Digital Journalism

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: www.tandfonline.com/journals/rdij20

Can Visuals Facilitate or Detract Attention to Text? Examining the Effects of the Amount and Type of Visuals on Attention to Digital Longforms

Zijian Harrison Gong, Yani Zhao & Kerk Kee

To cite this article: Zijian Harrison Gong, Yani Zhao & Kerk Kee (21 Jan 2025): Can Visuals Facilitate or Detract Attention to Text? Examining the Effects of the Amount and Type of Visuals on Attention to Digital Longforms, Digital Journalism, DOI: 10.1080/21670811.2024.2439362

To link to this article: https://doi.org/10.1080/21670811.2024.2439362

Published online: 21 Jan 2025.

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2024.2439362

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Can Visuals Facilitate or Detract Attention to Text? Examining the Effects of the Amount and Type of Visuals on Attention to Digital Longforms

Zijian Harrison Gong, Yani Zhao and Kerk Kee

College of Media and Communication, Texas Tech University, Lubbock, TX, USA

ABSTRACT

Digital longforms have been viewed as more engaging and pushing the boundaries of digital journalism by including a large amount of visuals, yet it remains unclear whether the inclusion and placement of visuals in digital longforms attracts or diverts attention to textual information. Drawing on the theoretical framework of the Cognitive Load Theory, two eye-tracking experiments were conducted to examine the effects of the number of visuals, the type of visuals, featured image inclusion, and issue involvement on attention to digital longforms. The results suggested increasing the number of visuals increased attention to textual information. However, changing the number of visuals did not affect all readers uniformly such that issue involvement and featured image inclusion moderated the effects of the number of visuals on attention. The type of visuals also affected selective attention, such that infographics were the most effective in grasping readers' attention than photos and banners. However, such advantage of infographics occurred at the expense of diverting attention away from the text following infographics.

KEYWORDS

Digital longforms; visual attention; visual communication; experiment; reader engagement

Multimodal longforms present opportunities for using various types of visuals to help readers absorb information, prevent fatigue, and maintain interest throughout the article (Greussing and Boomgaarden 2019). One of the main differences between traditional print news and digital news is the abundance of visuals in digital longforms. Digital longforms are text-based news stories with a focus on using multiple compelling visual elements to maintain readers' attention (Jacobson, Jacqueline, and Gutsche 2016; Greussing and Boomgaarden 2019). As digital longform articles no longer have the page and layout restrictions of traditional print news, digital news outlets often prioritize visual elements and include multiple types of visuals such as photos, infographics, and illustrations to capture audiences' attention.

However, the excessive use of visuals could be distracting and reduce attention to textual content because longforms require readers to toggle between textual information and various types of visuals (Duffy 2022). The presence of multiple visuals

CONTACT Zijian Harrison Gong Harrison.gong@ttu.edu

can disrupt the intended flow of information, reducing the effectiveness of text-based storytelling. The captivating yet non-essential or even extraneous visuals, referred to as seductive details (Garner, Gillingham, and White 1989), could advertently detract readers from crucial content. A meta-analysis of 39 experimental effects indicated difficulties in understanding and remembering the main concepts when learners were attracted to seductive details presented alongside crucial information (Rey 2012). A more recent analysis of 58 studies on the effects of seductive details suggested that including such details can harm learning, although the extent of this impact varies depending on factors like the type of images used, the mode of presentation (dynamic or static), and the placement of seductive details (Sundararajan and Adesope 2020).

Prior empirical evidence also suggested that texts and pictures have fundamentally different cognitive functions (Schnotz and Wagner 2018). Specifically, the initial processing of an article focuses on developing a broad understanding of the subject to use for further message elaboration (initial mental model construction), whereas the adaptive mental model after the initial processing involves switching focus between text and pictures (Hochpöchler et al. 2013). Schnotz and Wagner (2018) studied how readers process information that includes both text and images and found that during the initial phase of reading and comprehension, constructing a mental model relies more on text than on pictures. Conversely, adaptive mental model elaboration relies more on pictures than on text (Schnotz and Wagner 2018). Moreover, more switches between textual information and pictures occurred during the initial model construction compared to the adaptive model elaboration phase (Schnotz and Wagner 2018). Findings from another study corroborated that texts offer clearer guidance for building initial mental models than pictures, while pictures are better at helping adaptive mental models by allowing easy access to specific information needed for comprehension (Zhao et al. 2020).

In the context of journalism, empirical research also identified that the integration of different components in digital longforms significantly affected interface evaluation, meaningful learning, and involvement with the content (Greussing and Boomgaarden 2019). However, whether and how visuals affect the function text element in digital longforms remains unclear, especially on the aspect of selective attention to different message elements. Longforms utilize the layout space to arrange various semiotic resources, resulting in different types of digital news packages such as text-flow, image-flow, page-flow, and individual semiotic modes (Hiippala 2017). The most prevalent layout type among these is text-flow, which is characterized as a linearly unfolding written narrative composed of paragraphed text (Hiippala 2017). Several theoretical models were proposed to explain the impact of image-text integration on reading comprehension. A large body of research that focused on the multimedia context applied the Cognitive Load Theory (CLT; Sweller 1988; Sweller, Ayres, and Kalyuga 2011) to specify the relationship between working memory, selective attention, and cognitive load when processing information. When reading digital longform articles, readers need to allocate their limited working memory resources effectively to encode the information and construct a schema for learning. Selective attention helps in this process by prioritizing goal-related information that enters working memory, which temporarily holds information such as words and their meanings (Sweller 1988). However, individuals may not always be able to effectively direct attention and manage cognitive load effectively because the cognitive load required to process the information is imposed by both the inherent complexity of the content (intrinsic cognitive load) and how the material is presented (extraneous cognitive load). When both intrinsic and extraneous cognitive load are high due to the presence of an excessive number of images, fewer cognitive resources are available for germane cognitive processing, which is the working memory resources that are devoted to learning.

Based on the CLT, if a digital longform article contains an excessive number of visuals, both processing the visuals and understanding the relationship between visuals and textual information would increase the extraneous cognitive load. Studies applied the CLT framework observed the split-attention effect, which showed distraction and hindered comprehension occurred when textual and pictorial information are presented together but spatially separated (e.g., Chandler and Sweller 1991; Ginns 2006). Additionally, people are "cognitive misers" that will not allocate all the available attentional resources to a news story (Lang 1990). Readers may make assumptions about the content solely based on attention-grabbing images and skip the text. It is worth noting that theoretical discussions on the relationship between visuals and textual information integration remain inconclusive. Contrary to the CLT, the cognitive theory of multimedia learning (Mayer, Heiser, and Lonn 2001) suggests that media content can improve comprehension and audience engagement by combining text and pictures. Empirically, the image-text relationship has been studied in the context of learning materials such as textbooks (Sandqvist 1995) and newspaper reading (Holšánová, Holmberg, and Holmqvist 2005). Most of these studies suggested images help facilitate comprehension when they reinforce or demonstrate information given in the text or provide additional information.

These seemingly discrepant findings suggested that visual-text integration could be the key to the effective use of visuals in digital longforms. However, the critical barriers to evaluating the effectiveness of visual-text integration in digital news are 1) there is no consensus on whether a large number of visuals would transfer or deprive attention to textual information, 2) a lack of research that examined the visual-text ratio (i.e., how many visuals should be included in a digital longform) as most prior research focused on the content or rhetorical connection between images and text, and 3) few empirical evidence demonstrated how different types of static visuals (e.g., infographics, illustrations, and photos, etc.) affect attention as they often occur together in digital news. Additionally, most prior research used message recall or comprehension to index attention without directly measuring visual attention allocation. This research expands existing literature on how visuals in digital news affect attention by dissecting the structure of digital longforms to explore effective ways to organize and integrate visual and textual content. Two eye-tracking experiments were conducted to systematically examine the impact of the number and type of visuals on attention allocation.

Visuals in Digital Longforms

Inspired by the New York Times' groundbreaking and successful "Snow Fall: The Avalanche at Tunnel Creek" coverage, journalists hold an optimistic perspective on longform stories as an innovative form of digital journalism (Pavlik and Bridges 2013; Greussing and Boomgaarden 2019). The layout of digital longforms usually includes bold headlines and a featured photo, link lists, main news frames, illustrations, photos, drop quotes, advertisements, and testers/briefs. Among these elements, visual elements received the most research attention because they can capture initial attention in online news reading, demonstrate the content, facilitate attention to the text, and/or provide an esthetic touch to the website (Holmberg, Holšánová, and Holmqvist 2006). In web-based longforms, visuals such as illustrations, infographics, and photos are considered macro-level reading supports for comprehension (Liu, Mao, et al. 2019). Barnes (2023) emphasized the importance of encapsulation in simplifying complex events into brief visual and textual representations, thereby making complex information more accessible and relatable. However, she also underscored that maintaining a balanced and equal relationship between text and visual elements, known as parity, is essential for creating compelling narratives (Barnes 2023). Readers still need to stay focused on textual information if they were to read digital longforms because they are much longer and more detailed than typical news articles or blog posts. Among various visual elements, the number of visual elements, the use of featured photos, and the type of visual elements are the main factors investigated in our research due to their ability to affect attention and cognitive processes.

The Number of Visuals and Cognitive Load

Using multiple visuals is an integral part of digital longform news. Some of the multimedia learning research suggested images help facilitate comprehension when they reinforce or demonstrate information given in the text or provide additional information (e.g., Barnes 2023). However, including multiple visuals may overwhelm and distract readers. The cognitive load theory suggests that our working memory can only process a limited amount of information at a time (Sweller, Ayres, and Kalyuga 2011). The germane cognitive load—the portion of working memory devoted to constructing and automating schemas for learning—is especially limited in its capacity when dealing with new information (Sweller, Ayres, and Kalyuga 2011). When a large number of visuals are included in a digital longform article, our brains struggle to prioritize and process the news, leading to a decreased ability to focus on textual information. The split-attention effect occurs when readers have to mentally integrate multiple forms of information to develop a coherent understanding of the material (Pouw et al. 2019). This mental integration takes working memory capacity and exhausts cognitive load. As a coping strategy, readers might reduce attention to textual information or visuals to mitigate cognitive load.

From the CLT perspective, cognitive load is determined by how much cognitive resources viewers allocate to the message, the easiness of integrating new information into existing knowledge networks (intrinsic cognitive load), and the cognitive resources required to process the message (intrinsic and extraneous cognitive load). In digital longforms, readers first need to construct initial mental models to form a preliminary understanding of the content, followed by the development of adaptive mental models to refine and expand the initial mental model as they progress through the article (Schnotz and Wagner 2018; Zhao et al. 2020). They also must process different types of content, such as text and images, and integrate these elements effectively. This dual processing raises both intrinsic load, which pertains to the inherent complexity of the content, and extraneous load, which involves the effort required to manage and integrate these diverse information sources. Consequently, exhausted cognitive load occurs when the resource required to process the message exceeds the resource allocated to the message, leaving less cognitive resources for essential processing (Sweller, Ayres, and Kalyuga 2011).

Within the context of journalism, high cognitive load hinders key outcomes such as attention to textual information, and ultimately, comprehension and memory of the content. Because the cognitive load is determined by both cognitive resources required by the message and cognitive resources allocated to the message, we first discussed message factors (visuals and texts) that affect the intrinsic and extraneous cognitive load and then explained the impact of issue involvement on cognitive resource allocation. Next, the effects of featured photos and visual types are discussed.

Texts, Visuals, and Cognitive Load

In digital longforms, texts present fact-based information and require readers to allocate a certain amount of cognitive effort because they can only be coded using symbolic coding (Paivio 1991). Meanwhile, visuals require less cognitive effort to encode because they are encoded through both analogue coding (i.e., the code of perceptual features) and symbolic coding (i.e., coding based on mental representations; for review, see Paivio 1991). Readers usually prioritize visuals over textual information because visuals can be encoded and retrieved more easily than texts. This phenomenon is labeled as the "picture superiority effect" in prior research (Hockley 2008), suggesting that when visuals and texts compete for attentional resources, visuals will receive most of the attentional resources and even divert attention away from texts, especially when multiple visuals are presented in the message. For example, Merle, Callison, and Cummins (2014) found attention and memory to graphical information detracted attention to textual information when infographics are included in online news.

The effects of visual-text integration on cognitive load can also be explained from the layout design perspective. Visuals are usually integrated into digital longforms using the top and bottom placement by inserting a visual between two paragraphs (for example, see Figure 1). Images that supplement the text (e.g., illustrations and photos) or are unrelated to the text (e.g., advertisements) require readers to split their visual attention between different sources of information, which can exhaust the visual information-processing channel (Mayer, Heiser, and Lonn 2001). Mentally integrating the images with the text may result in an increased load on working memory and unnecessarily divert cognitive resources. Corroborated by visual attention research, the E-Z reader model (Reichle, Rayner, and Pollatsek 2003) also suggests that attention is allocated to one word at a time during reading. Eyes move across the visual field in a series of rapid jumps called saccades, and fixations occur between saccades (Jacob 1995). Visual processing occurs during fixations and is suppressed during saccades. A saccadic movement requires about 24 milliseconds to execute, and during this process, the lexical processing will continue using information from the previous fixation location until the new fixation starts (Jacob 1995). When

Figure 1. Sample stimulus.

interruption during this saccadic movement occurs (i.e., an image placed between two paragraphs), the lexical process will be interrupted, resulting in a pause and/or regressions in reading. When multiple visuals are placed between paragraphs or sentences, viewers have to go through the images when reading the news, potentially disrupting the lexical process and thus increasing the extraneous load to connect the text with visuals.

Issue Involvement and Cognitive Load

Previous studies identified that perceived issue involvement—the extent to which an attitudinal issue is considered personally important—affects the amount of resources allocated to the message (Petty and Cacioppo 1986; Lang 1990). For readers who perceive the topic of longform news as highly self-related and instrumental, cognitive resource allocation is increased as personally relevant messages elicit the automatic allocation of attentional resources to relevant information (Lang 1990). In this case, including a large number of visuals may not distract readers from visually attending to textual information due to this increase in cognitive resource allocation. By contrast, when the issue covered in the news has low perceived relevance, readers allocate limited cognitive resources that focus on processing visuals and reduce attention to textual information due to the visual superiority effects.

Similarly, based on dual process models such as the Elaboration Likelihood Model (Petty and Cacioppo 1986), the central route of information processing involves message elaboration and requires readers to systematically evaluate the content, which demands more cognitive resources. By contrast, peripheral processing is usually guided by heuristic cues or surface characteristics when recipients do not examine the information thoroughly due to a lack of motivation and/or ability to process a message (Petty and Cacioppo 1986). Being on the lower end of message elaboration and process thoroughness, the peripheral route relies on a cursory evaluation of the message, which demands less cognitive resources.

Thorough processing *via* the central route places a high cognitive load because readers need to handle complex arguments (high intrinsic load) and actively integrate new information with existing knowledge. Since people are "cognitive misers" (Lang 1990), they will not devote all their cognitive resources to a news story. Readers may rely on simple peripheral cues rather than analyzing the textual information comprehensively because the presence of peripheral cues such as images and photos places less demand on cognitive load. From the perspective of CLT, the presence of a large number of peripheral cues such as images can make readers less likely to process the information centrally because these cues increase the effort required to process more elements and establish connections between these elements to comprehend the information (i.e., increased extraneous load). This added cognitive burden negatively affects readers' ability to process essential information, especially when peripheral cues (e.g., banner ads) do not contribute to schemata acquisition and learning (Debue and van de Leemput 2014). Empirical evidence also supported that visuals are more likely to divert attention away from textual information among readers who are not motivated to dispense enough cognitive resources to the news story. For example, Lee and Kim (2016) investigated how infographics affect the processing of news stories and found adding graphics to the news increased news elaboration only among participants with a high level of issue involvement. This distraction effect of visuals observed in prior research (e.g., Lee and Kim 2016; Merle, Callison, and Cummins 2014) should be exacerbated when a news article includes a large amount of visuals.

In short, digital longforms with multiple visuals may lead to two different types of processing scenarios. For readers who perceive the focal topic in the news as highly relevant and important (i.e., high issue involvement), a high level of resource allocation warrants the processing of both visual and textual information. By contrast, when readers perceive the focal topic in the news as having low importance, they may not allocate enough cognitive resources to the message. The lack of resource allocation results in reduced attention to texts and increased attention to visuals, especially when multiple visuals are presented in a news article. As such, we propose the following hypotheses:

H1: Increasing the number of visuals will increase attention to visuals (H1a) and reduce attention to textual information (H1b).

H2: Issue involvement will moderate the effect of the number of visuals on attention to textual information, such that the effect of visuals diverting attention from texts will be more significant among participants with low issue involvement.

Featured Image and Attention

A featured image (also known as a "lead photo" or "lead image") is the primary image typically placed at the top of a news article (Economou 2006). Featured images usually illustrate the main theme of a news story and/or demonstrate the most newsworthy moment. While having a featured image could enhance the visual appeal of an article, it may also prevent readers from reading the entire article. Readers are cognitive misers who tend to conserve mental efforts and avoid the effortful processing of news articles (Lang 1990). When a featured photo is prominently displayed at the top of an article, readers may assume the content of the article solely based on the featured image because a provocative or powerful lead photo can set strong expectations and serve as a preview of the article. As a result, the featured image could prompt readers to either skip the article entirely or only skim it, missing out on important details or nuances in the text. Prior research on text-image integration suggested that images serve as valuable scaffolds in the initial stages of reading an article, aiding readers in quickly forming mental models for better understanding. However, after this initial phase, images offer limited support for systematic processing of the content (Schnotz and Wagner 2018). An experiment that manipulated the visual types (static or animated) and the narrative position of the visuals (placed within the first, second, or third paragraph of a four-paragraph story) indicated that textual information directs the development of mental models by determining the sequence in which the story's events should be processed. Visuals enhance the text by highlighting the essence of the situation depicted, and proper placement of visuals significantly enhances the story's comprehensibility. Parity is achieved when visuals are positioned alongside text at specific narrative points that help facilitate cross-referential connections (Barnes and Sontag 2024). Journalism research also found participants read faster on digital media than reading on papers yet with shallower processing of the text (Lenhard, Schroeders, and Lenhard 2017). While Lenhard and colleagues did not focus on the impact of visuals on reading comprehension, their study suggested information modality affected cognitive resource allocation, and the cognitive miser effect may be more prominent when people consume digital content. Based on the aforementioned empirical evidence, the inclusion of featured images in digital longform may prevent readers from reading the entire article, and we propose the following hypothesis and research question:

H3: The inclusion of a featured image will reduce visual attention to textual information.

RQ1: Will the inclusion of a featured image moderate the impact of the number of visuals on attention to textural information?

Visual Type and Attention

Prior research categorized the role of text-adjacent visuals as decorative, representational, interpretational, organizational, and transformational when examining how images influence comprehension when reading texts (Levin et al. 1987). Studies have focused on whether explicit markers (e.g., arrows, links, or numerical notes) or implicit markers (e.g., in-text reference) can change the text-images relation from unidirectional to bidirectional (e.g., Holšánová, Holmberg, and Holmqvist 2005). Considering the definitions of the various types of visuals, there may be some degree of overlap between the content of visual and textual representations. Redundancy between the visuals and text occurs when they are combined (Levin et al. 1987). Research further specified image redundancy as the extent to which the image added to a written text contains the same content (Schnotz 2005). The current research compared infographics, photos, and banner ads as three types of typical visuals featured in digital longforms.

Specifically, infographics provide succinct textual (e.g., keywords) and visual (e.g., diagrams) information to schematize the summary of the content (Gilbert 2010). As infographics schematize and summarize the written content, it is plausible to conclude that an infographic is a detailed diagram and has a higher level of redundancy with written text than photos, which mainly provide context, visualization, or backdrop of the text. Although more studies are needed to directly compare different visual types in terms of visual-text redundancy, some studies demonstrated that a simplified diagram with fewer irrelevant elements is less likely to distract readers compared to a detailed diagram (Cromley et al. 2013). From the perspective of cognitive load theory, it is possible that infographics, which have a higher level of redundancy with the accompanying textual information, are more likely to require more cognitive resources and lead to cognitive overload. Photos, on the other hand, are often made with the express intention of illustrating a specific notion or idea. As a result, photos are more narrowly focused, share less redundancy with texts, and thus require less resources to process. Therefore, photos might distract readers less from the paired text than the infographics.

Compared to infographics and photos, banner advertisements share the lowest level of redundancy with textural information. Additionally, readers may be able to identify banner ads immediately and bypass such information. However, the split-attention effect occurs when readers have to mentally integrate multiple forms of information to develop a coherent understanding of the material, which suggests banner ads still may distract readers even if they are irrelevant to the text (Liu, Mao, et al. 2019). Therefore, we propose the following hypotheses and research question:

H4: Infographics will receive more attention than photos (H4a), yet textual information under infographics will receive less attention than texts after photos (H4b).

RQ2: How will attention to banner ads affect attention to textual information presented after banner ads?

Method

Two experiments were conducted to address the hypotheses and research questions. The first experiment tested 1) whether increasing the number of visuals reduces attention to textual information, and 2) the moderating effects of featured images and issue involvement on attention to texts (H1-3 and RQ1). Experiment 2 focused on the impact of visual types (i.e., infographics, photos, and banners) on attention to textual information presented after these visuals (H4 and RQ2).

Experiment 1 Design, Procedure, and Stimuli

A 3 (number of visuals: one v. three v. five) X 2 (featured image: presence v. absence) between-subjects eye-tracking experiment was conducted with 156 participants. Participants were recruited from a large public U.S. Southwest University and surrounding communities (M_{age} = 22, SD=2.12; Female = 64.20%). Data from four participants were excluded from the analysis due to inaccurate calibration and software malfunction.

Data was collected in individual sessions. During the lab session, participants sat in front of a desktop computer where a small eye tracker mounted at the bottom of the screen directs infrared light toward the participant's eyes to generate corneal reflections. The Tobii X2-120 eye-tracking apparatus and software then continuously recorded the distances between the corneal reflection and the center of the pupil to capture the gaze point on screen. The sampling rate is 120 Hz. Each digital longform article was presented in the middle of the screen without other web design elements (e.g., buttons and external links) in the visual fields. After a practice session, participants were randomly assigned to one of the six experimental conditions. Participants were allowed to browse the stimuli at their own pace, which mimics natural reading conditions and helps gather data that more accurately reflects how individuals read digital longforms in real-world settings. The stimuli did not contain any hyperlinks, and participants were only able to scroll up or down the webpage when reading the articles. Upon completion of the eye-tracking session, participants finished a posttest that measured their issue involvement and demographics.

Stimuli

Four articles related to humans' impact on the environment from a local collaborator's (a regional PBS station) website were selected. The four articles have roughly the same length and serve as a repetition to avoid the idiosyncratic effect of a particular article. The word counts of the articles range from 865 to 984 words (4504 – 4915 characters without spaces). All conditions contained the same four articles, but different numbers of visuals were inserted into each article based on the experimental condition. A featured image was counted as a visual when manipulating the number of visuals. For example, in the "three visuals + featured image" condition, the featured image was followed by two additional visuals inserted throughout the article. The selected visuals and articles were first archived, and we built a mock website using the archived images/articles to eliminate the potential effects of source credibility, website design, and layouts on attention. All webpages were hosted on the same mock website. Each webpage was paired with a white background and used the same layout and font size (see Figure 1).

In the single image conditions, the same image was placed either at the beginning of the article as the featured image (one visual with a featured photo condition) or after the first paragraph (one visual with no featured photo condition). In the three visuals condition, each article contained one photo, one infographic, and one banner ad. In the five visuals condition, each article contained two photos, two infographics, and one banner.

Measurements

Experiment 1 gauged the impact of the number of visuals and the featured image on *total fixation durations*. The central premise of the eye-tracking method is that gaze fixations reveal what element in the visual field is being processed cognitively, whereas fixation durations reveal the amount of attentional resource allocation to these elements (King et al. 2019). In this study, a fixation is defined as the moment when the gaze point remained stationary on the screen for over one-tenth of a second. To conceptualize the relationship between fixation and cognitive processes, we adopted the eye-mind relationship proposed by Rayner (1998), which suggested that gaze fixations can be used as a proxy for attention because our eyes move to and fixate on areas of interest within a visual field to help visually gathered information processed in visual working memory. The location and duration of gaze fixations indicate the allocation of attentional resources as the eyes fixate on particular points in the visual field when the brain actively processes information presented there (Rayner 1998). Additionally, because "additional time spent on perception is not used to examine the secondary element, but to reexamine the most important elements" within a visual presentation (Yarbus 1967, 193), this research used total fixation duration to indicate the total amount of attentional resource allocated to an area of interest (i.e., visuals or textual information).

Because the stimuli included articles focusing on environmental issues, *issue involvement* was measured using the New Environmental Paradigm (NEP) scale (Dunlap et al. 2000). This scale included 15 items to systematically evaluate environmental concerns and the perceived severity of environmental issues (e.g., "If things continue on their present course, we will soon experience a major ecological catastrophe."). Participants evaluated each scale item using a 7-point scale (*1 = Strongly Disagree*; *7 = Strongly*

Agree). All scale items were averaged to generate a single measure of issue involvement, in which a higher value indicates a higher level of issue involvement.

Experiment 1 Results

The data was analyzed using the SPSS Process Macro due to its ability to apply bootstrapping methods to help assess indirect effects. SPSS Process also allows model customization by specifying different types of mediators and moderators for tailored analyses. A moderated mediation model was constructed to assess the effect of the number of visuals, issue involvement, and featured image on attention to textual information (see Figure 2 for model configuration). The number of images served as the predictor (X), attention to visuals was included as the mediator (M), issue involvement and featured image served as moderators, and visual attention to textual information served as the outcome variable (Y). Because the number of visuals is a three-level nominal level variable, we applied indicator coding using the one image condition as the reference group and recoded number of images into two variables in the model (w1: one visual versus three visuals; w2: one visual versus five visuals). Direct effect results are summarized in Table 1, and indirect effects are summarized in Table 2.

The direct effect of the number of visuals on attention to textual information was not significant. However, the indirect effect of the number of visuals on attention to textual information through attention to images was significant in all conditions (see Table 2). The results using contrast coding (w1 and w2) reviewed that while increasing the number of visuals from one to five increased attention to visuals, attention to textual information was not affected (see Table 1). H1a was supported, and H1b was rejected. A higher level of issue involvement significantly increased attention to textual information but not visuals. Interestingly, issue involvement did not moderate the direct effects of the number of visuals on attention to texts. Instead, issue involvement moderated the indirect effect of the number of visuals on attention to textual information, and this moderation effect was conditioned by the number of visuals. When the number of visuals increased from one to three, the number of visuals showed a significant positive indirect effect on attention to textual information through increased attention to images. This indirect effect was *more pronounced* among participants with high issue involvement. By contrast, when the number of visuals increased from

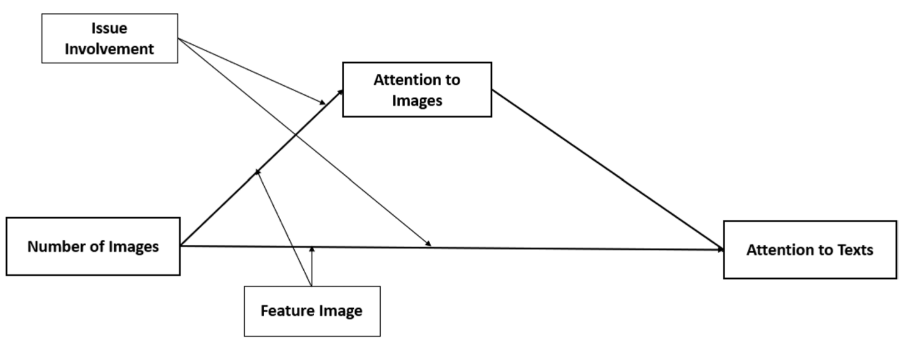

Figure 2. Experiment 1 model.

Table 1. Moderated mediation regression model results.

	Attention to visuals *B* (SE)	Attention to textual information *B* (SE)
Constant	4.149(6.210)	−6.710(36.341)
w1 (one visual v. three)	1.808(8.080)	−73.804(47.215)
w2 (one visual v. five)	**20.691(7.965)***	1.153(47.622)
Issue involvement	−.303(1.291)	**15.078(7.541)***
w1 × issue involvement	.758(1.652)	6.727(9.660)
w2 × issue involvement	−2.677(1.640)	−9.962(9.671)
Feature image	−1.232(1.932)	−11.285(11.303)
w1 × feature image	1.527(2.659)	**42.129(15.553)****
w2 × feature image	−1.032(2.661)	25.124(15.550)
Attention to visuals		**3.123(.489)*****
R^2	.502	.582
Adjusted R^2	.252	.339

*$p < .05$; **$p < .05$; ***$p < .001$.

Table 2. Indirect effect of number of images on attention to textual information.

Moderator variable	*B* (SE)	LLCI	ULCI
Without feature image			
Number of visuals (w1: one v. three) → attention to images →attention to textual information			
Low issue involvement[a]	14.521 (6.583)	2.078	28.474
Medium issue involvement[b]	16.670 (5.636)	6.564	28.977
High issue involvement[c]	19.057 (6.365)	8.080	33.083
With feature image			
Number of visuals (w1: one v. three) → attention to images →attention to textual information			
Low issue involvement[a]	19.292 (5.786)	9.323	32.050
Medium issue involvement[b]	21.461 (5.678)	11.716	33.629
High issue involvement[c]	23.828 (7.229)	11.581	39.655
Without feature image			
Number of visuals (w2: one v. five) → attention to images →attention to textual information			
Low issue involvement[a]	33.274 (9.118)	17.800	53.796
Medium issue involvement[b]	25.610 (6.470)	14.617	39.891
High issue involvement[c]	17.249 (6.213)	5.920	30.659
With feature image			
Number of visuals (w2: one v. five) → attention to images →attention to textual information			
Low issue involvement[a]	30.051 (8.509)	16.589	49.980
Medium issue involvement[b]	22.387 (6.498)	11.600	37.081
High issue involvement[c]	14.027 (7.146)	1.413	29.783

[a]Conditional indirect effect of number of images when issue involvement is at + 1 SD.
[b]Conditional indirect effect of number of images when issue involvement is at the mean.
[c]Conditional indirect effect of number of images when issue involvement is at - 1 SD.

one to five, the number of visuals also exerted a positive indirect effect on attention to texts *via* increased attention to images, yet this effect was *less significant* among participants with high issue involvement (See Table 2).

Hypothesis 3 predicted the inclusion of a featured image would reduce attention to textual information. As seen in Table 1, the inclusion of a featured image did not significantly reduce attention to textual information. However, the indirect effect of the number of visuals on attention to texts was moderated by featured image inclusion. As seen in Figure 3 and Table 2, when a digital longform only included one image, placing the image at the beginning (as the featured image) reduced attention to textual information more than placing the image after the first few paragraphs. However, when the number of visuals increased to three, participants spent more time reading texts in articles with featured images than in articles without featured images. Finally, when the number of visuals increased to five, the inclusion of a featured image decreased the attention to texts compared to the three visuals condition.

H3 was partially supported in the one visual condition but not the three or five visuals conditions.

Experiment 1 Summary

The results suggested that the number of visuals affected attention to texts through attention to visuals. Additionally, the indirect effects of the number of visuals on attention were moderated by issue involvement and featured image inclusion. When the number of visuals increased from one to three without the presence of a featured image, attention to visuals benefited attention to texts, and this effect was more significant among participants with high issue involvement than participants with low issue involvement. However, when the number of visuals increased from one to five, increasing the number of visuals promoted attention to texts more among participants with low issue involvement than participants with high issue involvement. This finding suggests both the number of visuals and issue involvement moderated the effect of visuals on attention to texts. Specifically, while increased attention to visuals did not divert attention from texts, the beneficial effects of the number of visuals on attention to text were conditioned by issue involvement. When a moderate number of visuals were featured (i.e., increasing from one to three visuals), visuals were able to capture attention and facilitate attention to texts, especially among readers with high issue involvement. However, when a large number of visuals were included (i.e., increasing from one to five visuals), participants with low issue involvement allocated more attention to texts than participants with high issue involvement.

The effect of featured image inclusion also moderated the impact of the number of visuals on attention. As seen in Figure 2, when a longform article only presented one visual, placing the visual as the featured image exerted a more significant effect in reducing attention to texts compared to placing the visual after the first paragraph. This finding is consistent with our prediction that readers may assume the content of the article solely based on the featured image, which prompts readers to skim the content and thus reduce attention to the texts, especially without the presence of other visuals in the longform article. When a moderate amount of visuals were

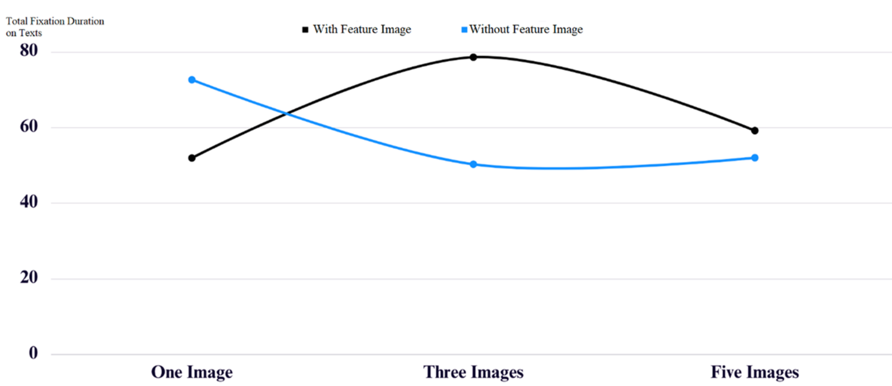

Figure 3. The effect of feature image inclusion and number of visuals on attention to texts.

presented (i.e., the three visuals condition), featured images with other visuals can create a visual narrative that complements the written text, and thus increases attention to textual information. However, when a large number of visuals are presented, the amount of visual and textual information overloads readers' cognitive capacity, and readers started processing longform articles by relying on peripheral cues such as featured images and other visuals, resulting in reduced attention to textual information.

Experiment 2 Design, Procedure, and Stimuli

The procedure of experiment 2 was the same as in experiment 1 except for the following differences. First, experiment 2 applied a within-subject design to examine the impact of different types of visuals on attention to textual information immediately following these visuals. All participants read the same four longform articles. Each article included three visuals (i.e., an infographic, a banner, and a photo). Second, the total fixation duration within each visual and the immediate paragraph following each visual were recorded separately. Forty-five participants (M_{age} = 23.50, 57.78% Females) were recruited to participate in this experiment, and data from two participants was excluded due to software malfunction. Data was collected in individual sessions, and the four articles were presented in a random order.

Experiment 2 Results

A pair of 3 (image type: infographics v. photos v. banner ads) X 4 (repetition) repeated-measures ANOVAs were conducted to examine the effect of image type on attention using attention to images and attention to texts as dependent variables, respectively. The effect of image type on attention to images was significant ($F(1.06, 43.58) = 50.48$, $p < .001$, $\eta_p^2 = .55$). Infographics ($M = 7.61s$, $SD = .96s$) received more attention than photos ($M = 1.54s$, $SD = .20s$) and banner ads ($M = .947s$, $SD = .13s$). The effect of image type on attention to texts accompanying the visual was also significant ($F(1.55, 42.05) = 9.45$, $p = .001$, $\eta_p^2 = .19$). Texts following infographics ($M = 18.93s$, $SD = 2.59s$) received less attention than texts after photos ($M = 23.57s$, $SD = 2.84s$) and texts after banner ($M = 20.20s$, $SD = 2.58s$). As predicted, while infographics received more visual attention than photos, texts presented after infographics received less attention than texts presented after photos. H4a and H4b were both supported. The effects of banner ads on attention are more similar to photos than infographics, as texts presented after banners received less attention than texts presented after photos but more attention than text after infographics.

Experiment 2 Summary

The second experiment focused on the impact of different visual types, including infographics, photos, and banner ads, on audiences' attention to textual content. The results suggested infographics were the most effective in grasping readers' visual attention, followed by photos and banner ads. However, such attentional advantage occurred at the expense of diverting attention away from text following infographics.

The findings presented here are in line with the Cognitive Load Theory and picture superiority effect, which suggests that when a high level of visual-text redundancy occurred, visual elements receivee more attention and reduce attention to texts.

Discussion

The use of visuals in digital longforms has become increasingly popular in modern journalism. However, it is crucial to consider how these visuals affect audiences' attention to textual content. A large body of prior research focused on the differences between static and animated visuals on attention and comprehension, or on the placement of visuals within text (e.g., Barnes and Sontag 2024; Merle, Callison, and Cummins 2014). However, the current study contributes to this body of research by exploring two novel aspects: the impact of the number of visuals and the effects of visuals that are not directly relevant to the textual information, such as banner ads. By examining these factors, this research provides a deeper understanding of how various visual elements affect attention allocation to and retention of textual content.

Study 1 examined the number of visuals on attention allocation, as well as how issue involvement and featured images moderate such effect. The findings indicated that the number of visuals on attention to textual information was moderated by both individual-level (i.e., issue involvement) and message-level (i.e., featured images) factors, confirming that cognitive load is affected by both allocated resources and the message's required resources. Changing the number of visuals did not affect all readers uniformly. People with high issue involvement did not appear to be affected when the number of visuals increased from one to three, yet the beneficial effects of visuals on attention to texts started diminishing when the number of visuals increased from one to five. People with lower issue involvement paid greater attention to visuals and text information when the number of images increased due to the large number of peripheral cues that facilitate peripheral processing. By contrast, increasing the number of visuals made it more difficult for high issue involvement participants to connect visuals with text and to centrally process the information. In addition, it is also possible that people who have high issue involvement already possess pertinent information and experiences about digital longform material. The significant proliferation of graphics helps reduce the intrinsic cognitive load and facilitates the rapid comprehension of the primary content. This finding suggests that high issue involvement may guard against the negative effects of high cognitive load caused by increasing the number of visuals in digital longforms but only to a certain extent.

Moreover, lead images had an interesting moderating effect on the impact of the number of visuals on visual attention to texts. When a digital longform only presented one image, placing the image at the beginning of the article (as the featured image) reduced attention to text more than placing the image after the first paragraph. However, when the number of visuals increased to three, participants spent more time reading texts in articles with featured images than in articles without featured images, yet such positive effect of featured images disappeared when the number of visuals increased to five. Theoretically, these findings contribute to the CLT by further clarifying the cognitive load dynamics between the initial mental model construction and the adaptive mental model. An article begins with a featured image that could create an initial mental model built around the image, possibly causing a delay in integrating textual information and affecting adaptive mental model construction. When the image is placed after the first paragraph, readers begin constructing their mental model based on the text, leading to an initial focus on the narrative and a more text-centered mental model. This finding also corroborates prior research that the construction of a mental model during the initial phase of message processing relies more on text (Schnotz and Wagner 2018). Additionally, the findings suggested that featured images could play a significant role in the development of adaptive mental models by providing a visual anchor that structures readers' understanding and integration of subsequent information. When only three visuals were presented with a featured image, the text-image integration was balanced, and the mental model could adapt to both textual and visual information. However, when five visuals are included, the transition from the initial mental model to adaptive mental model construction could be continuously disrupted by the frequent appearance of new visuals. As a result, readers could not develop adaptive mental models efficiently to integrate visuals with texts, and the phenomenon of visuals diverting attention from texts occurred. Practically, content creators need to be strategic when deciding the placement and number of visuals in digital longform news. Presenting a large number of visuals may not be the best strategy to maintain readers' attention and interest, even when readers are engaged with the focal topic. Rather, journalism practitioners should be mindful of the number and types of visuals they use and consider their potential impact on cognitive overload and attention allocation.

Understanding the impact of different visual types on audiences' attention to textual content is also crucial for creating effective and engaging multimedia content. Findings from our second experiment have important implications for journalists and content creators who aim to engage readers with digital longform articles. While infographics are useful in conveying complex information, their distracting effect on textual content should be taken into account. The study suggests that placing important textual content under banners or photos rather than under infographics might improve attention and comprehension of textual information. Alternatively, infographics could replace some statistical or descriptive textual information to reduce the infographic-text redundancy. Additionally, writers could also emphasize the differences between infographics and the accompanying texts before showing infographics to reduce visual-text redundancy. Webpage layout designers can also use arrows or other visual cues to connect infographics with related textual information to reduce the need for readers to mentally integrate multiple sources of information.

Limitation and Direction for Future Research

While our study provided valuable insights into the impact of the number and type of visuals on attention to digital longforms, it is essential to acknowledge certain limitations in our research design. These limitations point out directions for future research to expand our work and offer a more comprehensive understanding of the role of visuals in digital content.

First, while the controlled manipulation of number of visuals allowed us to draw conclusions about the relationship between visual quantity and attention allocation patterns, it may not accurately reflect real-world scenarios where visual content is often determined by various factors, including content type, user preferences, and available resources. Future research should explore more organic ways to operationalize the number of visuals to better capture the complexities of digital longforms. Excluding videos and focusing on static images enhanced the internal validity of the experimental design but may affect its external validity. Future research should consider investigating the impact of video content and how it interacts with images and textual information. Additionally, stimuli in this research excluded hyperlinks to keep participants focused on the article without the opportunity to click away. Future research could investigate whether the inclusion of hyperlinks can help enhance comprehension. Finally, we primarily conceptualized visuals as serving a decorative role in digital journalism as seductive details by emphasizing the importance of textual information, aligning with the perspective of previous research (e.g., Greussing and Boomgaarden 2019). However, it is essential to recognize that visuals are capable of conveying information as well, and they may even become the most memorable message elements in digital longform stories. Future research should delve deeper into the cognitive and emotional impact of visual messaging and explore how visuals can be optimized to effectively deliver content and enhance the overall reader experience and comprehension.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

Barnes, S. 2023. "On the Use of Encapsulation, Parity, and Visual Storytelling in Graphic Medicine." In *Health Humanities in Application. Sustainable Development Goals Series*, edited by C. Riegel, K.M. Robinson. Cham: Palgrave Macmillan.

Barnes, S., and J. Sontag. 2024. "On the Use of Narrative Structure and Encapsulation for Health-Related Stories Contained within Digital News Packages." *Journalism Practice* 18 (3): 658–682. https://doi.org/10.1080/17512786.2022.2049011

Chandler, P., and J. Sweller. 1991. "Cognitive Load Theory and the Format of Instruction." *Cognition and Instruction* 8 (4): 293–332. https://doi.org/10.1207/s1532690xci0804_2

Cromley, J. G., B. W. Bergey, S. Fitzhugh, N. Newcombe, T. W. Wills, T. F. Shipley, and J. C. Tanaka. 2013. "Effects of Three Diagram Instruction Methods on Transfer of Diagram Comprehension Skills: The Critical Role of Inference While Learning." *Learning and Instruction* 26: 45–58. https://doi.org/10.1016/j.learninstruc.2013.01.003

Debue, N., and C. van de Leemput. 2014. "What Does Germane Load Mean? An Empirical Contribution to the Cognitive Load Theory." *Frontiers in Psychology* 5: 1099. https://doi.org/10.3389/fpsyg.2014.01099

Duffy, A. 2022. "Joining the Dots: The Literacies of Multimodal Longform Journalism." *Digital Journalism:* 1–19. https://doi.org/10.1080/21670811.2022.2069585

Dunlap, R. E., K. D. Van Liere, A. G. Mertig, and R. E. Jones. 2000. "New Trends in Measuring Environmental Attitudes: Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale." *Journal of Social Issues* 56 (3): 425–442. https://doi.org/10.1111/0022-4537.00176

Economou, D. 2006. "The Big Picture: The Role of the Lead Image in Print Feature Stories." In *Mediating Ideology in Text and Image: Ten Critical Studies*, edited by I. Lassen, J. Strunck, T. Vestergaard, 211–233. Amsterdam: John Benjamins.
Garner, R., M. G. Gillingham, and C. S. White. 1989. "Effects of "Seductive Details" on Macroprocessing and Microprocessing in Adults and Children." *Cognition and Instruction* 6 (1): 41–57. https://doi.org/10.1207/s1532690xci0601_2
Gilbert, P. 2010. "An Introduction to Compassion Focused Therapy in Cognitive Behavior Therapy." *International Journal of Cognitive Therapy* 3 (2): 97–112. https://doi.org/10.1521/ijct.2010.3.2.97
Ginns, P. 2006. "Integrating Information: A Meta-Analysis of the Spatial Contiguity and Temporal Contiguity Effects." *Learning and Instruction* 16 (6): 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001
Greussing, E., and H. G. Boomgaarden. 2019. "Simply Bells and Whistles?: Cognitive Effects of Visual Aesthetics in Digital Longforms." *Digital Journalism* 7 (2): 273–293. https://doi.org/10.1080/21670811.2018.1488598
Hiippala, T. 2017. "The Multimodality of Digital Longform Journalism." *Digital Journalism* 5 (4): 420–442. https://doi.org/10.1080/21670811.2016.1169197
Hochpöchler, U., W. Schnotz, T. Rasch, M. Ullrich, H. Horz, N. McElvany, and J. Baumert. 2013. "Dynamics of Mental Model Construction from Text and Graphics." *European Journal of Psychology of Education* 28 (4): 1105–1126. https://doi.org/10.1007/s10212-012-0156-z
Hockley, W. E. 2008. "The Picture Superiority Effect in Associative Recognition." *Memory & Cognition* 36 (7): 1351–1359. https://doi.org/10.3758/MC.36.7.1351
Holmberg N., Holsanova J., and K. Holmqvist. 2006. "Using Eye Movement Measures to Describe Readers' Visual Interaction with Newspapers." *Conference Proceedings EARLI SIG 2, Text and Graphics Comprehension, August 30 – September 1, 2006*, University of Nottingham.
Holšánová, J., N. Holmberg, and K. Holmqvist. 2005. "Tracing Integration of Text and Pictures in Newspaper Reading." *Lund University Cognitive Studies* 125: 1–19.
Jacob, R. J. 1995. "Eye Tracking in Advanced Interface Design." *Virtual Environments and Advanced Interface Design* 258 (288): 2.
Jacobson, S., M. Jacqueline, and R. E. Gutsche. 2016. "The Digital Animation of Literary Journalism." *Journalism* 17 (4): 527–546. https://doi.org/10.1177/1464884914568079
King, A. J., N. Bol, R. G. Cummins, and K. K. John. 2019. "Improving Visual Behavior Research in Communication Science: An Overview, Review, and Reporting Recommendations for Using Eye-Tracking Methods." *Communication Methods and Measures* 13 (3): 149–177. https://doi.org/10.1080/19312458.2018.1558194
Lang, A. 1990. "Involuntary Attention and Physiological Arousal Evoked by Structural Features and Emotional Content in TV Commercials." *Communication Research* 17 (3): 275–299. https://doi.org/10.1177/009365090017003001
Lee, Eun-Ju, and Y. W. Kim. 2016. "Effects of Infographics on News Elaboration, Acquisition, and Evaluation: Prior Knowledge and Issue Involvement as Moderators." *New Media & Society* 18 (8): 1579–1598. https://doi.org/10.1177/1461444814567982
Lenhard, W., U. Schroeders, and A. Lenhard. 2017. "Equivalence of Screen versus Print Reading Comprehension Depends on Task Complexity and Proficiency." *Discourse Processes* 54 (5–6): 427–445. https://doi.org/10.1080/0163853X.2017.1319653
Levin, R. C., A. K. Klevorick, R. R. Nelson, S. G. Winter, R. Gilbert, and Z. Griliches. 1987. "Appropriating the Returns from Industrial Research and Development." *Brookings Papers on Economic Activity* 1987 (3): 783. https://doi.org/10.2307/2534454
Liu, C., Mao, Z., Liu, A.-A., Zhang, T., Wang, B., and Y. Zhang. 2019. "Focus Your Attention: A Bidirectional Focal Attention Network for Image-Text Matching." *Proceedings of the 27th ACM International Conference on Multimedia*, 3–11.
Liu, T. S., Y. Liu, and C. D. Chen. 2019. "Meaningfulness is in the Eye of the Reader: Eye-Tracking Insights of L2 Learners Reading e-Books and Their Pedagogical Implications." *Interactive Learning Environments* 27 (2): 181–199. https://doi.org/10.1080/10494820.2018.1451901
Mayer, R. E., J. Heiser, and S. Lonn. 2001. "Cognitive Constraint on Multimedia Learning: When Presenting More Materials Results in Less Understanding." *Journal of Educational Psychology* 93 (1): 187–198. https://doi.org/10.1037/0022-0663.93.1.187

Merle, P. F., C. Callison, and R. G. Cummins. 2014. “How Arithmetic Aptitude Impacts Attention, Memory, and Evaluation of Static versus Dynamic Infographics in Online News.” *Electronic News* 8 (3): 177–197. https://doi.org/10.1177/1931243114557595

Paivio, A. 1991. “Dual Coding Theory: Retrospect and Current Status.” *Canadian Journal of Psychology / Revue Canadienne de Psychologie* 45 (3): 255–287. https://doi.org/10.1037/h0084295

Pavlik, J. V., and F. Bridges. 2013. “The Emergence of Augmented Reality (AR) as a Storytelling Medium in Journalism.” *Journalism & Communication Monographs* 15 (1): 4–59. https://doi.org/10.1177/1522637912470819

Petty, R. E., and J. T. Cacioppo. 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer.

Pouw, W., G. Rop, B. De Koning, and F. Paas. 2019. “The Cognitive Basis for the Split-Attention Effect.” *Journal of Experimental Psychology. General* 148 (11): 2058–2075. https://doi.org/10.1037/xge0000578

Rayner, K. 1998. “Eye Movements in Reading and Information Processing: 20 Years of Research.” *Psychological Bulletin* 124 (3): 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Reichle, E. D., K. Rayner, and A. Pollatsek. 2003. “The E-Z Reader Model of Eye-Movement Control in Reading: Comparisons to Other Models.” *The Behavioral and Brain Sciences* 26 (4): 445–476. https://doi.org/10.1017/s0140525x03000104

Rey, G. D. 2012. “A Review of Research and a Meta-Analysis of the Seductive Detail Effect.” *Educational Research Review* 7 (3): 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

Sandqvist, C. 1995. “Från 50-tal till 80-tal. Om form och innehåll i historieläroböcker. I:Läroboksspråk.” *Om språk och layout i svenska läroböcker*. Ord och stil 26. Uppsala. s. 124–188.

Schnotz, W. 2005. “An Integrated Model of Text and Picture Comprehension.” In *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer, 49–69. UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.005

Schnotz, W., and I. Wagner. 2018. “Construction and Elaboration of Mental Models through Strategic Conjoint Processing of Text and Pictures.” *Journal of Educational Psychology* 110 (6): 850–863. https://doi.org/10.1037/edu0000246

Sundararajan, N., and O. Adesope. 2020. “Keep It Coherent: A Meta-Analysis of the Seductive Details Effect.” *Educational Psychology Review* 32 (3): 707–734. https://doi.org/10.1007/s10648-020-09522-4

Sweller, J. 1988. “Cognitive Load during Problem Solving: Effects on Learning.” *Cognitive Science* 12 (2): 257–285. https://doi.org/10.1207/s15516709cog1202_4

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

Yarbus, A. L. 1967. “Eye Movement during Perception of Complex Objects.” In *Eye Movement and Vision* (B. Haigh, Trans.), edited by A. L. Yarbus, 171–211. New York: Plenum Press.

Zhao, F., W. Schnotz, I. Wagner, and R. Gaschler. 2020. “Texts and Pictures Serve Different Functions in Conjoint Mental Model Construction and Adaptation.” *Memory & Cognition* 48 (1): 69–82. https://doi.org/10.3758/s13421-019-00962-0