Nordic Journal of Language Teaching and Learning

Zooming into Digital Multimodal Reading Comprehension and Cognitive Load

Abdu Al-Kadi
University of Bergen
abdu.al-kadi@uib.no

Abstract
This study, leveled on multimodality and cognitive load, was based on an assumption that multimodal ensembles impact reading selection and comprehension. Data were collected from a sample of 40 learners at the entry-level of a License program through two multimodal reading tasks (RTs) in a workshop on digital reading at a Jordanian University. The two RTs included eight multimodal versions: five formats in RT1 and three in RT2 to assess reading selection and comprehension. Results of RT1 showed that learners who opted for reading with a few multimodal elements answered the comprehension questions relatively faster and better than those who read texts with heavy or no multimodality. Results of RT2 showed that the group with a reduced load of elements that combined moving images and soundtrack outperformed those who experienced moving images per se and the group that read a combination of all the modes in focus: writing, sound, and moving images. The findings of both RTs stand to reason for adding multimodal elements moderately to plain written texts to enhance readers' engagement in reading and comprehension. Results also suggest better reading comprehension when cognitive load is medium than when it is low or high, regardless of the mode. In closing, the study calls attention to some limitations that should be considered when interpreting the results. It also suggests topics for further research.

Keywords: *Comprehension, Digital Reading, Cognitive Load, Multimodality, Reading Selection*

1. Introduction

In a digitally saturated world, the clichéd view that foreign language learning requires exposure to the target language is complicated by the quality and quantity of exposure that digital technology has made possible (Blake, 2016). Digital technology, which has streamlined daily reading now, usher in exposure that Ting (2013) referred to as *multimodal input*. It has substantially changed learners' reading (Brun-Mercer, 2019; Kriaučiūnienė & Papaurėlytė, 2021). It involves, besides written texts, a range of oral, aural, and video materials (Blake, 2016; Parker et al., 2024). Reading such materials requires a grasp of semiotic potentials in multimodal content (Kriaučiūnienė & Papaurėlytė, 2021; Lim & Toh, 2020; Serafini, 2014). This change has been examined within the scope of learning motivation (Suwancharoen, 2016), strategies (Chou et al., 2019; Lee & Krone, 2019; Lim & Toh, 2020; Parker et al., 2024), and learners' perceptions (Alejandra et al., 2018; Anari et al., 2019; Bao, 2017; Pellicer-Sánchez, 2022; Singer & Alexander, 2017; Varaporn & Sitthitikul, 2019; Warman, 2017). Nevertheless, digital ubiquity continues to offer reading opportunities that outpace the existing digital reading research volume.

Most existing studies, based on individual views, value illustrative photos, emojis, sound, and animations thatmake text reading faster and more readily comprehensible (O'Brien & Voss, 2011; Petrová & Nemec, 2019; Serafini, 2014). Knorr (2019) argued that in a multimodal text, readers' eyes land first on the dominant element, be it an image, infographic or such like that highlights a focal

 Nordic Journal of Language Teaching and Learning, Vol. 14, No. 1, 2026. ISSN: 2703-8629. https://doi.org/10.46364/njltl.v14i1.1447

point. Multimodality supporters believe that multimodal elements mirror daily reading materials and foster comprehension (O'Brien & Voss, 2011; Petrová & Nemec, 2019; Singer & Alexander, 2017) and word retrieval paths (Sankó, 2006). Those who favor digital reading construe digital readers as faster and better comprehenders who employ efficacious strategies for skimming and scanning sizeable information and skipping irrelevant elements in the content they read (Brun-Mercer, 2019; Singer & Alexander, 2017). Based on this argument, digital reading and comprehension are positively correlated.

However, this postulation has been challenged by an oppositional stance. Several authors stand against heavy reliance on digital and multimodal reading (Brun-Mercer, 2019; Chou et al., 2019; Tran et al., 2013). Debaters of this strand assume that digital reading barely enhances comprehension and, hence, the digital reading modes, e.g., online e-books, newspapers, blogs, wikis, and emails, encourage reading at the surface with a less profound understanding of texts (Chou et al., 2019; Singer & Alexander, 2017). That is, digital readers are poor comprehenders— an argument based on concerns related to the cognitive abilities of learners to apperceive and manage sizeable digital reading materials they are exposed to (Mayer & Moreno, 2010). The advocates of this argument persist that reading digitally does not necessarily lead to thorough or substantial reading comprehension because normal individuals have a specific brain capacity for processing information. Such debaters believe that changing reading mediums does not necessarily alter reading comprehension for digital readers' brain architecture remains unchanged (Brun-Mercer, 2019; Chou et al., 2019). Digital-first readers still need frequent learning and practice in reading and reasoning comprehensibly (Meneses et al., 2018). According to Goodwin et al. (2020), screens may inhibit the sustained attention needed for reading long texts. That is, digital reading may result in slower reading because it requires more strategies and thus takes more time. Relative to that, Waxler and Hall (2011) wrote of how the brain reads, arguing that the brain changes due to incessant expansion to encompass the whole human sensory system. This argument has given rise to what counts as the digital brain, which engages with new digital devices in the screen era. The brain, before the massive digital content, interacted with printed materials and it functioned accordingly.

The two arguments boiled down into two assumptions that spur a big deal of research: (a) multimodal elements add ancillary functions to a plain text and facilitate comprehension, (b) there is no superiority of digital reading over reading traditionally. The first assumption supports the idea that those who read digitally comprehend content faster than non-digital readers. Supporters of this assumption contend that multimodal reading enhances learners' visual literacy (Suwancharoen, 2016) by facilitating words' retrieval paths and comprehension (Bao, 2017; Singer & Alexander, 2017; Varaporn & Sitthitikul, 2019; Warman, 2017). Nevertheless, this is challenged by other research findings, as in Pellicer-Sanchez (2022) which showed that the participants who read the print versions outperformed those who read digitally. Paper-oriented readers reported being less likely to multitask than those who read digitally. The second assumption encapsulates an argument that digital reading is associated with multimodal elements that add extra cognitive load. Arguably, the extra load affects sustained attention and thus has deleterious effects on comprehension (Goodwin et al., 2020; Mayer & Moreno, 2003). Hence, it is important to add to the body of previous research by diving into the impact of multimodal elements on readers' selection of texts and comprehension.

The present study intends to extend evidence of the potential of multimodal elements and their impact on reading selection and comprehension. Reading selection here is conceptualized as the learners' decision about which multimodal configuration they prefer to engage with. In the study at hand, it refers to learners' voluntary choice of one reading format among the options provided: paper-based, PDF-based, and online/URL. The study explores how digital verbal, and audio-visual elements change conventional reading selection and comprehension of learners who do not use English natively. Their processing of multimodal input involves not only coordinating multiple semiotic modes but also managing linguistic demands in a non-native language. Therefore, understanding how

multimodal load influences comprehension in this group offers direct insights into EFL reading pedagogy and the design of multimodal learning materials.

1.1 Research Questions

This study addressed the following questions:
1. Does reading mode based on text layouts affect reading selection and comprehension?
2. Does reading mode involving different cognitive loads (low, mid, and high) affect students' reading comprehension?

2. Review of literature

Reading, described as a process of word recognition and comprehension, basically includes a series of cognitive activities: "word decoding, lexical access, syntactic processing, inference generation, reading strategies…and post-reading activities" (Magliano et al., 2007, p.109). However, the technological possibilities have made reading more multimodal. Indeed, multimodality includes even the font types and letter sizes through hyperlinked texts and many more multimodal ensembles (Abraham & Farias, 2017; Petrová & Nemec, 2019). Gradually and rapidly, reading has become broadly categorized as 'print reading' vs 'digital reading' (Knorr, 2019; Pellicer-Sánchez, 2022; Petrová & Nemec, 2019; Serafini, 2014). While earlier definitions emphasized reading as word recognition and comprehension (Magliano et al., 2007), contemporary multimodality research conceptualizes reading as meaning-making across modes, and 'multimodal reading' denotes the interpretive engagement with ensembles of modes, not only written text, with different modes interacting to shape comprehension.

Previous multimodality and digital-literacy studies note that readers' format choices, such as preferring paper, screen, or image-rich layouts, reflect their expectations of ease, comfort, and cognitive load (Brun-Mercer, 2019; Singer & Alexander, 2017). Thus, learners' selection of a reading mode can itself be considered part of the reading process, influencing how they subsequently comprehend the text.

Importantly, digital reading should not be conflated with multimodal reading. Digital reading is less inclusive than multimodal reading alone (Khadka & Lee, 2019). Both print and digital reading can be multimodal. Many contemporary print genres, such as posters, textbooks, newspapers, and flyers, include rich visual, spatial, and typographic resources, digital texts may at times be predominantly verbal. Hence, multimodality is not inherent to the digital medium, and the distinction made in this study reflects the specific reading formats operationalized in the two reading tasks rather than a universal divide between print and digital reading. In this study, the term *multimodal reading* refers to the presentation of written text with static and dynamic pictures and auditory input to engage readers (Pellicer-Sánchez, 2022). In this sense, reading' is understood in a broad, multimodal sense consistent with social semiotic theory (Kress, 2010). It is not limited to the decoding of written words but the interpretive processing of meaning across modes, including images, sound, movement, and alphabetical texts. To ensure conceptual consistency, the term 'reading' is used throughout the paper to denote meaning-making across modes rather than exclusively decoding alphabetic text. Therefore, multimodal reading describes reading situations involving more than one semiotic mode. It involves attending to and integrating semiotic resources that jointly construct meaning, and written text is not assumed to have privileged status over other modes.

2.1 Multimodal Meaning-Making

Given the new modes and mediums of information retrieval and sharing, the notion of *comprehension* is increasingly discussed with reference to meaning-making. It results from not only words but also all things through which individuals make meaning or make meaning more

engaging (Kress, 2010; Lankshear & Knobel, 2006). Meaning-making, a synergy between alphabetic text and other elements in the text, is essential in digital literacy (Brosseuk & Downes, 2024; Petrová & Nemec, 2019). The digital age has inaugurated a new era of multiple forms of meaning-making. It is a truism that even younger generations are socialized in a world where communication is accomplished through several meaning-making modes, such as sending digital audio, visual, or text message on the same device (Brosseuk & Downes, 2024; Muela et al., 2024; Parker et al., 2024). Digital readers attend to images, videos, and PowerPoint slides, among others, to conceive meaning and shape perceptions of life and reality (Kress, 2010). Grounded in the socio-semiotic theory, pictorial signs such as emojis and emoticons integrated into texts impart meaning to the recipient, and the use of such semiotic resources of the twenty-first century helps individuals represent meaning in multiple modes (Kessler, 2024; Kress, 2010; Serafini, 2014).

While digital environments often foreground multimodal ensembles, print-based reading materials similarly rely on complex visual and textual arrangements. Multimodality, therefore, cannot be equated with digitization. The present study acknowledges this overlap but focuses on specific instantiations of print and digital formats used in RT1 and RT2. Multimodality researchers emphasize not only the properties of individual modes but also the relationships between them in meaning-making (Barthes, 1977; van Leeuwen, 2005). Modes may interact through elaboration, complementarity, anchorage/relay, contrast, or even contradiction, and these relationships shape how readers interpret multimodal ensembles. Importantly, images may underscore, extend, modify, or counter verbal text, and different types of images (e.g., realistic photos, diagrams, drawings) afford different meaning potentials. In this study, although a full multimodal discourse analysis is beyond the scope, the mode combinations in the first reading task and the second reading tasks (see data collection section) represent specific interaction patterns, e.g., images designed to elaborate rather than contradict text, and this contextualization informs how our findings should be interpreted.
This challenges the long-dominant view that language is the only full means of communication. Now communication is viewed as an experience of verbal and non-verbal interaction. The focus on multimodality differs from the pure linguistics focus that marked pre-multimodality research. In the current multimodality age, successful communication occurs via multimodal ensembles (Dressman, 2019). For instance, photos, music, painting, typography, and sculpture communicate messages in certain ways. From the semiotic standpoint, any symbol, image, sound or suchlike contributes to meaning-making. This idea received considerable attention and theorized ‘learning by design’ (Cope & Kalantzis, 2015; Lee & Krone, 2019), which involves learners in multimodal content creation (Diamantopoulou & Ørevik, 2022; Parker et al., 2024; Skulstad, 2023). It goes beyond conventional reading and writing literacies and brings new elements that multimodally make meaning of texts (Abraham & Farias, 2017).

Indeed, digital technology has already altered the reading paths and textual mindset, allowing reading across mediums instead of reading from papers (Parker et al., 2024). Mitsikopoulou (2022) viewed this alteration as an extension of the boundaries: from language-centered to semiotic frameworks. Petrová (2022) studied the impact of the medium (traditional, paper-based vs. screen-based) on reading and comprehension, explaining how digital technologies and readers’ experience and abilities are incorporated into textual meaning. In a similar vein, Goodwin et al. (2020) noted that reading differs based on modes or even across the tools in each mode. Other researchers persuasively argued that what matters in digital or print reading is not the medium that holds texts but their meaning (Kessler, 2024; Singer & Alexander, 2017). As mentioned above, multimodality implies a diversity of meaning conveyors, e.g., words, images, and sound – all depend on each other to create an overall integrated and holistic meaning (Dressman, 2019; Kessler, 2024; Ting, 2013). These carriers of meaning, in Serafini’s (2014) words, “involve several semiotic resources that represent and communicate meaning potentials” (p. 2).

In this sense, the multimodal elements shape the meaning of what we read and comprehend, not only what goes on in reading classes. Given the linguistic, audio, spatial, visual, and gestural elements of multimodality, it is hard to mark a sharp line between what is multimodal and non-multimodal. It is reasonable to view it as a continuum of multimodal elements. All texts can be, more or less, multimodal. Cohn (2020) viewed a website as an example of a word-heavy multimodal text, and space in a paragraph as an illustration of light modality that signifies meaning. In Cohn's words, "the space before and after the paragraph indicates that one thought is about to begin while another thought ends" (p. 31). To Cohn, even texts described as plain texts inevitably include spaces, font size, and color, which all account for multimodal meaning.

2.2 Cognitive Load and Multimedia Learning Theory

The theory of multimedia learning rests on deeper learning that involves information presented in both text and graphics, assuming the existence of two channels for learning: auditory and visual (Mayer & Moreno, 2003). When information is presented in the two channels, chances are that readers comprehend texts faster and better. The theory has five principles that work together to reduce cognitive overload. Although the two information processing channels in the human mind boost comprehension, the cognitive load demands more time to process information successfully (Mayer & Moreno, 2003). Activating multiple modes simultaneously in the reader's mind, arguably, slows down the comprehension process.

The cognitive load theory, originated in problem-solving research by John Sweller in the 1980s, stimulated a great deal of thought. The theory has been accounted for in digital reading as an instructional design theory (Mayer & Moreno, 2003, 2010). It echoes how individuals process information in short and long-term memory. It is well known that during learning, information, before passing into the long-term memory, remains in the working memory until processed sufficiently (Van Gog et al., 2010). The question is, what happens to this process when information floods through electronic vessels, primarily the Internet and web-based Apps that outpace the average capacity of the working memory?

A raft of researchers including O'Brien and Voss (2011) questioned information processing in the brain: How much of the reading do readers comprehend, and how much of this interaction is considered? Anderson (2003) and Mohseni1 et al. (2020) contend that cramping texts with additional elements such as images, hyperlinks, and soundtracks accounts for reading distortion and thus undercuts sensing large quantities of texts, which is, according to Sweller et al. (2011), a matter of cognitive overload. Assumingly, cognitive load minimizes the memory function and thereby stymies complete comprehension. Sweller et al. noted three types of cognitive loads– intrinsic, extraneous, and germane– that adversely affect comprehension. The first type refers to the complexity of new information, the second to distracting and unnecessary information, and the third to linking new and current information.

Critics of digitally mediated multimodal texts claim that extra elements in digital reading mediums maximize cognitive load (Zumbach & Mohraz, 2008). An example of unnecessary cognitive load is reading from a screen in PowerPoint presentations because of the repetition of written information, either on paper or on slides (Singer & Alexander, 2017). Zumbach and Mohraz (2008) alleged that poor material design, noisy reading ecology, and perhaps the teacher's complex instructions contribute to increased cognitive load. This is important to bear in mind for designers of teaching materials such as PowerPoint presentations: how to design them and use them to make the different modes work together for comprehension and avoid unnecessary cognitive load. Brun-Mercer (2019) and Mayer and Moreno (2010) opined that cognitive load may also result from superfluous information that technology adds to the conventional reading format. This is supported by findings from eye-tracking research on reading online, wherein texts are associated with ads and hyperlinks to related topics (Brun-Mercer, 2019; Pellicer-Sanchez, 2022). Brun-Mercer (2019) reported that even

if online readers ignore hyperlinks, they are often sidetracked by ads and additional text on the same webpage or the online text they read.

2.3 Multimodal Texts and Comprehension

Multimodal texts, defined as a configuration of linguistic, audio, spatial, visual, and gestural elements, comprise more than one mode (Kress, 2010; Petrová & Nemec, 2019; Suwancharoen, 2016). Modes can be written or spoken texts, moving or still images, sound or videos (Kress, 2010; Kriaučiūnienė & Papaurėlytė, 2021). With this concept in mind, multimodal texts may be picture books (Muela et al., 2024) or webpages (Brun-Mercer, 2019), and such texts normally have font color and sound effects, emojis, and other symbols in print and electronic layouts. Kriaučiūnienė and Papaurėlytė (2021) observed that most texts have now become a combination of "a multimodal ensemble of speech, sounds, images and moving elements on the screen" (p. 83). According to Abraham and Farias (2017), visual components such as images and diagrams are usually added to written texts for their meaning potential in supporting "deficits in print knowledge or conventions" (p. 63).

Comprehension has long been defined without considering the digital effects. It has been simply understood as the "ability to construct meaning from a given written text" (Lems et al., 2010, p.170), varying according to the reading purpose and the texts involved. The rise of multimodal reading mode has contributed to a new definition that takes in learners' background and linguistic knowledge whatever the reading mediums (paper-based or on-screen texts) (Goodwin et al., 2020; Mayer & Moreno, 2003). Multimodal reading researchers have tapped into verbal and visual elements and their effects on students' critical reading (Varaporn & Sitthitikul, 2019), comprehension (Warman, 2017), and vocabulary retention (Anari et al., 2019; Sankó, 2006). Other voices reported text-layout-based changes in comprehension (Brun-Mercer, 2019; Magliano et al., 2007). For instance, Varaporn and Sitthitikul (2019) speculated that multimodality enhances comprehension and motivates learners to read critically. In their study, learners who were tasked with multimodal activities performed better than those who performed monomodal activities per se. In another similar experimental study, Warman (2017) revealed that the group with multimodal reading outperformed the control group who experienced monomodal reading only. Likewise, Bao (2017) showed that multimodality positively affected 55 learners who underwent an experiment reading multimodal texts (over 16 weeks) compared to their counterparts (n=55) who were not exposed to the same reading materials. For EFL readers, multimodal elements may either scaffold limited linguistic knowledge or compound cognitive demands, making it essential to examine multimodality within an L2-learning context.

Along the line, Anari et al. (2019) studied the effectiveness of multimodal texts on comprehension at an intermediate level using a pre-post testing research design. Findings exhibited that multimodal reading impacted learners' performance in reading and vocabulary retention. Likewise, Baharani and Ghafournia (2015), in an experimental study with EFL learners, revealed that reading texts with multimodal elements was perceived as more entertaining. In a related context, Meneses et al. (2018) examined the impact of multimodal texts on EFL learners' reading skills and vocabulary in which low and high multimodal texts were used. Results reveal non-significant differences between experimental and control groups. At the same time, high-multimodal scaffolding texts were significant boosters of comprehension on the part of the participants who were categorized as low-skilled comprehenders.

As the multimodal approach engages learners' senses to make meaning, it is important to point out that comprehension does not ensue in isolation from readers' surroundings, which are now full of multimodality and semiotics, learners' multimodal background and knowledge shape their overall understanding of the world (meaning-making) (Cope & Kalantzis, 2015; Diamantopoulou & Ørevik, 2022; Kessler, 2024; Kress, 2010; Lankshear & Knobel, 2006; Serafini, 2014). Learners' language

skills develop over large numbers of such multimodal resources, and engagement with texts basically depends on "readers' interests, preferences, habits…context or social purpose of reading" (Kriaučiūnienė and Papaurėlytė, 2021, p. 84). To echo Lems et al.'s (2010) delineation, "when the prerequisite skills are in place … reading becomes an evolving interaction between the text and the background knowledge of the reader" (p. 170). Reading from this angle is a multifaceted process, requiring a thorough requisite that involves not only word recognition and comprehension but also coordination of words and meaning as well as motivation (Suwancharoen, 2016). Along with that, some factors might play roles in comprehension as to text size, screen resolution, backlighting, luminance contrast, prior knowledge of the participants, cultural competencies and demographics (Serafini, 2014).

In conclusion, learners' digital repertoire is foundational for understanding multimodal digital texts. Despite a wealth of prior research, the existing research volume has not kept up with the widespread availability of material in digital formats and modes. This is a driving motive to advance research on multimodal reading and its effect on reading selection and comprehension. Departing from the multimodal theory (Kress, 2010), and cognitive load theory (Sweller et al., 2011; Moreno & Park, 2010), the present study unravels the digital affordances that, assumingly, engage students in reading and enhance reading comprehension more than paper-wise reading. Drawing on these theoretical tenets, this study is leveled on multimodality and cognitive load. It is built on previous studies on multimodality (visual images, videos, and audio tracks) and how these modes assist learners in comprehending texts and, as well, increase the cognitive load that hinders successful comprehension.

3. Method

This study adopted a quantitative descriptive research approach to objectively measure reading selection and comprehension under multimodal and cognitive load conditions. Working with data that could be counted or translated into numbers, it can be described as a micro-investigation into learners' selection of texts and comprehension. It was structured around a learning situation that presents comprehension of texts with multimodal elements and potential cognitive load.

3.1 Participants

Data were collected from a sample of forty learners of the entry level of a License program at a Jordanian university where the author has undertaken a two-year fellowship. The participants' English proficiency was not formally measured beforehand but that their placement in the program and prior coursework suggest a shared entry-level EFL competence, which aligns with the purpose of the study. As first-year undergraduates, their comprehension processes reflect early-stage L2 reading development, making them an appropriate population for exploring how multimodal input supports or hinders L2 meaning-making. During their study, the participants attended face-to-face and online classes through LMS: Moodle and Teams. This cohort of undergraduates in their first year of enrollment, aged between 18 and 21, were accessed in a workshop on digital reading that the Faculty of Arts and Educational Sciences organized during their first semester of the program. In the workshop, they were tasked to read two texts: Reading Task 1 (RT1) and Reading Task 2 (RT2) to assess their reading selection and comprehension across eight multimodal versions, as explained in this section.

3.2 Data Collection

Data were collected through two RTs involving eight reading modes: five in RTs and three in RT2. The first yielded data corresponding to the first research question and the second reading to the second research question.

3.2.1 Reading Task 1

This reading activity was used to examine whether changing text layouts alter learners' selection of texts and comprehension. In this activity, a 253-word story, retrieved from an online reading resource called *English for Students* based on its Lexile Level and readability level (Fry Readability Graph), was prepared in five layouts to measure the selection of preferred reading formats and comprehension through five comprehension questions. This RT was based on the premise that while some readers engage with written text first, others give primacy to the visual elements. The paper-based plain text selection was rationalized by evidence from eye-tracking research on using images in reading that, despite defenders, learners, while reading texts with images, tend to center most of their attention on the text, not on images (Serrano & Pellicer-Sánchez, 2022. The other layouts were informed by Pellicer-Sanchez' (2022) premise that sound, pictures, and animations associated with textual content foster comprehension.

RT1 measures reading selection by documenting which of the five available formats each participant voluntarily chose before engaging with the text. The selection variable is analyzed alongside comprehension to explore how preferred multimodal layouts relate to readers' processing of the text. The illustrative images in RT1 were selected to function as elaborative cues, i.e., to support or specify elements already present in the written text, rather than to introduce contrasting or unrelated content. Thus, the image–text relationship in RT1 is one of elaboration, consistent with van Leeuwen's (2005) classification of intermodal relations. The distinctions made in RT1 between 'paper,' PDF,' and 'URL' formats should not be read as general distinctions between print and digital cultures. These categories reflect only the specific configurations used in this experiment and not inherent differences in multimodality across media.

Since the focus of this RT1 was to examine the effect of paper-oriented and PDF modes versus the URL mode on reading selection and reading comprehension, all intervening variables unrelated to the modes were held constant across the three paper-based modalities based on Cohn (2020): font type (Times New Roman), font size (12), and font color (black/white) were the same in all the three paper-based reading modes. The first is a paper-based plain text and the second is a text with a colored illustrative picture, both fit into an A4-size paper. The third is a PDF file (without an illustrative picture) fitted into a one-page PDF, the fourth is a PDF page with an illustrative colored picture, and the fifth is a URL to the text (one full window on a computer screen). As for reading digitally, because the efficacy of digital reading may be affected by screen resolution, backlighting, and luminance contrast, on top of screen size (Serafini, 2014) these were also held constant for the two (PDF vs. webpage) digital reading conditions.

3.2.2 Reading Task 2

This reading task was used to measure the participants' comprehension across three formats in which the cognitive load ranged from weak to heavy (also dubbed low vs high) multimodality, determining the order of modes: 1. images only, 2. text and sound, 3. text, image, and sound, implying varying cognitive load across the three versions— graded multimodality from weak to heavy (Figure 1). In the absence of accurate measurements of multimodality in a multimodal discourse, making the distinction between heavy and weak multimodality rather blurred. In this reading task, multimodality was based on the gradual addition of elements of the reading modes.

The text was adopted from *Our Health and Sports* website. Because health and sports are a shared topic for most of us, it was selected for this RT in three multimodal conditions. It presented to the respondents three multimodal layers. First, it was presented as moving images. Then, a deliberate

background track sound was added to the moving images. After that, the moving images with sound and text were spliced together. Although the first two RT2 versions (moving images and moving images with sound) did not include written text, they are still considered multimodal reading conditions. This follows multimodality theory, which views the processing of meaning from visual and auditory modes as part of reading-as-meaning-making, even when alphabetical text is absent (Kress, 2010; Serafini, 2014). Thus, RT2 evaluates comprehension across multimodal meaning-making conditions, only one of which includes written text.

In RT2, the sequential addition of modes (moving images → sound → written text) reflects layered multimodal relations in which each added mode contributes additional meaning rather than contradicting earlier modes. While this design does not explore all possible mode–mode relationships, it reflects one common pattern of multimodal orchestration found in educational media. Likewise, the multimodal configurations in RT2 represent particular design choices for the experimental task rather than assumptions that digital texts are uniformly more multimodal than print texts.

The length of the text in the three formats/modes extended over five minutes at normal speed. The informants' digital readership was examined by evaluating their reading comprehension. They recorded their answers to the comprehension questions using a pre-prepared answer sheet.

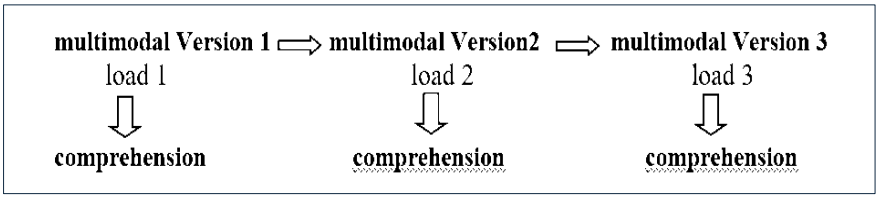

Figure 1. Variability of cognitive load in the three multimodal versions

3.3 Procedures

Before the study, approval was granted from the deanship of the Faculty of Arts and Educational Sciences on December 19, 2022, to integrate the two RTs into the workshop program. Also, psychometric features of the reading activities, their difficulty and readability levels were checked they fit university students at the entry level. In its initial form, the first reading activity was presented to three specialists in curricula, teaching methods, and educational psychology to ensure the task measures what it was set for and ensure the estimated time required for the task was sufficient. The reading was prepared as one page only because reading more than one page in PDF format or website requires scrolling up and down (i.e., flipping between pages), which would probably prolong the reading time. Its difficulty on the Lexile Level measurement was attested, and it was acceptable to determine the appropriateness of the reading materials—neither too easy nor too difficult for learners at a college-entry level. The text was also piloted on a sample of seven learners with features similar to those of the actual sample to check that the readability level is adequate for learners at the entry-level. On implementation of RT1, each participant read the text in the format they selected individually and answered the comprehension questions. During this activity, the researcher closely observed the participants' selection of texts, took notes, and set up the timer to calculate the time spent on the task. Likewise, Preparing RTs involved the consultation of a psychologist and two experts in curriculum design to check that the information load in the task suits the objectives of the study. A validity form was prepared after designing the initial draft. It was presented to three arbitrators to solicit their comments, after which some amendments were made in terms of design, order and merge of some images, and background track sound. During this activity, the researcher closely observed the participants and took notes.

Both RTs attend to comprehension across multimodal text versions. They involved five reading comprehension questions prepared along with their answer keys separately. The questions, presented in different modes (in print, PDF vs. webpage) corresponding to their reading layouts, were multiple-choice questions about gist, detail, and inference. They were scored as (5x20=100) by three qualified scorers who participated in organizing the workshop.

4. Results

Touching on reading selection and comprehension under multimodal conditions, the overall aim of the study was to explore how multimodal elements influence readers' text selection and comprehension and the extent to which multimodal elements with varying cognitive loads engage learners in reading and enhance comprehension. The data collected from RT1 and RT2 were analyzed quantitively. The numeric data represents the participants' scores on the comprehension questions and the time they spent on each reading condition. The mean scores and standard deviations were calculated using MS Excel. Results of the two RTs were assembled, presented, and discussed.

4.1 Results of RT1

Corresponding to the first research question, the RT1 results relate to the participants' selection of reading mode based on text layouts and how that affects comprehension. Based on their preferences of selection, the participants were divided into three categories: paper-based, PDF-based, and online readers. Their distribution across the five RT1 reading modes is shown in Figure 2. The comprehension outcomes and reading times for each mode are presented in Table 1. As shown, modes involving moderate multimodality (text + image) produced higher mean scores and faster reading times than plain text or heavily multimodal online versions. The paper-oriented readers read the text with and without an illustrative image accompanying the text. Again, the PDF-enabled reading represented a digital reading medium—both modes represented a weak version of multimodality. Sixteen participants (representing 40% of the sample) preferred to read the text online– the most multimodal version of the other formats in question, as online reading includes ads, hyperlinks, and other multimodal ensembles that come with online reading materials and webpages. In this task, 16 (out of 40) participants read the text online using the URL, and 13 (out of 40) read it in a PDF format—both online and pdf readers (29 out of 40) were considered digital readers (also called screen-minded readers), whereas 11(out of 40) preferred to read from papers.

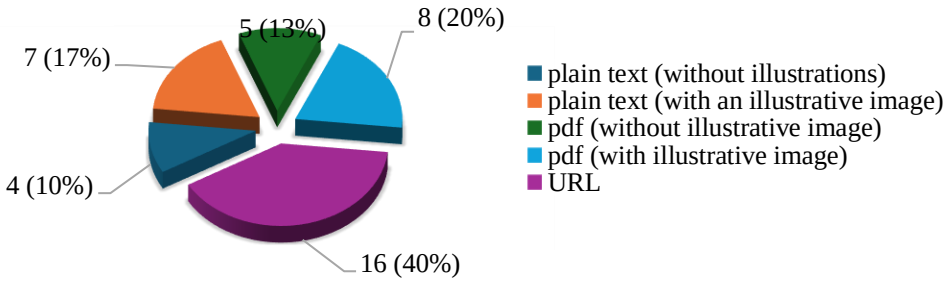

Figure 2. Participants' distribution in RT1

Regarding the time invested in reading, both online and paper-oriented reader categories finished reading the text and submitted their answers to the comprehension questions within the average time allotted to the assigned task (10 minutes). Nevertheless, only half of those who read electronically (either online or in PDF) answered all the questions correctly, i.e., their comprehension was less compared to paper-based readers' responses. Also of note, learners in the fifth category in which

readers opted for reading the text online using a URL needed more time to read the text and answer the comprehension questions. It denotes that online readers encounter content that stops on their way (e.g., ads and hyperlinks) while reading a specific text. Considering the five reading formats, it can be argued that moderate multimodality minimizes reading time and enhances comprehension whereas excess of multimodality likely distorts readers' attention.

The mean scores and standard deviations of the responses to the comprehension questions were calculated and arranged in Table 1. In the table, the highest mean scores belong to the second and fourth categories (mean scores: 83.700 and 85.500 respectively) with the least reading time (4.32/4.11 minutes). Noticeably, these two categories underwent reading versions of weak multimodality. Both included a text with an illustrative image: one was on paper, and the other in a PDF format. According to the theory of multimodality, individuals learn more from words, pictures, and other elements than from words alone.

Table 1. *Mean scores, Std., and duration of comprehension (RT1)*

	Reading Modes	Average Time	Mean scores	Std. Deviation
1	Category 1 (plain text without an image)	6.33 mins	77.800	11.272
2	Category 2 (plain text with an illustrative image)	4.32 mins	83.700	9.9000
3	Category 3 (PDF reading without image)	5.17 mins	75.700	10.1767
4	Category 4 (PDF reading with image)	4.11 mins	85.500	4.4032
5	Category 5 (online reading)	7.12 mins	74.100	8.6081

4.2 Results of RT2

Corresponding to the second research question, the second RT's results dwell on digital reading with varying cognitive loads—from weak to heavy, based on the gradual addition of elements of the reading modes (as explained above). Multimodality in the three categories (Figure 3) was determined as low, medium, and high. The participants were tasked to read the text in three ways. The first included reading moving images only; the second moving images with sound. In the third, the moving images and sound were supplemented with relevant printed text. Based on this reading, the participants were distributed into three categories (Figure 3). The majority, contrary to prior assumptions, fall in the second category; the majority managed to answer the comprehension questions in the second step—using moving images and track sound. Perhaps, the readers with moving images only (category 1) had insufficient input to digest the content, similar to readers in the first category in RT1, who opted to read plain text without any illustrations (either on paper or in a PDF format).

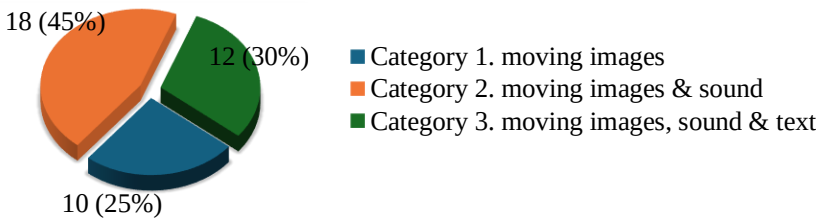

Figure 3. Participants' distribution in RT2

Figure 3 illustrates how participants performed across the three RT2 multimodal conditions, and Table 2 presents the mean comprehension scores and standard deviations for each condition. Under the reading conditions in focus, the three categories' comprehension scores were calculated and arranged in Table 2. The one-mode category (moving images only) had lower mean scores (μ=80.100) than those which involved several modes (Category 2 and 3). Nevertheless, Category 3, which involved more elements than Categories 1 and 2, also had a lower overall mean (μ=83.500). The highest overall mean score (μ=86.00) belongs to the second category, which involves moderate multimodality. Relying on these mean scores, the learners in question performed better in the second reading format which involved moderate multimodality. This mainly concurs with Kessler's (2022) advocation for minimal multimodality in reading, as it "minimally supports or supplements linguistic instruction" (p. 553).

Table 2 Mean scores of the three groups' comprehension

Modes	Mean scores	Std. Deviation
Category 1 (moving images)	80.100	6.2441
Category 2 (moving images & background music)	86.000	7.14920
Category 3 (moving images, background music, text)	83.500	6.13279

5. Discussion

The results demonstrate how varying multimodal loads influence L2 reading comprehension across controlled task conditions. Data in Table 1, for instance, suggest that adding multimodal elements to plain texts facilitates comprehension and reduces reading time, which is in alignment with Pellicer-Sanchez's (2022) contention that adding "static pictures to written texts supports reading comprehension" (p. 3). The results also align with Baharani and Ghafournia's (2015) assertion that reading words with pictures enables learners to "read content more effectively and remember it better" (p. 168). Some other voices in the literature supported using words and images complementarily, arguing that both modes, not obligatorily though, convey meaning more subtly (Knorr, 2019; Kriaučiūnienė & Papaurėlytė, 2021). Nonetheless, this should be interpreted with caution. Informed by eye-tracking research, pictures in some cases predict neither recall nor comprehension of the content (Pellicer-Sanchez, 2022). In Pellicer-Sanchez's (2022) study, most of the learners' attention was allocated to the text, not the pictures. Arguably, allocating visual attention to images does not necessarily imply fast and better comprehension.

It is important to note, however, that multimodality theory does not assume that adding modes automatically facilitates comprehension. The effects of multimodal ensembles depend on the specific intermodal relations they instantiate (e.g., elaboration, complementarity, or contrast) and on how readers integrate the modes. The findings of this study reflect only the specific image–text–sound configurations used in the two RTs and should not be generalized to all multimodal combinations.

The fifth category in RT1, which involved online reading tied with some ads, links, and other modes, is theoretically based on the cognitive load tenet (Van Gog et al., 2010) and eye-tracking research (e.g., Pellicer-Sanchez, 2022; Serrano & Pellicer-Sánchez, 2022) according to which cramming texts with multimodal elements distract readers from comprehensive understanding of reading. In this respect, Singer and Alexander (2017) argued that intensive digital multitasking results in more superficial processing of texts. Online readers, who toggle among several websites and modes (Sankó, 2006), cannot help attending to more than one window simultaneously, and the windows are of various types– images, hypertexts, audio, videos, and suchlike. Switching to multiple activities and

links during reading has been reported to lack deep thought during text processing (Brun-Mercer, 2019; Singer & Alexander, 2017) and may impair critical reading. Brun-Mercer (2019) speculated that online readers are likely diverted from their focus by ads and other elements on the same webpage, regardless of whether they click the hyperlinks.

As for the RT2 results, it makes sense that more than one mode in a reading task results in better reading comprehension, but additional elements pose a challenge to the cognitive load. For instance, images tied to texts "support comprehension and facilitate learning" (Pellicer-Sánchez, 2022, p. 2). Despite the common sense that dynamic visuals attract more attention than static images, adding images to words is assumed to be redundant because words and images convey the same content in different ways (Kress, 2010). Either way, dynamic/animations or static images, in effect, require more time to digest content. Based on the multimedia learning theory (Mayer & Moreno, 2003), the more cognitive load, the more challenging the comprehension. Pellicer-Sanchez (2022) pointed out that the visuals in a text are additional elements that consume more time for text processing. According to Mayer and Moreno (2003), extra elements to texts "increase the amount of incidental processing to the extent that the learner devotes some cognitive capacity to processing" the content (p. 45). In this regard, Pellicer-Sánchez (2022) advocates that the multiple input sources impact the underlying cognitive processes involved in the multimodal input and the attentional demands.

Tying the findings of both RTs substantiates prior investigations of multimodality and cognitive load effects on reading comprehension (Alejandra et al., 2018; Anari et al., 2019; Singer & Alexander, 2017; Varaporn & Sitthitikul, 2019; Warman, 2017). The results of RT1 and RT2 together suggest that reading comprehension is better when cognitive load is moderate than when it is low or high, regardless of the mode. Images seem to help the most, at least in RT1. Based on the cognitive load theory (Sweller, 1988), a few elements in reading are better than several multimodal ensembles because the add-ons require more time and more thinking to understand content. Hence, excessive multimodal elements (heavy multimodality) in a text, contrary to expectations, distort the focus of reading to variant topics (Sankó, 2006), especially if added to a text without a clear purpose.

If we accept the multimodality theory, it stands to reason that comprehension is a matter of meaning-making, and it can be argued that learners construct meaning out of different modes of representation (Kriaučiūnienė & Papaurėlytė, 2021; Lim & Toh, 2020; Sankó, 2006). The assumption that sound, pictures, and animations associated with a written text foster comprehension (Pellicer-Sanchez, 2022) had support in reading texts with some illustrations or modes that focus on meaning-making (Kessler, 2022, 2024; Suwancharoen, 2016). All modes, including words, have affordances for making meaning. This corresponds to the assumption that underpins a text combined with multimodal ensembles (e.g., sound, pictures, and animations) fosters comprehension. The results should therefore be interpreted with awareness that comprehension arises from how learners coordinate meaning across modes within the specific intermodal relationships present in the RTs, rather than from multimodality in an abstract sense.

Cohn (2020) opines that even in plain text, space signifies meaning. For example, when splitting paragraphs into individual units, "the space before and after the paragraph indicates that one thought is about to begin while another thought ends" (p. 31). Because multimodality is viewed as a continuum and the distinction between heavy and weak multimodality is blurred, the relatively few instances of multimodalities in the texts need more evidence to reason for a significant difference between those who read multimodally and those who read plain texts, those who read from a paper and those who read screen-mediated texts. The results should therefore be interpreted within this multimodal definition of reading, where comprehension depends on how learners integrate information across modes, not solely on written text.

Taken as a whole, the findings imply rethinking the concept of digital reading in light of emerging literacies, cognitive load, and multimodality, which may offer important insights for curriculum reform. Here, "digital reading" is understood not as the opposite of traditional print reading but as

one set of reading practices shaped by technological, multimodal, and pedagogical conditions. Thus, the implications of this study relate to these specific configurations rather than to a generalized digital-versus-print distinction.

The findings should be interpreted in the sense that multimodality characterizes both print and digital reading, and that the contrasts observed in this study arise from the specific reading configurations used in the RTs, not from inherent differences between print and digital media. Educators may draw on Mayer and Moreno's (2003, 2010) cognitive load theory, which informs teachers and course developers about the extraneous load in their teaching ecology— be it classroom or virtual online learning sessions. Materials that contain confusing and, perhaps, unnecessary information (intrinsic load) including the teachers' talk or excessive multimodal ensembles should be reduced to the extent that enables learners to digest and react. The digital generation now attends to reading digitally, which may require reading strategies that, in some way, differ from paper-based reading strategies.

The study arouses interest in and discussion of multimodal reading, requiring learners to develop strategies to comprehend multimodal ensembles and visual images. These findings must also be interpreted within an EFL context, where learners' linguistic proficiency shapes how effectively they can integrate multimodal cues; multimodality may support comprehension when linguistic load is high, but it may also compound cognitive demands for less proficient readers. These findings carry implications for EFL teaching, as they show that moderate multimodality provides effective scaffolding for L2 readers, while excessive multimodality can interfere with linguistic processing and comprehension. For EFL learners, the findings suggest that moderate multimodal scaffolds (e.g., supportive images or audio) can enhance comprehension, while high multimodal load may strain cognitive resources and reduce processing efficiency. These insights inform classroom material design and digital reading instruction.

6. Conclusion

The study examined the role of multimodal texts on learners' cognitive abilities insofar as reading selection and comprehension are concerned. To put it bluntly, the findings explain that adding multimodal elements to plain texts suggests more engagement in reading and comprehension. However, reading comprehension is better when cognitive load is medium than when it is low or high, regardless of the mode. Across both reading tasks, the strongest comprehension occurs under moderate multimodality. These empirical patterns indicate that adding appropriate multimodal elements, rather occurs or excessive ones, can meaningfully support readers' engagement and comprehension. For the field of language teaching, the study highlights that EFL learners benefit most from multimodal materials that balance visual and auditory support without overwhelming cognitive resources. These insights can guide teachers, curriculum designers, and digital-platform developers in constructing multimodal reading experiences that enhance, rather than hinder, L2 comprehension.

Impressive though this may seem, the study endured some limitations that should be pointed out. Given the short time frame of this study, as it was part of a workshop, the convenient sampling of relatively small sample size was the way to go for the data collection, and this makes inferential statistical analysis, for testing the statistical significance among different reading modes rather impractical. These limitations limit the generalizability of the findings, making them suggestive and tentative rather than conclusive. There is room for experimental research design to strengthen evidence on the impacts of multimodal input on comprehension. Future research may consider possibilities for sequencing the reading modes other than the sequence adopted in this study. More information about the participants' preferred reading formats and comprehension could be garnered

using some other tools such as interviews to consolidate findings. For measuring multimodality and cognitive load, more sensitive measurements including the task-invoked pupillary response could be adopted to measure the effects of image-rich or word-heavy multimodal texts on comprehension.

Acknowledgments

I am grateful to the Dean of the Faculty of Arts and Educational Sciences and the Head of the Department of English at PU for their cooperation in organizing the reading workshop. I wish to thank the participants whose engagement in the reading tasks contributed to generating the data for this study. I am also indebted to Prof. Sigrid Johanna Håheim Ørevik for the valuable comments and feedback on the initial draft of this paper.

Funding

No funding was received to assist with the preparation of this manuscript.

Competing interests

The author reports there are no competing interests to declare.

Data Availability Statement

The data presented in this study are not publicly available but can be retrieved from the author on reasonable request under restricted conditions.

Ethical/institutional Approval

At the time the study was conducted, there was no ethics or institutional committee in place at PU where the study was conducted. Instead, formal approval was obtained from the dean of the Faculty of Arts and Educational Sciences.

References

Abraham, P. & Farias, M. (2017). Reading with eyes wide open: Reflections on the impact of multimodal texts on second language reading. *Íkala, Revista de Lenguajey Cultura, 22*(1), 57-70.

Anari, N., Abusaeedi, A. & Shariati, M. (2019). The effects of multimodality on reading comprehension and vocabulary retention among Iranian EFL learners. *Iranian Journal of English for Academic Purposes, 8*(4), 1-16.

Baharani, B. & Ghafournia, N. (2015). The impact of multimodal texts on reading achievement: A study of Iranian secondary school learners. *International Journal of Applied Linguistics & English Literature, 4(4),* 161-170.

Bao, X. (2017). Application of multimodality to teaching reading. *English Language and Literature Studies, 7*(3), 78-84.

Blake, R. (2016). Technology and the four skills. *Language Learning & Technology, 20*(2), 129-142.

Brosseuk, D., Downes, L. (2024). Listening to teachers talk about multimodality and multimodal texts: Considerations for the national English curriculum. *The Australian Journal of Language and Literacy*. https://doi.org/10.1007/s44020-024-00064-8

Brun-Mercer, N. (2019). Online reading strategies for the classroom. *English Teaching Forum, 57(4)*, 2-11. ERIC Number: EJ1236175

Chou, H., Chao, C. & Tu, C. (2019). An experimental study of digital reading experiences. *Asian Journal of Education and E-Learning*, *7*(2). 56- 62.

Cohn, J. (2020). Understanding visual rhetoric. In C. Lowe & P. Zemliansky (Eds.), *Writing spaces: Readings on writing* (pp.18-39). Parlor Press.

Cope, B. & Kalantzis, M. (2015). The things you do to know: An introduction to the pedagogy of multiliteracies. In B. Cope & M. Kalantzis, (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp.1-36). Palgrave Macmillan.

Diamantopoulou, S. & Ørevik, S. (2022). Multimodality in English language learning: The case of EAL. In S. Diamantopoulou & S. Ørevik (Eds.), *Multimodality in English language learning* (pp. 3-15). Routledge.

Dressman, M. (2019). Multimodality and language learning. In M. Dressman & R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 39-55). DOI: 10.1002/9781119472384.ch3

Goodwin, A.; Cho, S.; Reynolds, D. & Brady, K. (2020). Digital versus paper reading processes and links to comprehension for middle school students. *American Educational Research Journal, 57*(4). DOI: 10.3102/0002831219890300

Khadka, S. & Lee, J. C. (2019). Extending the conversation theories, pedagogies, and practices of multimodality. In S. Khadka & J. C. Lee (Eds.), *Bridging the multimodal gap: From theory to practice* (pp.3-13). University Press of Colorado, Utah State University Press.

Kessler, M. (2024). *Digital multimodal composing: Connecting theory, research and practice in second language acquisition*. Multilingual Matters.

Kessler, M. (2022). Concepts in ELT: Multimodality. *ELT Journal*, 552-554.

Knorr, A. (2019). Infographics: A powerful combination of word, image, and data. In L. Sabatino & B. Fallon (Eds.), *Multimodal composing: Strategies for twenty-first-century writing consultations* (pp. 97-109). Utah State University Press.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kriaučiūnienė, R. & Papaurėlytė, P. (2021). On the issue of reading multimodal texts. *International Journal of Multilingual Education,*17, 84-104.

Lankshear, C. & Knobel, M. (2006). *New literacies: Everyday practices and classroom learning* (2nd ed.). Open University press.

Lee, S. & Krone, J. (2019). Brochures: Helping students make good design decisions. In L. Sabatino & B. Fallon (Eds.), *Multimodal composing strategies for twenty- first-century writing Consultations* (pp. 51-66). Utah State University Press.

Lems, K., Miller, D. L. & Soro, M. T. (2010). *Teaching reading to English language learners*: Insights from linguistics. The Guildford Press.

Lim, F. V. & Toh, W. (2020). How to teach digital reading? *Journal of Information Literacy, 14*(2), 24-43.

Magliano, J. P., Millis, K. Ozuru, Y. & McNamara, D. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107-136). Lawrence Erlbaum Associates publishers.

Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52

Mayer, R. E. & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno & R. Br¨unken (Eds.), *Cognitive load theory* (pp. 131-152). CUP.

Meneses, A., Escobar, J. & Véliz, S. (2018). The effects of multimodal texts on science reading comprehension in Chilean fifth-graders: Text scaffolding and comprehension skills, *International Journal of Science Education, 40*(18), 2226-2244, DOI: 10.1080/09500693.2018.1527472

Mitsikopoulou, B. (2022). Integrating multimodal and digital literacies in foreign language policy design: Two examples from the Greek context. In S. Diamantopoulou & S. Ørevik (Eds.), *Multimodality in English language learning* (pp. 51-66). Routledge.

Moreno, R. & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno & R. Br¨unken (Eds.), *Cognitive load theory* (pp. 9-28). CUP.

Muela, D., Tabernero, R. & Hernández, L. (2024). Investigating primary pre-service teachers' perceptions of nonfiction picturebooks. *The Australian Journal of Language and Literacy*. https://doi.org/10.1007/s44020-024-00065-7

O'Brien, D., & Voss, S. (2011). Reading multimodally: What is afforded? *Journal of Adolescent and Adult Literacy, 55*(1), 75-78. https://doi.org/10.1598/JAAL.55.1.9

Parker, M., Mantei, J. & Kervin, L. (2024). Examining the opportunities for digital text production in the Australian Curriculum for the first years of school. *The Australian Journal of Language and Literacy*. https://doi.org/10.1007/s44020-024-00061-x

Pellicer-Sánchez, A. (2022). Multimodal reading and second language learning. *ITL - International Journal of Applied Linguistics, 173*(1), 2-17. DOI: https://doi.org/10.1075/itl.21039.pel

Petrová, Z. & Nemec, R. (2019). Changing reading paths in a digital age: What are the consequences for meaning-making? *Journal of Pedagogy,10*(2) 65-85. https://doi.org/10.2478/jped-2019-0007

Petrová, Z. (2022). Medium-specific aspects of digital reading and their impact on reading comprehension. *Human Affairs*, *32*(2), 134-144. https://doi.org/10.1515/humaff-2022-0011

Sankó, G. (2006). The effects of hypertextual input modification on L2 vocabulary acquisition and retention. In M. Nikolove & J. Horváth (Eds.), *University of Pécs Roundtable 2006: Empirical studies in English applied linguistics* (pp. 157-178). Pécs, Hungary: Lingua Franca Csoport

Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. Columbia University, Teachers College Press.

Serrano, R. & Pellicer-Sánchez, A. (2022). Young L2 learners' online information processing in a graded reader during reading-only and reading-while-listening conditions: A study of eye-movements. *Applied Linguistics Review*, *13*(1), 49-70. https://doi.org/10.1515/applirev-2018-0102

Singer, L. M. & Alexander, P. A. (2017). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education, 85*(1), 155-172. DOI: 10.1080/00220973.2016.1143794

Skulstad, A. S. (2023). *Text creation in English teaching*. Universitetsforlaget.

Suwancharoen, P. (2016). *The effects of multimodal texts on EFL students (MA thesis),* Srinakharinwirot University.

Sweller, J. Kalyuga, S. & Ayres, P. (2011). *Cognitive load theory*. Springer.

Ting, K. (2013). Input in multimodality for language learning. *The International Journal of Literacies, 19*(1), 47-56. DOI: https://doi.org/10.18848/2327-0136/CGP/v19i01/48795

Tran, P., Carrillo, R. & Subrahmanyam, K. (2013). Effects of online multitasking on reading comprehension of expository text. *Journal of Psychosocial Research on Cyberspace, 7*(2). doi: 10.5817/cp2013-3-2

Van Gog, T., Paas, F. & Sweller, J. (2010). Cognitive load theory: Advances in research on worked examples, animations, and cognitive load measurement. *Educational Psychology Review, 22(4),* 375-378.

van Leeuwen, T. (2005). *Introducing social semiotics: An introductory textbook*. Routledge.

Varaporn, L. & Sitthitikul, P. (2019). Effects of multimodal tasks on students' critical reading ability and perceptions. *Reading in a Foreign Language, 31(1), 81-108.*

Warman, L. A. (2017). *The effect of multimodal texts on students' reading comprehension.* Proceedings of the UR International Conference on Educational Sciences, 437-443. https://ices.prosiding.unri.ac.id/index.php/ICES/article/view/4808

Waxler, R. P. & Hall, M. P. (2011). *Transforming literacy: Changing lives through reading and writing*. BRILL.

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypertext reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, *24*, 875–887.