagogical Agents on multimedia instruction. *Journal of Educational Multimedia and Hypermedia, 23*(4), 359–384.
Rop, G., van Wermeskerken, M., de Nooijer, J. A., Verkoeijen, P. P., & van Gog, T. (2016). Task experience as a boundary condition for the negative effects of irrelevant information on learning. *Educational Psychology Review,* 1–25. https://doi.org/10.1007/s10648-016-9388-9.
*Ruf, T., & Ploetzner, R. (2014). One click away is too far! How the presentation of cognitive learning aids influences their use in multimedia learning environments. *Computers in Human Behavior, 38*, 229–239. https://doi.org/10.1016/j.chb.2014.06.002.
*Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction, 36*, 11–26. https://doi.org/10.1016/j.learninstruc.2014.11.002.
Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In C. Was, F. Sansost, & B. Morris (Eds.). *Eye-tracking technology applications in educational research* (pp. 143–165). Hershey PA, USA: Information Science Reference.
*Schmidt-Weigand, F. (2011). Does animation amplify the modality effect–or is there any modality effect at all? *Zeitschrift für Padagogische Psychologie, 25*(4), 245–256. https://doi.org/10.1024/1010-0652/a000048.
*Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010a). A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and Instruction, 20*(2), 100–110. https://doi.org/10.1016/j.learninstruc.2009.02.011.
*Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010b). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology, 24*(2), 226–237. https://doi.org/10.1016/10.1002/acp.1554.
*Schüler, A., Scheiter, K., & Gerjets, P. (2011). Does the modality effect in multimedia learning appear only with text containing spatial information? *Zeitschrift für Padagogische Psychologie, 25*(4), 257–267. https://doi.org/10.1024/1010-0652/a000050.
*Schüler, A., Scheiter, K., & Gerjets, P. (2012). Verbal descriptions of spatial information can interfere with picture processing. *Memory, 20*(7), 682–699. https://doi.org/10.1080/09658211.2012.693935.
*Skuballa, I. T., Fortunski, C., & Renkl, A. (2015). An eye movement pre-training fosters the comprehension of processes and functions in technical systems. *Frontiers in Psychology, 6*, 1–13. https://doi.org/10.3389/fpsyg.2015.00598.
*Stalbovs, K., Scheiter, K., & Gerjets, P. (2015). Implementation intentions during multimedia learning: Using if-then plans to facilitate cognitive processing. *Learning and Instruction, 35*, 1–15. https://doi.org/10.1016/j.learninstruc.2014.09.002.
Suvorov, R. (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing, 32*(4), 463–483. https://doi.org/10.1177/0265532214562099.
Sweller, J. (1999). *Instructional design in technical areas.* Camberwell: ACER Press.
Tabbers, H. K., Martens, R. L., & Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71–81.
Tabbers, H. K., Paas, F., Lankford, C., Martens, R. L., & Van Merriënboer, J. J. G. (2008). Studying eye movements in multimedia learning. In J.-F. Rouet, R. Lowe, & W. Schnotz (Eds.). *Understanding multimedia documents* (pp. 169–184). New York: Springer.
*Takacs, Z. K., & Bus, A. G. (2016). Benefits of motion in animated storybooks for Children's visual attention and story comprehension. An eye-tracking study. *Frontiers in Psychology, 7*, 1–12. https://doi.org/10.3389/fpsyg.2016.01591.
Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485–498. https://doi.org/10.1037/a0026609.
*Wang, C. Y., Tsai, M. J., & Tsai, C. C. (2016). Multimedia recipe reading: Predicting learning outcomes and diagnosing cooking interest using eye-tracking measures. *Computers in Human Behavior, 62*, 9–18. https://doi.org/10.1016/j.chb.2016.03.064.
*Williamson, V. M., Hegarty, M., Deslongchamps, G., Williamson, K. C., & Shultz, M. J. (2013). Identifying student use of ball-and-stick images versus electrostatic potential map images via eye tracking. *Journal of Chemical Education, 90*(2), 159–164. https://doi.org/10.1021/ed200259j.
Wu, W. H., Wu, Y. C. J., Chen, C. Y., Kao, H. Y., Lin, C. H., & Huang, S. H. (2012). Review of trends from mobile learning studies: A meta-analysis. *Computers & Education, 59*(2), 817–827. http://doi.org/10.1016/j.compedu.2012.03.016.
*Yang, F. Y., Chang, C. Y., Chien, W. R., Chien, Y. T., & Tseng, Y. H. (2013). Tracking learners' visual attention during a multimedia presentation in a real classroom. *Computers & Education, 62*, 208–220. https://doi.org/10.1016/j.compedu.2012.10.009.
*Zhao, F., Schnotz, W., Wagner, I., & Gaschler, R. (2014). Eye tracking indicators of reading approaches in text-picture comprehension. *Frontline Learning Research, 2*(5), 46–66. https://doi.org/10.14786/flr.v2i4.98.