Proceedings Series on Proceedings of Multidisciplinary Sciences, Volume 2, No. 1
International Conference of Business, Education, Health, and Scien-Tech (ICBENS)
ISSN: -

Integrating Multimodality in English Language Teaching (ELT): Junior High School Students' Perception in Digital Transformation Era

Amaliah Ramdani[1*], Suratman Dahlan[2], Samian Wulandari[3], A. Rasdiana Amsal Langgara[4]
[1,2,3]Postgraduate Program, Universitas Negeri Makassar, 90222, Indonesia
[4]SMP Negeri 42 Bulukumba, 92552, Indonesia

Corresponding Author Email: amaliah.ramdani@student.unm.ac.id

Abstract

The integration of multimodality in English Language Teaching (ELT) has received increasing attention in the context of digital transformation in education. This study explored junior high school students' perceptions of the integration of multimodal elements in their English classes through a mixed-method design. A total of 65 junior high school students participated in this study. The quantitative data were collected through close-ended questionnaire and analyzed descriptively, while the qualitative data collected from open-ended questionnaire were coded thematically. The researchers found that students perceived audio/video and visual design as the most effective multimodal elements in supporting learning, followed by voice and facial expressions, gestures and spatial position. Some challenges were identified, including linguistic difficulties, media-related problems, personal/concentration issues, technical problems. These findings highlight the potential of multimodality in enhancing the students' engagement and comprehension, while emphasizing the importance of careful media selection, provision of learning support, and effective classroom management.

Keywords: English_Language_Teaching; Multimodality; Students_Perception

1. Introduction

The development of digital technology has brought fundamental changes to the ways students learn foreign languages, including English. The current generation, often referred to as digital natives, is accustomed to interactive content and visual media [1]. This shift necessitates changes in English language teaching approaches in schools so that they align with students' everyday digital experiences. To meet the learning needs of digital natives, teaching practices need to move beyond traditional text-based methods and incorporate multimodal forms of communication.

In this context, the concept of multimodality, introduced by [2], becomes particularly relevant. Kress emphasized that communication is never monomodal but inherently multimodal, combining various semiotic modes such as text, sound, gesture, and image. Multimodality involves using more than one semiotic resource in constructing meaning (Jewitt, 2014). In language learning, it not only focuses on conveying messages through written text but also considers how voice, facial expressions, images, and even classroom layout contribute to meaning-making [3].

Building on this perspective, [4] argue that multimodal learning enables students to "read and write" across multiple and interconnected modes of communication. Consequently, students are expected not only to comprehend verbal language but also to develop the ability to critically interpret visual and nonverbal information. In the context of English language learning, the application of multimodal pedagogy allows teachers to design learning experiences that are more authentic, creative, and responsive to students' diverse needs [5].

Several previous studies have demonstrated the positive impact of implementing multimodality in EFL classrooms. For instance, [6] found that integrating multimodality in English classrooms enhances students' comprehension and helps them maintain focus during the learning process. Similarly, [7] revealed that the use of audiovisual media makes the learning process more interactive and facilitates the efficient delivery of materials, although certain technical challenges remain. [8] found that the use of multimodal grammar learning modules improves grammatical understanding and learner autonomy among middle school students. Furthermore, [9] discovered that the use of augmented reality in CLIL classes supports students in strengthening vocabulary acquisition and content comprehension.

The integration of multimodality influences not only the cognitive but also the affective aspects of learning. [10] found that high school students perceived multimodality as a factor that enhances their emotional engagement during learning. This finding is particularly important, as students' emotional engagement can reinforce their motivation and retention in foreign language learning. Nevertheless, the effectiveness of multimodal integration also depends on students' individual experiences and perceptions, providing a basis for deeper exploration of this dimension.

This study therefore focuses on exploring junior high school students' perceptions of multimodal integration in English language learning. Understanding how they perceive and respond to multimodal learning is crucial for designing effective multimodal pedagogy. Without such insights, multimodal integration in the classroom risks failing to meet students' actual learning needs. The main research question guiding this study is "how do junior high school students perceive the integration of multimodality in English language teaching?" in addition, this study will also explore the challenges that students experience when learning English through multimodal media. To address these research questions, this study employs a mixed-methods approach. By combining both quantitative and qualitative data, this study aims to provide a comprehensive understanding of how junior high school students experience and respond to multimodal learning in English classrooms. Thereby informing the design of more effective and engaging English language instruction.

2. Methodology

This study employed a mixed-method design. According to [11], mixed-method research is an approach that involves collecting and analyzing both quantitative and qualitative data, and then integrating the results to answer research questions comprehensively. This design was considered appropriate because it not only captures general trends in student responses through quantitative data, but also provides deeper insights into their experiences and challenges through qualitative data.

The study was conducted at SMP Negeri 42 Bulukumba, involving 65 student participants drawn from the eighth and ninth grades. The selection of participants was based on two considerations: (1) teachers at this school had already integrated multimodality into their English classes; and (2) the students had prior experience learning English through multimodal integration. These criteria ensured that the participants had relevant learning experiences with multimodal teaching

Two types of questionnaires were used as instruments in this study, namely a closed-ended questionnaire and an open-ended questionnaire. The closed-ended questionnaire, adapted from [12], focused on multimodal elements in English language learning. It consisted of 24 statements categorized into five areas: audio/video, visual design, voice and facial expression, gesture, and spatial position. A 4-point Likert scale was used to measure students' level of agreement with each statement. The open-ended questionnaire consisted of two questions designed to supplement the closed-ended data by providing exploratory insights, thus offering a more comprehensive picture of students' perceptions and the challenges they faced in learning English through multimodal methods. To avoid confusion, the questionnaire was written in Indonesian and distributed online via Google Forms.

The data collected through the questionnaires were analyzed descriptively to explore students' perceptions of multimodality in ELT. Responses from the closed-ended questionnaire were processed by calculating the percentage of student answers for each category. The results were then presented in table and figures to facilitate interpretation. Meanwhile, data obtained from the open-ended questionnaire were analyzed using [13] thematic analysis. All student statements were classified based on relevant themes. Finally, the result of quantitative data combined with the qualitative data to provide an in-depth interpretation of students' perceptions regarding multimodal integration in English classrooms, as well as the challenges they encountered.

3. Results and Discussion

3.1 Results

The Students' Perception of the Use of Multimodality in ELT

Based on result of the close-ended questionnaire, there are 24 questions representing the students' perspective on the strategy which is focused on five indicators, including audio/video, visual design, voice and facial expressions, gestures, and spatial position. This questionnaire involved 65 students as the respondent. The result summarized and presented descriptively in Table 1 below.

Table 1. Analysis of Close-Ended Questionnaire

Elements of Multimodality	Strongly Agree	Agree	Disagree	Strongly Disagree
Audio/Video	39,38	50,15	6,77	3,7
Visual Design	34,46	49,54	12,62	3,38
Voice and Facial Expression	39,7	42,46	12,92	4,92
Gestures	22,31	43,08	24,62	10
Spatial Position	25,85	43,38	20,31	10,46

The data in table 1 show generally positive results regarding students' perceptions of multimodal integration in English classroom. Among the five multimodal elements presented in the questionnaire, most students responded positively to the use of audio/video (39.38% strongly agree and 50.15% agree) and visual design (34.46% strongly agree and 49.54% agree). This was followed by voice and facial expressions, with 39.7% strongly agree and

42.46% agree. In contrast, gestures and spatial positions received higher percentages of disagreement compared to the other elements (gestures: 24.62% disagree and 10% strongly disagree; spatial positions: 20.31% disagree and 10.46% strongly disagree). These findings indicate that students prefer audio-visual elements more than non-verbal aspects such as gestures and spatial positioning.

The following section presents detailed results of each indicator within the multimodal elements from this questionnaire. The first element is audio/video, which consists of five indicators, and the results can be seen in Figure 1 below.

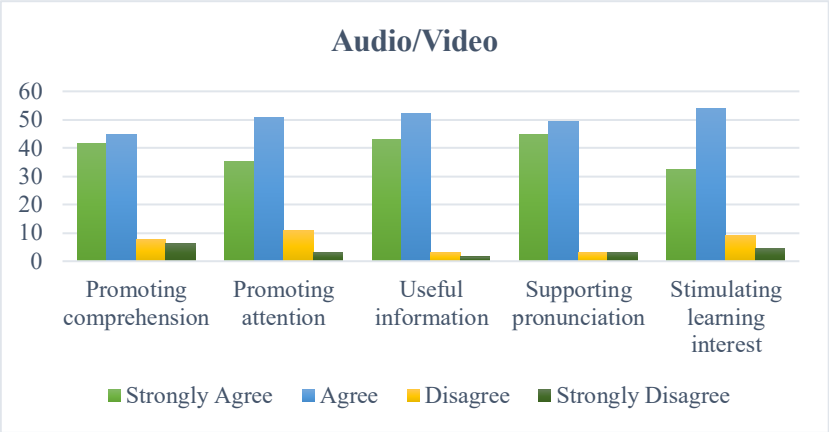

Figure 1. The students' perceptions of using audio/video in learning

Figure 1 above illustrates positive responses of the majority of students regarding the use of audio/video elements in English language learning. In terms of content comprehension, 41.54% of students strongly agree and 44.62% agree that these media helped them understand the lesson material. Most students also thought video/audio as effective in capturing attention (35.38% strongly agree; 50.77% agree) and providing useful information (43.08% strongly agree; 52.31% agree).

Furthermore, 44.62% of students strongly agree and 49.23% agree that it can support their pronunciation skills. Regarding motivation, 32.31% strongly agree and 53.85% agree that video/audio stimulated their interest in learning. These findings indicate that more than three-quarters of the students consistently responded positively to the integration of video/audio in English classes, while the proportion of students expressing disagreement or strong disagreement was relatively small across all indicators.

Next, the visual design element also consists of five indicators. The results can be seen in Figure 2 below.

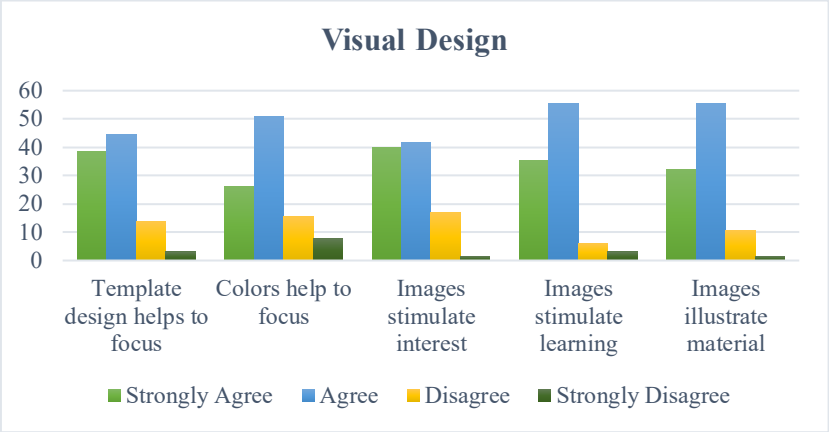

Figure 2. The students' perceptions of visual design in learning

Students' positive responses to the visual design element are illustrated in Figure 2 above. A total of 38.46% of students strongly agree and 44.62% agree that the template design helped them maintain focus, although 13.85% disagree. The use of colors in the PowerPoint was also considered fairly effective, with 26.15% strongly agree and 50.77% agree, despite a relatively higher proportion of disagreement compared to other aspects (15.38% disagree; 7.69% strongly disagree). Furthermore, images or photos in the slides were perceived as stimulating students' interest in learning, with 40% strongly agree and 41.54% agree.

Regarding comprehension, 35.38% of students strongly agree and 55.38% agree that images helped them understand the lesson content. A similar trend was observed for the function of material illustrations, with 32.31% strongly agree and 55.38% agree. Overall, these data indicate that visual design, particularly the use of images/photos, is perceived by students as a multimodal element that can support their focus, interest, and understanding of the lesson in English learning.

The third element is voice and facial expression, and the results can be seen in Figure 3 below.

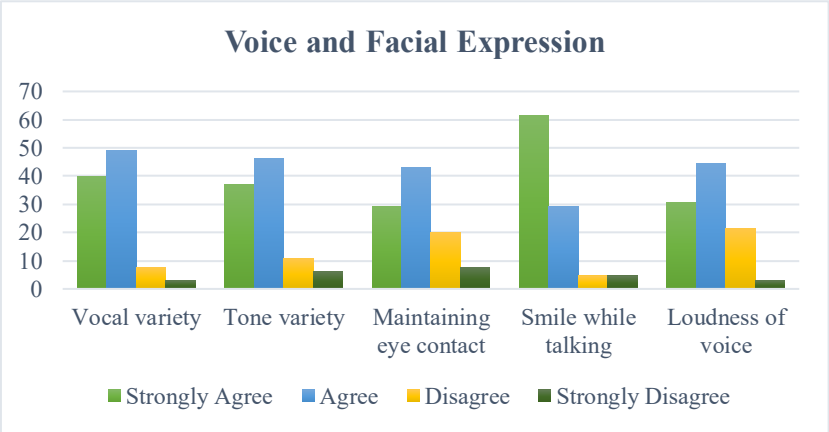

Figure 3. The students' perceptions of voice and facial expression in learning

The data in Figure 3 above indicate that students perceive the teacher's voice and facial expressions positively as important components of multimodality. A total of 40% of students strongly agree and 49.23% agree that the teacher's vocal variation supports learning. The use of pitch and intonation was also considered beneficial, with 36.92% strongly agree and 46.15% agree. Regarding eye contact, 29.23% strongly agree and 43.08% agree that it supports classroom interaction, although 20% disagree and 7.69% strongly disagree, indicating some differences in perception.

Interestingly, the teacher's smile while speaking received the most positive response: 61.54% strongly agree and 29.23% agree, while only around 9.24% disagree. Meanwhile, the clarity and volume of the teacher's voice were perceived relatively well, with 30.77% strongly agree and 44.62% agree, although 21.54% disagree. Overall, the teacher's facial expressions and voice use (particularly smiling, vocal variation, and intonation) are viewed as effective multimodal elements in enhancing students' comfort and engagement in English learning.

For the fourth element, which relates to gestures, there are four indicators. The students' responses can be seen in Figure 4 below.

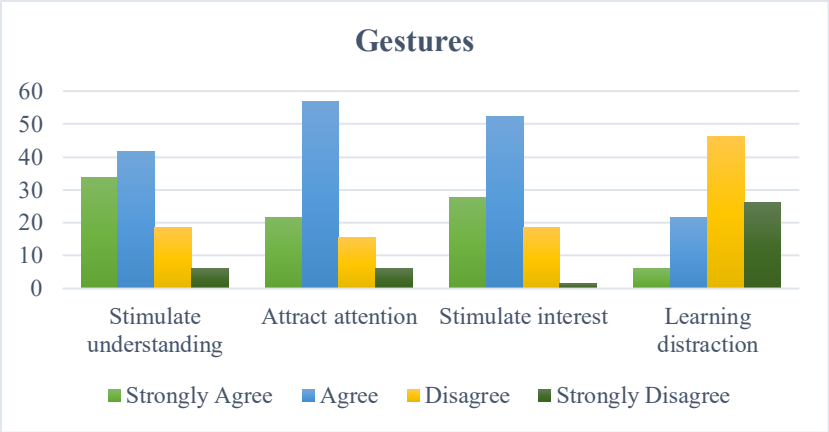

Figure 4. The students' perceptions of gestures in learning

The data illustrated in Figure 4 above show that 33.85% of students strongly agree and 41.54% agree that the teacher's hand gestures helped them understand the content being discussed, although 18.46% disagree and 6.15% strongly disagree. Regarding attention, 21.54% strongly agree and 56.92% agree that gestures captured their focus, while 15.38% disagree and 6.15% strongly disagree. Furthermore, 27.69% of students strongly agree and 52.31% agree that the teacher's hand movements could stimulate interest in learning, whereas 18.46% disagree and 1.54% strongly disagree.

Interestingly, when asked whether the teacher's gestures distracted them, the majority of students disagree: 46.15% disagree and 26.15% strongly disagree, while only 21.54% agree and 6.15% strongly agree. These findings indicate that, overall, the teacher's hand gestures are generally perceived as multimodal elements that aid comprehension, attract attention, and enhance learning interest rather than as distracting factors.

The final element relates to spatial position. This element consists of five indicators, and the students' responses can be seen in Figure 5 below.

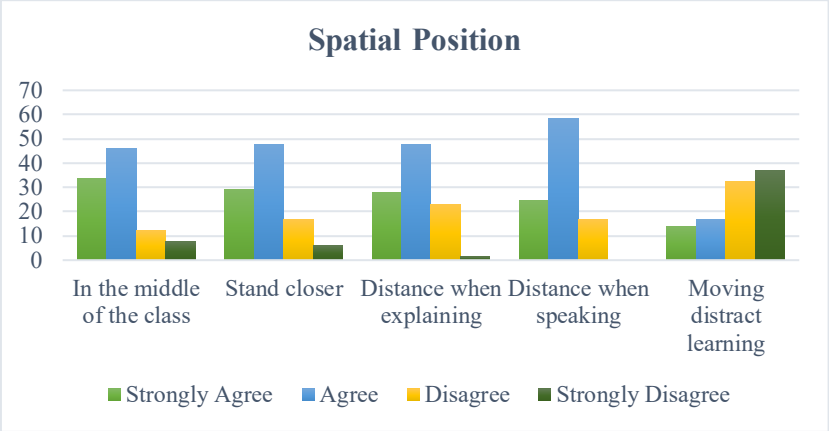

Figure 5. The students' perceptions of spatial position in learning

The data in Figure 5 above indicate that the teacher's position and mobility in the classroom are perceived by students as fairly important in supporting English learning. A total of 33.85% of students strongly agree and 46.15% agree that a teacher moving to the center of the class facilitates the learning process, although 12.31% disagree and 7.69% strongly disagree. When the teacher stands closer to the students, 29.23% strongly agree and 47.69% agree, while 16.92% disagree and 6.15% strongly disagree. Similarly, 27.69% strongly agree and 47.69% agree that a closer distance during explanation improves comprehension, although 23.08% disagree.

Interestingly, 24.62% strongly agree and 58.46% agree that sitting or standing near students while speaking has a positive impact, with only 16.92% disagree. Regarding the statement that the teacher's mobility during teaching distracts students, the majority disagreed: 32.31% disagree and 36.92% strongly disagree, while only 16.92% agree and 13.85% strongly agree. These findings suggest that a teacher's proximity and active movement in the classroom are generally perceived as enhancing interaction and focus, rather than as distractions.

Students' Challenges in the Integration of Multimodality in ELT

The analysis of the open-ended questionnaire reveals several challenges experienced by students when multimodal media are integrated into English language teaching. These problems can be categorized five main themes: linguistic difficulties, media-related problems, personal/concentration issues, technical problems, and no reported difficulties. The data distribution can be seen in Figure 6 below.

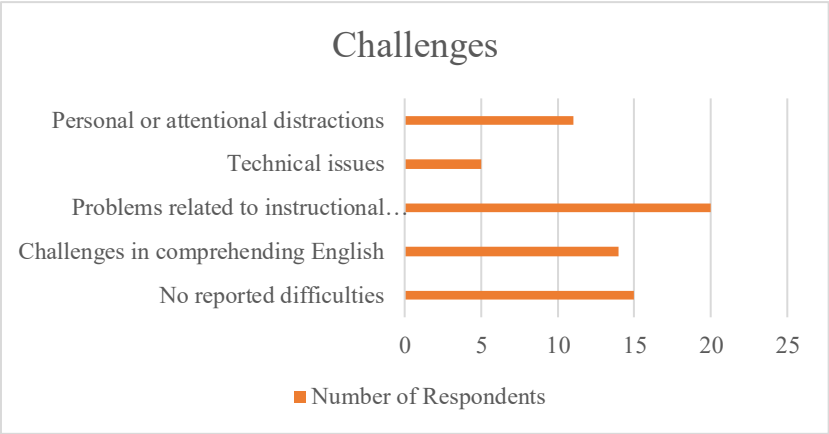

Figure 6. The Challenges Faced by Students

To begin with, several students reported that their difficulties were not related to the media or the teacher's methods, but rather to their limited knowledge of English itself. They struggled with understanding vocabulary or translating sentences. As one student explained, *"belum sepenuhnya mengetahui kata dan kalimat bahasa Inggris"* (DP), while another added, *"tidak mengetahui arti dari Bahasa Inggris"* (AK). Overall, 14 students mentioned linguistic difficulties as a major barrier.

Another challenge concerns the way media were presented in class. Students felt that videos and audios were sometimes too fast or unclear, and PowerPoint slides were either too brief or overloaded with text. For instance, one student admitted, *"Saya kesulitan memahami video/audio karena terlalu cepat tanpa teks"* (NIA), and another mentioned, *"kendalanya yaitu media yang di putarkan biasanya terlalu cepat"* (RZN). This category was reported by 20 students, indicating that media design and delivery play a crucial role in the effectiveness of multimodal integration.

In addition, concentration issues and internal factors also emerged. Some students found it difficult to stay focused due to personal reasons such as feeling sleepy, lacking attention, or being distracted by noisy classmates. As expressed by EM, *"kendala saya kadang tidak bisa konsentrasi saat teman-teman berbicara terus dan ribut"*, while AII added, *"kadang² tidak mendengarnya guru karena tidak fokus"*. This theme was noted by 11 students, highlighting how psychological and environmental conditions can shape learning experiences.

Beyond these internal challenges, students also pointed to external barriers related to technology and infrastructure. Problems such as unstable internet connections, limited devices, or technical errors occasionally disrupted learning. One student mentioned, *"Kesulitan memahami bahasa Inggris yang cepat, masalah teknis, keterbatasan waktu"* (NA), while another simply said, *"jaringan jelek"* (AF). A total of 5 students mentioned technical issues, showing that access to reliable resources is crucial for fully benefiting from multimodal strategies.

Interestingly, despite the challenges outlined above, a number of students stated that they experienced no significant difficulties at all. On the contrary, they expressed enjoyment in learning with multimodal integration, reporting that it made lessons more engaging and motivating. As NA noted, *"bagi saya tidak ada kendala karena adanya berbagai media tersebut saya merasa tertarik"*. This positive perception was reported by 15 students, indicating that, when implemented thoughtfully, multimodal strategies can enhance student engagement and support a positive learning experience.

In conclusion, while students identified several barriers ranging from linguistic limitations to technical problems, the overall perception toward multimodal learning remains positive. This indicates that multimodal strategies are not only relevant but also beneficial to be implemented in English language classrooms in the digital era.

3.2 Discussions

The findings of this study reveal that junior high school students generally hold positive perspective toward the integration of multimodality in English language teaching. Among the five multimodal elements examined, audio/video and visual design were perceived as the most effective components in supporting learning, while gesture and spatial position received lower levels of agreements. This indicates that students can learn better when teachers combine sounds, images, and written texts rather that using only one mode of communication. In line with this, [14] said that learning through multiple channels/mediums can make the language input easier to understand and remember.

Another important aspect that emerged was the role of emotional engagement in the learning process. The students' responses shows that the teacher's voice, intonation and even the facial expression play a crucial role in shaping the students' motivation during the learning process. This aligns with the *social semiotic* perspective of [15], which views nonverbal cues such as tone and expression as socially meaningful modes that influence interpersonal relationship. Moreover, [16] agree that emotional engagement is a vital factor in maintaining attention and participation in teaching and learning process. Therefore, multimodal teaching not only facilitates students' understanding through multiple sensory channels but also fosters affective engagement, encouraging them to interact more actively with the learning process.

Although gestures and spatial position received less enthusiastic responses compared to audio-visual elements, they were still recognized as meaningful components of classroom communication. [17] said that gestures are extensions of thought and language, serving as visual aids that reinforce meaning. In the Indonesian EFL context, classroom teaching tends to emphasize verbal and visual resources rather than embodied modes of communication [18]. Consequently, the role of gestures and spatial position may not yet be fully optimized as part of multimodal pedagogy.

The qualitative findings also shed light on several challenges faced by students during multimodal integration. Linguistic difficulties emerged as one of the most reported barriers, indicating that limited English proficiency may reduce the effectiveness of multimodal input. [19] *Input Hypothesis* explained that comprehensible input is necessary for language acquisition; when multimodal materials are linguistically demanding, they may hinder rather than facilitate

understanding. Other difficulties include problems related to media design (e.g., videos played too fast, PowerPoint slides containing excessive text) and technical constraints such as unstable internet connections or limited access to devices. These findings resonate with [20], who observed that the success of multimodal learning in Indonesian schools depends greatly on technological readiness and the quality of instructional media.

Despite these challenges, some students expressed that learning English through multimodal approaches made lessons more enjoyable and motivating. This indicates that well-designed multimodal pedagogy can accommodate the characteristics of the digital generation, who are accustomed to interactive and media-rich environments [1]. However, the effectiveness of multimodality is not guaranteed simply by the use of technology. [4] that teachers need multimodal literacy, which is the ability to design, interpret, and coordinate various modes of communication in a meaningful way. Without this, multimodal teaching risks becoming superficial or fragmented. Therefore, teacher professional development should prioritize training in multimodal design and classroom coordination to ensure that media selection, language level, and interactivity are aligned with students' learning needs and contexts.

In summary, this study emphasizes that multimodal integration can enhance cognitive and affective dimensions in English language learning. Audio-visual materials have been proven effective in improving student comprehension and engagement, while teacher expressiveness contributes to emotional engagement. However, linguistic barriers, media design, and infrastructure limitations remain significant challenges. To maximize the benefits of multimodal pedagogy, ELT practitioners need to develop balanced strategies, integrate various modes, adapt to students' linguistic levels, and utilize technology wisely. Ultimately, effective multimodal teaching requires not only creative media but also pedagogical sensitivity to how different modes work together to shape meaning and learning experiences.

4. Conclusion

This study shows that integrating multimodality in English language teaching can enhance both students' understanding and engagement. Audio-visual materials were found to be the most effective, while teacher expressiveness supports emotional involvement. Challenges such as limited English proficiency, media design issues, and technical constraints were also identified. To maximize the benefits, teachers need to carefully design and coordinate multiple modes, adapt to students' linguistic levels, and use technology effectively. However, this study is limited to exploring students' perceptions and does not examine the direct impact of multimodal approaches on learning outcomes. Future research should investigate the effects of multimodal integration on students' achievement and overall language performance.

5. Acknowledgement

The author would like to express sincere gratitude to all the students who participated in this study and shared their valuable insights. Special thanks are extended to the school teachers and staff, especially the English teacher, for their support in facilitating data collection and distributing the online questionnaires to the students. The author also appreciates the guidance and constructive feedback from reviewers and colleagues, which greatly improved the quality of this manuscript.

6. References

[1] M. Prensky, "Digital natives, digital immigrants part 2: Do they really think differently?," *Horiz.*, vol. 9, no. 6, pp. 1–6, 2001.

[2] G. Kress, *Multimodality: A social semiotic approach to contemporary communication*. routledge, 2009.

[3] J. Bezemer and C. Jewitt, "Multimodal analysis: Key issues," *Res. methods Linguist.*, vol. 180, 2010.

[4] B. Cope and M. Kalantzis, "'Multiliteracies': New literacies, new learning," in *Framing languages and literacies*, Routledge, 2013, pp. 105–135.

[5] T. A. Al Fajri, "Pentingnya penggunaan pendekatan multimodal dalam pembelajaran," *Waskita J. Pendidik. Nilai Dan Pembang. Karakter*, vol. 2, no. 1, pp. 57–72, 2018.

[6] E. Luthfiyah and D. Mulatsih, "Exploring Senior High School Students' Perceptions of Multimodality in ELT," *J. English Lang. Teach. Linguist.*, vol. 10, no. 2, pp. 261–277, 2025.

[7] E. Daulay, "A case study in the use of audio-visual media for teaching listening skills in junior high school EFL classrooms: Perceived effectiveness and challenges," *J. English as a Foreign Lang.*, vol. 15, no. 1, pp. 135–159, 2025.

[8] S. M. Abdelhalim, "From traditional writing to digital multimodal composing: promoting high school EFL students' writing self-regulation and self-efficacy," *Comput. Assist. Lang. Learn.*, pp. 1–30, 2024.

[9] M. Fernandez-Alcocer and J. Belda-Medina, "Augmented Reality's Impact on English Vocabulary and Content Acquisition in the CLIL Classroom," *Appl. Sci.*, vol. 15, no. 19, p. 10380, 2025.

[10] S. H. Fitri, H. E. Putra, S. Suyono, E. Elfi, and M. A. Afriandi, "Students' Perceptions of the Benefits of Multimodal Texts Used by English Teachers at SMA Uswatun Hasanah Padang Panjang," *Ahmad Dahlan J. English Stud.*, vol. 12, no. 1, pp. 58–73, 2025.

[11] J. W. Creswell and V. L. P. Clark, *Designing and conducting mixed methods research*. Sage publications, 2017.

[12] J.-E. Peng, "The roles of multimodal pedagogic effects and classroom environment in willingness to communicate in English," *System*, vol. 82, pp. 161–173, 2019.

[13] V. Braun and V. Clarke, "Successful qualitative research: A practical guide for beginners," 2013.

[14] R. E. Mayer, H. Lee, and A. Peebles, "Multimedia learning in a second language: A cognitive load perspective," *Appl. Cogn. Psychol.*, vol. 28, no. 5, pp. 653–660, 2014.

[15] T. Van Leeuwen, *Introducing social semiotics*, vol. 490. Routledge London, 2005.

[16] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Rev. Educ. Res.*, vol. 74, no. 1, pp. 59–109, 2004.

[17] S. Kang and B. Tversky, "From hands to minds: Gestures promote understanding," *Cogn. Res. Princ. Implic.*, vol. 1, no. 1, p. 4, 2016.

[18] N. A. Drajati, L. Tan, S. Haryati, D. Rochsantiningsih, and H. Zainnuri, "Investigating English language teachers in developing TPACK and multimodal literacy," *Indones. J. Appl. Linguist.*, vol. 7, no. 3, pp. 575–582, 2018.

[19] S. D. Krashen, "The input hypothesis: Issues and implications," *(No Title)*, 1985.

[20] M. Munawarah, N. Khotimah, A. Widiastuti, and A. T. Wijayanti, "Analysis of Junior High School Students' Learning Media Preferences: Case Study in Bener Meriah Regency," *J. Pemberdaya. Masy.*, vol. 4, no. 2, pp. 399–413, 2025.