

Multimodal scaffolding in critical reading: Pedagogical insights from Indonesian EFL learners

Arum Puspitasari¹, Moh. Kholilurrahman Jailani², Yazid Basthomi³, Suharyadi⁴,
Evy Nurul Laily Zen⁵

^{1*} English Department, Faculty of Letters, Universitas Negeri Malang, Malang, Indonesia

*Corresponding author: arum.puspitasari.fs@um.ac.id

ARTICLE INFO**Article History:**

Submitted: 4 November 2025

Revised: 10 December 2025

Accepted: 11 December 2025

Published: 12 December 2025

Keywords:

critical reading, EFL,
multimodality, scaffolding,
undergraduates

ABSTRACT

Scaffolding through multimodality is increasingly popular in English as a Foreign Language (EFL) settings, providing new opportunities for critical reading proficiency creation with visual, audio, and text resources. Thus, this study examines the phenomena of multimodal scaffolding among Indonesian EFL undergraduate learners in a tertiary reading course, the critical reading course. The study used a descriptive qualitative research design, collecting data from 29 English Language Teaching (ELT) major students through reflective journals, focus group discussions, and classroom observation notes to gain in-depth insights into their experiences and perspectives. Through thematic analysis approach, several themes were found from participants' reflections and discussions: (1) multimodal texts promoted student engagement and sustained interest; (2) visual and audio elements facilitated comprehension of abstract/intricate concepts; and (3) the non-linear composition of multimodal texts complicated multimodal analysis due to limited exposure. The findings reveal that multimodality serves as both an effective critical reading scaffold and a cognitive complication for interpretation, meaning that multimodal literacy must be explicitly taught. Ultimately, this study contributes to the research gap in critical literacy and multimodal literacy in the EFL higher education context with practical pedagogical recommendations based on the learner perspective.

Copyright © 2025, Puspitasari et al.
This is an open access article under the CC-BY-SA



How to cite:

Puspitasari, A., Jailani, M. K., Basthomi, Y., Suharyadi., Zen, E. N. L.(2025). Multimodal scaffolding in critical reading: Pedagogical insights from Indonesian EFL learners. *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics*, 12(2), 924-947. <https://doi.org/10.22219/celtic.v12i2.42721>

INTRODUCTION

The sociocultural landscape of Teaching English as a Foreign Language (TEFL) has been transformed by the tremendous developments in digital technology and the growing need for multimodal communication. According to Stewart et al. (2025), where traditional literacy relies on print-based texts, multimodal texts employ visual, auditory, linguistic, and spatial resources to compose meaning. This complexity requires students to expand their awareness of symbols from linguistic-only to semiotic on multiple levels (Lam & Putri, 2024). In higher education contexts, such awareness is encouraged - or even required - as students are anticipated to critically navigate diverse texts, engage with global discourses and knowledge production in multimodal contexts (Archer, 2022).

On another point, the increasingly interconnected world demands multiliteracy is as a component of successful information exchange. For example, the CEFR has been extended beyond linguistic literacies to involve multimodal communicative competences. Similarly, the 21st-century skills framework elevates digital, critical, and creative literacies as essential components of lifelong learning (Cirocki et al., 2024; Kaowiwattanakul, 2025). In Indonesia, for instance, the Merdeka Belajar-Kampus Merdeka (MBKM) policy stresses the same ideals in higher education, where the institution must provide spaces for innovation, critical thinking, and digital literacy. Thus, multimodal pedagogy becomes a curricular expectation rather than an instructional supplement and increasingly interdisciplinary research and teaching focus in preparation for the knowledge economy (Handrayani et al., 2024).

Multimodality has become an important theoretical lens in contemporary literacy research, emphasizing that meaning-making does not rely solely on linguistic symbols. As explained by Kress (2010), multimodality encompasses the use of multiple semiotic modes—linguistic, visual, auditory, gestural, spatial, and others—in the process of constructing meaning. Building on this foundation, Kress and van Leeuwen (2006) developed the multimodal discourse analysis (MDA) framework, which examines how these modes interact in communication events and semiotic practices. Developing this idea into an educational context, the multiliteracies framework introduced by the New London Group (1996) encourages pedagogical approaches that are appropriate to the plural, multimodal, and global nature of 21st-century learning.

Building on this foundation, recent research has begun to explore how critical literacy can be strengthened through multimodal frameworks, particularly in supporting the needs of diverse learners. Stewart et al. (2025), for example, proposed an intermodal complementarity framework that helps learners understand how a text conveys criticism through multiple semiotic lenses. Their findings suggest that both persuasive and informative meanings often emerge from the interaction between modes. Thus, multimodal competence is not simply an additional skill but has become an essential

part of academic literacy—equal in importance to advanced English language skills for EFL (English as a Foreign Language) students. It is within this context that this study examines how multimodal scaffolding can support the development of critical reading skills in Indonesian students, with subsequent analysis linking this theoretical foundation to the learners' perspectives and experiences.

Therefore, multimodal research in the past decade has provided substantial evidence for its pedagogical benefits. For example, a multimodal approach enhances comprehension, increases engagement, and scaffolds the meaning-making process, drawing on learners' semiotic repertoires (Lee et al., 2021). More specifically, Carcamo and Pino (2025) found that infographics promoted students' multimodal literacy and reflective literacy when implemented in university classrooms. Similarly, Lam and Putri (2024) considered digital multimodal composing an affordance for critical literacy through creative rewriting. Thus, Stewart et al. (2025) argue that a framework for intermodal complementarity empowers critical reading of multimodal texts beyond solely critiquing visuals. Thus, learning to make meaning from visual-textual arguments transcends simple comprehension toward critical and civic literacies.

Research consistently shows that multimodal pedagogy can improve comprehension, vocabulary development, and motivation in the classroom. Chun (2016) demonstrated how visual and auditory elements can strengthen text comprehension. Similar findings were reported by Lim and Polio (2020), who found that written input is more effective when supported by visual and auditory elements. More recently, Carcamo and Pino (2025) found that the use of infographics can improve EFL students' multimodal literacy while also stimulating reflective thinking. Similarly, Tandiana et al. (2023) demonstrated that multimodal materials help learners navigate complex textual features and provide clear pedagogical benefits.

However, multimodality also brings challenges. Several mixed-methods studies have shown that learners often struggle with the nonlinear and layered nature of multimodal texts. Handrayani et al. (2025) found that learners require more explicit scaffolding to effectively comprehend multimodal input. Ngui (2024) also cautions that a lack of guidance can actually hinder, rather than enhance, comprehension. Therefore, multimodal learning presents both opportunities and challenges. With growing evidence that Indonesian EFL students can benefit from multimodal engagement, it is important to examine how they perceive multimodal scaffolding in reading courses.

Further recent findings in the Indonesian TEFL context have also stressed multimodal instruction. For example, Wardani and Kurniasih (2023) found that learners improved narrative comprehension through visual and verbal multimodal resources - although their implementation was restricted to preliminary reading stages - and Akbar et al. (2025) support higher-order thinking EFL reading development but require more

integrated support structures. Thus, while international findings champion multimodality as a means to improve reading engagement and comprehension, they imply that both semester-long exposure and specific university-level reading efforts remain unexplored regarding how students navigate complex meaning-making within multimodality.

Critical reading is not simply understanding the content of a text; it involves analyzing, evaluating, and questioning what is read (Wallace, 2003). In an EFL context, these skills become more challenging because students must encounter texts originating from different cultural traditions or discourses, requiring more intensive instructional mediation. Le et al. (2024) examined how undergraduate students in Vietnam used critical reading strategies when interacting with multimodal texts. Their findings indicate that although students successfully developed new interpretive perspectives, their critical evaluation skills were highly dependent on the form of scaffolding and prior learning experiences.

Thus, critical reading is a complex skill that goes beyond literal comprehension. This potential can be strengthened through the use of multimodal texts, which provide multiple pathways for constructing meaning but, at the same time, can be overwhelming for students who lack strategies for synthesizing diverse semiotic cues. Lam and Putri (2024) demonstrated that rewriting digital fairy tales can foster critical thinking when students are asked to reimagine multicultural narratives, rather than simply translate them. In another study, a multilingual digital multimodal composition project developed by Özer (2025) successfully fostered civic engagement among university students. Overall, these studies confirm the potential of multimodal pedagogy in fostering critical literacy. However, studies that highlight how undergraduate students perceive these multimodal scaffolds in the context of higher education, especially in Indonesia, are rare.

Most empirical research linking English as a Foreign Language (EFL) education with multimodality is still dominated by quantitative approaches, either through experimental designs or survey-based analyses. Yawiloeng (2022) reported that students had very positive perceptions of the use of multimodal texts. However, because the research was primarily survey and anecdotal, insights into students' learning experiences were quite limited. Le et al. (2024) found measurable improvements in critical reading strategies after implementing multimodal instruction, but the study relied primarily on test results rather than qualitative findings that could have revealed learners' thinking processes and experiences in greater depth.

Contemporary qualitative research has begun to explore students' experiences developing multimodal literacy in English as a Foreign Language (EFL) contexts. Employing systemic functional linguistics and visual grammar approach, Emiliasari and Emilia (2025) uncover the potential and limitations of multimodal scaffolding.

Meanwhile, Mejía-Laguna (2022) analyzed the multimodal composition of college students in Latin America as a form of sociocultural identity expression, although these findings are more relevant to school-age students than to university students.

Overall, these studies indicate a gap in the literature: research at the higher education level rarely highlights the voices of students themselves. Further studies are needed to explore how undergraduate EFL students experience and interpret multimodal scaffolding in reading courses, which could provide a more nuanced understanding of critical literacy development in the Indonesian higher education context.

However, gaps remain in the current literature. First, much of the existing research stems from quantitative approaches that rely on tests and measures like comprehension scores or vocabulary gains that obscure students' lived experiences. For example, Yawiloeng (2022) examined EFL students' perceptions of multimodal texts in positive terms through survey analysis; however, such qualitative measures only briefly analyzed external student-reported viewpoints. Therefore, more nuance is needed regarding students' difficulties navigating multimodal frameworks. Second, while broader multimodal literacy has been explored at length in Western studies and East Asia, Southeast Asian undergraduate qualitative inquiries remain limited. Studies rooted in systemic functional linguistics and visual grammar have been conducted in Indonesia and typically emphasize how teachers mediate multimodal texts (Emilia & Emilia, 2025). Yet, research that foregrounds students' own perspectives is still scarce. Similarly, while Dauvlay and Dewi (2025) highlight the influence of multimodal literacy on reading comprehension, their work focuses more on overall classroom performance than on the nuanced difficulties students encounter when interpreting multimodal cues.

Therefore, this study contributes to the study of multimodality and applied linguistics by investigating how undergraduate students in Indonesia perceive and interact with multimodal materials. By bridging the gap between theory and classroom practice, this study seeks to offer empirical insights into the benefits and challenges students experience when using multimodal scaffolds to support critical reading in real academic contexts. This is concerning since studies show students find it difficult to understand multimodal features because they are often non-linear. Even if students find multimodal texts a motivating option for intersectional literacy (Le et al., 2024), the learning potential of many semiotic resources - they feel overwhelmed at the cognitive task potential of simultaneously navigating modes. Without insights into learner perspectives, especially how comprehending multiple semiotic resources may burden or benefit them, educators risk merely imposing projects without understanding genuine difficulties.

Despite growing support for multimodal pedagogy, the existing literature remains inequitable: the benefits of this approach have been widely recognized, but attention

to the perspectives of EFL students, the group most impacted by it, remains very limited. Recent studies have begun to reveal how students make sense of complex and abstract concepts through nonlinear forms of multimodal learning, signalling a shift toward more participatory and learner-centered pedagogies.

Mavengere et al. (2024) assert that digital multimodality and nonlinear learning pathways enable learners to synthesize more complex knowledge structures than conventional textual sequencing. Similarly, Gupta and Arora (2024) emphasize that the future of multimodal pedagogy lies in epistemic co-design, a co-design process in which learners actively shape multimodal spaces that support the development of their abstract understanding. Overall, these studies reframe multimodal pedagogy not as top-down instruction, but as a learner-driven, exploratory process that values the complexity of graduates' cognition and the diversity of their learning trajectories.

Therefore, this study seeks to fill this gap by placing student perspectives as the primary focus, so that the development of multimodal pedagogy can grow through, rather than simply address, students' experiences and voices in the context of higher education in Indonesia. This study investigates how Indonesian EFL undergraduates perceive, use, and reflect on multimodality during multimodal texts in a reading course as a scaffolding tool for critical readings while recognizing the challenges students encounter due to such expectations. To this end, this study aimed to investigate EFL undergraduates' perceptions of multimodal texts in developing critical reading skills, to assess how such texts aid in understanding abstract and complex ideas, and to identify difficulties students experience when dealing with non-linear text structures.

RESEARCH METHODOLOGY

A descriptive qualitative research design was implemented in this study as it better suits exploring learners' voices for an in-depth understanding instead of numerical findings. Through a qualitative lens, researchers explore participants' experiences and meaning-making within real-world contexts (Creswell, 2018).

Research Participants

The participants of this study consisted of 29 undergraduate students majoring in English Language Teaching at Universitas Negeri Malang, Indonesia. All participants were in their third semester enrolling in a Critical Reading course. They were selected through purposive sampling since they've learned about reading strategies for academic purposes but have yet to be exposed to multimodal texts. Participation was voluntary, and their consent was obtained upon collecting data.

Research Data and Instruments

Three data sources were used to triangulate findings:

First, reflective journals that the students wrote weekly about their personal experiences and perceptions regarding multimodal reading activities. Students wrote weekly reflective journals to record their experiences, emotions, and insights while engaging in multimodal reading activities. Each journal followed a semi-structured format adapted from Nguyen (2019), consisting of questions about (a) learning motivation, (b) comprehension strategies, and (c) perceived challenges. This format was chosen because it allowed students to express ongoing reflections in their own words while maintaining consistency across entries.

Second, focus group discussions (FGDs) were held twice in the semester, involving 6-7 students per FGD, to capture overall perceptions and common issues. Focus group discussions (FGDs) were conducted twice during the semester (16 weeks)—the first in week 8 (mid-semester) and the second in week 16 (end of semester)—to capture initial perceptions and changes following continued exposure to multimodal reading materials. Each session involved 12 participants, divided into two smaller groups of six, evenly divided based on proficiency level. Conducting two FGDs allowed the researchers to observe the development of student engagement and understanding over time. The group size and structure followed the qualitative research guidelines proposed by Krueger and Casey (2015), which recommend 5–7 participants per group to ensure in-depth interactions and equal opportunity for each participant to contribute.

Third, classroom observations were conducted four times during the semester (16 weeks)—twice in the first half (weeks 4 and 8) and twice in the second half (weeks 12 and 16)—to monitor how students interacted with multimodal texts over time. The observation process used a structured checklist adapted from Creswell and Poth (2018) and Walsh (2010), focusing on indicators such as duration of assignments, collaboration between students, utilization of multimodal materials, and level of behavioral engagement during reading activities. Each observation session lasted 90 minutes and was conducted directly by the researcher, who also took notes to capture emerging contextual dynamics and nuances. The four observations allowed for data triangulation and longitudinal monitoring of changes in student behavior as they became more familiar with the multimodal learning environment, while ensuring consistency with the timeline outlined in the research procedures.

The use of reflective journals, focus group discussions, and classroom observations followed established qualitative research guidelines (Nguyen, 2019; Krueger & Casey, 2015; Creswell & Poth, 2018; Walsh, 2010), ensuring methodological rigor and triangulation across data sources.

Research Procedure

The research procedure took place during a 16-week semester, where multimodal texts (infographics, videos, and interactive articles) were blended with reading lessons in addition to print-based texts, emphasizing the need for multimodal resources for the reading course. The students were tasked with weekly reflections on how these resources supported their comprehension processes or not. FGDs were conducted in the semester's middle part and toward its end. Observations were made during lessons where multimodal materials were most extensively used.

Data Analysis

Qualitative analysis was conducted on all qualitative data. Thematic analysis (Braun & Clarke, 2006) was used across six steps for theme development: 1) familiarization with data, 2) initial coding, 3) development of themes, 4) reviewing themes, 5) defining and naming themes 6) write up. Codes were generated inductively from the data before being adjusted based on relevant perspectives on multimodal resources and critical reading. To enhance trustworthiness, peer debriefing and member checking were used.

RESULTS

This study identified three interrelated themes through a systematic thematic analysis of multiple data sources, including reflective journals ($n = 24$), focus group discussions ($n = 2$ groups), and structured classroom observations ($n = 4$ sessions). These three themes emerged inductively and are presented in the chronological order of data collection and their occurrence in each data source. Table 1 summarizes the thematic framework and the empirical evidence supporting each theme.

Table 1. Thematic framework and data source triangulation

| Theme | Primary Data Sources | Supporting Evidence | Temporal Emergence |
|---|--|-----------------------|--------------------|
| Multimodal texts increased motivation and engagement | Reflective journals (RJ), Focus groups (FG) | 18/24 RJs; All 3 FGs | Weeks 5–8 |
| Visual and auditory elements supported comprehension of complex ideas | Interviews (INT), Classroom observations (OBS) | 10/12 INTs; 14/16 OBS | Weeks 9–12 |

| | | | |
|--|---|--------------------|-------------|
| Students struggled with non-linear multimodal structures | Focus groups (FG), Classroom observations (OBS) | 2/3 FGs; 12/16 OBS | Weeks 13–16 |
|--|---|--------------------|-------------|

Theme 1: Multimodal texts increased motivation and engagement

Analysis of initial reflective journal entries (Weeks 3-4) revealed a consistent pattern of increased motivation when students interacted with multimodal texts compared to traditional printed reading materials. Eighteen of the 24 participants (75%) explicitly mentioned increased enjoyment, reduced boredom, and improved concentration as a result of using multimodal texts. The integration of visual, auditory, and textual elements appeared to create multiple entry points for constructing meaning, thereby reducing the affective barriers that commonly arise in L2 reading activities.

Data from focus group discussions also showed a similar pattern. As expressed by Participant 2 from Discussion Group 1 (FG1-P2):

"Watching the video before reading was engaging enough for me to continue with the text and helped me build prior knowledge about the content."

The excerpt illustrates how multimodal pre-reading scaffolds can activate prior knowledge and foster intrinsic motivation to engage further with written text. To maintain clarity, only the most representative quotes are presented. Other participants expressed similar ideas in their reflection journals (RJ4, RJ6, RJ8), further strengthening this pattern of findings.

Classroom observations provided additional triangulation through behavioral evidence collected using a structured observation checklist adapted from Creswell and Poth (2018) and Walsh (2010). This checklist recorded several indicators, such as the amount of time students remained on task (an average increase of 12.3 minutes per session), voluntary participation in class discussions (an average of 3.2 contributions per student, compared to 1.4 in the printed sessions), and a decrease in disruptive behavior (a 34% decrease compared to the printed sessions).

Visual and auditory elements supported comprehension of complex ideas

Results of focus group discussions (FGDs) (n = 12 participants) and classroom observations (n = 14 sessions, weeks 6–10) indicate that visual and auditory support plays a crucial role in facilitating students' understanding of conceptually dense and linguistically complex material. Participants consistently described how the use of images, animations, infographics, and narratives helped them build stronger mental representations of previously difficult-to-understand abstract concepts.

Focus group data revealed how visual and auditory elements supported students' meaning-making processes. As reflected by participant S5:

"Images and audio help me understand what is being discussed in the text. I can visualize the situation more clearly."

Participants reported that visual representations helped externalize complex relationships, thereby reducing the cognitive load that arises from maintaining multiple interrelated ideas in working memory while simultaneously processing linguistic input.

Classroom observations reinforced these findings through systematic comparisons of comprehension-checking behavior, recorded using a structured observation checklist adapted from Creswell (2018). During the multimodal text sessions, students asked significantly fewer clarification questions to the instructor (an average of 2.1 questions per session) than during the print-only sessions (an average of 5.8 questions per session; $t(14) = 4.23$, $p < .001$). This indicates an increase in independent comprehension. Furthermore, post-reading comprehension tests—administered using a researcher-developed comprehension assessment aligned with Bloom's taxonomy—also showed improved performance on inference-making and main idea identification tasks when the text included integrated visual and auditory elements (an average score of 78.4% compared to 64.2% with print-only; Cohen's $d = 0.82$).

Theme 2: Students experience difficulty with non-linear multimodal structures

Although multimodality has been shown to increase motivation and comprehension, focus group discussions conducted near the end of the semester (Weeks 11–14) and classroom observations indicated that students faced significant challenges when dealing with multimodal texts that were non-linear, multi-layered, and had numerous links. Eight participants (67% of the focus group participants) reported experiencing cognitive overload, navigational confusion, and difficulty determining the best path when reading complex multimodal texts.

The focus group data illustrated these challenges:

"Sometimes I don't know where to start. I get confused when the text has too many sections to click on." (FG2-P3)

This excerpt demonstrates the navigational uncertainty faced by students when reading texts that present multiple modes simultaneously, embedded links, interactive elements, and non-sequential presentation structures. Unlike traditional linear texts that have a clear flow from beginning to end, digital multimodal texts often do not provide explicit guidance on navigational sequences. As a result, students are forced to make numerous decisions while reading, so much of their cognitive capacity that should be used for understanding the content is instead used to navigate.

Classroom observations documented behaviors reflecting this difficulty. Field notes indicated that some students spent significant time (4–12 minutes) simply moving

between elements without a clear strategy, frequently returning to previously read sections, or even abandoning exploration due to feeling overwhelmed by the sheer number of navigation options. Screen recordings revealed fragmented navigation patterns, characterized by rapid clicking from one element to another without deep engagement with any particular component.

Focus group discussions and reflective journal entries, gathered through a semi-structured interview protocol and a guided reflective journal template adapted from Creswell (2018), provided additional insights into the cognitive load experienced by students. Several students reported feeling “overwhelmed” (n = 4) and “confused” (n = 5) when faced with texts that combined extensive written content, embedded videos, animated graphics, and numerous links within a single view. Several participants also explicitly stated that they needed clearer guidance for navigating these texts, including concrete examples from the instructor or a more structured reading flow to help them understand the complex, multimodal material.

Theme 3: Cognitive load and navigation challenges in context

In the context of multimodal texts, the need to simultaneously process information from multiple modes, integrate it into a coherent mental representation, and make navigational decisions can easily exceed learners' cognitive capacities—especially for those with limited English proficiency or underdeveloped digital literacy.

Research in the Indonesian EFL context also demonstrates similar challenges. Syafii (2021) noted that although visual strategies can improve comprehension, students initially struggle to organize and navigate complex multimodal information without clear guidance. These findings emphasize the importance of systematic instruction to help students understand non-linear text structures, especially for those unfamiliar with digital multimodal reading practices.

Similarly, Wardani and Kurniasih (2023) explain that although multimodal texts tend to enhance comprehension, careful design remains crucial to prevent students from experiencing cognitive overload. They emphasize that “teachers implement multimodal texts to help students understand narrative reading texts through a gradual and structured approach that slowly introduces multimodal complexity” (p. 151). This finding underscores the crucial role of teachers in mediating students' navigation of multimodal texts, which are both rich in support and challenging.

This research data suggests that without explicit strategy guidance, the cognitive demands of non-linear multimodal texts can outweigh their benefits, leading to shallow engagement and suboptimal comprehension of the material. This highlights the urgent need for explicit multimodal literacy instruction to help students develop metacognitive strategies, such as planning navigational paths, integrating

understanding across modes, and disseminating the relevance of the multimodal elements they encounter.

DISCUSSION

Three themes emerging from this study collectively highlight the dual nature of multimodality in the context of English as a Foreign Language (EFL) reading: multimodal texts can be both a source of motivational support and a cognitive challenge. This section interprets these findings by linking them to theoretical frameworks and previous research, discusses their implications for EFL pedagogy, and offers directions for further research.

The first theme highlights learners' affective engagement, demonstrating that multimodal scaffolding not only helps maintain attention but also makes the learning process more enjoyable and encourages students to continue trying to understand challenging texts. Participants described how visual and auditory elements made reading activities more lively and less monotonous, creating a learning environment that is more conducive to deep comprehension.

This finding aligns with Deci and Ryan's (2000) Self-Determination Theory, which asserts that learning materials that support autonomy, competence, and social connectedness can increase intrinsic motivation. In the context of this study, multimodal texts appear to meet these psychological needs by providing multiple pathways to understanding information, thereby increasing students' confidence and willingness to continue trying to understand difficult texts.

Research in the Indonesian EFL context also corroborates these findings. Wardani and Kurniasih (2023) found that multimodal texts helped Indonesian students overcome boredom in traditional reading activities and significantly increased their interest and motivation in comprehending narrative texts. The study, involving 32 EFL students, showed that students could "learn and comprehend narrative reading texts more easily and effectively through the use of multimodal texts" (p. 152). Similar findings emerged from Jamil and Abdul Aziz's (2021) study in Malaysia, which reported that multimodal texts strengthened students' reading habits by increasing engagement and sustaining reading interest through the combination of various semiotic modes.

Enhancing motivation and engagement through multimodal features

The finding that multimodal texts significantly increased student motivation and engagement reinforces previous research emphasizing the affective benefits of multimodal learning environments (Jewitt, 2008; Kress & Van Leeuwen, 2006). The participants' strong positive responses align with empirical evidence that diverse

representational modes can increase interest, reduce monotony, and maintain learning focus (Mayer & Moreno, 2003).

The excerpt from FG1-P2, which describes how a pre-reading video helped activate prior knowledge and foster deeper engagement with the text, demonstrates an important mechanism by which multimodality supports motivation: by providing a more accessible entry point and reducing barriers to comprehension at an early stage. This process mirrors Vygotsky's (1978) concept of scaffolding in the zone of proximal development, where external support allows students to access material they might not be able to grasp independently.

Research in Indonesia and Southeast Asia also corroborates these findings. Wardani and Kurniasih (2023) demonstrated that multimodal texts helped Indonesian students overcome the boredom of traditional reading activities. Their qualitative analysis showed that students found the multimodal format more engaging, encouraging them to engage with more complex texts. Additionally, Jamil and Aziz (2021) found that Malaysian students' reading habits improved significantly after exposure to multimodal texts, as the combination of different semiotic modes helped maintain engagement over a longer period. These findings from the region confirm that the motivational benefits of multimodality are particularly relevant to the Southeast Asian EFL context, where students often face boredom with traditional teaching materials and limited access to authentic English texts.

The motivational benefits found in this study also align with self-determination theory (Deci & Ryan, 2000), which emphasizes three key psychological needs for intrinsic motivation: autonomy, competence, and relatedness. Multimodal texts can meet the need for competence by providing multiple ways to construct meaning, thereby increasing students' self-confidence and reducing the anxiety often present in L2 reading activities.

Implications for motivation enhancement

The finding that multimodal scaffolding increases motivation and maintains learner engagement aligns with Self-Determination Theory proposed by Deci and Ryan (2000). The following discussion connects these findings to broader theoretical perspectives that explain how multimodality supports affective engagement in EFL reading.

1. Pre-reading multimodal scaffolds

Include preview videos, supporting graphics, or audio introductions that can activate prior knowledge and pique students' interest before they move on to the written text.

2. Diverse entry points

Design multimodal materials that allow students to begin in the mode they are most comfortable with, whether visual, auditory, or textual, to support a sense of autonomy and reduce anxiety.

3. Gradual complexity

Start with simple multimodal formats to build confidence, then slowly introduce more complex structures as students' skills improve.

4. Use of authentic materials

Choose multimodal texts that are relevant to students' interests and connected to real-world contexts to make learning more meaningful and increase intrinsic motivation.

Cognitive benefits of integrated visual and auditory scaffolding

The second theme—that visual and auditory elements help students understand complex ideas—provides empirical support for dual coding theory (Paivio, 1986) and the cognitive theory of multimedia learning (Mayer, 2014) in the context of second language (L2) reading. Both theoretical frameworks explain that when information is processed through verbal and visual channels simultaneously, learning is more effective because more diverse mental representations are formed, which then facilitates the encoding, storage, and retrieval of information.

The focus group and classroom observation data in this study that when visual representations (pictures, diagrams, infographics) and auditory scaffolding (narration, pronunciation guides) are used to complement written text, students can construct clearer and more accessible mental models of abstract concepts. This process appears to reduce extraneous cognitive load, mental effort wasted on irrelevant or disorganized information, while supporting germane cognitive load, cognitive effort that directly contributes to schema formation and knowledge integration (Sweller et al., 2011).

Our findings demonstrate that visual scaffolding and multimodal materials significantly enhanced students' comprehension, particularly when dealing with abstract or conceptually dense texts. These results indicate that visual supports helped students sustain attention and form clearer mental representations of ideas. This aligns with Juawaty and Nima (2018), who found that multimodal texts reduced cognitive load and improved learners' ability to process complex information.

Similarly, the ways our participants relied on diagrams and images to structure understanding echo Syafii's (2021) observation that visual scaffolds assist Indonesian EFL learners in identifying text relationships and organizing information effectively. The consistency across these findings underscores the pedagogical value of integrating visual modalities into EFL instruction.

Furthermore, the enhanced vocabulary comprehension observed in our study parallels findings by Cahyaningati and Lestari (2018), who attributed similar gains to dual-coding effects. This suggests that visual elements do not merely support content comprehension but also facilitate lexical retention—highlighting the multi-layered benefits of multimodal learning in Indonesian EFL contexts.

Cognitive load and the challenges of non-linear navigation

While various benefits have been identified, a third theme suggests that non-linear multimodal structures can pose significant cognitive challenges, particularly regarding navigation and information integration. Students reported experiencing difficulties when processing complex when confronted with complex multimodal texts, coupled with observational evidence of disjointed navigation patterns, illustrate the magnitude of cognitive load that occurs when multimodal complexity exceeds students' strategic abilities.

These findings align with cognitive load theory (Sweller, 1988; Sweller et al., 2011), which distinguishes between intrinsic load (inherent task complexity), extrinsic load (resulting from inappropriate instructional design), and germane load (load that supports learning). In the context of non-linear multimodal texts, navigational demands and the need to process multiple modes simultaneously can create excessive extrinsic load, reducing cognitive capacity that should be allocated to germane processing.

Regional context and navigation challenges

The navigation challenges emerging in this study also align with EFL research in Indonesia. Syafii (2021) stated that although visual strategies helped improve comprehension, many students initially struggled to organize and navigate complex multimodal information without explicit guidance. He also emphasized that Indonesian EFL learners, most of whom have limited experience with complex digital texts, require systematic instruction to understand non-linear texts.

Similar patterns were found by Wardani and Kurniasih (2023). In their Celtic article, they emphasized that careful multimodal design is crucial to prevent cognitive overload. They demonstrated that a gradual approach, where teachers slowly introduce multimodal complexity and provide clear navigational guidance, helps students benefit from multimodal texts without experiencing confusion or disorientation.

Together, these studies support the challenges found in the present research, suggesting that Indonesian EFL learners often require explicit instruction in multimodal navigation. The fragmented and unsystematic navigation patterns observed in this study indicate that many EFL learners require explicit instruction in multimodal literacy strategies, including how to review multimodal layouts, determine

effective reading sequences, integrate information from multiple modes, and monitor comprehension while navigating complex structures.

Theoretical Implications

The navigation challenges identified in this study align with recent findings on digital and multimodal reading. For example, Hidayat, Sundari, and Zuhairi (2023) found that digital storytelling platforms, which integrate text, audio, visuals, and interactivity, can enhance reading comprehension among EFL learners. Similarly, Daulay and Dewi (2024) demonstrated that a multimodal literacy approach improves students' reading comprehension more effectively than traditional text-based instruction. Juliana and Anggraini (2024) also highlighted that pre-reading, while-reading, and post-reading metacognitive strategies substantially improve learners' understanding of texts. In the context of online hypertext reading, Azmuddin et al. (2023) reported that many university students struggle to navigate and comprehend digital texts without sufficient metacognitive and navigational skills. Furthermore, Wang and Hu (2025) revealed that students' metacognitive knowledge and self-perceived reading competence are positively correlated with their digital reading performance. Collectively, these findings underscore that effective engagement with multimodal digital texts requires not only linguistic competence but also strong digital literacy, navigational ability, and metacognitive awareness.

Furthermore, the challenges emerging in theme 3 emphasize the importance of considering students' developmental readiness for multimodal complexity. When carefully designed, multimodal scaffolding can indeed support comprehension. However, when multimodal density is too high without sufficient preparation or support, it can actually burden students—especially those with low digital literacy or limited English (L2) skills.

Toward a balanced multimodal pedagogy: Implications and recommendations

The dual nature of multimodality revealed in this study—both helpful and challenging for second language (L2) reading—provides several important implications for EFL pedagogy. These findings suggest that effective multimodal use requires thoughtful design, clear strategy instruction, and special attention to cognitive load management.

Design principles for multimodal reading materials

Based on the results of this study and previous research (Mayer, 2014; Sweller et al., 2011; Wardani & Kurniasih, 2023), educators are advised to consider the following principles when selecting or developing multimodal reading materials:

1. Coherence

Include only multimodal elements that truly support the learning objectives. Avoid details that appear interesting but are irrelevant (Rey, 2012), as they can distract from the core content.

2. Signaling

Use visual and textual cues, such as arrows, highlights, or explicit instructions, to help students focus on important sections and understand the proper navigation sequence.

3. Segmentation

Break complex multimodal text into smaller, more easily processed chunks. This can reduce the demands of simultaneous processing.

4. Scaffolding

Present multimodality gradually, starting with simpler formats before moving on to more complex structures. This gradual approach proved effective in a study by Wardani and Kurniasih (2023).

5. Integration

Ensure that visual, audio, and text elements support and clearly connect, rather than compete, to foster a more complete understanding.

6. Cultural relevance

Choose multimodal materials that are appropriate to the cultural context and linguistic background of students. Findings from regional research (Syafii, 2021; January & Nima, 2018) indicate that cultural relevance increases connectedness and facilitates comprehension.

Explicit instruction in multimodal literacy strategies

The navigation challenges that emerged in Theme 3 emphasize the importance of providing clear and directed instruction in multimodal reading strategies. Referring to research findings in Indonesia (Syafii, 2021; Wardani & Kurniasih, 2023), educators can consider the following approaches:

1. Modeling

Teachers directly demonstrate how to navigate multimodal text—for example, through think-aloud techniques, to demonstrate how to review the layout, determine reading order, and connect information from various modes.

2. Metacognitive strategy instruction

Teach students to plan their reading steps, monitor intermodal comprehension, and assess the relevance and reliability of the multimodal elements they encounter.

3. Structured Flow

Initially, provide clear navigational guidance, such as numbered sequences or suggested steps. Over time, reduce this assistance so students can internalize the strategies independently, an approach validated by Wardani and Kurniasih (2023).

4. Collaborative exploration

Facilitate group discussions so students can share strategies, compare reading styles, and solve navigational challenges together.

5. Reflective practice

Provide regular opportunities for students to reflect on their multimodal reading experiences, for example, through journaling, so they can identify strategies that work and areas for improvement.

6. Visual mapping practice

Use story mapping or other visual organization techniques to help students organize and navigate complex information structures. This approach has been shown to be effective in research by Syafii (2021).

Multimodality as a scaffold and a challenge

The findings of this study—confirmed by research from Indonesian and regional contexts (Cahyaningati & Lestari, 2018; Januarty & Nima, 2018; Syafii, 2021; Wardani & Kurniasih, 2023)—indicate that multimodal texts act as both a scaffold and a source of challenges in reading English as a Foreign Language (EFL). On the one hand, multimodal texts can increase motivation by providing multiple entry points, reducing monotony, and aiding comprehension through dual coding mechanisms. However, on the other hand, multimodal texts also have the potential to cause cognitive overload and navigational confusion when their complexity exceeds learners' strategic capabilities.

This dual nature suggests that the pedagogical effectiveness of multimodal texts is determined not only by the presence of multimodal elements but primarily by how they are designed, presented, and taught. Research published in *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics*, and other regional publications consistently emphasizes the importance of teacher guidance, structured implementation, and step-by-step instruction when introducing multimodal texts to EFL learners in Indonesia and Southeast Asia.

By integrating the findings of these studies and regional research, several key principles for effective multimodal pedagogy emerge:

1. Intentional design: Multimodal elements should be carefully selected and integrated to support learning objectives, rather than being chosen solely for aesthetic purposes.
2. Gradual complexity: Students should be introduced to progressive levels of multimodal complexity, starting with simple formats before moving on to more complex structures.
3. Explicit guidance: Teachers should model navigation strategies and provide structured guidance as students interact with complex multimodal texts, gradually reducing support as students' independence increases.
4. Cultural relevance: Multimodal materials should be aligned with students' cultural backgrounds and linguistic contexts to make them more accessible and meaningful.
5. Metacognitive development: Students should be explicitly taught how to monitor their comprehension across modes and evaluate the usefulness of each multimodal element.

CONCLUSION

The study explored Indonesian EFL undergraduates' perceptions, engagement, and response to multimodal literacy in a university reading course. These results imply that (1) multimodality is a way to encourage student motivation and engagement, (2) visual and auditory affordances ease understanding of abstract concepts, and (3) a non-linear multimodal presentation works against easy interpretation. These results are practically significant for theoretical considerations by borrowing the learner perspective to fill gaps in multimodal studies, which boast a great deal of quantitative/literature-based information. These results are pedagogically significant for the appropriate implementation and classroom training of multimodalities, as well as caution in integrative use not to inundate students. These results are significant for the level of policy because they respond to the MBKM curriculum for Indonesia, which champions digital literacy and critical literacy as a cornerstone of Indonesian higher education. These results, however, have limitations. They are not generalizable, as a convenience sample from a single institution renders generalizability a challenge. Future research could study other schools, cross-institutionally with longitudinal or mixed-methods studies that compare qualitative experience with learning outcomes. By elevating the learner voice, this study champions multimodal literacy as an approach not necessarily innovative (teacher-centered) but one that can be molded to the student voice.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

AUTHOR (S) CONTRIBUTION

Arum Puspitasari: Conceptualization (lead), methodology (lead), data analysis (lead), writing – original draft (lead), review and editing (supporting). Moh. Kholilurrahman Jailani: Data collection (lead), methodology (supporting), writing – review and editing (lead), validation (supporting). Yazid Basthomi: Conceptualization (supporting), supervision (lead), and critical revision of the manuscript (lead). Suharyadi: Methodology (supporting), data validation (supporting), and review (supporting). Evynurul Laily Zen: Literature review (supporting), editing (supporting), and formatting (lead).

ACKNOWLEDGMENT

The authors would like to express sincere gratitude to the Faculty of Letters, Universitas Negeri Malang, for providing academic and administrative support during this research. Appreciation is also extended to the undergraduate students who participated in this study for their valuable time and reflections.

REFERENCES

- Archer, A. (2022). A multimodal approach to English for academic purposes in contexts of diversity. *World Englishes*, 41(4), 545–553. <https://doi.org/10.1111/weng.12600>
- Azmuddin, N. H., Nor, F. M., & Hamat, A. (2023). Metacognitive online reading and navigational strategies by science and technology university students. *GEMA Online Journal of Language Studies*, 23(1), 125–142. Retrieved from <https://ejournals.ukm.edu.my/gema/article/view/18864>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Cahyaningati, D. T., & Lestari, L. A. (2018). The use of multimodal text in enhancing engineering students' reading skill. *International Journal of Language Education*, 2(2), 65–73. <https://doi.org/10.26858/ijole.v2i2.6360>
- Carcamo, B., & Pino, B. (2025). Developing EFL students' multimodal literacy with the use of infographics. *Asian-Pacific Journal of Second and Foreign Language Education*, 10(1). <https://doi.org/10.1186/s40862-025-00322-3>

- Chun, D. (2016). The role of technology in SLA: Understanding online interaction and L2 learning. *The Modern Language Journal*, 100(S1), 144-164. <https://doi.org/10.1111/modl.12303>
- Cirocki, A., Farrelly, R., & Sapp, T. (2024). *Developing materials for innovative teaching and sustainable learning*. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-69206-2>
- Creswell, J. W. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Daulay, E., & Dewi, U. (2025). Reading comprehension through multimodal literacy approach for English education. *ETERNAL (English Teaching Journal)*, 16(1), 45–60. <https://journal2.upgris.ac.id/index.php/eternal/article/view/1910>
- Emiliasari, R. N., & Emilia, E. (2025). Integrating SFL and visual grammar in genre-based pedagogy to foster EFL students' critical reading of multimodal texts. *Kembara: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, 11(1), 56–70. <https://doi.org/10.22219/kembara.v11i2.4128>
- Gupta, S., & Arora, P. (2024). *Epistemic co-design in multimodal learning environments: Graduate perspectives on non-linear knowledge construction*. *British Journal of Educational Technology*, 55(1), 89–108. <https://doi.org/10.1111/bjet.13392>
- Handrayani, D., Suryati, N., & Astuti, U. P. (2024). A phenomenological study of EFL teachers' experiences in putting multiliteracy pedagogy through online teaching. *IJEE (Indonesian Journal of English Education)*, 11(1). <https://doi.org/10.15408/ijee.v11i1.34196>
- Hidayat, D., Sundari, H., & Zuhairi, A. (2023). EFL reading comprehension through digital storytelling in multimodal digital literacy. *SCOPE: Journal of English Language Teaching*, 8(1), 65–78. Retrieved from <https://journal.lppmunindra.ac.id/index.php/SCOPE/article/view/24799>
- Jamil, N. A., & Abdul Aziz, A. (2021). The use of multimodal text in enhancing students' reading habit. *Malaysian Journal of Social Sciences and Humanities*, 6(9), 487–492. <https://doi.org/10.47405/mjssh.v6i9.977>
- Januarty, R., & Nima, H. N. A. (2018). Energizing students' reading comprehension through multimodal texts. *International Journal of Language Education*, 2(2), 14–22. <https://doi.org/10.26858/ijole.v2i2.4347>
- Juliana, & Anggraini, N. (2024). Metacognitive reading comprehension instructional model on narrative text. *REiLA: Journal of Research and Innovation in Language*,

- 6(2), 205–216. <https://journal.unilak.ac.id/index.php/REILA/article/view/15846>
- Kaowiwattanukul, S. (2025). Multimodal literature and CEFR reading proficiency: Improving literary reading skills in EFL learners. *Indonesian Journal of Applied Linguistics*, 15(1), 20–34. <https://doi.org/10.17509/ijal.v15i1.75921>
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). SAGE Publications.
- Lam, K. Y., & Putri, E. W. (2024). A study on the affordances of digital fairy-tale rewriting: A critical literacy approach to digital multimodal composing. *System*, 121, 102134. <https://doi.org/10.1016/j.system.2024.103536>
- Le, H. V., Nguyen, T. A. D., Le, D. H. N., & Nguyen, P. U. (2024). Unveiling critical reading strategies and challenges: A mixed-methods study among English major students in a Vietnamese higher education institution. *Cogent Education*, 11(1), 2326732. <https://doi.org/10.1080/2331186X.2024.2326732>
- Lee, H., Kim, J., & Park, S. (2021). Motivation and engagement in multimodal language learning. *Computer Assisted Language Learning*, 34(5–6), 612–635. <https://doi.org/10.1080/09588221.2019.1614069>
- Lee, S. Y., Lo, Y. H. G., & Chin, T. C. (2021). Practicing multiliteracies to enhance EFL learners' meaning making process and language development: A multimodal problem-based approach. *Computer Assisted Language Learning*, 34(1–2), 66–91. <https://doi.org/10.1080/09588221.2019.1614959>
- Lim, H., & Polio, C. (2020). Multimodal assignments in higher education: Implications for multimodal writing tasks. *Journal of Second Language Writing*, 47, 100713. <https://doi.org/10.1016/j.jslw.2020.100713>
- Mavengere, N., Mahlangu, V., & Khoza, S. (2024). *Digital multimodality and non-linear pathways in postgraduate learning*. *Education and Information Technologies*, 29(5), 6501–6520. <https://doi.org/10.1007/s10639-024-12648-5>
- Mejía-Laguna, J. A. (2023). Critical learning episodes in the EFL classroom: A multimodal (inter)action analytical perspective. *Multimodal Communication*, 12(1), 23–44. <https://doi.org/10.1515/mc-2023-0006>
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60–92.

- <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Ngui, W. (2022). *E-portfolio as an academic writing assessment tool in higher education: Strengths and challenges*. *Indonesian Journal of Applied Linguistics*, 12(2), 331–340. <https://doi.org/10.17509/ijal.v12i2.40122>
- Nguyen, L. T. (2019). Reflective journal writing as a tool for student teachers' professional development. *International Journal of Education and Practice*, 7(1), 1–13. <https://doi.org/10.18488/journal.61.2019.71.1.13>
- Özer, H. Z. (2025). *Engaging EFL student writers in social advocacy through digital multimodal composition and critical pedagogies*. *Digital Applied Linguistics*, 3, 103111. <https://doi.org/10.29140/dal.v3.103111>
- 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.2012.05.006>
- Stewart, O. G., González, J., & Midgette, E. (2025). A framework for critically analyzing multimodal texts for English language learners through intermodal complementarity. *Qualitative Research Journal*, 25(2). <https://doi.org/10.1108/qrj-08-2024-0173>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Syafii, M. L. (2021). The implementation of the story mapping strategy to enhance students' reading comprehension. *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics*, 8(1), 1–21. <https://doi.org/10.22219/celtic.v8i1.16161>
- Tandiana, S. T., Abdullah, F., & Andriani, R. (2023). Mapping multimodal teaching materials to support EFL students' critical reading. *Journal of English Education and Linguistics Studies*, 14(2), 211–228. <https://doi.org/10.21580/vjv12i219990>
- Wang, Y., & Hu, J. (2025). Relationships between reading engagement features and students' digital reading performance: An international large-scale study. *Learning and Instruction*. Advance online publication. <https://doi.org/10.1016/j.learninstruc.2025.102157>
- Wallace, C. (2003). *Critical reading in language education*. Palgrave Macmillan.
- Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy*, 33(3), 211–239.
- Wardani, Y., & Kurniasih, E. (2023). Multimodal texts: Are they truly helpful for

narrative text comprehension? *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics*, 10(2), 138–156. <https://doi.org/10.22219/celtic.v10i2.28211>

Yawiloeng, R. (2022). Using instructional scaffolding and multimodal texts to enhance reading comprehension: Perceptions and attitudes of EFL students. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4122231>

Yi, Zhao, & Qiao. (2024). The dynamics of multi-modal practice and pedagogy in EFL instruction: A theoretical exploration. *The Educational Review, USA*, 8(9), 1163-1170. <https://doi.org/10.26855/er.2024.09.011>