p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

p-ISSN 2252-7427 e-ISSN 2541-3635
Indonesian EFL Journal
AISEE
The Association of Indonesia Scholars of English Education

AI-ENHANCED MULTIMODAL READING: EXPLORING ITS IMPACT ON EFL LEARNERS' READING COMPREHENSION AND ENGAGEMENT IN MADRASAH

Hasna Fitri Labibah[1*], Januarius Mujiyanto[2], Dwi Rukmini[3], Whidiyanto[4]
[1,2,3,4]*Doctoral Program of Language Education, Faculty of Languages and Arts, Universitas Negeri Semarang, Central Java, Indonesia*

APA Citation: Labibah, H. F., Mujiyanto, J., Rukmini, D., Whidiyanto. (2025). AI-enhanced multimodal reading: exploring its impact on EFL learners' reading comprehension and engagement in madrasah. *Indonesian EFL Journal*, *11*(3), 693-704. https://doi.org/10.25134/ieflj.v11i3.72

Received: 15-08-2025 Accepted: 02-10-2025 Published: 30-10-2025

Abstract: This study investigates the impact of AI-enhanced multimodal reading on EFL learners' comprehension and engagement in an Indonesian madrasah context. Employing an explanatory sequential mixed-methods design, the research combined a pre-experimental one-group pretest–posttest design with semi-structured interviews. Thirty eleventh-grade students at MAS Al Falah, Tulang Bawang, participated in the quantitative phase, while six students were purposively selected for the qualitative phase. The intervention, conducted over four weeks, integrated DALL·E-generated visuals into reading tasks covering topics such as Indonesian environmental figures, digital literacies, and personal identity. Quantitative findings revealed a significant improvement in learners' comprehension, with mean scores rising from 75.07 (pretest) to 81.73 (posttest), and an N-Gain of 0.269 indicating low-to-moderate progress. Qualitative analysis further explained these gains, showing that students attributed their improvement to visual scaffolding, heightened motivation, stronger identity connection, and collaborative meaning-making, although some expressed challenges in interpreting AI-generated images. The integration of both strands through joint display confirmed that cognitive, affective, and sociocultural dimensions collectively shaped learning outcomes. The study concludes that AI-enhanced multimodal reading can enrich EFL pedagogy by promoting comprehension and engagement, while also highlighting the need for scaffolding and digital literacy development. Future research should extend to longitudinal designs, larger populations, and comparative analyses of different AI tools to sustain and generalize the benefits of multimodal learning.
.Keywords: AI; madrasah; reading; comprehension; multimodal; mixed method

INTRODUCTION

In recent years, multimodal approaches have increasingly attracted attention in the field of language education, reflecting a growing recognition that literacy extends beyond traditional print-based practices (Andriani et al., 2022; Hafner, 2015; Jovazino Bastos Medrado Costa, 2023). Scholars argue that the integration of multiple semiotic resources such as images, gestures, and digital media, enhances learners' meaning-making processes and fosters deeper engagement with texts (Kress, 2019; Li et al., 2025; Yang & Cui, 2024). The rise of artificial intelligence (AI) has further expanded these possibilities, offering new ways to generate visual, auditory, and interactive materials that support comprehension and creativity (Lee et al., 2025; Rudik & Onyshchuk, 2024; Santoso & Mangkuluhur, 2024). These developments suggest that multimodal pedagogy is entering a transformative phase shaped by emergent technologies.

Despite this promise, current research on multimodal literacy has largely concentrated on static or teacher-designed resources, with limited attention to AI-driven tools in classroom practice (Chen, 2018; Hafner, 2015; Park et al., 2022; Wyatt, 2015). While previous studies confirm that multimodal approaches can enhance comprehension, motivation, and learner identity,

however, only a few studies have combined these areas by examining AI-generated multimodality in religious educational contexts such as Indonesian madrasahs. There is limited research on how students in underrepresented contexts, such as Indonesian madrasahs, engage with AI-generated multimodal reading tasks (Andriani et al., 2022; Maspul, 2025; Santoso & Mangkuluhur, 2024). The absence of empirical evidence in these contexts highlights the need to investigate not only the effectiveness of AI-enhanced multimodal reading but also its pedagogical and cultural relevance (Kundu & Bej, 2025; Lee et al., 2025).

Equally important, learner engagement and identity remain underexplored within AI-mediated multimodal environments. Engagement, encompassing behavioural, emotional, and cognitive dimensions, is central to literacy development (Derakhshan, 2025; Huang, 2025; Kim et al., 2025). Studies show that multimodal resources can sustain learners' interest and motivation, while also enabling them to connect personal experiences with academic content (Brugar et al., 2018; Maspul, 2025). However, without adequate scaffolding, multimodal interventions risk creating confusion rather than comprehension (Chen, 2018; Fälth et al., 2023; Hafner, 2015). What distinguishes AI from other multimodal tools is its capacity to generate adaptive, context-sensitive materials that may reshape how learners perceive their engagement and identity in classroom interactions (Chatzichristofis et al., 2025; Nyrup & Robinson, 2022). This raises critical questions about whether AI can uniquely support learners in negotiating both cognitive and affective dimensions of literacy.

Furthermore, Indonesian madrasahs represent a unique setting where EFL learning intersects with religious, cultural, and institutional practices, creating distinct dynamics compared to mainstream schools (Andriani et al., 2022; Fälth et al., 2023; Park et al., 2022). These institutions often face resource constraints and limited exposure to advanced technologies, yet they embody strong traditions of literacy and identity formation that significantly shape learner participation (Brugar et al., 2018; Hafner, 2015). Exploring how AI-enhanced multimodal reading operates in such contexts can therefore provide valuable insights into how global pedagogical innovations intersect with localized educational realities (Chen, 2018; Park et al., 2022; Santoso & Mangkuluhur, 2024). This perspective acknowledges that technology integration in education is not neutral but deeply shaped by context (Brugar et al., 2018; Ifinedo & Kankaanranta, 2021).

To address these gaps, the present study is designed to explore how AI-enhanced multimodal reading contributes to literacy practices in a madrasah context. Three research questions guide the inquiry: to what extent does AI-enhanced multimodal reading improve EFL learners' reading comprehension in Class XI MAS Al Falah, Tulang Bawang, Lampung? How do learners perceive their engagement when experiencing AI-generated multimodal reading tasks in the classroom? And finally, in what ways do the observed gains in comprehension align with learners' qualitative experiences of engagement and identity connection? By addressing these questions, the study not only responds to the limited scholarship on AI and multimodality in EFL but also contributes to broader debates on how digital technologies can transform reading practices in underrepresented educational contexts.

METHOD

This study employed an explanatory sequential mixed-methods design by Creswell & Plano Clark (2023) to investigate the impact of AI-enhanced multimodal reading on EFL learners' reading comprehension and engagement. This design allowed for the identification of measurable learning gains in the quantitative phase, followed by an in-depth exploration of students' lived experiences in the qualitative phase.

In Phase 1 (quantitative), a one-group pretest–posttest design was implemented to measure changes in students' reading comprehension before and after the intervention. The participants consisted of 30 eleventh-grade students from MAS Al Falah, Tulang Bawang – Lampung, selected using total population sampling. The reading comprehension instrument consisted of 20 items covering literal, inferential, and evaluative comprehension, validated by two EFL experts. Besides, the treatment involved AI-enhanced multimodal reading activities using DALL·E to generate visual representations of reading passages. Students engaged with the AI-generated images as prompts for comprehension exercises, discussion, and guided writing. The intervention was conducted over four weeks, with two sessions per week, each lasting 90 minutes.

The intervention lasted four weeks with two 90-minute sessions per week. Each session followed a three-stage cycle. In the pre-reading

p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

stage (20 minutes), the teacher selected short English passages and generated images using DALL·E based on predetermined criteria such as clarity, textual relevance, and cultural appropriateness. Students then predicted the text content and brainstormed interpretations of the visuals. In the while-reading stage (40 minutes), students engaged with the passages through comprehension tasks (literal, inferential, evaluative) supported by the AI-generated images, discussing how visuals complemented textual meaning. In the post-reading stage (30 minutes), students reflected on their learning, composed short written responses, and shared how the images shaped their understanding and personal or religious connections. This structure ensured that students were not passive viewers of AI outputs but active participants who interpreted, discussed, and created meaning from the multimodal prompts.

A reading comprehension test, adapted from national curriculum standards and validated by two EFL experts, was administered before and after the intervention. The test consisted of 20 items covering literal, inferential, and evaluative comprehension. The test was in multiple-choice format with four options per item. The instrument showed high reliability, with Cronbach's alpha = 0.87, confirming consistency of measurement.

In Phase 2 (qualitative), semi-structured interviews were conducted with six students selected through maximum variation sampling to represent different levels of comprehension improvement and engagement. The interviews explored learners' experiences interacting with AI-generated visuals, their motivation and engagement, perceived benefits, and challenges in multimodal reading activities. The overall research procedure is visualized in figure 1 below to illustrate the sequential mixed-methods design more clearly.

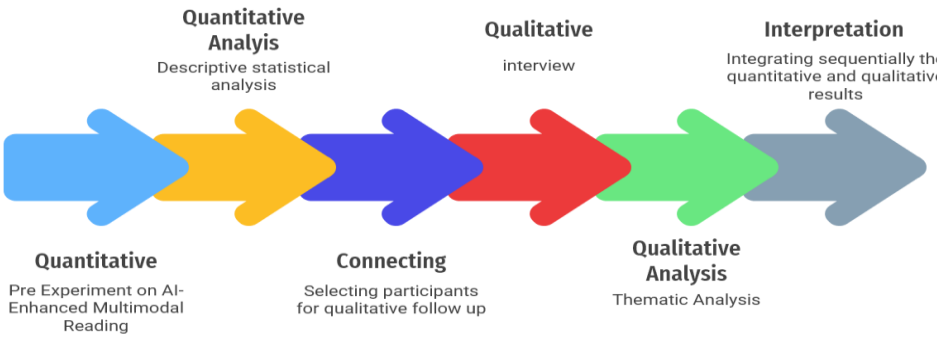


Figure 1. *Applied explanatory sequential mixed-methods design*

The participants were 30 eleventh-grade students from MAS Al Falah, Tulang Bawang, Lampung**.** Total population sampling was used for the quantitative phase, while purposive sampling based on maximum variation ensured a rich diversity of perspectives in the qualitative phase.

Quantitative data were analysed using descriptive statistics and paired-sample t-tests to examine significant differences between pretest and posttest scores (Wallen & Fraenkel, 2013). Qualitative data were transcribed and analysed thematically by Braun & Clarke (2006). Instead of relying on software, coding was conducted manually using spreadsheets and printed transcripts. First, the researchers read and reread the transcripts to achieve familiarization. Initial codes were generated inductively and deductively, and relevant excerpts were highlighted and categorized in tables. Codes were then collated to form candidate themes, which were iteratively refined through discussion among the researchers. To enhance credibility, two researchers independently coded the data, and intercoder reliability was checked through comparison and consensus-building, reaching substantial agreement (Cohen's kappa = 0.81). Member checking with participants further ensured accuracy of interpretation.

For the quantitative phase, instrument validity was confirmed through expert judgment, and reliability analysis yielded Cronbach's alpha = 0.87. For the qualitative phase, trustworthiness was enhanced using triangulation, member checking, and peer debriefing. These steps ensured credibility and interpretive depth, allowing meaningful integration of findings from both phases.

RESULTS AND DISCUSSION

Quantitative phase

The first phase of this study investigated the impact of AI-enhanced multimodal reading on

students' reading comprehension using a one-group pretest–posttest design with 30 participants. Quantitative data were obtained from pretest and posttest assessments to determine changes in students' performance following the intervention. Furthermore, the normalized gain (N-Gain) was calculated to assess the relative improvement in learning outcomes. The descriptive statistics presented below provide a comparative summary of pretest, posttest, and N-Gain scores, offering a clear picture of students' progress.

Before conducting the paired-sample t-test, the distribution of difference scores was examined using the Shapiro–Wilk test, which indicated that the data did not significantly deviate from normality ($p > .05$). This confirmed that the assumptions for applying a parametric test were met.

Table 1. *Summary of quantitative findings*

Statistics	Pre-test	Post-tests	N-Gain
N	30	30	30
Minimum	60	70	0.14
Maximum	85	91	0.46
Mean	75.07	81.73	0.269
Std. Deviation	7.688	6.063	0.08352
Variance	59.099	36.754	0.007
Skewness	-0.577	-0.420	0.581

The analysis involved 30 participants who completed both the pretest and posttest. The pretest scores ranged from 60 to 85, while the posttest scores improved to a range of 70 to 91. This indicates that both the lower and upper bounds of student performance increased after the intervention, suggesting a consistent improvement across participants.

In terms of averages, the mean score rose from 75.07 (pretest) to 81.73 (posttest), reflecting a gain of approximately 6.66 points. This improvement demonstrates that the intervention had a positive effect on students' performance. Moreover, the standard deviation decreased from 7.688 in the pretest to 6.063 in the posttest, implying that the scores became more homogeneous and learning outcomes were more evenly distributed among participants.

The normalized gain (N-Gain) yielded an average value of 0.269, which is generally categorized as a low to moderate improvement. Despite not reaching the "high" category, this gain is meaningful, particularly given the relatively short duration of the intervention. Finally, the skewness values suggest that both pretest (-0.577) and posttest (-0.420) distributions are slightly left-skewed, meaning most students scored above the mean, which is a favorable trend.

Qualitative findings

To complement the quantitative results, six students representing different levels of comprehension improvement were interviewed. In this study, engagement was operationalized as students' behavioural attention (e.g., focus during tasks), emotional responses (e.g., interest, enjoyment, or boredom), and cognitive investment (e.g., effort to interpret and connect text and visuals). Interview questions were designed to elicit reflections on these three dimensions. The semi-structured interviews explored how students engaged with AI-enhanced multimodal reading, how DALL·E-generated visuals shaped their understanding, and what challenges they encountered. Overall, the responses revealed positive engagement and meaningful learning experiences, though with some variation in confidence and digital literacy.

Table 2. *Summary of qualitative findings*

Theme	Sample Student Quotes	Interpretation
Enhanced Comprehension	- S1: "The AI image made me see it more clearly, like a story, not just words." - S2: "When reading about environmental leaders, the picture helped me understand faster." - S3: "I could remember details better when I saw the visuals."	DALL·E visuals supported deeper understanding of abstract or unfamiliar concepts.
Motivation & Engagement	- S1: "Usually reading is boring, but with AI pictures, I feel curious to continue reading." - S2: "It was exciting because every picture looked different." - S3: "I paid more attention in class since we used the AI images."	The novelty of AI increased interest and sustained student engagement.
Identity Connection	- S1: "The picture helped me imagine myself	Visual prompts strengthened

	in the text. It feels more connected to my life." - S2: "In the 'My Identity' topic, I could really see myself in the story." - S3: "It felt personal, not just another reading passage."	personal relevance and self-connection to the texts.
Digital Literacy Challenges	- S1: "Sometimes the AI picture looks strange, but the teacher explained, so I could understand." - S2: "One image was confusing, but after discussion I understood better." - S3: "I was not sure how the picture related, but my friends helped me."	Some students faced confusion, highlighting the need for guidance and scaffolding.
Perceived Learning Gains	- S1: "In the test, some questions were easier because I remembered the picture from the reading." - S2: "I think my score improved because I could connect the text with the visuals." - S3: "The images stayed in my mind, so I answered more confidently."	Students linked their comprehension improvement to the multimodal learning support.

The interviews reveal how students experienced AI-enhanced multimodal reading in ways that go beyond what the quantitative scores alone can explain. First, students consistently described how the visuals made texts easier to understand. One student noted, *"The AI image made me see it more clearly, like a story, not just words"* (S1). Another added, *"When reading about environmental leaders, the picture helped me understand faster"* (S2). Such comments suggest that the images functioned as cognitive supports, reducing abstraction and aiding recall during comprehension tasks.

Second, the sense of motivation and engagement emerged strongly. As one student put it, *"Usually reading is boring, but with AI pictures, I feel curious to continue reading"* (S1). Another explained, *"It was exciting because every picture looked different"* (S2). These reflections indicate that the integration of visuals not only captured attention but also sustained interest across different reading sessions. Yet, it is worth noting that this heightened engagement may partly stem from the novelty of the technology.

Third, the identity connection reported by learners is particularly significant. One participant reflected, *"The picture helped me imagine myself in the text. It feels more connected to my life"* (S1). Another shared, *"In the 'My Identity' topic, I could really see myself in the story"* (S2). These statements reveal that AI-generated visuals did more than illustrate the texts, they personalized the reading experience and allowed students to connect academic content to their lived realities.

Fourth, challenges were also voiced. Some students found certain images confusing, as one admitted, *"Sometimes the AI picture looks strange, but the teacher explained, so I could understand"* (S1). Another confirmed, *"One image was confusing, but after discussion I understood better"* (S2). These accounts highlight that while AI visuals can be powerful, they are not always self-explanatory and require teacher mediation and peer support.

Finally, several students explicitly linked their test performance with the visual prompts. One remarked, *"In the test, some questions were easier because I remembered the picture from the reading"* (S1), while another said, *"I think my score improved because I could connect the text with the visuals"* (S2). These self-reported perceptions reinforce the idea that the multimodal approach supported not only engagement but also measurable learning gains.

While the interviews generally revealed positive engagement, it should be acknowledged that part of this heightened enthusiasm may reflect a novelty effect, the initial excitement of interacting with AI-generated materials. Such effects can temporarily increase attention and motivation but may diminish over time if the tool becomes routine.

Phase integration of quantitative and qualitative findings

The integration phase is essential in an explanatory sequential mixed-methods design, as it connects the measurable outcomes from the quantitative phase with the experiential insights gathered through the qualitative phase. To achieve this, the study employed a joint display strategy, as recommended by Creswell & Plano Clark (2023), which allows researchers to align statistical results with narrative themes in a single

framework. Through this approach, the quantitative evidence of gains in reading comprehension is directly juxtaposed with students' reflections on how AI-enhanced multimodal reading shaped their engagement and understanding. Such integration not only highlights points of convergence but also reveals the explanatory power of qualitative findings in interpreting the numerical trends.

Table 3. *Joint display of quantitative and qualitative findings*

Quantitative Results	Qualitative Findings (Student Voices)	Meta-Inferences (Integration)
Mean score improved from 75.07 (pretest) to 81.73 (posttest); N-Gain = 0.269 (low-to-moderate).	"In the test, some questions were easier because I remembered the picture from the reading" (S1). "I think my score improved because I could connect the text with the visuals" (S2).	The modest but significant improvement is partly explained by students' reliance on AI-generated visuals as memory aids.
Standard deviation decreased (from 7.688 to 6.063), showing more homogeneous scores.	"My friends and I could understand better after discussing the AI pictures together" (S3). "The teacher explained when the picture was confusing, so we all learned" (S2).	The reduction in score variability aligns with collaborative meaning-making and teacher scaffolding during multimodal tasks.
Positive skewness reduced slightly (from -0.577 to -0.420), suggesting more balanced distribution.	"Usually reading is boring, but with AI pictures, I feel curious to continue reading" (S1). "It was exciting because every picture looked different" (S2).	Increased motivation and attention likely contributed to more consistent performance across the group.
Test confirmed significant gain ($p < 0.001$).	"The picture helped me imagine myself in the text. It feels more connected to my life" (S1). "In the 'My Identity' topic, I could really see myself in the story" (S2).	The statistical significance is supported by students' deeper identity connection with texts, fostering meaningful learning.

The integration of findings demonstrates a clear alignment between the quantitative improvements and the qualitative experiences reported by students. The rise in mean scores from 75.07 to 81.73, supported by a significant *p* value, reflects measurable learning gains that students themselves attributed to the scaffolding effect of visuals. As one participant explained, *"In the test, some questions were easier because I remembered the picture from the reading"* (S1), while another added, *"I think my score improved because I could connect the text with the visuals"* (S2). These accounts suggest that AI-generated images did more than decorate the texts; they functioned as cognitive tools that enhanced recall and comprehension.

The reduction in standard deviation further indicates that student performance became more consistent after the intervention. This statistical trend resonates with the students' reflections on peer and teacher support during the multimodal sessions. As one learner stated, *"My friends and I could understand better after discussing the AI pictures together"* (S3). Such collaborative meaning-making appears to have narrowed achievement gaps, leading to more homogeneous outcomes across the group.

Finally, the qualitative theme of motivation provides insight into why test performance improved consistently across the sample. Students highlighted their increased engagement, with one noting, *"Usually reading is boring, but with AI pictures, I feel curious to continue reading"* (S1). This heightened motivation likely encouraged greater attention and persistence, thereby reinforcing the statistical evidence of overall improvement. In sum, the integrated analysis reveals that the quantitative gains were not isolated numerical shifts but were deeply intertwined with learners' lived experiences of visual scaffolding, collaboration, and enhanced motivation.

In short, the findings suggest that AI-enhanced multimodal reading supported comprehension and engagement in meaningful ways. However, the possibility of novelty effects highlights the need for longitudinal research to examine whether such engagement is sustained once the technology becomes familiar**.**

p-ISSN 2252-7427, e-ISSN 2541-3635
https://ieflj.uniku.ac.id/pub

The findings of this study demonstrate that AI-enhanced multimodal reading improved learners' comprehension and engagement, although the quantitative gains remained in the low-to-moderate category. This moderate outcome may reflect contextual constraints specific to the madrasah setting. For instance, limited access to digital resources meant that students' prior exposure to AI tools was minimal, requiring additional time for acclimatization rather than deep comprehension practice (Alazemi, 2024; Khan et al., 2024). Furthermore, variations in digital literacy levels created uneven participation, as some learners needed basic guidance in interpreting AI-generated visuals before engaging with higher-order comprehension tasks (Robinson et al., 2020; Yi & Siqian, 2025). The short duration of the intervention (four weeks) also restricted opportunities for sustained practice, which may have capped the learning gains. These factors suggest that while the intervention was effective, the extent of improvement was tempered by infrastructural and readiness-related challenges.

Another consideration is the broader pedagogical culture within madrasah contexts, which often emphasizes text-based learning and memorization. Such traditions may shape how students engage with multimodal innovations, leading to cautious adoption rather than full integration into learning practices (Haule et al., 2024; Zou et al., 2025). Studies have shown that institutional norms and teacher beliefs can significantly mediate the uptake of digital and multimodal pedagogies (van Twillert et al., 2020; Zhang & Peng, 2025). This indicates that the moderate N-Gain observed in this study is not only a reflection of technological limitations but also of systemic and cultural factors that shape classroom practice.

This outcome resonates with previous research showing that multimodal approaches offer cognitive and affective benefits, yet their impact varies depending on design, scaffolding, and learner readiness (Andriani et al., 2022; Brugar et al., 2018; Hafner, 2015). In the present study, the moderate N-Gain may reflect contextual constraints in the madrasah setting, including limited technological resources, uneven levels of digital literacy, and restricted exposure to AI-mediated tasks, factors that have been shown to influence the effectiveness of multimodal interventions (Goldenthal et al., 2021; Morgado et al., 2025; Namaziandost & Rezai, 2024). The integration of statistical results with students' reflections therefore provides a nuanced account of both potential and limitations of multimodal pedagogy in EFL contexts.

One of the most salient findings concerns the role of visualization in scaffolding comprehension. Students frequently described AI-generated images as aids for grasping difficult passages, particularly when the texts addressed environmental figures and digital literacies. This aligns with evidence that multimodal resources function as semiotic mediators, bridging abstract content with concrete representations (Fälth et al., 2023; Hafner, 2015; Kress, 2019; H. Liu et al., 2022). Such scaffolding operates through dual coding, where verbal and visual inputs reinforce one another, leading to deeper comprehension and long-term recall (Andriani et al., 2022; Fälth et al., 2023; van Nooijen et al., 2024).

The data also highlight the motivational dimension of multimodal learning. Students perceived reading as more engaging and less monotonous when accompanied by AI visuals. This echoes studies emphasizing how multimodal tasks foster curiosity, enjoyment, and persistence in language learning (Fälth et al., 2023; Hafner, 2015; Kim et al., 2025). Importantly, motivation emerged as a mediating factor between multimodal activities and comprehension gains, suggesting that affective engagement plays a crucial role in sustaining cognitive effort (Chen, 2018; H. Liu et al., 2022; Park et al., 2022). However, the risk of novelty effects should be considered, as learners' enthusiasm may diminish if multimodal tools are not integrated within a sustainable pedagogical framework (Fälth et al., 2023; Jasmine et al., 2025; Z. Liu et al., 2025).

Another notable theme is the connection to identity. Students reported that multimodal prompts in the My Identity text helped them relate personal experiences to academic readings, thereby creating a sense of ownership over meaning-making. This finding affirms the importance of voice and agency in multimodal pedagogy, where learners position themselves as active participants rather than passive consumers (Andriani et al., 2022; Hafner, 2015). The interplay between multimodal resources and learner identity also suggests that AI can serve as a tool for personalization, reinforcing cultural and self-relevance in EFL instruction (Kim et al., 2025; Liu et al., 2022; Park et al., 2022).

Despite these strengths, challenges were evident in the form of digital literacy gaps. Several students expressed confusion when AI-generated images diverged from textual meaning,

underscoring the importance of scaffolding. Prior studies similarly caution that without explicit guidance, multimodal tasks may overwhelm learners or produce misinterpretations (Andriani et al., 2022; Fälth et al., 2023; Hafner, 2015). This highlights the indispensable role of teachers in mediating multimodal and AI-driven learning environments, ensuring that students develop critical digital literacies for interpreting and evaluating diverse resources (Chen, 2018; Park et al., 2022; Santoso & Mangkuluhur, 2024).

The integration of quantitative and qualitative strands strengthens the explanatory power of the study. The statistical evidence of comprehension gains was directly supported by students' accounts linking improved test performance to AI-generated visuals (Pellas, 2025; Xu et al., 2025). This convergence mirrors calls in mixed-methods and multimodal research for holistic perspectives that capture both outcomes and processes (Archibald, 2018; Creswell & Plano Clark, 2023). By aligning performance data with experiential insights, the study demonstrates how multimodal literacy enhances not only measurable comprehension but also learner engagement, identity, and collaboration (Kim et al., 2025; H. Liu et al., 2022; Santoso & Mangkuluhur, 2024).

An important implication concerns the sustainability of AI-enhanced multimodal reading in classroom practice. While short-term gains and heightened motivation were evident, the relatively modest N-Gain suggests that long-term integration strategies are needed. This aligns with literature warning against overreliance on technological novelty and advocating for systematic incorporation of multimodal practices into curricula (Harun & Singh, 2024; Park et al., 2022; Santoso & Mangkuluhur, 2024). Effective integration requires balancing innovation with pedagogy, ensuring that learners receive both stimulating resources and guided instruction (Andriani et al., 2022; Hafner, 2015; H. Liu et al., 2022).

Finally, the study contributes to broader debates on the role of multimodality in EFL literacy education. By demonstrating that AI-generated visuals can scaffold comprehension, increase engagement, and foster identity connection, it affirms the transformative potential of multimodal pedagogy (Kim et al., 2025; Santoso & Mangkuluhur, 2024; Wyatt, 2015). At the same time, the findings caution against ignoring learner diversity and digital literacy challenges, reminding educators that human mediation remains central. Thus, AI-enhanced multimodal reading should be seen not as a replacement for pedagogy but as a complementary tool that, when carefully scaffolded, expands the semiotic resources available to learners (Andriani et al., 2022; Chen, 2018; Hafner, 2015).

CONCLUSION

This study demonstrated that AI-enhanced multimodal reading has the potential to improve EFL learners' comprehension and engagement in a madrasah context. The quantitative findings showed statistically significant gains, while the qualitative results illuminated the mechanisms driving these outcomes, including visualization, motivation, identity connection, and collaboration. Together, these strands highlight the pedagogical value of integrating AI-generated visuals into reading instruction, affirming that multimodality can enrich both cognitive and affective dimensions of learning when carefully scaffolded.

Nevertheless, the study is not without limitations. The intervention was conducted with a single group of 30 students, in a relatively short duration of four weeks, and relied primarily on one AI tool (DALL·E) to generate visuals. These factors limit the generalizability of the results and suggest several avenues for future inquiry. Subsequent studies could adopt longitudinal designs to examine whether motivational effects persist over time, compare different AI-driven multimodal tools, or involve larger and more diverse student populations. Furthermore, future researchers may investigate how multimodal interventions interact with varying levels of digital literacy and explore strategies for sustainable integration into curricula. By addressing these gaps, future work can strengthen the evidence base and advance the transformative role of AI in multimodal literacy education.

REFERENCES

Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, *14*(1), 44. https://link.springer.com/article/10.1186/s40468-024-00319-8

Andriani, D., Ningsih, A. W., Shara, A. M., & Shinoda, K. (2022). The nexus between reading visual texts and scientific non-visual texts in a High School in Indonesia: The nexus between reading visual texts and scientific non-visual texts. *International Journal of Curriculum and*

Instruction, *14*(1), 183–195. https://www.ijci.net/index.php/IJCI/article/view/452

Archibald, M. M. (2018). Integrating the arts and mixed methods research: a review and a way forward. *International Journal of Multiple Research Approaches*, *10*(1). https://doi.org/10.29034/ijmra.v10n1a23

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brugar, K. A., Roberts, K. L., Jiménez, L. M., & Meyer, C. K. (2018). More than mere motivation: Learning specific content through multimodal narratives. *Literacy Research and Instruction*, *57*(2), 183–208. https://doi.org/10.1080/19388071.2017.1351586

Chatzichristofis, S. A., Tsopozidis, A., Kyriakidou-Zacharoudiou, A., Evripidou, S., & Amanatiadis, A. (2025). Designing an AI-Supported Framework for Literary Text Adaptation in Primary Classrooms. *AI*, *6*(7), 150. https://doi.org/10.3390/ai6070150

Chen, C. W. (2018). Developing EFL students' digital empathy through video production. *System*, *77*, 50–57. https://doi.org/10.1016/j.system.2018.01.006

Creswell, J. W., & Plano Clark, V. L. (2023). Revisiting mixed methods research designs twenty years later. *Handbook of Mixed Methods Research Designs*, *1*(1), 21–36. https://doi.org/10.4135/9781529614572.n6

Derakhshan, A. (2025). EFL students' perceptions about the role of generative artificial intelligence (GAI)-mediated instruction in their emotional engagement and goal orientation: A motivational climate theory (MCT) perspective in focus. *Learning and Motivation*, *90*, 102114. https://doi.org/10.1016/j.lmot.2025.102114

Fälth, L., Brkovic, I., Kerestes, G., Svensson, I., Hjelmquist, E., & Tjus, T. (2023). The effects of a multimodal intervention on the reading skills of struggling students: An exploration across countries. *Reading Psychology*, *44*(3), 225–241. https://doi.org/10.1080/02702711.2022.2141399

Goldenthal, E., Park, J., Liu, S. X., Mieczkowski, H., & Hancock, J. T. (2021). Not all AI are equal: Exploring the accessibility of AI-mediated communication technology. *Computers in Human Behavior*, *125*, 106975. https://doi.org/10.1016/j.chb.2021.106975

Hafner, C. A. (2015). Remix culture and English language teaching: The expression of learner voice in digital multimodal compositions. *Tesol Quarterly*, *49*(3), 486–509. https://doi.org/10.1002/tesq.238

Harun, R. N. S. R., & Singh, A. K. J. (2024). A systematic review of multimodal learning in higher education. *Innovative Pedagogical Practices for Higher Education 4.0*, 288–303. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003400691-18

Haule, D. B., Shaker, T., & Nkwera, G. Y. (2024). Assessing the influence of multimedia instruction on knowledge transfer in higher education: Multimodal learning experiences. *Indonesian Journal of Education and Pedagogy*, *1*(3), 142–160. https://doi.org/10.61251/ijoep.v1i3.77

Huang, M. (2025). Student engagement and speaking performance in AI-assisted learning environments: A mixed-methods study from Chinese middle schools. *Education and Information Technologies*, *30*(6), 7143–7165. https://link.springer.com/article/10.1007/s10639-024-12989-1

Ifinedo, E., & Kankaanranta, M. (2021). Understanding the influence of context in technology integration from teacher educators' perspective. *Technology, Pedagogy and Education*, *30*(2), 201–215. https://doi.org/10.1080/1475939X.2020.1867231

Jasmine, M. P., Silvhiany, S., & Inderawati, R. (2025). Exploring Young Learnersâ€™ Needs in Multimodal Text Design for Integrating Sustainable Education into ELT. *Journal of Languages and Language Teaching*, *13*(2), 785–797. https://doi.org/10.33394/jollt.v13i2.13775

Jovazino Bastos Medrado Costa, P. (2023). *Multimodal Digital Literacy Practices: Perspectives of L2 Academic Writing Instructors*. https://scholarcommons.sc.edu/etd/7485

Khan, M. A., Kurbonova, O., Abdullaev, D., Radie, A. H., & Basim, N. (2024). Is AI-assisted assessment liable to evaluate young learners? Parents support, teacher support, immunity, and resilience are in focus in testing vocabulary learning. *Language Testing in Asia*, *14*(1), 48. https://link.springer.com/article/10.1186/s40468-024-00324-x

Kim, J., Yu, S., Lee, S.-S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. *Journal of Research on Technology in Education*, 1–18. https://doi.org/10.1080/15391523.2025.2456043

Kress, G. (2019). Transposing meaning: translation in a multimodal semiotic landscape. In *Translation and multimodality* (pp. 24–48). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780429341557-2

Kundu, A., & Bej, T. (2025). Transforming EFL teaching with AI: A systematic review of

empirical studies. *International Journal of Artificial Intelligence in Education*, 1–34. https://link.springer.com/article/10.1007/s40593-025-00470-0

Lee, G., Shi, L., Latif, E., Gao, Y., Bewersdorff, A., Nyaaba, M., Guo, S., Liu, Z., Mai, G., & Liu, T. (2025). Multimodality of ai for education: Towards artificial general intelligence. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2025.3574466

Li, D., Xia, S., & Guo, K. (2025). *Using AI to enhance digital multimodal composing: EFL learners' semiotic decision-making, self-efficacy, enjoyment, and continuance intention*. https://doi.org/10.64152/10125/73636

Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Liu, Y., Jain, A., & Tang, J. (2022). Trustworthy ai: A computational perspective. *ACM Transactions on Intelligent Systems and Technology*, *14*(1), 1–59. https://doi.org/10.1145/3546872

Liu, Z., Lin, G., Tan, H. L., Zhang, H., Lu, Y., Gao, X., Yin, S. X., Sun, H., Goh, H. H., & Wong, L. H. (2025). SingaKids: A Multilingual Multimodal Dialogic Tutor for Language Learning. *ArXiv Preprint ArXiv:2506.02412*. https://doi.org/10.48550/arXiv.2506.02412

Maspul, K. A. (2025). ChatGPT as an Educative and Pedagogical Tool: Perspectives and Prospects in International Schools in Indonesia. *Innovative Technologica: Methodical Research Journal*. https://doi.org/10.47134/innovative.v4i2.137

Morgado, E., Leonido, L., Pereira, A., & Gouveia, L. B. (2025). *Technology-Mediated Education: impact of AI on the main distance learning modalities*. http://hdl.handle.net/10400.14/53824

Namaziandost, E., & Rezai, A. (2024). Artificial intelligence in open and distributed learning: Does it facilitate or hinder teaching and learning? *The International Review of Research in Open and Distributed Learning*, *25*(3), i–vii. https://doi.org/10.19173/irrodl.v25i3.8070

Nyrup, R., & Robinson, D. (2022). Explanatory pragmatism: a context-sensitive framework for explainable medical AI. *Ethics and Information Technology*, *24*(1), 13. https://link.springer.com/article/10.1007/s10676-022-09632-3

Park, H., Darko, J., Deshpande, N., Pandey, V., Su, H., Ono, M., Barkley, D., Folsom, L., Posselt, D., & Chien, S. (2022). Temporal multimodal multivariate learning. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3722–3732. https://doi.org/10.1145/3534678.3539159

Pellas, N. (2025). Effects of Generative AI Feedback and Interactive Video Assessment on Student Learning Achievement in Philological Content Creation Courses. *Journal of Educational Computing Research*, 07356331251372800. https://doi.org/10.1177/07356331251372800

Robinson, L., Schulz, J., Blank, G., Ragnedda, M., Ono, H., Hogan, B., Mesch, G. S., Cotten, S. R., Kretchmer, S. B., & Hale, T. M. (2020). Digital inequalities 2.0: Legacy inequalities in the information age. *First Monday*, *25*(7). https://dx.doi.org/10.5210/fm.v25i7.10842

Rudik, I. V, & Onyshchuk, I. Y. (2024). *Artificial inteligence tools for developing educational resources: enhancing digital learning experience for teachers and learners*. https://dspace.uzhnu.edu.ua/jspui/handle/lib/64493

Santoso, W., & Mangkuluhur, B. D. (2024). Pre-Service English Teachers' Perceptions Towards The Use of Dialogical Learning as Reflective Practice: A Case Study. *SALEE: Study of Applied Linguistics and English Education*, *5*(2), 493–507. https://doi.org/10.35961/salee.v5i2.1401

van Nooijen, C. C. A., de Koning, B. B., Bramer, W. M., Isahakyan, A., Asoodar, M., Kok, E., van Merrienboer, J. J. G., & Paas, F. (2024). A cognitive load theory approach to understanding expert scaffolding of visual problem-solving tasks: A scoping review. *Educational Psychology Review*, *36*(1), 12. https://link.springer.com/article/10.1007/s10648-024-09848-3

van Twillert, A., Kreijns, K., Vermeulen, M., & Evers, A. (2020). Teachers' beliefs to integrate Web 2.0 technology in their pedagogy and their influence on attitude, perceived norms, and perceived behavior control. *International Journal of Educational Research Open*, *1*, 100014. https://doi.org/10.1016/j.ijedro.2020.100014

Wallen, N. E., & Fraenkel, J. R. (2013). Experimental Research. In *Educational Research* (pp. 279–306). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781410601001-15

Wyatt, M. (2015). Using qualitative research methods to assess the degree of fit between teachers' reported self-efficacy beliefs and their practical knowledge during teacher education. *Australian Journal of Teacher Education (Online)*, *40*(1), 117–145. https://www.taylorfrancis.com/chapters/edit/10.4324/9781410618894-37

Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2025). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology*, *56*(4), 1463–1487. https://doi.org/10.1111/bjet.13530

Yang, C., & Cui, F. (2024). Multimodal mediation in translation and communication of Chinese

museum culture in the era of artificial intelligence. *Corpus-Based Studies across Humanities*, *1*(1), 51–77. https://doi/10.1515/csh-2023-0020

Yi, W., & Siqian, W. (2025). Digital Literacy Education Among Students: Status, Pathways, and Implications in the Current Digital Era. *International Journal of Education and Teaching Zone*, *4*(1), 66–85. https://doi.org/10.57092/ijetz.v4i1.385

Zhang, Y., & Peng, J. (2025). Implementing digital multimodal composing in tertiary EFL curriculum: a focus on teachers' beliefs. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2025.2497499

Zou, Y., Kuek, F., Feng, W., & Cheng, X. (2025). Digital learning in the 21st century: trends, challenges, and innovations in technology integration. *Frontiers in Education*, *10*, 1562391. https://doi.org/10.3389/feduc.2025.1562391