Ishel
International Seminar on Humanity, Education, and Language
Universitas Negeri Jakarta | Online Conference |
eISSN 3124-9396
Volume 2 , July 2026

THE INTEGRATION OF VISUAL GRAMMAR INTO GENRE-BASED APPROACH TO DEVELOP STUDENTS' MULTIMODAL VIEWING SKILLS

Alan Jaelani[1], Liliana Muliastuti[2], Samsi Setiadi[3]
Applied Linguistics, Universitas Negeri Jakarta, Indonesia
Email: alan.jaelani@mhs.unj.ac.id[1]; liliana.muliastuti@unj.ac.id[2]; syamsi.setiadi@unj.ac.id[3]

ABSTRACT
In contemporary EFL classrooms, students increasingly encounter multimodal texts that combine images, written language, layout, color, and digital visual resources. However, viewing skills are often insufficiently addressed because English instruction remains largely focused on verbal literacy. This study investigated the integration of visual grammar into genre-based approach to improve students' viewing skills. An embedded mixed-methods design with a quasi-experimental non-equivalent pre-test–post-test control group was employed. The participants were 72 Grade 9 junior high school EFL students from a public junior high school in Bogor, Indonesia, divided into experimental and control groups. The experimental group received instruction through the teaching and learning cycle integrated with visual grammar while the control group received regular instruction. Data were collected through viewing skills tests, classroom observations, questionnaires, interviews, and students' learning artifacts. The findings showed that the experimental group improved from 56.42 to 76.83, with a mean gain of 20.41 and a moderate N-Gain of 0.47, while the control group improved from 55.81 to 65.28, with a mean gain of 9.47 and a low N-Gain of 0.21. An independent-samples post-test comparison indicated a significant advantage for the experimental group, $t(69.66) = 6.64$, $p < .001$, Hedges' $g = 1.55$. Qualitative findings further showed that students perceived the model as clear, engaging, and useful, although abstract visual grammar concepts such as salience, framing, perspective, and communicative purpose required repeated modelling. The study suggests that integrating visual grammar into GBA offers a structured and pedagogically meaningful model for teaching, practising, and assessing multimodal viewing skills in junior high school EFL classrooms.

Keywords: EFL learning; genre-based approach; multimodal literacy; viewing skills; visual grammar

INTRODUCTION
In the 21st century, communication has evolved into increasingly multimodal forms, where meaning is constructed through the integration of linguistic, visual, spatial, and digital resources. This transformation has been driven by rapid technological advancement and the proliferation of digital media which require learners to engage with complex multimodal texts in both academic and everyday contexts. As a result, multimodal literacy has become a crucial competence in contemporary education particularly in English as a Foreign Language (EFL) settings (Carcamo & Pino, 2025). Learners are no longer expected to rely solely on traditional language skills but must

also develop the ability to interpret and critically evaluate visual and multimodal representations.

One key component of multimodal literacy is viewing skills, which refer to the ability to understand, interpret, analyze, and evaluate visual and multimodal texts (Brunfaut & Kormos, 2025). Recent studies have emphasized that viewing is not a passive activity but an active meaning-making process that involves interpreting the interplay between visual and verbal elements (Liu et al., 2025). However, despite its importance, viewing skills remain underdeveloped in many EFL contexts, particularly at the junior high school level, where instruction tends to prioritize reading and writing over multimodal competencies. Research indicates that integrating multimodal resources into language instruction can enhance students' engagement and comprehension (Xiong et al., 2022). Multimodal instructional approaches provide multiple pathways for understanding which allows learners to access meaning through various semiotic modes (Singer et al., 2025). Furthermore, multimodal learning environments have been shown to support higher-order thinking skills, such as analysis and evaluation, by encouraging learners to interpret complex relationships between different modes of communication (Shen & Han, 2026). Despite these benefits, classroom practices often fail to systematically incorporate multimodal analysis resulting in superficial engagement with visual materials.

The genre-based approach (GBA) has been widely recognized as an effective pedagogical framework in EFL education. Rooted in systemic functional linguistics, GBA emphasizes the relationship between language, context, and social purpose, guiding students through structured stages of learning (Anderson, 2025). While GBA has been successfully applied to improve students' reading and writing skills, its implementation has largely focused on monomodal texts, limiting its relevance in addressing the multimodal demands of contemporary communication (Kessler & Casal, 2024). In parallel, Kress & van Leeuwen's (2020) visual grammar has emerged as a theoretical framework for analyzing how images convey meaning through representational, interactive, and compositional elements. This framework provides analytical tools that can help learners understand how visual texts construct meaning and communicate messages (Elmiana & Shen, 2025). However, despite its potential, visual grammar is rarely integrated into classroom instruction in a systematic manner. There remains a gap between theoretical advancements in multimodal discourse analysis and their practical application in EFL pedagogy.

Previous studies on multimodal literacy have explored various aspects, including image-text relationships (Chen et al., 2025; Zhang et al., 2025), digital multimodal composition (Herzog et al., 2025; Wünsch-Nagy, 2025; Xia, 2024), and multimodal assessment (Beltrán-Palanques, 2024; Fjørtoft, 2020; Jiang et al., 2024). However, much of this research has been descriptive and has not resulted in pedagogical models that can be readily implemented in classroom settings. Moreover, many studies have been conducted in higher education contexts leaving a gap in research focusing on younger learners. These limitations imply the need for pedagogical innovation that integrates multimodal theory with classroom practice. Despite growing recognition of the importance of multimodal literacy, several critical gaps remain in both research and practice. First, EFL pedagogy continues to be dominated by monomodal approaches that prioritize linguistic features while neglecting visual and multimodal dimensions of meaning. This imbalance limits students' ability to interpret multimodal texts, which are increasingly prevalent in digital and educational contexts. Second, although multimodal

eISSN 3124-9396
Volume 2 , July 2026

instructional approaches have verified positive effects on learning outcomes, their implementation is often fragmented and lacks a coherent pedagogical framework. Teachers may use images or videos in their instruction; however, these resources are rarely accompanied by explicit strategies for analyzing visual meaning. Consequently, students may engage with multimodal texts at a surface level without developing deeper analytical skills. Third, the integration of visual grammar into language instruction remains limited. Existing studies have primarily focused on theoretical discussions and isolated applications, without embedding visual grammar within a structured instructional model. This lack of integration prevents learners from fully understanding how visual and linguistic elements interact to create meaning. In response to these gaps, this study aims to investigate the integration of visual grammar into genre-based approach and its impact on EFL students' viewing skills. The research questions are formulated as follows:

1. How is visual grammar integrated into the teaching and learning cycle of genre-based approach in EFL classroom practice?
2. To what extent does the integration of visual grammar into genre-based approach improve students' viewing skills?
3. What are students' perceptions of the implementation of visual grammar-integrated genre-based instruction?

This study gives contribution to the field of EFL education by offering a practical and theoretically grounded model for integrating multimodal literacy into classroom instruction. The study also provides empirical evidence on the effectiveness of multimodal pedagogy, addressing the gap between theory and practice in language education.

LITERATURE REVIEW

Multimodal Literacy and Multimodal Pedagogies

The rapid expansion of digital communication has transformed literacy practices from predominantly monomodal to inherently multimodal. Contemporary texts integrate linguistic, visual, spatial, and audio resources, requiring learners to engage in complex meaning-making processes (Nathues et al., 2025). This shift has led to increasing recognition of multimodal literacy as a core competence. Multimodal literacy refers to the capacity to engage with and construct meanings through multiple forms of representation. According to Lim et al. (2022), multimodal texts incorporate a variety of semiotic resources, including written language, speech, gesture, images, audio, video, graphs, charts, and physical objects. These semiotic modes may function independently or interact collectively within communicative ensembles to establish connections across contexts, spaces, and temporal dimensions (Pun & Cheung, 2023). Multimodal literacy involves the technical ability to interpret, create multimodal texts and the development of semiotic awareness that enables learners to understand how meaning is shaped through the interaction of multiple modes in contemporary communication.

In response to the growing significance of multimodal communication, researchers have developed pedagogical metalanguages and theoretical frameworks to support the teaching and analysis of multimodal texts. These include frameworks addressing print media, film, image–text relations, and digital multimodal composing.

Scholars such as Unsworth (2021), Lim et al. (2022) and others have contributed extensively to the conceptualization of intersemiotic relations and multimodal meaning-making in educational contexts. Additional models, such as the semiotic modes framework and multimodal pedagogic text models (Oyama, 2021), have further facilitated the formal incorporation of multimodal literacy into classroom instruction.

The term multimodal pedagogies refer to instructional approaches through which teachers design learning experiences that support students' multimodal literacy development (Tseng et al., 2025). Multimodal pedagogies involve strategic decisions regarding the selection, organization, and sequencing of representational modes for specific curricular purposes (Lim et al., 2022; Tseng et al., 2025). These pedagogical practices also include opportunities for students to produce multimodal compositions that integrate various semiotic resources. The enactment of multimodal pedagogies requires teachers to orchestrate diverse representations and explanatory elements into coherent instructional experiences while employing meaning-making resources appropriately to enhance students' learning processes.

Research has shown that the integration of multimodal texts can increase students' engagement, comprehension, and retention in language learning contexts (Liu et al., 2025; Valente et al., 2024). Multimodal composition activities were found to foster higher-order critical and analytical thinking among lower-progress learners, extending beyond the limitations of traditional literacy practices (Shen & Han, 2026). Similar findings were observed in studies involving English as a Foreign Language (EFL) learners, where multimodal composing activities enhanced learners' motivation, confidence, digital empathy, and opportunities for self-expression (Lei & Zhang, 2024; Tseng, 2025). Students also responded positively to classroom environments that provided diverse meaning-making options through multimodal pedagogical designs.

Viewing Skills in EFL Contexts

Viewing refers to the active and critical process of interpreting meaning from visual and multimodal texts. It requires learners to examine how visual, verbal, spatial, gestural, and sometimes auditory modes work together to produce meaning (Hiippala, 2021). In this sense, viewing should be understood as a higher-order literacy skill. It requires students to notice visual signs, identify relationships among semiotic resources, infer implied meanings, interpret cultural references, and evaluate communicative purposes. Eye-tracking research on multimodal texts further suggests that readers need cognitive flexibility when engaging with multimodal materials because they must shift attention across textual, visual, and sometimes auditory elements while determining which information is most relevant for comprehension (Chen & Weninger, 2025). Thus, students with strong viewing skills do not merely observe visual texts at a surface level. They analyze how meaning is constructed through layout, salience, color, gaze, framing, image-text relations, cultural symbols, and implicit messages.

An analytical framework for understanding how visual and verbal modes interact is offered by Martinec & Salway (2005). Their model explains image-text relations through two major dimensions: status and logico-semantics (see figure 1). Status concerns the degree of dependency between image and text. In an equal relation, image and text may be independent, meaning that each mode can communicate meaning autonomously, or complementary, meaning that both modes depend on each other to construct a unified multimodal meaning. In an unequal relation, one mode becomes more dominant than the other. The image may be subordinate to the text when it mainly

eISSN 3124-9396
Volume 2 , July 2026

illustrates or supports verbal information, while the text may be subordinate to the image when it functions to clarify, label, or elaborate visual elements.

The dimension of logico-semantics explains the type of semantic relation that connects image and text. Drawing on Halliday's systemic functional grammar, Martinec & Salway (2005) classify these relations into expansion and projection. Expansion occurs when one mode develops, extends, or enriches the meaning of another mode. It includes elaboration, extension, and enhancement. Elaboration happens when image and text restate, specify, or exemplify similar meanings. Extension occurs when one mode adds new but related information to the other. Enhancement provides circumstantial meanings, such as time, place, cause, reason, or purpose. Projection occurs when one mode represents speech, wording, thought, or idea expressed through another mode, such as speech in comics, captions that project a speaker's voice, or visual sequences that represent imagined meanings.

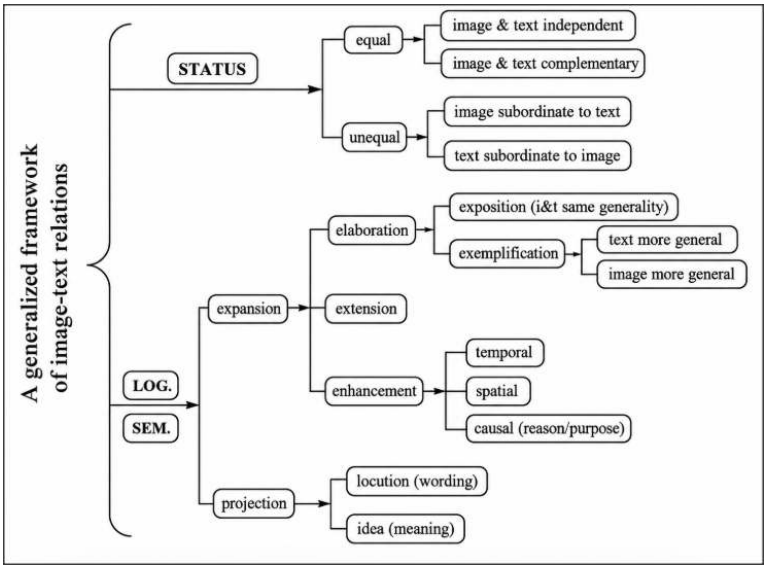

Figure 1. Image-text relations and logico-semantics (Martinec & Salway, 2005)

The framework is used for the study of viewing skills as it provides a systematic way to analyze how students interpret multimodal texts. By using the model, viewing can be assessed through students' ability to identify visual elements and also through their ability to explain how image and text relate to each other, whether one mode dominates the other, and how both modes jointly construct meaning (Brunfaut & Kormos, 2025; Kaderoğlu, 2026; Nguyen & Habók, 2024).

Visual Grammar as a Framework for Multimodal Meaning-Making

Visual grammar, developed by Kress & van Leeuwen (2006, 2021), provides a systematic framework for analysing how images construct meaning. Rooted in systemic functional linguistics, it extends the metafunctional principles of ideational, interpersonal, and textual meaning into the visual domain (Ma & Jiang, 2025). The representational metafunction concerns how visual elements depict participants, actions,

events, symbols, and circumstances. The interactive metafunction concerns how images position viewers through gaze, social distance, angle, and perspective. The compositional metafunction concerns how visual elements are arranged through layout, information value, salience, and framing. These categories are useful because they make visual meaning visible, nameable, and teachable for EFL learners.

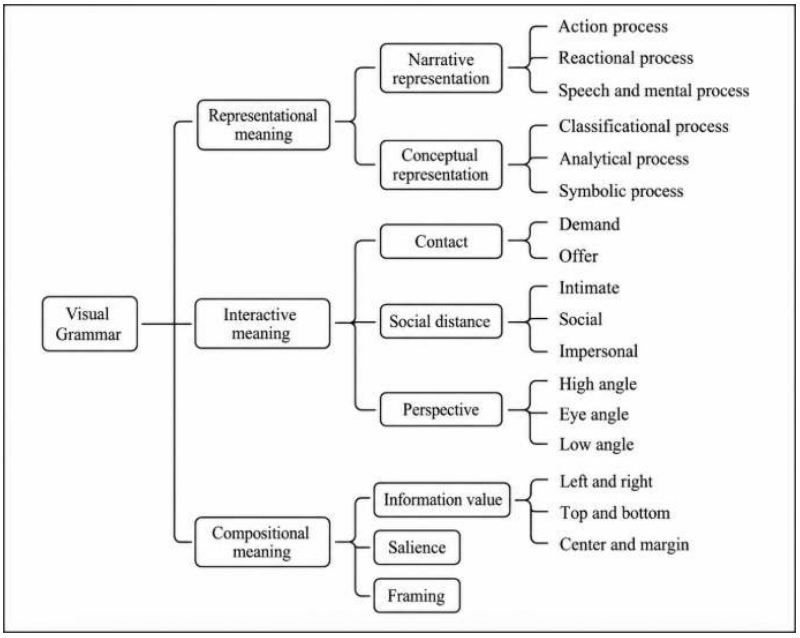

Figure 2. The Framework of Visual Grammar (Kress & van Leeuwen, 2006)

Visual grammar provides valuable analytical tools for developing students' viewing skills. Learners can move beyond surface-level interpretation and engage in deeper analysis of how meaning is constructed visually. However, despite its theoretical robustness, visual grammar has not been widely integrated into classroom practice. Many studies remain at the level of discourse analysis, with limited translation into pedagogical strategies. Consequently, there is a need to embed visual grammar within instructional models that can guide teachers and learners in systematically analyzing multimodal texts.

Genre-Based Approach in EFL Pedagogy

The Genre-Based Approach (GBA) has been widely adopted in EFL education as a framework for teaching language as a social semiotic system. Grounded in Systemic Functional Linguistics, GBA emphasizes the relationship between language, context, and social purpose (Anderson, 2025). It views texts as goal-oriented social processes that are structured according to recognizable stages (Sultan et al., 2025). A central feature of GBA is its pedagogical cycle, which typically consists of four stages: Building Knowledge of the Field (BKOF), Modeling of the Text (MoT), Joint Construction of the Text (JCoT), and Independent Construction of the Text (ICoT). In the BKOF stage, learners develop background knowledge and familiarize themselves with the context of the text. The MoT stage involves analyzing model texts to understand their structure and language features. In the JCoT stage, learners

eISSN 3124-9396
Volume 2 , July 2026

collaboratively construct texts with teacher guidance, while the ICoT stage requires learners to produce texts independently.

GBA has proven effective in improving students' reading and writing skills by providing structured scaffolding and explicit instruction. However, its application has traditionally focused on monomodal texts, particularly written language. As communication becomes increasingly multimodal, there is a growing need to adapt GBA to address the integration of visual and other semiotic modes. This adaptation requires not only incorporating multimodal texts into instruction but also developing analytical frameworks that can guide students in interpreting and producing multimodal meaning.

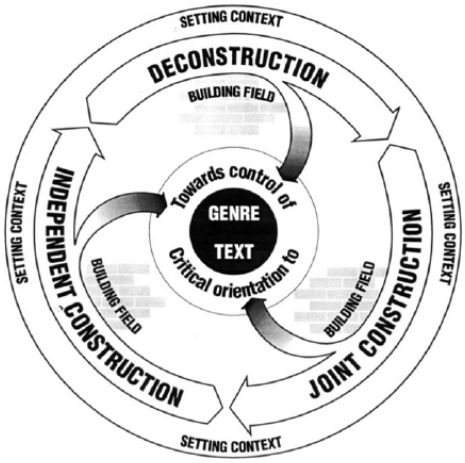

Figure 3. Teaching and Learning Cycle (Macnaught et al., 2013)

The integration of visual grammar into the genre-based approach offers a coherent bridge between multimodal theory and classroom practice. Visual grammar provides the analytical metalanguage for examining how images represent experience, position viewers, and organise meaning (Aleksić-Hajduković, 2025), while the genre-based approach provides a scaffolded pedagogical cycle through which such knowledge can be gradually developed (Pham & Bui, 2021).

METHODOLOGY

Research Design

This study employed an embedded mixed-methods design with a quasi-experimental non-equivalent pretest-posttest control group. This design was used because the study aimed to examine both the quantitative improvement of students' viewing skills and the qualitative process of implementing visual grammar within genre-based approach. The quantitative component measured the effect of the instructional model, while the qualitative component explained how the model was enacted in classroom practice and how students responded to it. Mixed-methods design enables researchers to integrate numerical and descriptive data to obtain a more comprehensive understanding of educational intervention outcomes (Creswell & Creswell, 2018).

Participants and Sampling

The participants were 72 Grade 9 EFL students from a public junior high school in Bogor, Indonesia. Two intact classes were selected because the school did not permit random assignment of individual students. One class was assigned as the experimental group (n = 36), while the other was assigned as the control group (n = 36). This class-based assignment is consistent with the quasi-experimental non-equivalent group design. The two classes were comparable at the beginning of the study, as indicated by their similar pre-test mean scores: 56.42 for the experimental group and 55.81 for the control group. Two English teachers were also involved to support classroom implementation and observation.

Data and Data Source

The data consisted of students' viewing skills scores, classroom observation notes, questionnaire responses, semi-structured interview data, and students' multimodal learning artefacts. Quantitative data were obtained from pre-test and post-test viewing skills scores and from the student perception questionnaire. Qualitative data were obtained from classroom observations, interviews, and students' worksheets or learning products. The learning materials focused on procedural multimodal texts as this genre commonly combines sequential language, images, icons, numbers, arrows, layout, and visual emphasis, making it suitable for teaching image-text relations and visual grammar.

Instruments

The instruments used in this study were multimodal viewing skills test, analytic scoring rubric, observation checklist, student perception questionnaire, semi-structured interview guide, and a learning artefact analysis sheet. The viewing skills test was administered as a pre-test and post-test to measure students' ability to interpret visual and multimodal texts. The test covered five dimensions: visual recognition, visual grammar analysis, image-text interpretation, inferential meaning-making, and critical evaluation. The analytic rubric was used to score students' responses based on these dimensions. The observation checklist documented the implementation of each stage of the genre-based approach, while the questionnaire measured students' perceptions of clarity, engagement, usefulness, and learning difficulty. The interview guide explored students' learning experiences, and the artefact analysis sheet was used to examine students' worksheets and multimodal analysis tasks. Viewing skills in this study were conceptualised as a multidimensional construct, operationalised as shown in Table 1.

Table 1. Operationalisation of Multimodal Viewing Skills

Dimension	Indicator	Assessment Focus
Visual recognition	Identifying visual elements such as participants, objects, colour, symbols, captions, and layout	Accuracy of observation
Visual grammar analysis	Explaining representational, interactive, and compositional meanings	Use of visual evidence and metalanguage
Image-text interpretation	Explaining relationships between images and written language	Quality of intersemiotic explanation
Inferential meaning-making	Inferring implied, symbolic, emotional, or cultural meanings	Reasoned inference based on evidence
Critical evaluation	Evaluating purpose, bias, reliability, perspective, and persuasion	Depth of critique and justification

eISSN 3124-9396
Volume 2 , July 2026

Through these indicators, students' viewing skills are assessed by both their ability to recognize visual features and their capacity to explain how visual and verbal modes interact, infer implicit meanings, and critically evaluate the purpose and credibility of multimodal communication.

Data Collection Procedure

Data collection was conducted in five stages. First, the researcher obtained permission, prepared the instructional materials, developed the instruments, and trained the teachers or observers to use the observation checklist and rubric. Second, both the experimental and control groups completed the viewing skills pre-test to establish their initial competence. Third, the experimental group participated in visual grammar-integrated GBA instruction, while the control group received regular EFL instruction with comparable procedural text topics. Fourth, classroom observations were conducted to document how each stage of the teaching and learning cycle was implemented. Fifth, both groups completed the post-test, and the experimental group completed the questionnaire and semi-structured interviews.

Table 2. Instructional Intervention Design

No.	GBA Stage	Main Focus	Learning Activity	Output
1	BKoF	Genre context and visual recognition	Students identify the purpose, audience, topic, and visible elements of multimodal procedural texts	List of visual and verbal elements
2	MoT	Representational meaning	Teacher explains how actions, participants, symbols, and settings create meaning	Guided analysis worksheet
3	JCoT	Image-text relations	Teacher and students interpret how images, captions, icons, and numbers work together	Collaborative interpretation
4	ICoT	Complete viewing skills	Students independently analyse a new multimodal text using the rubric	Individual analysis task

Data Analysis Procedure

Quantitative data were analysed using descriptive and inferential statistics. Descriptive statistics, including mean scores, standard deviations, mean gains, and N-Gain scores, were used to describe students' viewing skills before and after the intervention. Pre-test equivalence and post-test group differences were examined through independent-samples t-tests using group means, standard deviations, and sample sizes. The magnitude of the post-test difference was interpreted using Hedges' g because both groups had equal but relatively modest sample sizes. Qualitative data from observations, interviews, and students' artefacts were analysed thematically by identifying recurring patterns related to implementation, engagement, perceived usefulness, and learning difficulty (Braun & Clarke, 2022). Quantitative and qualitative findings were then integrated to explain whether the model improved students' viewing skills and how the model functioned in classroom practice.

Validity and Reliability

Instrument validity was established through expert judgement involving specialists in EFL pedagogy, multimodal literacy, and educational assessment. The

experts reviewed the relevance, clarity, representativeness, and age-appropriateness of the test items, questionnaire statements, observation indicators, interview prompts, and analytic rubric. Content validity was strengthened by aligning each instrument with the research objectives and the five dimensions of multimodal viewing skills. Reliability was addressed through internal consistency analysis for the questionnaire and test components, and through inter-rater checking for rubric-based responses. The scoring rubric was also discussed by the raters before use to minimise scoring inconsistencies. These procedures were necessary because multimodal literacy assessment requires clear construct representation and systematic scoring criteria (DeVellis, 2017).

RESULTS AND DISCUSSION

The Integration of Visual Grammar into Genre-Based Approach

The integration of visual grammar into the GBA was organised through the teaching and learning cycle. The model guided students from contextual awareness to independent multimodal interpretation. Classroom implementation showed that visual grammar and image-text relation analysis were not introduced as abstract theories but as practical tools for reading multimodal procedural texts. Students were gradually guided to identify visible elements, analyse visual meaning, interpret image-text relations, infer implied meanings, and evaluate communicative purposes.

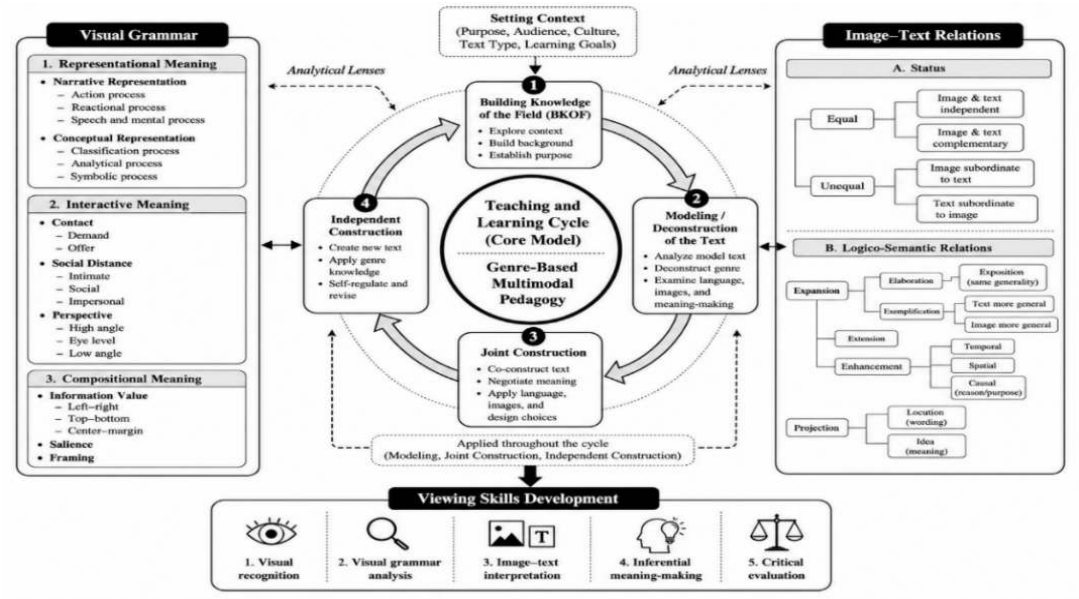

Figure 4. The integration of visual grammar into GBA

Visual grammar and image-text relations were embedded within classroom practice rather than treated as separate theoretical content. Through this integration, students learned to read multimodal texts by examining how images, written language, layout, colour, numbering, arrows, icons, and other visual elements work together to create meaning. The model began with setting the context. At this stage, students were introduced to the purpose, audience, cultural background, text type, and learning goals of the multimodal text. This step was necessary because visual texts cannot be interpreted only by naming visible objects. Students needed to understand why the text

eISSN 3124-9396
Volume 2 , July 2026

was produced, who it addressed, and how its genre shaped meaning. Context therefore became the foundation for deeper analysis.

The first main stage was Building Knowledge of the Field. Students explored the topic and activated prior knowledge by identifying participants, objects, setting, colours, symbols, captions, layout, and other visual features. The second stage was Modeling or Deconstruction of the Text. This was the central phase for introducing representational, interactive, and compositional meanings. Students also learned image-text relations by identifying whether images and written texts were equal, complementary, independent, or unequal, and whether one mode elaborated, extended, enhanced, or projected another mode. The third stage was Joint Construction. Students applied visual grammar collaboratively with teacher support, discussed multimodal texts, negotiated interpretations, and justified their answers using visual evidence. The fourth stage was Independent Construction. Students independently identified visual elements, analysed visual grammar features, interpreted image-text relations, inferred implied meanings, and evaluated communicative purposes.

Table 3. Empirical Indicators of Model Implementation

GBA Stage	Observed Teacher Support	Student Learning Activity	Viewing Skill Developed
BKOF	Guiding students to notice visible elements	Listing participants, objects, captions, colour, layout, and symbols	Visual recognition
MoT	Modelling visual grammar categories	Analysing action, gaze, angle, salience, framing, and layout	Visual grammar analysis
JCoT	Prompting collaborative interpretation	Explaining image-text relations using evidence	Image-text interpretation and inference
ICoT	Providing an independent task and rubric	Producing individual multimodal interpretations	Inference and critical evaluation

The implementation confirmed that viewing skills did not develop merely through exposure to images. Students needed repeated modelling, guided questions, peer discussion, and explicit metalanguage to move from surface observation to analytical interpretation. The model therefore positioned visual grammar as the main lens for understanding how images construct meaning and positioned GBA as the scaffolding structure through which that lens could be learned gradually.

Improvement of Students' Multimodal Viewing Skills

The improvement in students' multimodal viewing skills was measured through pre-test and post-test scores administered to both the experimental and control groups. The test assessed five dimensions: visual recognition, visual grammar analysis, image-text interpretation, inferential meaning-making, and critical evaluation. As presented in Table 4, both groups improved from pre-test to post-test. However, the experimental group showed a substantially higher improvement than the control group. The experimental group increased from 56.42 to 76.83, resulting in a mean gain of 20.41 points, whereas the control group increased from 55.81 to 65.28, with a mean gain of

9.47 points. The small pre-test difference of 0.61 points indicates that both groups had relatively comparable initial viewing skills before the intervention.

Table 4. Descriptive Statistics of Students' Viewing Skills

Group	N	Pre-test Mean	SD	Post-test Mean	SD	Mean Gain	N-Gain	Category
Experimental Group	36	56.42	7.85	76.83	7.12	20.41	0.47	Moderate
Control Group	36	55.81	8.12	65.28	7.64	9.47	0.21	Low

The N-Gain result supports the descriptive pattern. The experimental group obtained an N-Gain score of 0.47, categorised as moderate, while the control group obtained an N-Gain score of 0.21, categorised as low. This indicates that explicit instruction in visual grammar and image-text relations provided more meaningful learning gains than regular EFL instruction. The pre-test equivalence and post-test difference were further examined using independent-samples t-tests based on the reported means, standard deviations, and sample sizes.

Table 5. Independent-Samples Comparison of Viewing Skills Scores

Comparison	Mean Difference	t	df	p	Effect Size	Interpretation
Pre-test: experimental vs. control	0.61	0.32	69.92	.747	d = 0.08	Groups were comparable before treatment
Post-test: experimental vs. control	11.55	6.64	69.66	< .001	Hedges' g = 1.55	Experimental group outperformed control group

The improvement across the five dimensions of viewing skills is shown in Table 6. All dimensions improved in both groups, but the experimental group consistently achieved higher gains. The highest gain appeared in image-text interpretation, which increased from 54.15 to 77.80, with a gain of 23.65 points. This suggests that the model was particularly effective in helping students explain how images and written language jointly construct meaning. The second highest gain was found in visual grammar analysis, which increased from 52.30 to 75.60, with a gain of 23.30 points. This indicates that students became more able to analyse representational, interactive, and compositional meanings in multimodal texts. Visual recognition also improved substantially, increasing from 61.25 to 82.10, with a gain of 20.85 points. This was the highest post-test score among all dimensions in the experimental group, indicating that students became more attentive to visible elements such as participants, objects, symbols, colours, setting, captions, and layout. Inferential meaning-making increased from 56.80 to 74.90, with a gain of 18.10 points. The lowest gain appeared in critical evaluation, which increased from 57.60 to 73.75, with a gain of 16.15 points. This result suggests that evaluating bias, perspective, reliability, and communicative purpose remained the most demanding dimension and required more intensive scaffolding.

Table 6. Improvement across Viewing Skills Dimensions

Dimension	Exp. Pre-test	Exp. Post-test	Gain	Control Pre-test	Control Post-test	Gain
Visual Recognition	61.25	82.10	20.85	60.70	68.40	7.70
Visual Grammar Analysis	52.30	75.60	23.30	51.85	62.15	10.30
Image-Text Interpretation	54.15	77.80	23.65	53.90	64.25	10.35
Inferential Meaning-Making	56.80	74.90	18.10	56.25	64.70	8.45
Critical Evaluation	57.60	73.75	16.15	56.35	66.90	10.55

The control group also improved across all dimensions, but the gains were smaller. This pattern suggests that regular EFL instruction may support general comprehension, but it does not provide systematic analytical tools for interpreting visual and multimodal meaning. In contrast, the experimental model enabled students to move from simple visual observation to more analytical, inferential, and evaluative engagement with multimodal texts.

Students' Perceptions of the Implementation of the Model

Students' perception data were obtained from the questionnaire and semi-structured interviews with selected students from the experimental group. The questionnaire focused on four aspects: clarity of instruction, learning engagement, perceived usefulness, and learning difficulty. Most students gave positive responses to the model. In learning engagement, 83.3% of students selected either agree or strongly agree. Similarly, 83.4% of students gave positive responses to perceived usefulness, and 77.8% gave positive responses to clarity of instruction. These findings indicate that the model was generally acceptable and understandable for junior high school EFL students.

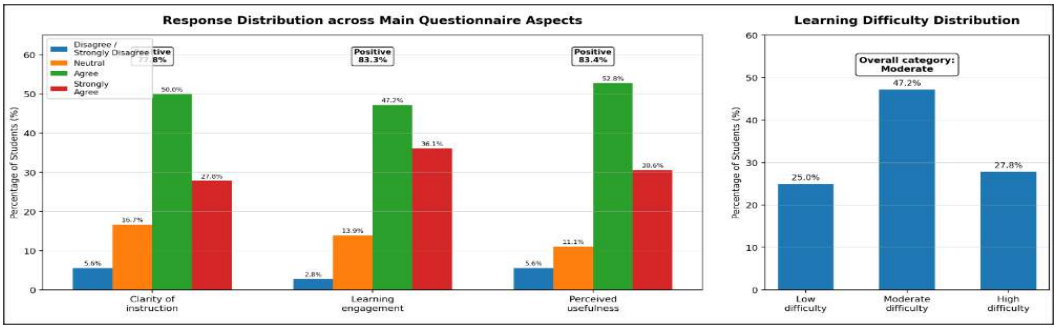

Figure 5. Distribution of Students' Questionnaire Responses

The distribution data show that most students perceived the model positively in terms of engagement and usefulness. Nevertheless, the learning difficulty data suggest that some students still struggled with the analytical metalanguage of visual grammar. This difficulty does not reduce the pedagogical value of the model; rather, it indicates that visual grammar needs to be introduced gradually through repeated modelling, guided practice, concrete examples, and simplified classroom language.

Table 7. Qualitative Themes from Students' Responses

Theme	Core Meaning	Evidence	Pedagogical Interpretation
Clarity of instruction	Students understood the stages of analysis more easily	Students reported that the steps helped them move from seeing to explaining	The GBA cycle made viewing instruction explicit
Learning engagement	Picture-based and collaborative tasks increased participation	Students enjoyed discussing posters and noticing details with peers	Multimodal texts connected learning with everyday visual experience
Perceived usefulness	Students became aware of image-text connections	Students recognised that images can clarify, extend, or strengthen verbal messages	The model supported intersemiotic interpretation
Learning difficulty	Abstract visual grammar terms were challenging	Students found salience, framing, hidden message, and purpose difficult	More modelling and simpler metalanguage are needed

The first theme emerging from the questionnaire and interview data was clarity of instruction. Students reported that the learning stages helped them understand how to analyze multimodal texts step by step. The teaching cycle provided a structured sequence in which students first explored the context, then analyzed model texts, worked collaboratively, and finally interpreted texts independently.

> *"Before this lesson, I usually only said what I saw in the picture. Now I know that the position, color, and the way a person looks at us can also give meaning."* (S5)
> *"The steps made it easier. First, we discussed the topic, then the teacher showed how to analyze the poster, and after that we tried it together."* (S11)

These excerpts show that students perceived the model as helpful because it made the process of viewing more explicit. The structured stages prevented students from relying only on personal impressions. Instead, they were guided to observe, analyze, interpret, and evaluate multimodal texts more systematically. The second theme was learning engagement. The questionnaire data showed that this aspect obtained the highest mean score (M = 4.36). Students reported that learning with picture-based texts was more interesting than learning only through written texts. The multimodal materials were perceived as closer to students' everyday experiences, especially because they frequently encounter images, captions, posters, and digital content in daily life.

> *"The lesson was more interesting because we did not only read sentences. We looked at posters and pictures, then discussed what they meant."* (S14)
> *"I liked the group discussion because my friends noticed things that I did not see. It helped me understand the picture better."* (S22)

Students were not positioned as passive viewers but as active interpreters who discussed and negotiated meanings with their peers. This suggests that multimodal pedagogy can increase classroom participation when students are given clear tasks and opportunities to explain their interpretations. The third theme was perceived usefulness. Students stated that the model helped them understand the relationship between images

eISSN 3124-9396
Volume 2 , July 2026

and written language. They became more aware that images and texts do not always repeat the same meaning. In some cases, the image supports the text; in other cases, the image adds new information, clarifies the message, or creates a stronger emotional effect.

> *"I learned that the picture and the sentence are connected. Sometimes the picture explains something that the words do not say directly."* (S7)
> *"When I looked at the picture, I could understand the message better because I learned to see the title, picture, number, and color together."* (S18)

These responses show that students perceived the model as useful for developing image-text interpretation. The model helped them understand multimodal texts as integrated meaning-making resources rather than separate visual and verbal elements. This finding supports the quantitative result in which image-text interpretation showed the highest gain among the five dimensions of viewing skills. Although students responded positively to the model, they also reported several difficulties. The most common difficulty was related to abstract concepts in visual grammar, particularly salience, framing, perspective, bias, and communicative purpose. Students could identify visible elements more easily, but they found it more challenging to explain hidden meanings or evaluate the intention behind a multimodal text.

> *"The difficult part was explaining the hidden message. I could see the picture, but I was not always sure what the purpose was."* (S9)
> *"Words like salience and framing were new for me. I understood them better after the teacher gave examples."* (S27)

These excerpts indicate that the difficulty was not caused by the use of multimodal texts themselves, but by the analytical demand required to interpret them critically. This finding shows that students need repeated exposure to visual grammar concepts. The concepts should be introduced through simple examples before students are asked to evaluate more complex multimodal texts.

DISCUSSION

The findings show that integrating visual grammar into genre-based approach offers a structured and effective model for developing junior high school EFL students' viewing skills. The model does not simply add pictures, videos, or digital texts to English instruction, but turns visual and image-text analysis into an explicit learning process. By placing the GBA's teaching and learning cycle at the center of instruction, students are guided gradually from contextual understanding to model analysis, collaborative interpretation, and independent meaning-making. This supports the view that multimodality becomes pedagogically meaningful only when learners are explicitly guided to interpret how semiotic modes interact (Jiang et al., 2022; Lim et al., 2022).

The first major finding concerns the productive integration of visual grammar into the stages of genre-based pedagogy. In the modeling stage, students were introduced to representational, interactive, and compositional meanings. In Joint Construction, they collaboratively applied these categories to analyze multimodal texts. In Independent Construction, they used the same framework individually. This indicates that visual

grammar becomes more teachable when presented not as abstract theory, but as practical classroom metalanguage. This finding is consistent with studies emphasizing the need to connect multimodal theory with scaffolded classroom practice (Jiang et al., 2022; Xia, 2024).

The quantitative results confirm the positive effect of the model on students' viewing skills. The results indicate that the intervention produced stronger learning gains than regular EFL instruction. The large effect size further shows that the model was not only statistically significant but also pedagogically meaningful. This improvement can be explained by the explicit analytical support provided in the experimental class. Students were trained to identify visual elements, analyze visual grammar, interpret image-text relations, infer implied meanings, and evaluate communicative purposes. Viewing, therefore, was treated as a complex literacy construct involving visual, verbal, cognitive, and interpretive processes rather than a passive act of looking at images (Brunfaut & Kormos, 2025). This also supports Carcamo & Pino (2025) finding that multimodal resources, such as infographics, can support EFL learning when used as structured tools for comprehension and meaning-making. The dimensional analysis shows that the strongest improvement occurred in image-text interpretation, followed by visual grammar analysis. This suggests that the model effectively helped students understand how images and written language jointly construct meaning. Students became more capable of recognizing whether visual and verbal modes were complementary, independent, or unequal, and whether one mode elaborated, extended, enhanced, or projected another. Contemporary EFL texts increasingly require learners to process visual and verbal information simultaneously, a process that depends on the coordination of linguistic, visual, and higher-order cognitive skills (Liu et al., 2025).

The improvement in visual grammar analysis confirms the value of representational, interactive, and compositional categories as classroom metalanguage. Students moved beyond surface-level description and began explaining how meaning is shaped through action, gaze, angle, salience, layout, and framing. This finding aligns with Ma & Jiang (2025) argument that visual elements guide attention, structure meaning, and engage audiences. The lower gain in critical evaluation indicates that this dimension remained the most challenging. Evaluating communicative purpose, bias, perspective, and reliability requires students to move beyond recognition and interpretation toward judgment. Such skills demand higher-order thinking, cultural awareness, and repeated reflective practice. Thus, the lower gain should not be seen as a weakness of the model, but as evidence that critical viewing requires longer exposure and stronger scaffolding (Brunfaut & Kormos, 2025; Shen & Han, 2026).

Students' perspectives strengthened the quantitative findings. They perceived the model as clear, engaging, and useful. The use of picture-based texts and collaborative analysis encouraged participation and helped students recognize that color, position, gaze, layout, and image-text relations carry meaning. This supports Tseng's (2025) finding that multimodal teaching can motivate EFL learners by connecting learning with sensory, cultural, and personal resources. Even so, students still found concepts such as salience, framing, perspective, hidden message, and communicative purpose moderately difficult.

eISSN 3124-9396
Volume 2 , July 2026

CONCLUSION

This study concludes that integrating visual grammar into the genre-based approach is an effective and systematic model for developing junior high school EFL students' multimodal viewing skills. Through the stages of GBA, students were gradually guided to understand the context of multimodal texts, identify visual and verbal elements, analyse representational, interactive, and compositional meanings, interpret image-text relations, and evaluate communicative purposes independently. The findings show that the experimental group achieved higher improvement than the control group in visual recognition, visual grammar analysis, and image-text interpretation, while inferential meaning-making and critical evaluation also improved although they still required stronger scaffolding. Students also responded positively to the model because it made multimodal texts clearer, more engaging, and more useful for English learning. However, abstract visual grammar concepts such as salience, framing, perspective, bias, and communicative purpose need to be introduced gradually through repeated modelling, guided practice, and concrete examples. The study confirms that visual grammar-integrated GBA can help students move beyond surface-level observation toward more analytical and critical interpretation of multimodal texts, while future studies may involve broader participants, longer interventions, and different multimodal genres to strengthen the generalisability of the findings.

BIBLIOGRAPHY

Aleksić-Hajduković, I. (2025). Deconstructing a kaleidoscope of semiotic modes and audience perception of multimodality in advertising discourse. *Visual Communication*, *24*(3 Special Issue: Multimodality and Reception Studies). https://doi.org/10.1177/14703572251328364

Anderson, K. A. (2025). Integrative genre-based pedagogy: Enhancing social responsiveness in English medium of instruction and STEM education. *Journal of English for Academic Purposes*, *74*. https://doi.org/10.1016/j.jeap.2025.101483

Beltrán-Palanques, V. (2024). Assessing video game narratives: Implications for the assessment of multimodal literacy in ESP. *Assessing Writing*, *60*. https://doi.org/10.1016/j.asw.2024.100809

Braun, V., & Clarke, V. (2022). Thematic Analysis: A Practical Guide. *QMiP Bulletin*, *1*(33). https://doi.org/10.53841/bpsqmip.2022.1.33.46

Brunfaut, T., & Kormos, J. (2025). Assessing Multimodal Viewing-to-Write Constructs: Task Design, Performance, Processing, and Rating. *Language Assessment Quarterly*, *22*(4–5). https://doi.org/10.1080/15434303.2025.2596374

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

Chen, W., Yang, Y., Tian, Z., Chen, Q., & Liu, J. (2025). A review of multimodal learning for text to images. *Multimedia Tools and Applications*, *84*(10). https://doi.org/10.1007/s11042-024-19117-8

Chen, Y., & Weninger, C. (2025). The potential of eye tracking data to strengthen CDA' explanatory power: the case of multimodal critical discourse analysis of

advertising persuasion. *Critical Discourse Studies*, *22*(4). https://doi.org/10.1080/17405904.2024.2331641

Creswell, W. J., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative adn Mixed Methods Approaches. In *Journal of Chemical Information and Modeling* (Vol. 53, Number 9).

DeVellis, R. F. (2017). Scale Development: Theory and Applications (4th ed.). In *SAGE Publication*.

Elmiana, D. S., & Shen, S. (2025). Let Pictures Explain the World: Text–Image Narration in Picturebooks. *Journal of Research in Childhood Education*, *39*(3). https://doi.org/10.1080/02568543.2024.2367406

Fjørtoft, H. (2020). Multimodal digital classroom assessments. *Computers and Education*, *152*. https://doi.org/10.1016/j.compedu.2020.103892

Herzog, M., Kim, J. eun, & Park, E. (2025). Quality and predictors of L2 digital multimodal composition: Mixed-methods study. *System*, *133*. https://doi.org/10.1016/j.system.2025.103787

Hiippala, T. (2021). Distant viewing and multimodality theory: Prospects and challenges. *Multimodality & Society*, *1*(2). https://doi.org/10.1177/26349795211007094

Jiang, L., Li, Z., & Leung, J. S. C. (2024). Digital multimodal composing as translanguaging assessment in CLIL classrooms. *Learning and Instruction*, *92*. https://doi.org/10.1016/j.learninstruc.2024.101900

Jiang, L., Yu, S., & Lee, I. (2022). Developing a genre-based model for assessing digital multimodal composing in second language writing: Integrating theory with practice. *Journal of Second Language Writing*, *57*. https://doi.org/10.1016/j.jslw.2022.100869

Kaderoğlu, K. (2026). Aligning input and assessment: Pictorial measures in audiovisual vocabulary research. *Research Methods in Applied Linguistics*, *5*(1). https://doi.org/10.1016/j.rmal.2025.100288

Kessler, M., & Casal, J. E. (2024). English writing instructors' use of theories, genres, and activities: A survey of teachers' beliefs and practices. *Journal of English for Academic Purposes*, *69*. https://doi.org/10.1016/j.jeap.2024.101384

Kress, G., & van Leeuwen, T. (2006). *Reading images: the grammar of visual design*. Routledge.

Kress, G., & van Leeuwen, T. (2020). Reading Images: The Grammar of Visual Design. In *Reading Images* (3rd ed.). Routledge. https://doi.org/10.4324/9781003099857

Lei, Q., & Zhang, C. (2024). Using multimodal resources to design EFL classroom lead-ins—A multimodal pedagogical stylistics perspective. *Linguistics and Education*, *83*. https://doi.org/10.1016/j.linged.2024.101338

Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education*, *69*. https://doi.org/10.1016/j.linged.2022.101048

Liu, M., Zhang, L. J., & Zhang, D. (2025). Enhancing student GAI literacy in digital multimodal composing through development and validation of a scale. *Computers in Human Behavior*, *166*. https://doi.org/10.1016/j.chb.2025.108569

Liu, Y., Cheong, C. M., & Zhu, X. (2025). The Relationships Between Lower- and Higher-Level Cognitive Skills and Multimodal Reading Comprehension Among Fourth-Grade Students in the Digital Age. *International Journal of Applied Linguistics (United Kingdom)*. https://doi.org/10.1111/ijal.12794

Ma, Y., & Jiang, F. K. (2025). Guiding and engaging the audience: Visual metadiscourse in PowerPoint slides of Three Minute Thesis presentations. *English for Specific Purposes*, *77*. https://doi.org/10.1016/j.esp.2024.10.003

Macnaught, L., Maton, K., Martin, J. R., & Matruglio, E. (2013). Jointly constructing semantic waves: Implications for teacher training. *Linguistics and Education*, *24*(1). https://doi.org/10.1016/j.linged.2012.11.008

Martinec, R., & Salway, A. (2005). A system for image–text relations in new (and old) media. *Visual Communication*, *4*(3). https://doi.org/10.1177/1470357205055928

Nathues, E., van Vuuren, M., Endedijk, M. D., & Wenzel, M. (2025). Shape-shifting: How boundary objects affect meaning-making across visual, verbal, and embodied modes. *Human Relations*, *78*(3). https://doi.org/10.1177/00187267241236111

Nguyen, L. A. T., & Habók, A. (2024). Tools for assessing teacher digital literacy: a review. *Journal of Computers in Education*, *11*(1). https://doi.org/10.1007/s40692-022-00257-5

Oyama, R. (2021). 'Reading' Literature Through Drawing: A Multimodal Pedagogic Model for the EAL University Classroom. In *Multimodality in English Language Learning*. https://doi.org/10.4324/9781003155300-13

Pham, V. P. H., & Bui, T. K. L. (2021). Genre-based approach to writing in EFL contexts. *World Journal of English Language*, *11*(2). https://doi.org/10.5430/WJEL.V11N2P95

Pun, J. K. H., & Cheung, K. K. C. (2023). Meaning making in collaborative practical work: a case study of multimodal challenges in a Year 10 chemistry classroom. *Research in Science and Technological Education*, *41*(1). https://doi.org/10.1080/02635143.2021.1895101

Shen, Y., & Han, Y. (2026). Fostering undergraduates' critical thinking in digital multimodal composition with generative artificial intelligence. *Thinking Skills and Creativity*, *59*. https://doi.org/10.1016/j.tsc.2025.102045

Singer Trakhman, L. M., Alexander, P. A., & Fusenig, J. (2025). Reading in print and digitally: Profiling and intervening in undergraduates' multimodal text processing, comprehension, and calibration. *Learning and Individual Differences*, *118*. https://doi.org/10.1016/j.lindif.2025.102627

Sultan, Rapi, M., Suardi, & Ismail, A. (2025). Perceptions of TPACK in genre-based pedagogy: the role of socio-demographics and ICT competence among pre-service Indonesian language teachers. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2482410

Tseng, M. i. L., Tan, L., & Chou, T. L. L. (2025). Place-based digital composing practices in multimodal pedagogy: a designed-based approach to enhancing bilingual educators' professional development. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2025.2588674

Tseng, Y. H. (2025). Motivating rural EFL students in multimodal teaching. *Linguistics and Education*, *87*. https://doi.org/10.1016/j.linged.2025.101425

Unsworth, L. (2021). High school science infographics: Multimodal meaning complexes in composite image-language ensembles. *Pensamiento Educativo*, *58*(2). https://doi.org/10.7764/PEL.58.2.2021.9

Valente, D., Chennaz, L., Archambault, D., Négrerie, S., Blain, S., Galiano, A. R., & Gentaz, E. (2024). Comprehension of a multimodal book by children with visual impairments. *British Journal of Visual Impairment*, *42*(1). https://doi.org/10.1177/02646196231172071

Wünsch-Nagy, N. (2025). From multimodal space to digital multimodal text: Making choices in digital multimodal compositions inspired by museum visits in higher education. *Computers and Composition*, *75*. https://doi.org/10.1016/j.compcom.2024.102909

Xia, S. (2024). Incorporating digital multimodal composition in content teaching: A multimodal analysis of students' legal popularization videos. *Journal of Second Language Writing*, *66*. https://doi.org/10.1016/j.jslw.2024.101163

Xiong, T., Feng, D., & Hu, G. (2022). Researching Cultural Knowledge and Values in English Language Teaching Textbooks: Representation, Multimodality, and Stakeholders. In *Cultural Knowledge and Values in English Language Teaching Materials: (Multimodal) Representations and Stakeholders*. https://doi.org/10.1007/978-981-19-1935-0_1

Zhang, D., Wong, W. K., & Chew, I. M. (2025). A Comprehensive Review of Multimodal Visual Representation Learning: Tracing the Evolution from CNNs to Transformers and Beyond. *International Journal of Multimedia Information Retrieval*, *14*(4). https://doi.org/10.1007/s13735-025-00382-8