

9-30-2025

A Multimodal Discourse Analysis of Visual Representation in English Language Textbooks at the Elementary Level: A Social Semiotic Perspective

Amina Bashir

Department of English, University of Sahiwal, Sahiwal (57000), Pakistan

Aamir Aziz

Department of English, University of Sahiwal, Sahiwal (57000), Pakistan

Muhammad Imran

Educational Research Lab, Prince Sultan University, Riyadh, Saudi Arabia AND Department of English and Literary Studies, University of Management and Technology, Lahore, Pakistan, mimran@psu.edu.sa

Norah Almusharraf

Educational Research Lab, Prince Sultan University, Riyadh, Saudi Arabia

Follow this and additional works at: <https://kjhss.khazar.org/journal>

How to Cite This Article

Bashir, Amina; Aziz, Aamir; Imran, Muhammad; and Almusharraf, Norah (2025) "A Multimodal Discourse Analysis of Visual Representation in English Language Textbooks at the Elementary Level: A Social Semiotic Perspective," *Khazar Journal of Humanities and Social Sciences*: Vol. 28: Iss. 3, Article 1. Available at: <https://kjhss.khazar.org/journal/vol28/iss3/1>

This Original Study is brought to you for free and open access by Khazar Journal of Humanities and Social Sciences. It has been accepted for inclusion in Khazar Journal of Humanities and Social Sciences by an authorized editor of Khazar Journal of Humanities and Social Sciences.

ORIGINAL STUDY

A Multimodal Discourse Analysis of Visual Representation in English Language Textbooks at the Elementary Level: A Social Semiotic Perspective

Amina Bashir ^a, Aamir Aziz ^a, Muhammad Imran ^{b,c,*},
Norah Almusharraf ^b

^a Department of English, University of Sahiwal, Sahiwal (57000), Pakistan

^b Educational Research Lab, Prince Sultan University, Riyadh, Saudi Arabia

^c Department of English and Literary Studies, University of Management and Technology, Lahore, Pakistan

ABSTRACT

Textbooks serve as a significant and comprehensive resource for language learners. This study focuses on a Multimodal Discourse Analysis of Visual Representations found in English Language Textbooks, employing a Visual Social Semiotic Perspective. Utilizing the framework established by Kress & van Leeuwen (2020), which comprises Representational, Interactional, and Compositional modes, this research investigates the multimodal elements present within these textbooks to assess their role in enhancing comprehension through visuals. Adopting a qualitative descriptive methodology, the study analyzes a purposively selected sample of 158 images from elementary school textbooks. The findings reveal a disproportionate representation of males compared to females across the images, with non-human elements also appearing frequently. The data indicate that conceptual meaning is most predominant, where the symbolic category of meaning plays a critical role. In terms of the interactional mode, the analysis shows a greater prevalence of ‘offer’ images as opposed to ‘demand’ images, with a notable medium social distance. Furthermore, the attitudes conveyed through these images are characterized by horizontal frontal and vertical eye-level angles, which are more prominent. The color modality and contextualization in 6th and 8th-grade textbooks exhibit maximal values, while the 7th-grade textbooks reflect moderate modality values. The compositional mode reveals that the center-margin information value is the most significant, with salient framing types indicating integrated connections and disconnections through separation. In conclusion, the study highlights that incorporating visual images and multiple modes, alongside textual components, significantly contributes to language learning. This multimodal approach enhances the meaning-construction capabilities of language learners, underscoring the vital role that textbooks play in educational contexts.

Keywords: Multimodality, Multimodal discourse analysis, Visual social semiotics, English language textbooks, Visual representations

Received 26 June 2025; revised 9 November 2025; accepted 15 November 2025.
Available online 30 September 2025

* Corresponding author.
E-mail address: mimran@psu.edu.sa (M. Imran).

<https://doi.org/10.5782/2223-2621.1348>

2223-2621/© 2025 Published by Khazar Journal of Humanities and Social Sciences. This is an open access article under the CC BY 4.0 Licence (<https://creativecommons.org/licenses/by/4.0/>).

Introduction

Before the 20th century, esteemed writings across various genres were presented in isolated text formats. These works often lacked supporting illustrations and were characterized by monotonous layouts and dense, rigid page prints, resulting in a monomodal presentation (Imran et al., 2024). However, by the end of the 20th century, a significant transformation occurred. Mass media, along with historical, musical, scientific, and academic writings, began to incorporate colorful illustrations, modern layouts, and sophisticated typography. This evolution facilitated the emergence of multimodality, transcending the traditional boundaries of text (Abdullah & Amroun, 2017). In contemporary society, communication is inherently multimodal, with information being accessed through a variety of sources. Hence, to understand multimodal information, learners need both visual and verbal literacy, without which effective communication does not remain intact (Aziz et al., 2021; Aziz & Shakir, 2023; Cheung & Erduran, 2023; Imran & Almusharraf, 2023; Shehzadi et al., 2024). Kress & van Leeuwen (2020) emphasize that concepts expressed verbally can often be more effectively represented visually. Just as verbal expression depends on word choice and sentence structure, visual communication conveys meaning through the careful selection of colors, contrast, composition, and the arrangement of images and text (Paltridge, 2018; De Souza et al., 2021). For EFL learners, it is not easy to be proficient in the language through isolated text form. The multiple modes as images, illustrations, graphs, charts, videos, animations, etc, help understanding language better (Guo, 2023; El-Dakhs et al., 2022; Gull et al., 2020).

The term “Multimodality” was coined in the late 1990s based on Charles Goodwin’s article in 1998, which refers to the use of multiple modes or multiple literacies within one medium. The term is highly attributed to Kress and Leeuwen’s work on Visual Grammar based on *visual social semiotic theory*. They illustrate how visual elements—such as animals, locations, people, and objects—interact with verbal statements, thereby creating meaning through varying degrees of complexity (Kress & van Leeuwen, 2020, p. 1). It points out how visual communication takes place through the relationship between represented and interactive participants, leading to an in-depth analysis of visuals (Friedman & Ron, 2017). Bi (2019) points out that the rise of Multimodality underscores the importance and role of symbolic resources in communication rather than linguistic content. Ayoola et al. (2022) noted that visual communication and representations have been a part of human society for over five thousand years, and thus, the images can never be isolated from the text, just as facial communication cannot be separated from what is said and meant. Rajistha & Claria (2023) indicate that multimodality is largely a semiotic approach that helps in the meaning construction of the visual elements and semiotic acts by utilizing various dimensions of it. Semiotics, the study of signs, recognizes that signs can manifest in various forms, including auditory, tactile, and visual (Gualberto & Kress, 2019). Social Semiotics highlights human practices in a particular culture and social environment to make meanings and their understanding (Fitriana & Wirza, 2021, April). According to Ferdinand de Saussure, a Sign is a fundamental and highly valued concept in semiotics, which consists of a signifier, a signified, and an underlying idea (Linden, 2025). It is important to acknowledge that the meanings of signs can differ significantly depending on the social and cultural contexts in which they are analyzed. There has been limited research conducted on multimodality in Pakistan, particularly concerning the visual analysis of images within textbooks from a Multimodal Discourse (MD) perspective. This study represents a unique contribution to the field by examining how images in educational materials serve as representational tools.

Objectives of the study

The primary objectives of this research are as follows:

- To investigate the significance of the representative mode in enhancing the understanding of complex concepts found in textbooks.
- To analyze how the interaction between represented and interactive participants conveys meaning.
- To assess the role of textbook composition in effectively capturing learners' attention.

Research questions

This study seeks to address the following research questions:

1. In what ways does the representative mode contribute to understanding complex concepts in elementary-level textbooks?
2. How can the interactional mode be utilized to elucidate the relationship between depicted and interactive participants?
3. What is the significance of compositional mode in elementary-level textbooks in engaging English language learners?

Significance of research

The present study is designed to encompass multiple dimensions to cover. It has both theoretical insights and practical applications. Firstly, theoretical terms, this framework will serve as a valuable aid for the upcoming and new researchers regarding the MDA of textbooks. On the other hand, the practical implications of this study offer the development of critical thinking skills. It also strengthens the objective interpretations of the researchers.

Literature review

In Pakistan, language textbooks serve as a fundamental resource for teaching and learning a foreign language (Imran & Ain, 2020). Multimodality is highly involved in primary and elementary-level language textbooks to promote language education. Books are essential in language teaching, so their layout, manuscript, composition, and portrayal of information need to be appealing as well as thoughtful. A relevant study of [Makinde & Odili \(2023\)](#) employs social semiotic analysis on the Nigerian EFL book 'Awka Metropolis', and [Saeed et al. \(2019\)](#) employ SFL for a Comparative Translational Semiotic Study of a Novel 'River of Fire'. The study extends the same concept to explore the relationship between multimodal resources and educational materials; it combines various ways in which books are composed just to create a unified and cohesive meaning of the text. It also relies on Halliday's SFL (1978) and Kress and Leeuwen's MDA (2020) to analyze sixteen visuals. It contributes to the present findings on how the integration of images helps learners' holistic understanding of the language and meaning-making.

Recent research by [Qi \(2024\)](#) explores the significance of visual illustrations in Chinese EFL textbooks to investigate and emphasize their functions in textbooks. The most frequent and highly prevalent multimodal elements found in EFL textbooks are static visuals, animated images, tables & graphs. These multimodal elements assist the cognitive and pedagogical content. The findings indicate that the lack of multimodal elements in designing the textbooks affects learners' comprehension and understanding. Additionally, [Roekhan et al. \(2024\)](#) note that incorporating discursive practices into textbooks can help instill peace values in young learners, influencing their behavior and attitude

towards a language. As [Ahmad et al. \(2021\)](#) investigate gender bias represented through the various visual illustrations just to extend male dominance. These values hold great importance when adopted; the books portray the materials keeping in view the needs of learners. In interactive mode, medium shot and full-color saturation are dominant, while in compositional mode informative image is dominant. [Triacca \(2017\)](#) emphasizes that the primary aim of integrating visual imagery is to support learners' cognitive processes. This integration provides clear and comprehensible examples of abstract concepts and fosters a connection between learners' understanding of the text and their surroundings, such as environmental and cultural content. Additionally, [Arshad et al. \(2020\)](#) in a study explore the representation of culture in various textbooks of the Oxford series, concluding that textbooks promote hidden agendas of various cultures. Similarly, Chouhan (2025) relates discourse with Eco-linguistic studies on how climate change narratives are formed by the Ministry of Climate Change.

[Yang and Zhang \(2014\)](#) explain and analyze the representational meaning in English Editorial pieces utilizing MDA. They take up [Royce's \(1998\)](#) inter-semiotic approach, along with Kress and Leeuwen's MDA (2020) for the analysis. The editorials, in the English language, give almost an equal ratio of linguistic texts and visual illustrations. Given their role in addressing pressing national issues, the inclusion of visuals is critical alongside linguistic data to facilitate a comprehensive understanding of the situation. This method also aids in clarifying the main ideas conveyed within the editorial columns. [Spencer \(1924\)](#) asserts that the integration of visual images in the editorials effectively presents the facts, opinions, and logic in a concise manner, making the content more accessible to average readers. The textbooks not only assess language learning and pedagogy but also help transmit cultural and social attitudes and other global ideas. [Fitri et al. \(2023\)](#) investigate the role of multimodal source material in a language textbook titled 'Get Smart Plus'. It points out that the verbal mode is dominant and the visual elements are not present; at the same time, it mentions that multiple sources help in learners' cognitive development. In his study 'Windows to the World', [Risager \(2021\)](#) challenges that the old and traditional textbooks do not give the actual representation of the world. The things that approach the learner differently have a different impact on the mind; the material presented in a pictorial form delivers a message more effectively and quickly. [Pourrajab et al. \(2015\)](#) support their idea by noting that the students look at the images before they read the text or content because the visuals are more appealing. [Bashir et al. \(2025\)](#) explore the influences of Computer-Assisted Language Learning (CALL)-based multimodal pedagogical resources on learner motivation and willingness to communicate, concluding that such resources can significantly enhance English language acquisition. English is widely recognized as a Lingua Franca, as noted by [Jenkins \(2007\)](#).

Research methodology

A research methodology outlines the systematic process for conducting research and the necessary steps to arrive at a conclusion ([Mishra & Alok, 2022](#)). For this study, a Qualitative Descriptive Research Methodology has been employed. According to [Hart \(2020\)](#), this approach provides a comprehensive and clear description of the investigations, experiences, and perceptions undertaken, addressing the "who, what, and where" of events or experiences ([Kim et al., 2017](#)). Additionally, [Setiawan \(2021\)](#) asserts that "Descriptive Qualitative Research is aligned with Pragmatism." The data for this research were collected from the original PCTB textbooks utilized at the elementary school level, specifically focusing on three designated textbooks. Visual images from the chapters of these textbooks were selected, deliberately excluding those from exercises. These

images were then analyzed qualitatively using the MDA model proposed by [Kress & van Leeuwen \(2020\)](#), and the interpretations were presented descriptively. The frequencies and percentages were calculated manually; this study does not utilize any coding software.

Data collection

The population for this study encompasses all books published by the Punjab Curriculum and Textbook Board (PCTB) for grades 1 through 12, covering primary, middle, secondary, and higher secondary school levels. The sample was obtained using the Purposive Sampling Method, a non-probability sampling technique where participants are intentionally selected based on specific criteria, often referred to as judgmental or selective sampling. The data for this analysis has been taken from three English Language textbooks of the PCTB used as source material in elementary classes. Specifically, these textbooks are for grades 6th, 7th, and 8th. These books were available in both formats: as downloaded from the official website of E-Books PCTB in soft form and purchased from the market in hard form. The selection of these books was based on the presence of visual images, resulting in the collection of a total of 158 pictures that adhere to the established criteria.

Theoretical framework

The Multimodal Analytical framework utilized in the current study has its roots in discourse, as it is a discourse-oriented framework, and it studies the visual data collected from the selected books. MDA has three modes: representational, interactional, and compositional; the representational mode has two processes: narrative and conceptual. The former structure represents the dimensions of reality, actions, and events, taking place according to the transpatial orientation ([Han, 2025](#)). It explains what is represented, giving the relationship between the participants who are depicted in the images and the ‘action and reaction processes’ along with the categorization of the participants ([Khorina et al., 2023](#)). The action and reaction processes can be transactional or non-transactional. One of the main differences between these two structures is based on the presence and absence of vectors; the action process is based on the vectors formed by the direction of limbs, and the reaction process forms the vector formed by the direction of eye lines. The conceptual structure represents the association of participants based on their classification in overt or covert taxonomic forms ([Ouchra, 2025](#)), analytical structures, or the part-whole relationship, and the symbolic structures, attributive and suggestive ([McMurtrie, 2022](#); [Yang, 2025](#)). Other than the processes, there are participants in this mode, as human or non-human, which are further divided.

The contact is an imaginary relationship between the participants of the image and the viewers; it can be a demand or an offer. Social distance gives the framing of the images to portray the kind of social relations. It is directly related to the ‘Size of Frame’ and it has three main divisions: the close, medium, and long shot, which refer to the intimate, social, personal, and Impersonal relations, respectively ([Roberts et al., 2020](#)). Perspective means the attitude ([Byram et al., 2013](#); [Unsworth & Wheeler, 2002](#); [Kress & van Leeuwen, 2020](#)) of the viewers towards the depicted participants. It is the ‘point of view’ that implies the subjective or objective attitude towards the depicted participants in the image ([Kress & van Leeuwen, 2020](#), p. 129). Modality is particularly related to the desirability or permissibility of a visual image. It has two categories, colors and contextualization; colors are sensual as well as emotive ([Rahikummahtum et al., 2022](#)). The color modality is given in three respects: saturation, modulation, and differentiation. The contextualization gives the background information of the image; it can be maximum, minimum, or blank ([Fattah & Hussein, 2025](#)).

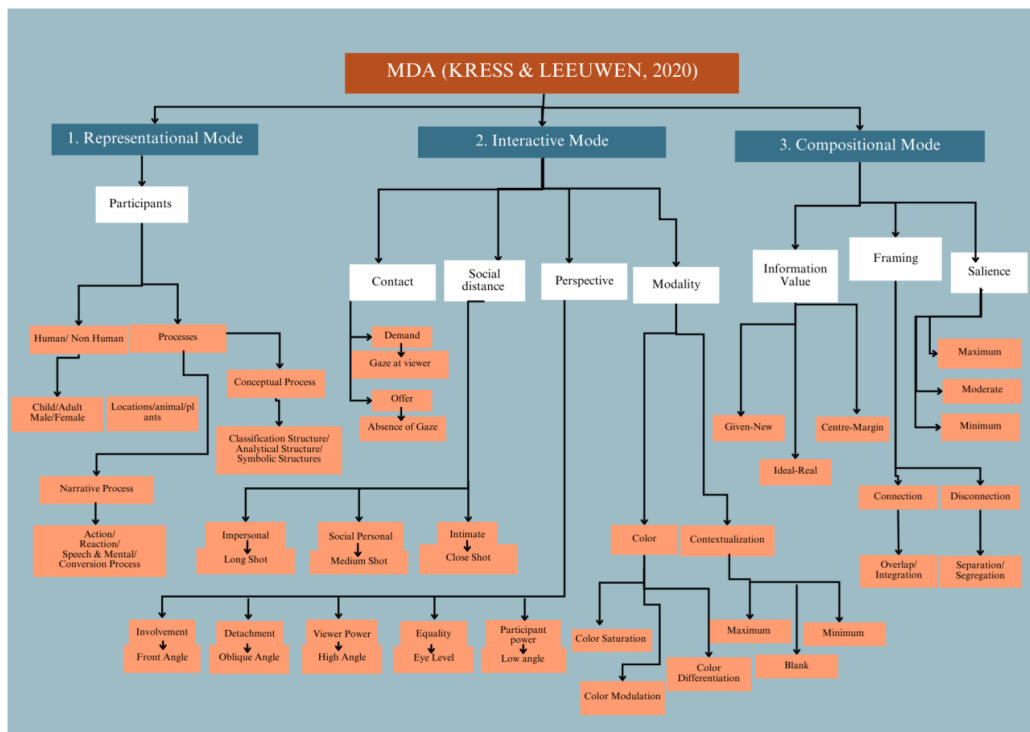


Fig. 1. Theoretical framework.

Note: This figure is Kress & van Leeuwen (2020) model of MDA.

As the name suggests, the compositional mode depicts (as shown in Fig. 1) how the visual and verbal elements are arranged and placed on the page, how the page is designed and composed. It has three categories as information value, framing, and salience. It shows if the elements on the page are connected or disconnected in the transpatial composition. The disconnection has two major domains: separation, which is a general division by framing lines, and segregation, which is an unjust and unnatural division (Anwar, 2018). Similarly, the connection is realized by the overlapping and integration of visual elements that make them coherent and united. The salience in composition plays a meaningful role in the visual and verbal elements, making them coherent (Kress & van Leeuwen, 2020, p.210). It grasps the attention of viewers in multiple ways and hierarchically arranges the elements according to their importance and significance concerning other elements (Novina, 2023). The ongoing study sheds light on the idea of how visual images play a role in effective communication (Heberle & Constanty, 2016), representing meaning (Cahyani & Fitriati, 2024), interacting with the viewers (Campbell et al., 2016), and composition (Chen & He, 2015) of textbooks play a vital role in meaning-making.

Data analysis

This study is qualitative descriptive in nature; the images are analyzed through a qualitative approach, and then the interpretation is written descriptively, for which Kress and Leewuen's model of MDA (2020) is utilized.

From Fig. 2, an Islamic setting is conveyed as the dress code of both participants depicts. Along with the human participants, the non-human participants are also portrayed: the



Fig. 2. Representational mode in the textbook of the 6th grade.



Fig. 3. Interactive mode in the textbook of the 6th grade.

utensils and food. The narrative and conceptual processes are present; in narrative one both the action and reaction processes are observed. The action narrative process here is transactional because both the actor and goal are present; the vector arising from the direction of their limbs, arms, or hands leads to the religious and spiritual idea of prayer and thankfulness, hence the actors are both male and female. In contrast, their goal is their prayer they are making before starting the meal. On the other hand, the reaction process here is also transactional because both the reactants and phenomena are present as the reactants are both the man and woman who do the looking, while the phenomena are the food they are looking at. In conceptual processes, the classification is overt because of the participant's same style, position, level, distance, etc. They are similar in each term except for their gender differences.

Fig. 3, in terms of interactive mode, serves as an offer because of the absence of gaze; the boy is not making any direct eye contact with the viewers. The contact suggests that the image is just imparting information, and no active participation is required from the viewers. It is a long shot image as a detailed background is given, and the boy is pictured as a whole figure. The visual is highly modulated, it looks natural, and there is a distinction in multiple colors and shades. The maximum context of the image is described, the setting appears to be a rural and countryside area, and it communicates the idea of environmental responsibility and care for the home and cleanliness.

Fig. 4 gives the information value in a centre-margin format; the human-like water droplet is the central figure conveying the actual meaning. The centre holds much importance, and the peripheries just give the information about the centre. A big water droplet is in the centre, it is the most significant element in the figure, and the surrounding bubbles



Fig. 4. Compositional mode in the textbook of the 6th grade.



Fig. 5. Representational mode in the textbook of 7th grade.

show the daily life uses of water and how to save it. The framing of the image is given in both connection and disconnection forms; disconnection is suggested through the white spaces present among the parts of the visual, while the connection is depicted through the overall coherence of the parts. The salience arranged elements to emphasize the important elements and send the less important elements to the background. The salience is maximum in the image. It conveys an important message symbolically; the image symbolizes the uses and precautions to prevent the loss of water unnecessarily.

Fig. 5 depicts two mountaineers, one male and one female, holding a Pakistani flag representing their national identity. Both participants are involved in a transactional action process as they are holding a flag and are engaged in an action suggested by the vectors formed; the flag is the goal of both participants. The flag of a country is a symbol of devotion to one's homeland and nation; it holds a central value in the figure. Due to the goggles, the eyes of the participants are not visible, so it is a non-transactional reaction process; there is no vector emerging from the eye-lines. In the conceptual process of representational mode, the kind of classification is overt because of equal distance, same position, and identity of the participants. The possessive attributes of the participants clearly show their profession and national identity. The figure gives vivid symbolic meaning in both respects: suggestive and attributive. To look at the symbolic meaning, this image gives both the symbolic attributive and suggestive meaning; the attributes give national pride and identity, and at the same time, the figure depicts the themes of gender equality, patriotism, and national pride.

In Fig. 6, two major figures are portrayed, one is demanding and the other is offering, according to the rule of contact. The picture portrays two army men, one is directly

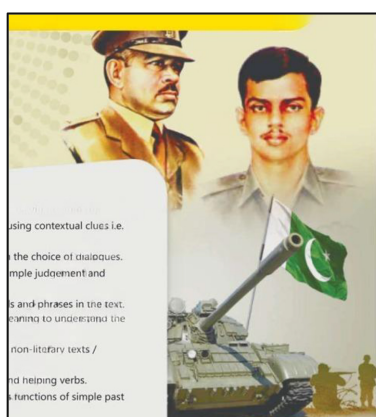


Fig. 6. Interactional mode in the textbook of the 7th grade.



Fig. 7. Compositional mode in the textbook of the 7th grade.

looking toward the viewers, showing gaze and connection, while the other participant is looking away, showing offer and disconnection. It depicts that the depicted participants require both active engagement and passive observation. The fire is a medium shot picture, because the full bodies of the participants are not illustrated, and no large background is represented. Horizontally, it is a frontal angle image, and vertically, an eye-level angle image. The frontal angle suggests active engagement, and the eye-level angle shows equality and neutrality. Overall color values are moderate; the saturation, differentiation, and modulation are all moderate due to non-sharp contrast and non-vibrant shades.

The information value in Fig. 7 is in top/bottom format, which gives the names of ideal/real. The title “Smog” mentioned above is the ideal part, which is a burning issue present worldwide. In contrast, the bottom part deals with some students in uniforms, wearing masks, going to school is the real thing, a practical problem and issue nowadays. The background of Minar-e-Pakistan narrows the context to Pakistan. The framing is given through an integrative connection; all the visual elements are strongly integrated, presenting a cohesive background, and there is no segregation or separation in framing. The salience is moderate in the figure; the school-going children are prominent against the hazy background, grasping viewers’ attention. Their habit of wearing masks in unclear and polluted surroundings suggests a common environmental problem of smog, as the



Fig. 8. Representational mode in the textbook of 8th grade.



Fig. 9. Interactional mode in the textbook of 8th grade.

title describes. This moderate salience is due to the smog, particularly the background of Minar-e-Pakistan, which highlights that this issue of environmental pollution is most common in big cities.

In Fig. 8, both human and non-human participants are represented. Some village men are sitting around an elderly man. The background of the image conveys a rural area social setting. Both action and reaction processes are present; the vector formed by the hand of the elder man has a particular goal, showing it as a transactional action process. The gaze of the elder man towards the other middle-aged men directs it towards the reaction transactional process; all three kinds of conceptual process are shown, the classification kind is overt, showing non-equal relations and values, it conveys the elder man's authority over the middle-aged and younger men. Their possessive attributes as dress code, turbans, style, etc, convey their social identification. Both types of symbolic meaning are presented; their attributes suggest their identity, the elderly man's distinct position and place show his authority and leadership. While the environment suggests their traditional and cultural values, the concept of "Panchayat" in villages.

In Fig. 9, no direct contact is established; the image is an offering image. The offers are just for the sake of observing and conveying information. It does not require any kind of active involvement. The figure is a medium shot image, the participants are portrayed as whole human figures, along with a proper setting, but no large background is conveyed. In horizontal perspective, the image is frontal, while in vertical perspective, it is an eye-level shot. The vertical eye-level angle suggests equality and neutrality of power; it shows everyone's equal and collective responsibility for keeping the environment clean. The modality, in both respects, is maximum; the colors are bright and vibrant, which is why they are differentiated. The objects in the figure are prominent, such as the garbage

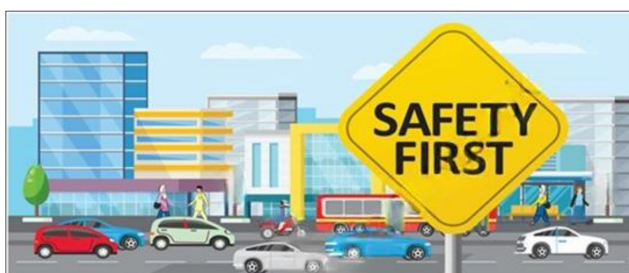


Fig. 10. Compositional mode in the textbook of 8th grade.

bags, trash, broom, and bin, which are all clear. The colors are fully modulated, making the scene more natural and real. The contextualization is maximum as the long shot of the image gives the detailed background of an unclean ground or public park, and the action of depicted participants conveying the idea of civic responsibility of keeping the environment and public places neat and clean.

The information value in [Fig. 10](#) is presented in a 'given-new' format. The left side, left to the signboard, represents the 'Given' as it presents the already known and understandable information. The surrounding presence of vehicles and high buildings suggests an Urban social setting. The right side is 'New', giving a big signboard that presents the conception of the importance of one's safety. The framing is connecting and slightly disconnecting; the elements are connected through integration and overlapping. All elements are closely integrated, and the signboard overlaps the background. The slight disconnection is presented through the signboard as it stands out separately in the figure. The salience of the image is maximum, the big yellow signboard on which "SAFETY FIRST" is written in black uppercase bold letters catches viewers' attention, and all the other elements are less dominant.

Results and discussion

Keeping in view the aim of the research, three research questions were developed, and each question corresponds with one mode of MDA. The RQ1 is related to the first mode; the aim of this mode is to describe the kind of participants and processes that are depicted in the figures. It helps in understanding how different and difficult concepts are deconstructed so easily. In the textbook of the 6th class, every selected picture shows both kinds of participants, human and non-human. The processes are important in conveying complex ideas, how the picture illustrates an idea. In this book, conceptual processes are more dominant than narrative ones. In the book of the 6th class, 50.4% of the images convey conceptual meaning, and 26% give the narrative process as shown in [Fig. 11](#).

In the book of 7th grade, 86.29% of participants are human; the overall distribution of participants is illustrated in [Fig. 12](#). In the book of 8th grade, the human participants are 73.19% prevalent, while in the other two books discussed above, non-human participants are more frequent in this book too; it contains 60.71% objects, 39.28% locations, and the animals/ plants are also present to some extent.

In terms of processes, the maximum number of images, about 45.08% show action processes, out of which 21.56% are transactional action and 23.52% are non-transactional. The reaction process is less dominant; 39.21% of images show reaction processes, 11.76% are transactional reactions, and 27.45% are non-transactional reaction processes. No mental or speech, or conversion processes are given in the book. In the book of the

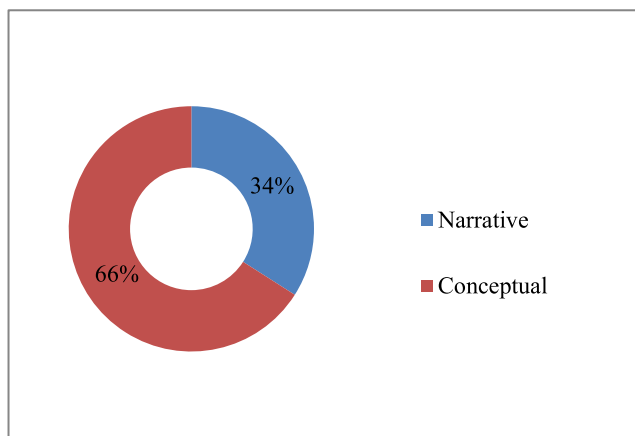


Fig. 11. Representational mode sub-categories (Book 6th Grade).

Note: The figure demonstrates the frequency of both narrative and conceptual processes by calculating the average of all the categories involved in both processes.



Fig. 12. Categories of participants (Book 7th Grade).

Note: The figure describes the frequency of categories of participants, both human and non-human.

7th class, the conceptual processes are more dominant than the narrative ones to make learners understand complex ideas. The categorization of conceptual processes is prevalent in Fig. 13.

In terms of processes, in the 8th-grade textbook, in the narrative process, 25% are action transactional and 23.21% are non-transactional action process, 12.5% and 32.14% are reaction transactional and non-transactional, respectively. The analytical process is given by possessive attributes, and symbolic suggestion is more significant in the book, as shown in Fig. 14

The RQ2 is related to the interactional mode of MDA; it focuses on the relationship between the depicted participants and the interactive participants. According to the contact, in each book, maximum images are 'offers' and only a few are 'demand-based' on the presence or absence of gaze; the percentage of demand or offer in each book is illustrated in the figure.

Considering the social distance, in all the books maximum number of figures are medium shot images, suggesting the personal social distance showing close social relationships among the participants.

The attitude of the image has two categories: horizontal and vertical. In 6th 6th-grade textbook, 58.51 % of figures show horizontal frontal perspective and 41.46% are oblique angle images. In vertical perspective, 58.53 % of images are eye-level, 21.95% figures are high angle, and the same number is for low angle. In the textbook of 7th, in horizontal perspective, the frontal angle is 68.62% present and the oblique angle is 21.37% prevalent.

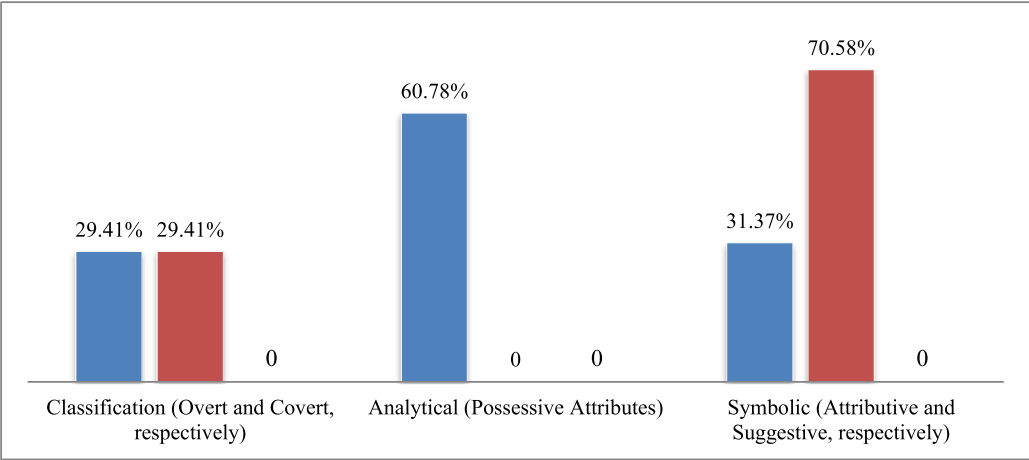


Fig. 13. Categories of conceptual processes in the book of 7th grade.
Note: The figure illustrates the presence and frequency of different categories of conceptual processes.

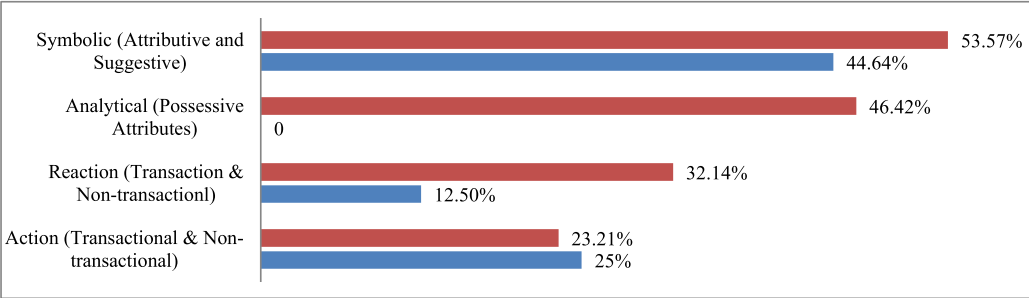


Fig. 14. Narrative process in 8th Grade.
Note: The figure illustrates the percentages of categories of the narrative process in the book of 8th grade.

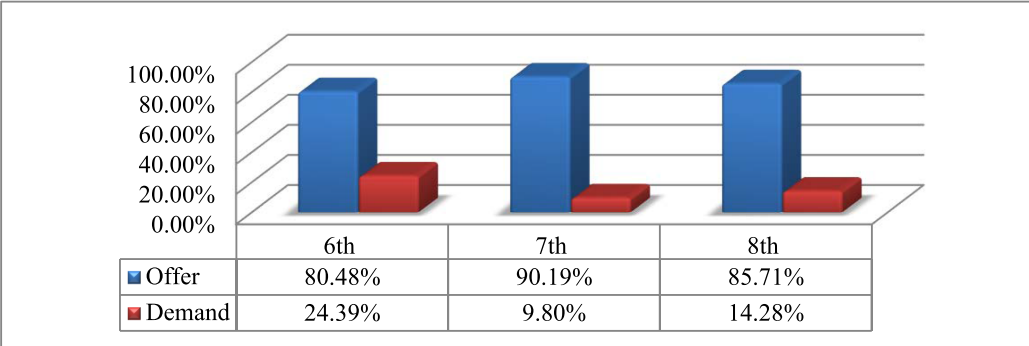


Fig. 15. Frequency of contact in all books.
Note: The figure illustrates the contact (demand and offer) of depicted participants with the viewers in all books.

In vertical perspective (as shown in Fig. 15), the high and low angles are both 9.80% each, and 80.39 of % images are eye-level. In the textbook of 8th, in horizontal perspective, 89.28 of % figures are frontal and only 10.71 of % images are oblique. In vertical perspective type, the eye-level images are 86.86%, high and low angle images are 8.92% and 5.35%,

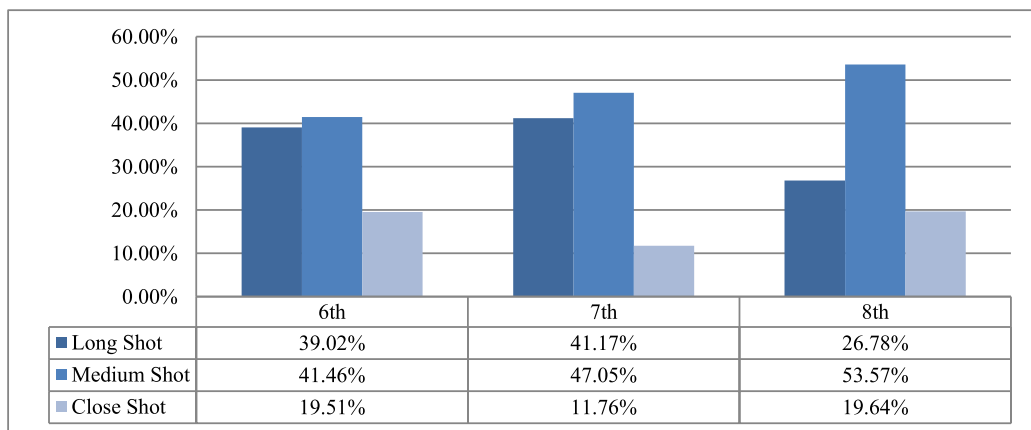


Fig. 16. Social distance in all elementary level school textbooks.

Note: The figure illustrates the social distance (Long, medium, and close shot) in the three elementary-level school textbooks.

respectively. In all textbooks, keeping in view the above given frequencies, horizontal frontal and vertical eye level angles are most dominant (as shown in Fig. 16).

The modality has two categories, colors and contextualization. The colors are given in three respects: color saturation, differentiation, and modulation. In the books of 6th and 8th grade, the colors are maximum in all three respects, while in grade 7th, the colors are moderate. In the textbook of 6th class, maximum saturation is given by 60.97% figures, 39.02% gives moderate color saturation, and there is no black and white image in this textbook. 58.53% figures show maximum color differentiation, moderate color differentiation is shown by 41.47% figures, and low differentiation is absent. Maximum color modulation is shown by 31.37% images, moderate modulation is noticed in 46.34% images. In the textbook of the 7th standard, the maximum color saturation is given by 31.37% images, moderate color saturation is shown by 62.74% images, and only 5.88% images depict low color saturation. In this book, color differentiation is maximum in 52.94% of % figures, 39.21% show moderate color differentiation, and only 7.84% of % images show the lowest value of differentiation. In this textbook, the modulation values are given as 17.64% images show maximum color modulation, a moderate modulation value is given by 58.82% figures, and only 23.52% images show a low modulation value. On the other hand, color saturation in the 8th-grade book is higher, with 55.53% in figures. However, moderate is lower than 12.5% and the remaining 33.92% images are without colors, i.e., in black and white. The contextualization in each textbook is demonstrated in Fig. 17.

The RQ3 is related to the compositional mode of MDA; it deals with the arrangement of visual elements in the figure to grasp learners' attention. The information value particularly deals with the placement of elements; in all three textbooks, the centre-margin type of information value is most significant, and the frequencies are illustrated in Fig. 18.

Frame lines and white spaces among the elements realize the framing in the visual illustrations. In the textbook of 6th class, the framing contains both connection and disconnection are present. The overlapping connection is 31.70% and integration is 41.46% present. In disconnection, separation is 21.95% present, and segregation is just 4.87%. In the textbook of 7th, overlapping is 23.52% and integration is 66.66% in the connection. Segregation is absent, while separation is 21.56%. In 8th grade, 83.92% of images show integration and only 10.71% give overlapping; no segregation is depicted, and 23.21% of figures show separation. The salience values in all three elementary school-level language textbooks are illustrated in Fig. 19.

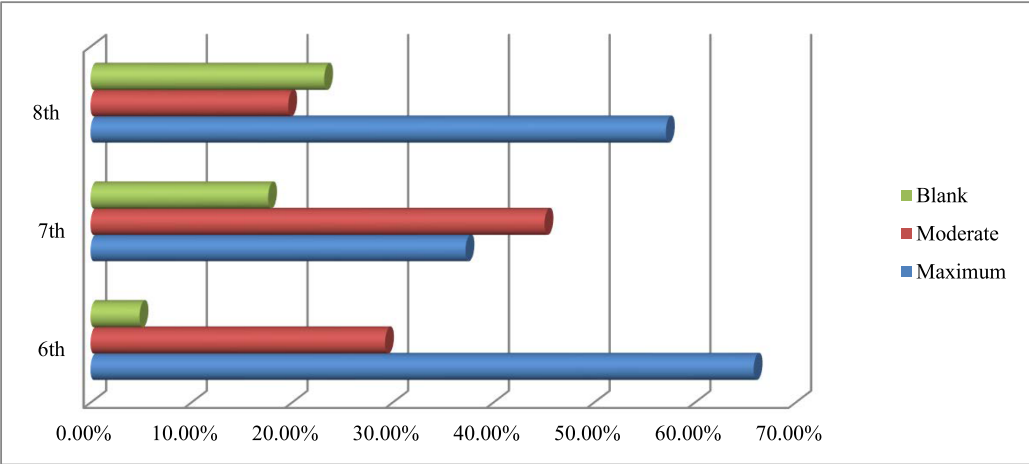


Fig. 17. Contextualization in all three elementary level school textbooks.
Note: The figure illustrates the contextual information in each book elementary school level textbook.

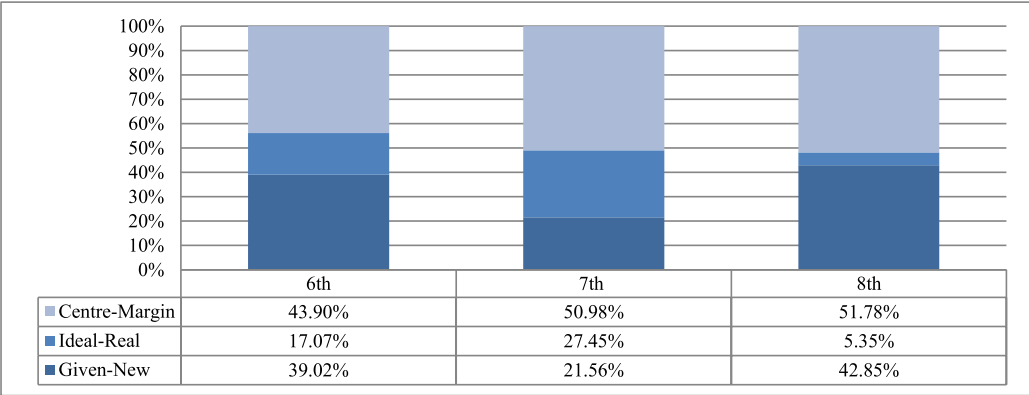


Fig. 18. Information value in all three elementary level school textbooks.
Note: The figure demonstrates the frequency of information value in all three elementary-level school textbooks.

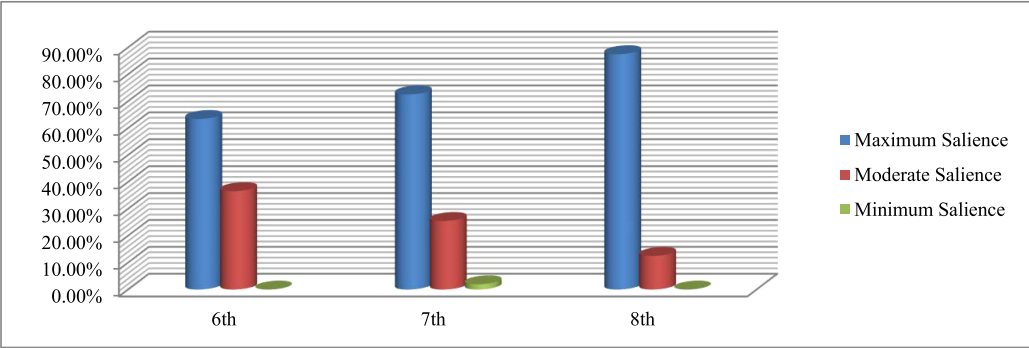


Fig. 19. Salience in all three elementary school-level textbooks.
Note: The figure illustrates the salience values in all three elementary school-level textbooks.

Conclusion

Multimodal language learning and Multimodal pedagogical content are of considerable importance since the end of the 20th century in all the fields of science and technology and in education, too. Presenting complex ideas and concepts using various modes of communication approaches to all types of learners. The most common multimodal elements in the textbooks are static images, animated images, flowcharts, graphs, and diagrams. These kinds of additional materials grasp learners' attention and make learning interesting, easy, and effective. The visual images particularly hold great importance for both the language teachers and learners. Multimodality, multimodal teaching, and learning multimodal literacy are all trending research topics globally as they are playing a significant role in curriculum designs for all levels and types of learners. The study examines the images illustrated in ELT textbooks utilizing [Kress & van Leeuwen \(2020\)](#) Model of MDA.

Regarding the aim of this research, which is to investigate the importance of visual representations from a VSS perspective, utilizing the lens of MDA of [Kress & van Leeuwen \(2020\)](#), in elementary school English language textbooks, several points are found. In all three elementary school-level language textbooks, in the representational mode, it is investigated that more concepts are presented conceptually, and in the conceptual processes, the symbolic kind of meaning is most prominent in conveying advanced and complex ideas. In the interactive mode of MDA, the images are illustrated as 'offers' and not 'demands'; medium social distance is most significant. The attitude of images, horizontal frontal is most prominent, and vertical eye-level angles are notable. In the color modality and contextualization, two books of 6th and 8th grade show maximum color values and context, while 7th 7th-grade textbook shows moderate values. With regard to the compositional mode, centre-margin information value is most frequent in all textbooks. In framing, connection through integration and disconnection through separation are highly significant. At last, the most considerable salience value in all the textbooks is maximum. The current study supports the idea of involving multimodal elements in pedagogical content and language textbooks to communicate meanings easily and effectively.

Recommendations

Language educators need to have a grip on multimodality and multimodal discourse so that it can help in understanding and importing the verbal concepts alongside supporting visuals. The integration of multiple modes within language textbooks helps in promoting multifaceted literacy, and it should be encouraged by teachers and learners both. The book designers must consider the pedagogical content for the particular level while designing books; they must do a needs analysis before designing. The multimodal construction of books can promote effective language learning. At the policy development level, the Curriculum boards, designers, and other educational authorities should incorporate multimodal elements; they can consider [Kress & van Leeuwen \(2020\)](#) MDA extracted from *Reading Images: Grammar of Visual Design* to enhance the semiotic richness of books. The visual images, when effectively integrated, can make the curriculum and syllabus more relevant to learners' needs.

Future researchers are recommended to explore various dimensions of this field, including the utilization of visual images, animations, videos, and other interactive elements. The data can be collected from different digital sources as animated content, e-platforms, and e-books. The Multimodal Approach can be expanded to the levels of Multimodal Critical Discourse Analysis (MCDA), Systemic Functional Multimodal Discourse Analysis

(SF-MDA). It can be practically linked with language learning classrooms to promote effective language learning.

Acknowledgement

The authors would like to thank Prince Sultan University for technical support.

Funding statement

This research received no external funding.

Conflict of interest

The authors declare no conflict of interest.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Author contribution

Amina Bashir and Aamir Aziz conducted the literature review, data collection, and multimodal analysis. Muhammad Imran and Norah Almusharraf designed and supervised the study, interpreted the findings, and finalized the manuscript.

References

- Abdullah, F. M. & Amroun, F. P. (2017). The impact of integrating modern technology in the law curriculum as an effective teaching strategy in higher education. *International Journal of Applied Business and Economic Research*, 15(25), 485–499.
- Ahmad, M., Mahmood, M. A., Siddique, A. R., & Asghar, S. A. (2021). ESL materials and gender representation: A corpus-based study of secondary school textbooks in Punjab, Pakistan. *Asian Women*, 37(1), 143–166.
- Anwar, R. (2018). The Use of Halliday's Interpersonal Metafunction in Creating Informative and Persuasive Advertisement: A Multimodal Analysis of AirAsia 2015 Commercial. In *International Conference on Language Phenomena in Multimodal Communication (KLUA 2018)* (pp. 129–134). Atlantis Press.
- Arshad, A., Shah, S. K., & Ahmad, M. (2020). Investigating cultural contents in English language teaching materials through textbook evaluation. *Journal of Language and Cultural Education*, 8(2), 127–145.
- Ayoola, M. O., Bankole, M. A., & Okorie, C. B. (2022). Text and Image Interplay in Selected Primary School English Textbooks in Ekiti State. *International Journal of Research and Innovation in Social Science*, 6(12), 312–318.
- Aziz, A. & Shakir, A. (2023). Exploring desirability and feasibility of learner autonomy: College English language teachers' beliefs in the Pakistani context. In *Linguistic Forum-A Journal of Linguistics* (Vol. 5, No. 2, pp. 85–102).
- Aziz, A., Mahmood, M. A., Ahmad, S., & Akbar, N. (2021). A corpus-based study of genre specific discourse: MA TEFL thesis abstracts. *Journal of Language and Linguistic Studies*, 17(S2), 884–898.
- Bashir, A., Aziz, A., Imran, M., & Almusharraf, N. (2025). Effect of CALL-based multimodal pedagogy on learner motivation and willingness to communicate in English: A study from university students' perspective. *Contemporary Educational Technology*, 17(2).
- Bi, M. (2019). Multimodal Discourse Analysis of News Pictures. *Theory & Practice in Language Studies (TPLS)*, 9(8).
- Byram, M., Holmes, P., & Savvides, N. (2013). Intercultural communicative competence in foreign language education: Questions of theory, practice and research. *The Language Learning Journal*, 41(3), 251–253.

- Cahyani, D. I. & Astuti, N. W. (2024). *English Teacher's Instructions In Implementing Bloom's Revised Taxonomy In Reading Practices To Foster Higher Order Thinking Skills (Hots) At Grade 10 Of Sma N 1 Mojolaban In The Academic Year 2023/2024* (Doctoral Dissertation, Uin Raden Mas Said).
- Campbell, C., Roberts, Y., Synder, F., Papp, J., Strambler, M., & Crusto, C. (2016). The assessment of early trauma exposure on social-emotional health of young children. *Children and youth services review*, 71, 308–314.
- Chen, Y. & He, W. (2015). Interpretation of the compositional meaning of movie posters from the perspective of multimodal discourse analysis. In *International Symposium on College Foreign Languages Education Reform and Innovation* (Vol. 56, No. 24, pp. 32–3263).
- Cheung, K. K. C. & Erduran, S. (2023). A systematic review of research on family resemblance approach to nature of science in science education. *Science & Education*, 32(5), 1637–1673.
- Chouhan, A. R., Saeed, F., & Abid, H. (2025). Reframing Climate Change Narratives in Pakistan: A Critical Ecolinguistic and Multimodal Discourse Analysis. In *Linguistic Forum-A Journal of Linguistics* (Vol. 7, No. 1, pp. 1–15).
- Colombo, M. & Afzaal, M. (2025). Revealing the ideology underpinning the European Union notion of resilient societies: A multimodal CDA approach. *Discourse & Society*, 09579265251365956.
- De Souza, R., Parveen, R., Chupradit, S., Velasco, L. G., Arcinas, M. M., Tabuena, A., ... & Ventayen, R. J. M. (2021). Language Teachers' Pedagogical Orientations in Integrating Technology in the Online Classroom: Its Effect on Students Motivation and Engagement. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(10), 5001–5014. <https://doi.org/10.17762/turcomat.v12i10.5268>
- Download E-Books | Punjab Curriculum and Textbook Board. (2022). Punjab.gov.pk. https://pctb.punjab.gov.pk/download_books
- El-Dakhs, D. A. S., Khan, S. K., & Al-Khodair, M. (2022). Do foreign language learners mine input texts for multiword expressions? The case of writing story retellings. *Ampersand*, 9, 100080.
- Fattah, S. A. & Hussein, A. S. (2025). Representational and Interactional Meanings of UNHCR Reports on Displaced Iraqis. *Alustath Journal for Human and Social Sciences*, 64(1), 36–17.
- Fitri, K., Hidayat, D. N., Husna, N., & Sufyan (2023). A. Multimodal Discourse Analysis Of Figures Used In The English Textbook "GET SMART PLUS".
- Fitriana, W. & Wirza, Y. (2021, April). An analysis of multimodal text in EFL textbook of secondary school in Indonesia in assisting students' text understanding. In *Thirteenth Conference on Applied Linguistics (CONAPLIN 2020)* (pp. 627–632). Atlantis Press.
- Friedman, A. & Ron, S. (2017). Unlocking the power of visual grammar theory: analyzing social media political advertising messages in the 2016 US election. *Journal of Visual Literacy*, 36(2), 90–103.
- Gualberto, C. & Kress, G. (2019). Contemporary landscapes of visual and digital communication: The interplay of social, semiotic, and technological change. *The Sage handbook of visual research methods*, 574–590.
- Gull, A., Imran, M., Yiunas, M., & Afzaal, M. (2020). A descriptive study of challenges faced by English language teachers in integrating information and communication technology (ICT) tools at elementary level in Pakistan. *International Journal of Advanced Science and Technology*, 29(8), 290–305.
- Guo, X. (2023). Multimodality in language education: implications of a multimodal affective perspective in foreign language teaching. *Frontiers in Psychology*, 14, 1283625
- Hafiz, A., Kurniawati, D., Aryanika, S., Verawati, H., & Firdaus, W. Y. (2024). Analyzing the Multimodal Discourse of the Tenth Grade English Work in Progress Textbook. *Journal of Linguistics and Social Sciences*, 2(1), 139–153.
- Halliday, M. A. K. (1978). Categories of the theory of grammar. *Word*, 17(2), 241–292.
- Han, X. (2025). A Multimodal Discourse Analysis of the Adapted Film To Live. *Research and Commentary on Humanities and Arts*, 3(1).
- Hart, C. (2020). Multimodal discourse analysis. In *Researching Discourse* (pp. 143–179). Routledge.
- Heberle, V. & Constanty, V. (2016). Meaning without words: analyzing the picture book *Bárbaro* through a social semiotic perspective. *Letras, Santa Maria*, 26(52), 89–109.
- Imran, M. & Ain, Q. (2019). Effects of non-native instructors' L1, beliefs and priorities on pronunciation pedagogy at secondary level in district Rajanpur, Pakistan. *Journal of Language and Cultural Education*, 7(2), 108–121.
- Imran, M. & Almusharraf, N. (2023). Language teaching in the linguistic landscape: Mobilizing pedagogy in public space: by David Malinowski, Hiram H. Maxim, and Sébastien Dubreil, Cham. The Social Science Journal, 1–3. <https://doi.org/10.1080/03623319.2023.2215116>
- Imran, M., Almusharraf, N., Abdellatif, M. S., & Ghaffar, A. (2024). Teachers' perspectives on effective English language teaching practices at the elementary level: A phenomenological study. *Heliyon*, 10(8).
- Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford University Press.
- Khorina, M., Sembiring, B., & Nugraha, D. G. (2023). Ideational Meanings in Results and Discussion Section of English Students' Final Project Reports Viewed from SF-MDA. *International Journal of English Language Studies*, 5(4), 24–31.
- Kim, H., Sefcik, J. S., & Bradway, C. (2017). Characteristics of qualitative descriptive studies: A systematic review. *Research in nursing & health*, 40(1), 23–42.
- Kress, G. & Van Leeuwen, T. (2020). *Reading images: The grammar of visual design*. Routledge.

- Linden, F. R. (2025). How 'empty' is the signifier 'the people'? Impasses of the poststructuralist approach in Ernesto Laclau's political ontology. *Journal of Political Ideologies*, 30(2), 327–351.
- Makinde, P. & Odili, P. (2023). A social semiotic analysis of selected basic pupils' English textbooks in Awka. *International Journal of Systemic Functional Linguistics*, 6(2), 62–69.
- McMurtrie, R. J. (2022). Observing, recording, visualising and interpreting visitors' movement patterns in art museums: A mixed method approach. *Multimodality & Society*, 2(2), 93–113.
- Mishra, S. B. & Alok, S. (2022). Handbook of research methodology. *EDUCREATION*.
- Novina, B. (2023). *Multimodal Analysis of English Language Teaching videos* (Doctoral dissertation, UIN RADEN INTAN LAMPUNG).
- Ouchra, H., Belangour, A., Erraissi, A., & Labied, M. (2025). New MDA Transformation Process from Urban Satellite Image Classification to Specific Urban Landsat Satellite Image Classification. *Journal of Environmental & Earth Sciences| Volume*, 7(01).
- Paltridge, B. (2018). Discourse analysis for the second language writing classroom. *The TESOL Encyclopedia of English language teaching*, 1–6.
- Pourrajab, M., Basri, R., Daud, S., M., & Asimiran, S. 2015. The resistance to change in implementation of total quality management (TQM) in Iranian schools. *The TQM Journal*, 27(5), pp. 532–543.
- Qi, J. (2024). Analyzing Image Meanings in Chinese EFL Textbooks: A Multimodal Perspective. *Theory & Practice in Language Studies (TPLS)*, 14(8).
- Rahikunmahtum, K., Nurkamto, J., & Suparno, S. (2022). The Pedagogical Potential of Visual Images in Indonesian High School English Language Textbooks: A Micro-Multimodal Analysis. *AL-ISHLAH: Jurnal Pendidikan*, 14(4), 5979–5990.
- Rajitha, I. G. N. A. & Claria, D. A. K. (2023). Implementing The Multimodal Communication Modes On Telegram Bot ELIA. In *WICSTH 2022: Proceedings of the 2nd Warmadewa International Conference on Science, Technology and Humanity, WICSTH 2022, 28–29 October 2022, Denpasar-Bali, Indonesia* (p. 69). European Alliance for Innovation.
- Risager, K. (2021). Language textbooks: windows to the world. *Language, Culture and Curriculum*, 34(2), 119–132.
- Roberts, F., Aziz, A. A., & Matore, M. E. E. M. (2020). Criteria of a Good ELT Textbook: Malaysian Teachers' Perception. *Universal Journal of Educational Research*, 8, 7433–7442.
- Roekhan, R., Suyitno, I., & Andajani, K. M. (2024). M., & Prastio, S. Discursive practices instilling the peace values for foreign learners in the BIPA textbook. *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan*, 43(1), 154–165.
- Royce, T. 1998. Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text. *JASFL Occasional papers*, 1(1), pp. 25–49.
- Saeed, S., Mahmood, M. A., & Hashmi, M. A. (2019). Comparative Translational Semiotic Analysis of 'The River of Fire' Systemic Functional Linguistic Approach. In *Linguistic Forum-A Journal of Linguistics* (Vol. 1, No. 2, pp. 14–23).
- Setiawan, A. (2021). Transitivity analysis of graduate casual conversation. In *English Language and Literature International Conference (ELLiC) Proceedings* (Vol. 4, pp. 313–319).
- Shehzadi, S., Aziz, A., & Bashir, A. (2024). Investigating the Interplay of Technology and Plagiarism: A Study of English Language Learners' Beliefs. *Journal for Social Science Archives*, 2(2), 469–485. <https://doi.org/10.59075/jssa.v2i2.76>
- Spencer, M.L. 1924. Editorial writing: Ethics, policy, practice. *Houghton Mifflin*.
- Triacca, S. (2017). Teaching and learning with pictures the use of photography in primary schools. In *Proceedings* (Vol. 1, No. 9, p. 952). MDPI.
- Unsworth, L., & Wheeler, J. (2002). Re-valuing the role of images in reviewing picture books. *Reading*, 36(2), 68–74.
- Yang, Y. (2025). Multimodal Discourse Analysis and Its Implications for English Learning in Chinese Secondary Schools. *New Explorations in Education and Teaching*, 2(11).
- Yang, J. & Zhang, Y. 2014. Representation meaning of multimodal discourse—A case study of English editorials in The Economist. *Theory and Practice in Language Studies*, 4(12), pp. 2564–2575.