The Influence of Multimodal Visual Methodologies on EFL University Students' Audio-Visual Comprehension, Verbal and Nonverbal Communication

Ibrahim Hassan Ali Al-jumaily [1], **Istabraq Tariq Jawaad Alazzawi** [2]

[1] University Of Anbar, Ramadi, Iraq

[2] University of Tikrit, Tikrit, Iraq

ABSTRACT

Background: In the last decades, there have been an increased use of multimodal teaching (text, images, audio, video) in EFL education, offering learners diverse input to suit different learning styles. However, how specific input types (e.g., dialects, registers, multimedia) affect learners with varying cognitive or cultural backgrounds remains unclear. Addressing this gap is essential for effective EFL instruction in today's multimodal learning environments.

Purpose: This study aims to examine how multimodal visual methodologies influence EFL students' development in audio-visual comprehension as well as verbal and nonverbal communication. By investigating these dimensions, the research seeks to fill a critical gap in the understanding of how integrated sensory modalities shape communicative competence in technologically enhanced learning environments. Furthermore, the study explores underlying psychological, social, and pedagogical factors that facilitate or hinder these outcomes, thereby providing tangible insights for designing more cognitively effective EFL curricula.

Method: The study's sample consists of 214 EFL university students. A mixed-mode descriptive research design was used. Audio-visual comprehension, verbal and nonverbal communication tests served as the quantitative data collection instruments. Repeated-measures ANOVA and t-test paired samples were conducted on the set of three test scores over time. In the qualitative phase, data were collected from 20 purposively selected students by using a semi-structured focus group interview and was analysed qualitatively based on thematic analysis.

Results: The findings demonstrated statistically significant improvements in students' audio-visual comprehension, verbal communication, and nonverbal communication across three repeated-measures assessments. For example, mean scores for audio-visual comprehension increased from 46.22 to 67.59 ($p < .001$, $\eta^2 = .535$). Similar gains were observed in verbal ($\eta^2 = .561$) and nonverbal communication ($\eta^2 = .559$). Qualitative data confirmed that students perceived the multimodal learning environment as psychologically engaging and socially supportive, highlighting increased mental readiness, reduced listening anxiety, and improved interpretation of body language.

Conclusion: The study highlights the critical role of linguistic and psychological factors in EFL, particularly emphasising the significance of audio-visual comprehension and both verbal and nonverbal communication. It also identifies three key variables (i.e., audio-visual comprehension, verbal, and nonverbal communication) that aid EFL students in enhancing their retention, long-term memory and confidence for independent English learning.

KEYWORDS

multimodal learning, EFL university students, audio-visual comprehension, verbal and nonverbal communication, subtitled video input, psychological factors in language learning

INTRODUCTION

In today's rapidly evolving digital landscape, the shift from traditional print-based media to digital platforms has transformed the ways in which discourse is practiced. This evolution reflects the concept of multimodal communication, in which meaning co-constructed based on interaction among different modes comprising text, image, sound, gesture and more (AbdulGhafoor & Challob, 2021). Multimodality, as a theoretical framework, explores how these modes work together under sociocultural influences (Jewitt & Kress, 2003; Norte Fernández-Pacheco, 2018). In education, multimodal ensembles—combinations of different modes—enhance learning by integrating visual, auditory, and kinesthetic elements (Love, 2019; Halliday & Hasan, 1985). In multimodal education, ensembles reflect a combination of modes that improve teaching and learning and articulate visual, aural and kinetic elements (Norte Fernández-Pacheco, 2018; Heimbürger, 2013). Multimodal approaches in EFL education have been revolutionizing language teaching by stepping away from text-based approaches and incorporating visual, auditory, as well as kinesthetic aspects, making intensive learning (Purwaningtyas, 2020; Ghoushchi et al., 2021). This is in accordance with Communicative Language Teaching (CLT) concepts which emphasize real and contextual communication (Richards & Rodgers, 2014). Utilizing images, videos, gestures, spatial arrangements, and appealing to a variety of senses supports learning and the ability to understand and produce language in natural settings (Van Leeuwen, 2020).

One of the key components of multimodal learning in EFL education is audio-visual comprehension, a core aspect of multimodal EFL learning, involves interpreting auditory and visual stimuli like videos, podcasts, and multimedia. It enhances real-life communication skills by exposing learners to authentic pronunciation, intonation, and nonverbal cues (Yi et al., 2024; Yeh, 2022). Such materials improve listening and speaking abilities while boosting engagement through contextualized, immersive scenarios (Chiriac, 2025; Putri et al., 2024). Research underscores their value in aiding comprehension via visual context and nonverbal decoding, crucial for EFL learners navigating cultural and linguistic nuances (Vandergrift & Goh, 2012; ED-DALI, 2024). However, challenges persist with fast-paced audio or lack of visual support (Huang, 2006; Fitria, 2024). Technology's role is crucial but needs to be mediated by the teacher to enhance the potential of multimodal tools for language learning. Furthermore, verbal communication (including the spoken and written word) is an indispensable vehicle through which messages can be delivered, thoughts can be expressed and interaction may be achieved. As pointed out by Ruswandi & Arief (2024) and Nofali & Gasim (2024), oral practice is a must when aiming to gain both fluency and accuracy, since it involves learners in active production of language. This active engagement is further sustained by conversational interaction that elicited feedback and clarification that support a dynamic learning setting (Ruswandi & Arief, 2024; Adhitya & Valiansyah, 2024). These interactions not only improve students' language proficiency but also install a sense of self-assurance, which is an important element in successful communication (Nofali & Gasim, 2024).

However, verbal communication does not function independently; it is deeply intertwined with nonverbal cues that shape meaning. For instance, a teacher's tone of voice can profoundly influence message reception (Abdikarimova et al., 2021; Sutiyatno, 2018; McDuffie et al., 2021). This coordination of verbal and nonverbal elements emphasizes the advantage of multi-modal approaches where various sensory inputs are combined to aid in understanding and comprehension. Studies have shown that video-based tasks (such as role- plays, simulations) introduce students to authentic use of language, assisting them to mimic the pronunciation and discourse patterns of native speakers (Hahl et al., 2025). Furthermore, visual learning tools like infographics and mind maps provide students with more coherence to structure and structure ideas synthesize and organize using a text, which connects the understanding with production.

It also emphasizes the significant role of nonverbal communication, such as gestures, expressions, eye contact, body volume and distance in enriching interaction and in backing verbal communication especially to EFL contexts. Because these features are cultural-bound, multimodal strategies (for example, video analysis and interactive activities) can be used to ecode and practise them (Pratolo, 2019; Ruswandi & Arief, 2024). Teachers may send non-verbal cues to help confirm the message, promote inclusiveness, and increase student engagement (Haneef et al., 2014; Burgoon, 2003; Bambaeeroo & Shokrpour, 2017). Studies have confirmed that facial expressions were one of the most influential nonverbal approaches, and verbal and non-verbal communications have a positive impact on the academic performance (Mikhael et al., 2022; Sutiyatno, 2018). Moreover, adjusting communication styles to cultural situation is important in EFL to avoid misconceptions and enhance the general communicative competence (Tussupova et al., 2017). The incorporation of nonverbal communication into language instruction encourages a more holistic understanding and genuine cross-cultural communication between individuals.

The integration of audio-visual comprehension, verbal, and nonverbal communication is essential because communication is inherently multimodal. In the real world, these modes are interconnected in which nonverbal signs frequently present themselves in support or supplement of verbal messages (Kress & Van Leeuwen, 2001; Ruswandi & Arief, 2024). For EFL learners, a combined use of the three modes is necessary for effective communication because auditory

and visual information presented together facilitates comprehension (Sueyoshi & Hardison, 2005). To concentrate on just one mode would be to get only a partial picture—one which ignores the oscillating relationships between modes, as well as their aggregate, holistic function in language acquisition. Here, of course, multimodal technologies including Smart Boards, projectors, Vocast (for audio-visual comprehension) and visual tools for verbal, nonverbal cues and the like would improve the EFL students' understanding and communication. They offer interactive, standardized and measurable way to evaluate progress made in listening, speaking, and body language which provides data driven evaluation of multimedia language teaching (Olelewe et al., 2023).

Despite the growing body of research supporting the effectiveness of multimodal methodologies in EFL education, there remain gaps in the literature that warrant further investigation. While many studies have addressed individual components of multimodal input, few have examined their integrative effect on communicative competence as a multidimensional construct, such as listening and speaking, there is a need for more comprehensive research that examines the interplay between audio-visual comprehension, verbal, and nonverbal communication. What remains insufficiently studied is how these three domains interact over time within a multimodally enriched EFL environment, and how learners' engagement and psychological readiness mediate this process. Addressing these gaps is essential for developing a more nuanced understanding of how multimodal methodologies can be tailored to meet the diverse needs of EFL learners.

On the basis of the above concepts, this study aims to examine how multimodal visual methodologies influence EFL students' development in audio-visual comprehension as well as verbal and nonverbal communication. By investigating these dimensions, the research seeks to fill a critical gap in the understanding of how integrated sensory modalities shape communicative competence in technologically enhanced learning environments. Furthermore, the study explores underlying psychological, social, and pedagogical factors that facilitate or hinder these outcomes, thereby providing actionable insights for designing more inclusive and cognitively effective EFL curricula. The study seeks to answer the following research questions:

RQ1: What is the influence of multimodal visual methodologies on EFL students' audio-visual comprehension, verbal and nonverbal communication among three rounds of testing?

RQ2: How can multimodal visual methodologies influence EFL students' audio-visual comprehension, verbal communication and nonverbal communication?

LITERATURE REVIEW

Multimodal Approach: The Integration of Diverse Modes in Educational Practices

The underlying theoretical basis for multimodal practices lies in socio-cultural theory (Vygotsky, 1978), which argues that cognitive development, is mediated by social interaction and culture tools. From this view, language development is a socially situated activity, involving shared activities with more knowledgeable others and the manipulation of cultural tools (tools in the broadest sense, such as texts, graphics, and technologies). A multimodal theory This principle is realised in practice as multimodal approaches that combine a range of cultural tools (i.e. visual, auditory and kinesthetic resources) to mediate reflection and learning. However, the possibility that learners may be differentially proficient in combining the two modes is not considered in the current investigation. And for some students processing too much audio, visual and kinesthetic input simultaneously can create cognitive dissonance and detract from rather than facilitate understanding.

Furthermore, these approaches are consistent with cognitive load theory (Sweller, 1988) presenting that when the information load is distributed over several sensory modalities, the efficiency of learning is increased. Multimodal technologies mitigate cognitive overload through multiple modes of representation (visual and auditory input) and learners can process linguistic input more effectively. Taken together, these theoretical perspectives provide an underpinning for the integration of multimedia in EFL teaching by suggesting its double significance in enabling social interaction and cognitive attention. These results support the emphasis in the present study on the ways in which multimodal scaffolding (e.g., subtitles, images, gestures) can support EFL learners' retention and comprehension of auditory and visual input.

According to multimodal academics, modern learners create meaning by applying several modes (written and spoken language, gesture, visuals, sound, and movement) in a fast-changing, diverse era filled with numerous semiotic and digital resources in their everyday lives (Jewitt & Kress, 2003; New London Group, 1996). This perspective underscores that knowledge is inherently multimodal, co-constructed, and performed or represented (Miller, 2007, p. 65). Consequently, students' ability to consume, understand, and produce multimodal literacies is crucial for both academic and social purposes (Jewitt & Kress, 2003; Yi, 2014). Multimodal approaches in education present content through visual, aural, and written formats, which can enhance learning performance. Research suggests that using various presenta-

tion modes makes learning easier and increases attention, particularly for lower-achieving students (Chen & Fu, 2003; Farías et al., 2014; Pintado & Fajardo, 2021). These findings align with the present study's focus on multimodal instruction as a tool for educational equity, as they demonstrate that differentiated presentation methods can reduce performance gaps. However, these studies primarily examine short-term engagement rather than long-term knowledge retention, which is a key concern in the current research.

Based on this idea, Mayer (2014) proposed the *multimedia effect*, which states that students learn more when words are accompanied by pictures than when words appear alone. This confirms the claim that the multichannel input facilitates understanding and long-term retention of the target language, an important factor at one of the EFL settings, where learners need to process information from number of sensory channels. Nevertheless, Mayer's research is based on cognitive processing in mainstream education that does not take into account the kind of language and culture difficulties experienced by EFL learners in their interpretation of modal information. Gilakjani et al. (2011) who claimed that visualizations could support learning by giving students an external representation of information, thus impelling deep cognitive processing and helping to maintain attention. Graphs and charts make information interesting and enticing and help simplify complex topics. These results are consistent with the focus of the current study on the potential of audio-visual stimuli to scaffold EFL learners' comprehension, specifically to lower cognitive load for verbal input processing. However, the research has not directly focused on non-verbal communication (gestures, visual symbolism), which is essential for EFL students to comprehend meaning in multimodal text. Kress (2010) further emphasized the importance of multimodal integration, arguing that it is now impossible to fully understand texts—even their linguistic components—without considering how other features, such as images and sounds, contribute to meaning. This reinforces the centrality of multimodal literacy in contemporary communication, particularly in digitally mediated EFL environments. This interconnectedness is amplified by computer technology, which seamlessly integrates text, audio, video, and images in meaning-making processes (Kress & van Leeuwen, 2006). Kress and van Leeuwen (2006) also introduced a visual design «grammar,» exploring how visual literacy can be integrated into education. This is particularly relevant to the present study's aim of developing visual methodologies for EFL instruction, as it offers tools to deconstruct visual rhetoric in learning materials. That said, their model primarily examines Western visual conventions, potentially marginalizing culturally diverse EFL learners who may interpret imagery differently.

A multimodal approach in the classroom emphasizes the strategic use of multiple modalities in authentic learning environments. Each modality serves as a resource for students' meaning construction, offering unique perspectives on phenomena that can challenge prior conceptions and provide tools for imagining and thinking (Kress, 2009). For instance, teachers often use gestures and voice to highlight images and other references during instruction, creating a dynamic interplay between modes (Kress, 2009). Kress (2010) also argued that gestures, drawings, voice, and physical objects interact in ways that enrich meaning construction. Each mode contributes uniquely: speech provides differentiation, blackboard images offer visual context, object manipulation creates a physical context, actions provide dynamic clarity, and textbook images serve as stable summaries. Repetition, synchronisation, similarity, and contrast further enhance cohesion. The selection of modes is purposeful, as each metaphorical journey is unique, and each mode builds meaning differently. Students must engage in distinct tasks to comprehend each mode, fostering deeper understanding. The dynamic interplay between modes (e.g., voice and gestures reinforcing images) aligns with the present study's focus on how synchronized multimodal input strengthens learners' ability to decode verbal and nonverbal cues in audio-visual materials.

Multimodal approaches utilize both verbal and nonverbal modes to represent content knowledge, thereby enhancing visual and sensory education. This approach not only promotes relevance and innovation but also improves course quality and diversifies academic programs (Maguire, 2005; Moreno & Mayer, 2007). By integrating multiple modes, educators can create richer, more engaging learning experiences that cater to diverse learner needs and preferences. This supports the argument that multimodal methodologies are crucial for fostering deeper comprehension and engagement in EFL contexts, as they leverage both auditory and visual stimuli to reinforce learning. However, these studies overlook the specific challenges EFL learners may face when processing multiple modes simultaneously, such as cognitive overload or cultural differences in interpreting nonverbal cues. Further research is needed to explore how different learner profiles (e.g., proficiency levels, cultural backgrounds) interact with multimodal input.

Multimodal Audio-Visual Comprehension

The progress in science and technology is transforming the landscape of second language (L2) listening teaching, which has evolved from the old-fashioned audio-only based instruction to modern multimodal audio-visual instruction. There's an essential emphasis on technology for both instruction and assessment. There is interactive content, real-time annotation and gesture-based interaction with Smart Boards, multitouches and data show tools (projectors/ visualizers) normalize visual oriented stimuli (multimodality). Taken together, these technologies offer a systematic and comprehensive model to investigate how multimodal approaches improve various communication skills among EFL students (Abdullah et al., 2020). This change has yielded a proliferation of academic literature on the effectiveness

of these techniques (Zhyrun, 2016; Namaziandost & Nasri, 2019; Arbab, 2020). Audio-visual input, with motion picture, sound, and text on a screen, appeals to vision and hearing, and has three metafunctions: image, writing and doing. Numerous studies have confirmed that information to both auditory and visual modalities is not only better received, but it also is better encoded in memory than that delivered to the auditory modality only (Surguladze et al., 2001; Campbell, 2008). Since listening comprehension is among other complex selection and information processing activities of the brain and higher cortical function is facilitated during the activation of several sensory channels. The onset of multimedia in classrooms has prompted an additional revolution in the teaching industry – the ability to make a simultaneous assessment of both auditive and visual information. These developments have now been supplemented by large-scale standardized high-stakes testing, such as TOEFL's iBT and China's CET-4 and CET-6 Internet-based exams (Wang at al., 2014). This supports the argument that multimodal input enhances cognitive processing by engaging multiple sensory pathways, which is critical for EFL learners who rely on both verbal and visual cues for comprehension. These studies do not account for longitudinal changes in multimodal comprehension. However, these studies overlook the specific role of nonverbal communication (e.g., gestures, facial expressions) in facilitating language acquisition, which is a key component of multimodal learning.

However, a need still persists for empirical research that focuses on the effects that technological tools and computer-based or digital materials of/the visual and auditory input combined resources may have on students' comprehension. Yet, many studies have also emphasized the role of both linguistic and extra-linguistic information in processing. For example, Sueyoshi and Hardison (2005) highlighted the importance of context-specific cues (e.g., facial expressions and gestures) for comprehension. This supports the argument that nonverbal elements in multimodal input are critical for comprehension, aligning with the present study's focus on how visual methodologies augment EFL learners' interpretive skills. However, this study overlooks the specific technological affordances (e.g., interactive digital platforms) that could further optimize these nonverbal cues for language learning.

Similarly, Ramírez and Alonso (2007) demonstrated that Spanish children achieved a deeper understanding of foreign language structures through engagement with English digital stories. Their findings also revealed that such audio-visual materials not only improved listening skills but also fostered other competencies, such as communication, as students were able to provide feedback in the target language after viewing the videos. These findings align with the present study's focus on multimodal input enhancing both comprehension and verbal communication. However, the study does not explore whether certain types of visual stimuli (e.g., animations vs. live-action videos) yield differential effects, a gap this research could address. Wagner (2010) further corroborated these findings through a comparative study, showing that an experimental group exposed to audio-visual input outperformed a control group that relied solely on auditory input in comprehension tests. This highlights the need for multimodal approaches in EFL Learning, notably for its role in retention. However, the study confined attention to comprehension scores and does not include the impact on nonverbal communication (e.g., gesture understanding) which is a central aspect of this study. Taken together, these studies highlight the multi-dimensional contributions of multimodal input to second language learning, and argue that complementary visual and auditory information provides a substantial boost to both understanding and remembering.

Multimodal Approach to Verbal and Nonverbal Communication

Verbal and visual communication are essential components of the learning process, each playing a distinct yet complementary role. In fact, visual communication enriches the active learning by involving learners in auditory and visual representations as well as promotes higher cognitive functions and application of learning in the real world (Khadimally, 2016). This is consistent with the claim that multimodal visual approaches can help to enhance EFL students' understanding of abstract concepts through the illustration of concrete visual materials. But what this research has failed to consider is the special difficulties which existence for EFL learner in dealing with reception of visual and verbal information at the same time. Because this is the foundation of Audio-visual comprehension. This can be clearly seen in the context of foreign language learning when images are used, such as in the 'Selfie Project', to prompt spoken or written language production (Victoria, 2021). These findings align with the present study's focus on how visual aids can enhance verbal production in EFL contexts, suggesting that multimodal tasks (e.g., digital storytelling or image-based reflections) may foster both linguistic and nonverbal expression. However, the limitations of verbal communication in diverse and complex contexts highlight the necessity of visual aids to improve understanding and engagement (Lennartsson, 2010). This reinforces the idea that EFL learners may rely heavily on visual scaffolding to decode meaning, particularly in linguistically demanding situations. Yet, the study does not explore how cultural differences in visual literacy might affect comprehension—a gap this research could address. Despite the predominance of verbal communication in education, the effective use of visual thinking is crucial for conveying messages and fostering critical thinking skills (Nuzzaci, 2019). This supports the argument that multimodal methodologies could empower EFL students to analyze and articulate ideas more critically. However, the study focuses on general education rather than language learners, leaving room for further investigation into how visual thinking strategies specifically benefit

EFL audio-visual processing. This interplay between verbal and nonverbal communication significantly influences the quality of learning experiences, underscoring the need for educators to develop proficiency in both forms (Wahyuni, 2018). These findings align with the present study's focus on multimodal pedagogy but do not examine the role of technology (e.g., videos, interactive graphics) in facilitating this interplay—an area this research could explore.

Moreover, having established the cognitive advantages of multimodal input, the following section examines its role in productive communication skills. Multimodal approaches to communication encompass both verbal and nonverbal elements, emphasizing their interdependence in conveying meaning. Research indicates that nonverbal cues, such as gestures, facial expressions, and prosody, play a crucial role in enhancing the understanding of verbal messages, particularly in contexts such as political discourse and media communication (Madella et al., 2023; Abduraximova, 2024; Harutyunyan, 2023). The evolution of multimodal studies has further expanded to include various media forms, highlighting the significance of visual components in enhancing the impact of verbal texts (Szawerna, 2023). This supports the argument that integrating visual and verbal modalities strengthens communicative effectiveness, which is crucial for EFL students developing audio-visual comprehension. However, this study overlooks the specific challenges EFL learners may face when processing multimodal input, such as cognitive overload or cultural differences in visual interpretation. These findings align with the present study's focus on how multimodal methodologies influence both verbal and nonverbal communication in EFL contexts. The neuroscientific perspective offered by Benetti et al. (2023) reinforces the idea that multimodal learning engages multiple cognitive processes, potentially enhancing retention and comprehension. For instance, studies show that learners who engage in multimodal communication experience fewer communication breakdowns and demonstrate improved language abilities (Bouchey et al., 2021).

In contrast to receptive modalities, productive and interactive modes such as gesture and prosody have received comparatively less attention. The multimodal approach is particularly relevant for students who are already accustomed to technology as part of daily life (Ngongo, 2022). This supports the argument that EFL learners, many of whom are digital natives, may benefit from instructional methods that align with their existing technological literacy, thereby facilitating engagement with multimodal texts. Multimodality refers to all verbal and nonverbal visual semiotic inputs that can be used to interpret dialogical associations in reading material (Herman et al., 2019). These findings align with the present study's focus on how visual and auditory semiotic resources interact to shape comprehension and communication in EFL contexts. As Baldry et al. (2020, p. 157) assert, "we live in a multimodal society," where individuals experience the world multimodality and construct meaning using words, visuals, gestures, sounds, and other resources. This perspective reinforces the idea that language learning should extend beyond traditional verbal instruction to incorporate diverse semiotic modes, which is central to this study's investigation of audio-visual comprehension. This perspective is supported by Bilfaqih and Qomarudin (2017), who argue that texts of various types are inherently multimodal, combining semiotic frameworks to create meaning in both standard and innovative ways. However, this study overlooks the specific challenges EFL learners may face when navigating multimodal texts, such as cognitive overload or cultural disparities in semiotic interpretation - a gap the current research seeks to address. Technological advancements further enhance text multimodality by simplifying the creation and dissemination of multimodal content, allowing learners to engage with diverse semiotic resources as a cohesive communication unit (Bezemer & Kress, 2016). This supports the argument that digital tools can scaffold EFL students' multimodal literacy, but it raises questions about whether current pedagogical practices adequately prepare learners to critically analyse and produce multimodal discourse, a key concern of this study.

In this environment, learners are placed in the role of "semiotic initiators and responders," not just attending to spoken language, but to producing and responding to, a range of texts and images as well as other multimodal resources (Coffin & Donohue, 2014). Frequently this creative mediation consent new readings or uses of the multimodal content that do not necessarily correspond to the original intentions of its design. Multimodal approaches in spoken and visual communication integrate linguistic (words), visual (gestures), and aural resources, demonstrating the interdependence of verbal and nonverbal communication. This is particularly evident in face-to-face communication, where gestures, facial expressions and prosody (intonation) contribute to the meaning and emotion of an utterance (Szawerna, 2023; Kenzhegaliev et al., 2023). Multimodal teaching, in the end, applies textual, visual, and auditory modes to enhance learning and equip students with powerful communication skills (Archvadze, 2023). While these studies highlight the benefits of multimodality, they do not sufficiently address potential challenges, such as cognitive overload in learners when processing multiple modes simultaneously or cultural differences in interpreting nonverbal cues. Additionally, there is limited discussion on how digital versus in-person multimodal interactions affect learning outcomes differently.

METHOD

Research Design

This study implemented a mixed-methods research design to answer its research questions. Proudfoot (2023) defines mixed methods research as the combination of both qual-

itative and quantitative threads as part of the same research program (although data analysis/ and or collection may be done separately and findings and inferences may be combined/ integrated through the integration of data. It seeks to establish an in-depth understanding of complex educational phenomena through the collection, analysis and synthesis of a range of data in relation to particular research questions (Creswell, 2024). The method can be particularly advantageous in educational research, as i t could contribute to the consistency of results and lead to a more informed viewpoint on topics, such as teaching practices and students' well-being (Fàbregues et al., 2024; Ercan et al., 2022). An explanatory sequential mixed methods design 'consists of first collecting quantitative data and then collecting qualitative data to help explain or elaborate on the quantitative results' (Creswell, 2017, p.542). The research was conducted according to an explanatory sequential research design, in which the findings obtained with quantitative data are examined in depth with qualitative methods and data. In this context, firstly, quantitative data were collected and analysed by using a repeated-measures ANOVA on the set of three test scores over time. Secondly, qualitative data collection processes and analysis were applied by using thematic analysis of the data collected by a focus group semi-structured-interview to enrich the findings obtained in the first step.

Participants

The study population of the quantitative method consists of randomly selected third-year EFL university students at Anbar University, College of Education for Humanities, and Tikrit University, College of Education for Humanities and College of Education for Women enrolled in the English departments, during the academic year 2023–2024. The total population consists of 373 students. In addition to this population, the researcher randomly selected and invited 214 EFL Iraqi students to participate in the study on the basis of their willingness and agreement. However, some students were excluded due to failure to comply with the study period (including withdrawal during the study), incomplete

responses, or non-compliance with test instructions, such as skipping sections or providing inconsistent responses. Only participants who fully adhered to the study protocols were included in the final analysis. The research's quantitative data method was determined by convenience sampling. Convenience sampling was selected to identify the nearest and easily accessible sample that the researcher can obtain. Convenience sampling was used due to logistical constraints (e.g., accessibility during the academic year). To mitigate bias, the sample was stratified by university (Anbar and Tikrit) and department (Humanities/Women), ensuring representation across institutional contexts. The researcher asked 214 students, who had the same educational and language background knowledge, to represent the sample by answering audio-visual comprehension, verbal communication and nonverbal communication tests.

The qualitative study sample consists of certain people and was determined through maximum variation sampling, one of the methods included in purposive sampling. Maximum diversity sampling aims to select a relatively small yet highly representative sample that captures the broadest possible range of perspectives within relevant population groups. This approach seeks to identify commonalities across diverse scenarios, thereby revealing the multifaceted nature of the issue under study (Yıldırım & Şimşek, 2006). In this study, maximum variation was achieved by incorporating participants with differing academic performance levels, age ranges, and degrees of digital literacy. Volunteering is crucial because the participants of the qualitative research will be drawn from the participants in the quantitative method (Creswell, 2017). For this reason, 20 students among the participants who supported the quantitative part were selected for the qualitative sample with the focus group semi-structured interview. Data were collected from three colleges to form the quantitative sample. In the qualitative study group, at least five students from each college were included to ensure maximum diversity. The gender breakdown of the interviewed students is 13 females and 7 males, with ages ranging from 19 to 22.

Table 1

The Demographic Data of the Study Participant of the Quantitative Dimension

Variable	Introductory Features	N	%
Sample	College of education for humanities- university of Tikrit	87	40.65%
	College of education for women- university of Tikrit	54	25.23%
	College of education for humanities- Anbar university	73	34.11%
Gender	Female	153	71.49%
	Male	61	28.50%
Age	19-21	178	83.17%
	22-24	36	16.21%

Instruments of Data Collection

Data collection was conducted using triangulation of mixed-mode design instruments, as outlined below.

Quantitative Research Instrument

The study employed three quantitative data collection instruments. The first, the Audio-Visual Comprehension Test, measured listening comprehension and inferential understanding through a modified version of Norte Fernández-Pacheco's (2018) test based on the British Council vodcast «English is Great» (https://goo.gl/xK4zYn). It included eight fill-in-the-blank items requiring exact words from the vodcast, six true/false questions on explicit facts, and one open-ended follow-up question analyzing speaker tone. The vodcast was selected for its authentic native-speaker dialogue, cultural neutrality, and alignment with B1 CEFR proficiency, with piloting (n=20) confirming appropriate difficulty after minor lexical adjustments. The second, the verbal communication test, evaluated written production via four tasks: five prompt-based dialogue continuations, one role-play scenario, one dialogue-writing task using rejoinders, and one structured topic-development exercise (renamed from «mental visualisation» for clarity). The third, the non-verbal communication test, assessed the identification and interpretation of gestures, facial expressions, and proxemics through 12 matching items (pairing images to nonverbal types), 10 multiple-choice questions (interpreting body language), and three short-answer analyses of emotional tone. All tests were refined for formal tone—replacing colloquialisms, adding exemplar responses, and standardizing scoring—to ensure replicability and alignment with language assessment best practices.

The second and third tests were collected and guided by the researcher. The repeated test design is used to test the effect, as indicated by Wiklund-Hörnqvist et al. (2014). It improves the ability to recall and learning outcomes at a significantly higher level. This method involves administering tests after particular periods and is more efficient than actions such as repeating.

To establish the face validity and content validity of the audio-visual, verbal and nonverbal communication tests, the researcher evaluated them with the help of a panel of 10 experts in the field of applied linguistics and methods of teaching. The experts' modifications and suggestions were taken into consideration in verifying the appropriateness and the final structure of the tests. Finally, the three tests were piloted to ensure the construct validity, language clarity, accuracy, practicality and reliability. According to Oluwatayo (2012), construct validity refers to whether the operational definition of a variable actually reflects the theoretical meanings of a concept. In other words, construct validity shows the degree to which inferences are legitimately made from the operationalisations in one's study to the theoretical constructs on which those operationalisations are based. Therefore, construct validity of all the instruments is established through item analysis such as item discrimination power, item difficulty level, the correlation coefficient between the item score and the total score of the test, the correlation coefficient between the item score and the component to which the item belongs, and internal correlation matrices. The results indicate that the audio-visual test items exhibited an appropriate level of difficulty, ranging between 0.41 and 0.67, while the discrimination power for all items fell within 0.31 to 0.53. For verbal communication, the difficulty level ranged from 0.42 to 0.58, with a discrimination power between 0.35 and 0.55. Similarly, nonverbal communication items demonstrated a difficulty level of 0.33 to 0.46 and a higher discrimination power of 0.43 to 0.63, which aligns with established benchmarks (Tang & Logonnathan, 2016).

Accordingly, the point-biserial correlation coefficient formula is used to calculate the correlation between the total score of the test, the binary score (intermittent) is used for the subject items and the Pearson correlation coefficient is used to find out the correlation between the items' scores and the total score for the 214 participants. Statistical analysis of audio-visual, nonverbal and verbal communication of the test's items reveals that all the correlation coefficient values are more significant than the critical value, which is 0.195 at 212 degrees of freedom and 0.05 significance level.

According to Franzen (2013, p. 15), one of the basic characteristics of a good instrument is reliability, which refers to the consistency or stability of scores values that an instrument elicits. The measurement of test–retest means that if the same respondents complete a test at two different points in time, then the responses should be stable and the set of results should be reproducible. With the use of the Pearson correlation coefficient to estimate reliability, or r-value, of responses, the r-value for the audio-visual comprehension scale is 0.91, that for the verbal communication scale is 0.87 and that for the nonverbal communication scale is 0.89, thus being indicators of good reliability because the values are higher than 0.70. Internal consistency of measure items is another way to test the reliability of this study. Cronbach's alpha is extensively used to determine a test's internal consistency (Franzen, 2013). Most internal consistency testing approaches treat each item as a separate measurement and the test as multiple measurements. The study's three measures (audio-visual comprehension scale, verbal communication scale and nonverbal communication scale) demonstrate high internal consistency, with values of 0.88, 0.92 and 0.85, respectively (Ravid, 2024). The high Cronbach's alpha (α) and test-retest (r) values (all above 0.85, except for one α at 0.85 and one r at 0.87) indicate strong reliability, while the difficulty (p) and discrimination (D) indices suggest appropriate item variability and effectiveness in distinguishing between high and low performers, confirming robust instrument quality.

Table 2

Psychometrics Properties of the Study Instruments

Test Name	Cronbach's Alpha (α)	Test–Retest (r)	Difficulty Index (p)	Discrimination Index (D)
Audio-visual Comprehension	0.88	0.91	0.41-0.67	0.31-0.53
Verbal communication	0.92	0.87	0.42-0.58	0.35-0.55
Nonverbal Communication	0.85	0.89	0.33-0.46	0.43-0.63

Qualitative Research Instrument

To triangulate quantitative findings, the researcher conducted semi-structured focus group interviews with 20 EFL university students divided into four groups (five participants each), each lasting 60–75 minutes, to explore the influence of multimodal visual methodologies on audio-visual comprehension, verbal communication, and nonverbal communication. Prior to implementation, a jury of experts evaluated the interview prompts for validity and relevance, leading to minor refinements, and a pilot test with five students confirmed the clarity and suitability of the questions. Core discussion prompts included questions such as, «How do multimodal materials influence your understanding of complex English-language content?» and «In what ways do these methodologies affect your confidence in verbal communication?», with flexibility for follow-up probes to explore emerging themes. Sessions were audio-recorded (with consent) and transcribed, ensuring rigorous qualitative data collection that complemented the study's quantitative approach while providing nuanced insights into learners' experiences with multimodal tools.

Procedures of Data Collection

The study extended for 60 days and included quantitative and qualitative data collection. For the quantitative data, the study passed through phrases such as 'before' (first test), 'during' (14 days after first test follow-up with a 1-hour lecture or more to provide feedback on the three tests) and 'after' (final tests after 14 days). The study employed various multimodal technological devices, including smartboards, to facilitate learning during the designated period. In the audio-visual comprehension assessment, a vodcast was utilised to present material for answering test questions. The verbal and nonverbal communication evaluations were conducted using visual imagery, which was effectively managed and articulated through PowerPoint presentations. The three important phases of quantitative data are as follows:

- In the 'before phase, the participants were required to complete three assessments (e.g. audio-visual comprehension, verbal and nonverbal communication) designed to evaluate their strengths and weaknesses across the three variables under investigation, utilising a smartboard to present the study questions.

- In the 'during' phase, the students received a lesson lasting one hour or more prior to the assessments. This instruction served as feedback regarding the variables being examined in the tests, allowing students to familiarise themselves with the material. The assessments were administered on consecutive days.

- In the 'after' phase, the stability of the information related to the study variables was assessed. This phase aimed to evaluate retention and both short-term and long-term memory over a specified duration.

The study rigorously controlled key variables to ensure consistency across all phases. Timing was standardized, with fixed intervals between the «before» (baseline), «during» (14-day follow-up with feedback), and «after» (final assessment) phases, and assessments were administered on consecutive days during the intervention period. Materials and technology were kept uniform, with smartboards used for presenting test questions, vodcasts for audio-visual comprehension tasks, and PowerPoint for verbal and nonverbal communication evaluations. Learning conditions were maintained through structured, multimodal instruction, including a minimum one-hour feedback lecture in the «during» phase, ensuring all participants received identical content and technological support. These controls minimized external variability, enhancing the reliability of the quantitative data collected.

As for qualitative data, the focus group semi-structured interview is used to gather information after the three phases of quantitative data. The interviews were conducted to enrich the findings of the study and explore the factors that affected English learning (e.g. psychological, social, increased student engagement and pedagogical factors). Notes were taken on nonverbal cues and group dynamics during the focus groups, as these can provide additional context to verbal responses.

Data Analysis Procedures

A multimodal transcription of the 'English is Great' vodcasts was conducted on the basis of the audio-visual comprehension test to illustrate the many communicative modes employed during the vodcasts. The analysis was conducted using the EUDICO Linguistic Annotator (ELAN) by which EUDICO means European Distributed Corpora Project. ELAN is sophisticated annotation software which can be used for annotating audio and video material. Produced at the Max Planck Institute for Psycholinguistics in Nijmegen, Netherlands, ELAN is an essential tool for linguists and other researchers who use video or audio recordings (Cheng, 2024). In this study, ELAN was utilized solely for instructional design purposes. The variety of modes present in the vod-

Figure 1

Procedures of the Study
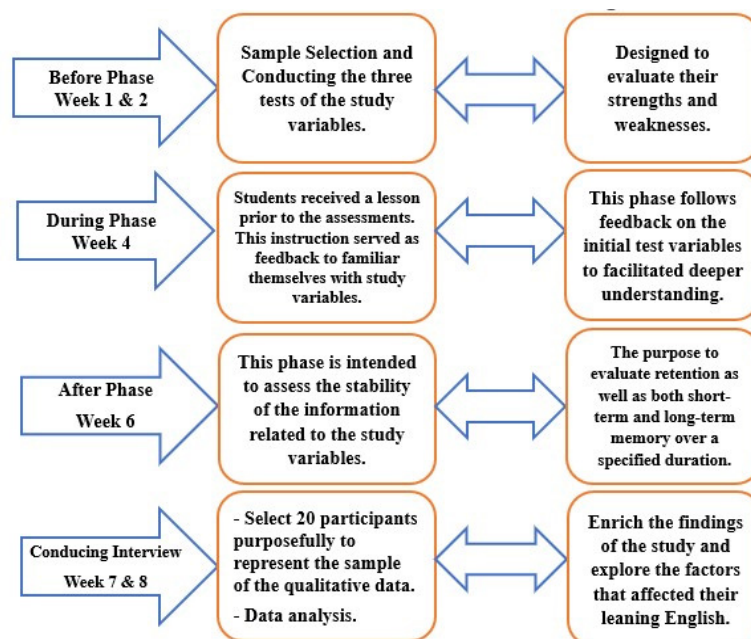
casts — such as participants' gestures, speech, written text, images and music — were categorised into different tiers. This classification allowed for the simultaneous display of precise annotations on the same screen. The teacher notified students that they will view a vodcast on the English language and thereafter respond to questions. To assuage any apprehension, the instructor told them that this was not an examination, but simply a routine audio-visual exercise. Subsequently, a comprehension assessment and a blank sheet of paper were administered to the students. They were allotted 90 seconds to review the test and anticipate the content they would be viewing. Following this reading session, students were directed to turn the test over to deter them from repeatedly referencing both the paper and the screen. While note-taking during the vodcast was not mandatory, students had the choice to engage in it if they desired. Following the elucidation of the directions, the class viewed 'English is Great' once. They were subsequently allotted three minutes to respond to the questions according to their comprehension. After three minutes, the students flipped their tests and rewatched the vodcast, adhering to the same protocol as previously established. Following the second viewing, they were allotted an additional three minutes to complete the test.

The images utilised in the verbal and nonverbal communication assessments were sourced from *Conversation Strategies: Pair and Group Activities for Developing Communicative Competence* by Kehe and Kehe (1994), as well as *Discussion Strategies: Beyond Everyday Conversation* by Kehe et al. (1998). The nonverbal communication referenced were derived from *100+ Body Language Tips* by Rosas (2010). As for the

data collected by audio-visual comprehension, verbal and nonverbal communication tests, quantitative data analysis procedures were followed through numerical statistical analysis by using Statistical Package for the Social Sciences version 26. Consequently, repeated-measures ANOVA was conducted on the set of three test scores over time. The Pearson correlation coefficient was employed to assess the relationship between students' performance across each test phase, enabling the calculation of effect sizes for each variable.

In the analysis of qualitative data, semi-structured focus group interviews were subjected to qualitative analysis using Creswell's (2012) thematic analysis steps. These steps are as follows: preparing and organising the data; exploring and coding the data; describing findings and forming themes; representing and reporting findings; interpreting the meaning of the findings; and ensuring the credibility and trustworthiness of the findings (p. 238). Denzin and Lincoln (2018) advised the use of two steps to verify the qualitative findings' objectivity and authenticity. Firstly, an external auditor conducts triangulation and reviews, ensuring the study's accuracy and compatibility by using multiple data gathering sources and instruments. Secondly, the external auditor, who is an experienced applied linguistics specialist, was asked to critically and deliberately review the initial draft of the qualitative analysis of the findings in all its key themes and subthemes. The external auditor then confirmed that the analytical themes were scientifically sound. These two steps ensured that the researchers' analysis was non-biased and scientifically compelling.

RESULTS

Quantitative Results

Results Related to Audio-Visual Comprehension Test

In order to find out *'the influence multimodal visual methodologies on audio-visual comprehension among EFL university students.'* The results represent sample participants (N = 214) from three colleges of education. The mean score and standard deviation were calculated to assess students' performance over time in the 'before', 'during' and 'after' phases. The mean scores of these tests in the three phases are 46.22, 56.58 and 67.59, respectively, as shown in Table 3.

To examine the variations in students' performance on the audio-visual comprehension test over time, repeated-measures ANOVA was performed to assess the statistical significance of the changes among the three tests over time.

Mauchly's test of sphericity indicated that the assumption of sphericity was not violated (P = 0.213). The partial eta squared (partial η2) was calculated to assess the variance in comprehension scores attributable to the three tests over time, as illustrated in Table 4.

According to Table 4, repeated-measures one-way ANOVA indicated a significant difference in the number of the three tests according to the 'before', 'during' and 'after' phases, $F_{(2, 426)}$ = 244.656, (p < .001). The obtained partial η2 was 0.535. The percentage of variance in the dependent variable is substantial and positively influenced, indicating that students achieve improved audio-visual comprehension results when multimodal visual images are utilised. In an educational context, a «significant effect» typically refers to a meaningful or important impact of an intervention, program, teaching method, or policy on student learning outcomes, engagement, or other educational measures. The analysis involved conducting three post hoc t-tests. This phase of the analysis aimed to assess the differences among

Table 3

Descriptive Statistics Table According to Students' Audio-Visual Comprehension Test Performance Over Time

AVC Phases	Mean	Std. Deviation
Before	46.22	10.65
During	56.58	9.475
After	67.59	9.90

Figure 2

Phases of Student Performance in Audio-Visual Comprehension Tests Over Time
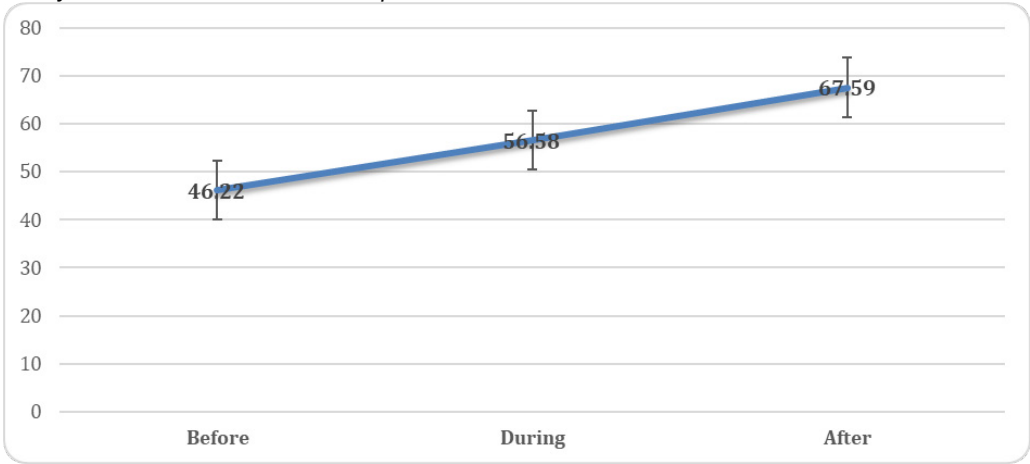
Table 4

Within-Subjects Effects Tests (Sphericity Assumed)

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Audio-visual	Sphericity Assumed	48875.629	2	24437.815	244.656	.000	.535
Error Audio-visual	Sphericity Assumed	42551.704	426	99.887			

the three pairs of conditions: before–during, before–after and during–after. Post hoc comparisons between conditions were conducted using three paired sample t-tests. A paired sample analysis revealed a significant difference between the before–during conditions, t (213) = −11.018, $p < .001$. A second paired sample t-test revealed a significant difference between the before–after measurements, t (213) = −20.899, $p < .001$. The third paired sample t-test revealed a significant difference between the during and after conditions, t (213) = −11.796, $p < .001$. All pairs exhibited a statistically significant difference at the $p < 0.05$ level. The reported effect size of $d = 1.072$ indicates a large practical difference between the compared conditions (before–during, before–after, and during–after). For example, if the mean score in the before condition was 50 with a standard deviation of 10, the during condition would average around 60.72, showing a substantial increase of ~10.72 points. Similarly, the after condition would be even higher if the effect is cumulative. This means that not only were the differences statistically significant ($p < .001$), but they also represented meaningful real-world changes, as an effect size above 0.8 is generally considered large. Thus, the findings demonstrate strong, practically important effects across all comparisons. These results indicate a clear progression in students' performance over time, with mean scores increasing significantly from «before» phase to the «during» and «after» phases. While no significant difference is found between the during and after phases. The lack of significant difference between the during and after phase may suggest that while students continued to benefit from

these multimodalities, the rate of improvement plateaued after initial exposure.

Results Related to Verbal Communication Test

Regarding *the influence of multimodal visual methodologies on EFL students' verbal communication among the three rounds of test.* Initially, the average score and standard deviation reveal the progression of students' performance over time using the terms 'before,' 'during' and 'after'. Table 5 shows that the respective mean scores before, during and after the three tests were 36.97, 53.95 and 56.14.

Repeated-measures ANOVA was performed on the three test results to assess the statistical significance of the differences observed over time. Mauchly's test of sphericity also revealed a violation of the sphericity assumption, with a P-value of 0.07. Huynh-Feldt epsilon[b] of 0.963 was used to correct an increase in the type I error rate. As shown in Table 6, partial $\eta2$ was used to figure out how much of the difference in verbal communication test scores can be explained by the three tests over time.

The repeated-measures one-way ANOVA F-value (1.926, 410.292) = 271.990; $p < .001$) demonstrated a notable disparity in the total number of tests related to the terms 'before', 'during' and 'after'. The partial $\eta2$ value was 0.561. The percentage of variation in the dependent variable is substantial and positive, suggesting that the multimodal visual image enhanced students' verbal communication. Analysis indicat-

Table 5

Descriptive Statistics Table According to Students' Verbal Communication Test Performance Over Time

VC Phases	Mean	Std. Deviation
Before	36.97	10.39
During	53.95	10.16
After	56.14	12.28

Figure 3

Phases of Student Performance in Verbal Communication Tests Over Time
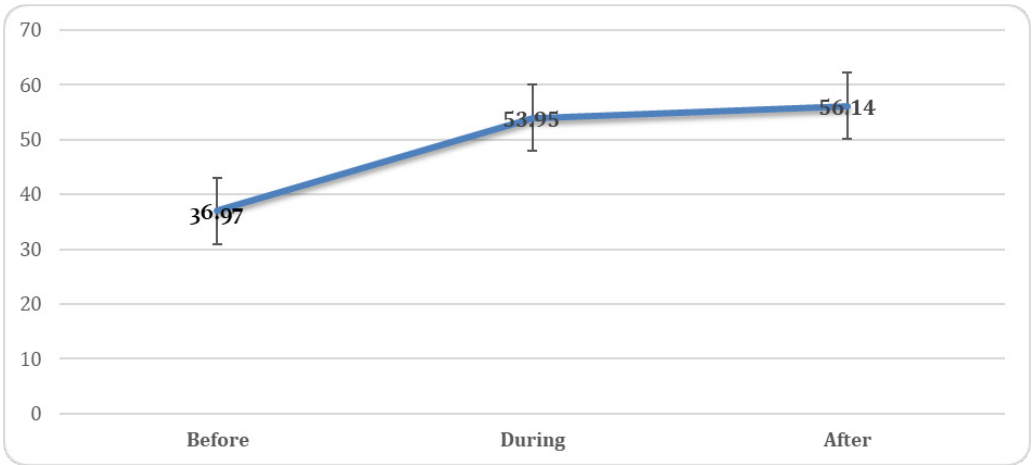
ed that each condition exhibited significant variability. The study utilised three post-hoc t-tests and analysed the differences among before–during, before–after and during–after conditions. Three paired sample t-tests were utilised for post hoc condition comparisons. A preliminary paired sample analysis indicated a significant difference between the before-during conditions, t (213) = −16.949, $p < .001$. A second paired sample t-test indicated a significant difference between the before-after conditions, t (213) = −20.982, $p < .001$. The third paired sample t-test indicated a significant difference between the during-after conditions, t (213) = −2.354, $p < .001$. All pairs exhibited significant differences ($p < 0.05$). A significant effect size $d = 0.707$ supported these results. Cohen's d is a measure of effect size that indicates how many standard deviations apart the means of two paired conditions are. A d of 0.707 falls between a medium and large effect, meaning the differences between conditions are practically meaningful, not just statistically significant. The results show that turn-taking dominates, suggesting that effective communication heavily depends on participants' ability to manage conversational turns, potentially enhancing engagement and understanding.

Results Related to Nonverbal Communication Test

In order to find out *'How do multimodal visual methodologies influence EFL students' nonverbal communication and what nonverbal cues that are strengthened by utilising multimodal approaches?'*, was addressed. The mean scores and standard deviations reveal the progression of students' performance over time using the terms 'before', 'during' and 'after'. Figure 4 shows that the respective mean scores before, during and after the three tests were 37.75, 54.37 and 62.24, as illustrated in Table 7.

Students' performance in the nonverbal communication test varied over time. Repeated-measures ANOVA was conducted on the three tests to assess the statistical signifi-

Table 6

Tests of Within-Subjects Effects (Huynh-Feldt)

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Verbal communication	Huynh-Feldt	66437.171	1.926	34490.333	271.990	.000	.561
Error Verbal communication	Huynh-Feldt	52028.162	410.292	126.808			

Table 7

Descriptive Statistics Table According to Students' Nonverbal Communication Test Performance Over Time

NVC Phases	Mean	Std. Deviation
Before	37.74	12.19027
During	54.37	10.15926
After	62.24	11.12358

Figure 4

Phases of Student Performance in Nonverbal Communication Tests Over Time
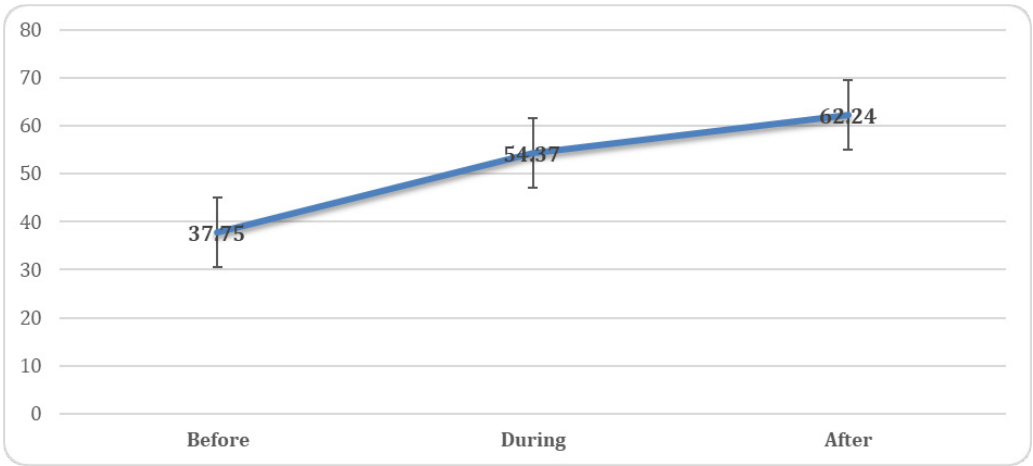
cance of the differences observed over time. Mauchly's test of sphericity revealed a violation of the sphericity assumption, with a P-value of 0.07. Huynh-Feldt epsilon[b] of 0.965 was used to correct an increase in the type I error rate. Quintana and Maxwell (1994) recommended using ε if ε is greater than 0.75. The partial η^2 was employed to determine the extent to which the variance in nonverbal communication test scores can be attributed to the three tests conducted over time, as illustrated in Table 8.

Repeated-measures one-way ANOVA F (1.931, 411.202) = 270.521; p <.001) showed a significant difference in the number of three tests based on the phrases 'before', 'during' and 'after'. The partial η^2 was 0.559. The percentage of variation in the dependent variable is significantly high and positive, suggesting that the multimodal visual image enhanced students' nonverbal communication. The analysis comprised three post hoc t-tests. This phase of the investigation analysed the differences among the three condition pairings: before–during, before–after, and during–after. Post hoc comparisons between conditions were conducted using three paired sample t-tests. A preliminary paired sample analysis revealed a significant difference between the before and during conditions, t (213) = −15.523, p < .001. A second paired sample t-test revealed a significant difference between the before and after intervention, t (213) = −20.901, p < .001. The third paired sample t-test revealed a significant difference between the during and after periods, t (213) = −8.085, p < .001. All pairs exhibited a statistically significant difference at the p < 0.05 level. The quoted results indicate that all pairwise comparisons (before vs. during, before vs. after, and during vs. after) showed statistically significant differences with very large effect sizes (Cohen's d = 1.00). The results indicate that there are significant differences in the performance of EFL university students across various types of nonverbal communication. Specifically: Facial Expressions are the most effective type of nonverbal communication, significantly outperforming all other types. Kinesics are more effective than kinesthetics, eye contact, proxemics, and artifacts but less effective than facial expressions. Kinesthetics, eye contact, and proxemics show similar effectiveness and are significantly better than artifacts. These findings indicate that facial expressions play a crucial role in nonverbal communication, followed by other types in the context of the performance test.

Qualitative Results

The qualitative results from the interview aligned with the previously mentioned quantitative ones, suggesting that the students experienced a high level of understanding. Thus, the integration of the multimodal visual images and the repeated testing should enrich students' audio-visual comprehension, verbal communication and nonverbal communication meaningfully. To explain further, the analysis of the interview extracts according to the four factors — psychological, social, increased engagement and pedagogical factors — revealed that the students expressed their memory encoding, as indicated in Table 9.

Students reported that the incorporation of audio-visual tools significantly influences listening comprehension and is crucial in contemporary language learning, based on four factors. Also, multimedia materials create engaging learning experiences that promote understanding by stimulating multiple senses, especially when included in instructional activities, and have a major beneficial impact on the growth of language proficiency in general and listening comprehension in particular. Students qualitatively emphasized that multimedia materials significantly improve learning by engaging multiple senses and creating immersive experiences, which aligns with the quantitative data showing a clear progression in comprehension scores from the before to the during and after phases. Together, these results demonstrate that multimedia tools play a crucial role in boosting comprehension, with their strongest impact occurring upon initial integration, followed by maintained—though not further increasing—advantages.

The students' responses to the interview questions show that visual representations greatly boost self-esteem and confidence when expressing their ideas. They enhance understanding and retention, encourage social connections in virtual spaces, and make narratives more captivating and relevant. Visuals that evoke emotions enhance the bond with the audience, resulting in greater engagement. Moreover, visuals enhance verbal communication and offer valuable feedback, thus improving verbal skills, as shown by the beneficial effects of educational resources that include visuals. Quantitatively, the prevalence of turn-taking suggests that managing conversational flow is key to effective dialogue, further supporting engagement and comprehen-

Table 8

Tests of Within-Subjects Effects (Huynh-Feldt)

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Nonverbal communication	Huynh-Feldt	66931.523	1.931	34670.063	270.521	.000	.559
Error Nonverbal communication	Huynh-Feldt	52699.810	411.202	128.160			

Table 9

Interview Extracts Indicating Students' Reference to "Audio-Visual Comprehension"

No.	Factors		Extracts
Audio-visual comprehension	Psychological factors	S01	"Ummm, I prefer visual images that make me more enjoyable/interesting/fun/exciting."
		S04	"I like Linking spoken words with visual cues supports vocabulary acquisition and is especially beneficial for language learners."
	Social factors	S05	"I feel comfortable when technologies employed such audio-visual that provide opportunities for realistic language learning experiences."
		S11	"I believe that audiovisual tools serve as catalysts for cross-cultural exchange and international cooperation."
	Increase students engagement	S18	"I feel more energetic when I watch the video."
		S08	"I enjoy how multimedia resources like audio-visual stimulate my senses and provide interesting learning opportunities that advance my comprehension."
		S03	"Compared to the audio track, I think I'm paying closer attention and being more focused."
	Pedagogical factors	S13	"In my opinion, using audiovisuals instead of only traditional audio helps the audience better understand the situation at hand."
		S02	"Visual materials contribute to an inclusive environment by accommodating different learning style."

Table 10

Interview Extracts Indicating Students' Reference to "Verbal communication".

No.	Factors		Extracts
Verbal communication	*Psychological factors*	S19	«Excellent, I favor visual representations that enhance my self-esteem and empower me to confidently share my thoughts and opinions.»
		S16	"I feel more comfortable talking about something when I use visual imagery, which improves my comprehension and memory."
	Social factors	S11	"I found that Visual images fosters verbal social interactions within online communities."
		S03	"Amazing, Visual elements facilitate my verbal storytelling and cultures norms by making the narrative more compelling and relatable."
	Increase students engagement	S20	"I like the way of Visual images that drive my willingness to engage in conversation."
		S06	"Wonderful, I feel that emotional stimuli, such as compelling images or videos, can create a strong connection with the audience, increasing the engagement through likes, shares, and comments."
	Pedagogical factors	S14	"I love educational textbooks' that employ visuals specially the book real speaking and listening improved my verbal skills."

sion. Together, these findings indicate that combining visual elements with structured interaction techniques creates a more dynamic and effective learning environment, where emotional resonance, cognitive clarity, and collaborative communication synergistically improve outcomes. This alignment underscores the value of multimodal approaches in education, leveraging both visual and conversational strategies to maximize participation and learning.

The qualitative results from the student interviews emphasized that the interplay of psychological, social, engagement, and pedagogical factors in nonverbal communication is vital for effective teaching and learning, enabling educators to refine their methods and enhance student performance. These findings align with the quantitative results, which revealed significant differences in EFL university students' performance across various types of nonverbal communication, with facial expressions emerging as the most effective, significantly outperforming other types. Kinesics ranked next in effectiveness, surpassing kinesthetics, eye contact, proxemics, and artifacts but remaining less impactful than facial expressions, while kinesthetics, eye contact, and proxemics showed similar effectiveness and were all significantly better than artifacts. Together, these results highlight the crucial role of facial expressions in nonverbal communication, followed by other types, further supporting

Table 11

Interview Extracts Indicating Students' Reference to "Nonverbal communication".

No.	Factors	Extracts
Nonverbal communication	Psychological factors	S06 "Visual images provided me with the opportunity to interpret nonverbal signals, such as the emotional tone of a lesson or the seriousness of a topic, which can impact my comprehension and retention of information."
		S12 "I enjoyed utilising visual images for learning nonverbal communication, which necessitates self-awareness and emotional intelligence, enabling me to comprehend signs such as gestures and facial expressions."
	Social factors	S14 "In my opinion, images help us comprehend how cultural norms influence nonverbal communication. Because Different cultures may perceive gestures, eye contact, and personal space in different ways."
		S10 "I enjoyed how positive nonverbal behaviours, such as maintaining eye contact or utilising open body language, may foster trust and better connections, which are necessary for effective teaching."
	Increase students engagement	S17 "I prefer to convey myself visually rather than verbally. For example, using thumbs-up or emoji reactions in online lectures enables me to swiftly express and grasp levels or emotional states."
		S20 "I like the way that visuals help me understand body language movements—such as nodding or fidgeting—and help me adjust my teaching strategies in real time to better meet my needs."
	Pedagogical factors	S09 "It exposes us to the effective use of visual resources, improves instructional clarity, and simplifies the delivery of complicated ideas."
		S07 "It was an excellent opportunity for me to learn nonverbal communication through the use of visual representations, such as gesturing to indicate transitions or utilising facial expressions to convey expectations."

the importance of integrating these components into teaching strategies to optimize student outcomes.

DISCUSSION

The results of this study demonstrate that multimodal visual methodologies enhance EFL learners' audio-visual comprehension, verbal fluency, and nonverbal communication. Students exposed to subtitled videos, contextual images, and gesture-based instruction consistently outperformed their peers in comprehension tasks. They also demonstrated a richer use of facial expressions and body language, confirming the integrative effect of multimodal input. These findings are consistent with Fay et al. (2013), who showed that gestures often precede or substitute for verbal expression and function as primary tools for conveying meaning. The positive outcomes extended across different learning preferences and were most evident when visual and verbal channels were activated simultaneously. Participants reported increased motivation, engagement, and communicative confidence, especially in tasks involving repeated exposure and varied media formats. While the study confirmed promising effects in verbal, written, and embodied modalities, some ar-

eas, such as vodcast-based movement tasks, would benefit from more rigorous statistical examination to clarify the role of non-linguistic cues in shaping performance.

These findings align with prior research on the benefits of multimodal instruction in language learning. Bairstow and Lavaur (2012) and Lin (2016) highlighted the contribution of L2 subtitles to listening comprehension, while Mohsen (2016) and Pardo-Ballester (2016) emphasized the role of audio-visual input in activating top-down strategies. Batty (2015), Lesnov (2017), and Hsieh (2020) further noted the importance of authentic multimodal contexts for receptive processing. Although Lin (2016) reported variable outcomes for L1 and L2 subtitles, her study lacked longitudinal scope. The present research addresses this gap by showing that sustained exposure to multimodal cues fosters improvement not only in comprehension but also in expressive and interactive skills over time.

Theoretical Integration

The observed learning gains can be understood through several theoretical frameworks. Cognitive Load Theory explains how distributing input across visual and auditory channels

reduces working memory overload. Working Memory Theory supports this view by showing that dual coding enhances retention and recall. Sociocultural theory adds another layer by framing visuals and gestures as scaffolding devices that support internalization of complex linguistic forms. The influence of learner-specific variables such as proficiency, anxiety, learning style, and cultural background also emerged as relevant. For example, subtitled videos were particularly effective for low-proficiency learners, and students from high-context cultures relied more on nonverbal cues. The effectiveness of instruction was moderated by media design quality, where clear and interactive resources supported engagement while poorly designed materials impeded learning. These findings reinforce Mayer's Cognitive Theory of Multimedia Learning (2005, 2009), which emphasizes dual-channel input, limited capacity, and active learning in second language acquisition.

Pedagogical Implications

The study suggests that multimodal instruction should be adapted to learner level and context. Subtitled videos in L1 may support beginners, while authentic L2 input can challenge advanced learners. Explicit instruction in nonverbal communication should be integrated into lesson planning to develop expressive fluency. Tasks such as vodcasts, storytelling, and gesture-based activities offer opportunities for multimodal practice. Repeated testing also contributed to vocabulary acquisition, consistent with Metsämuuronen (2013) and Vojdanoska et al. (2009). The emotional and social benefits of multimodal tasks were particularly visible in group activities and gamified formats. These observations align with Zhang and Zou (2022), Salamanti et al. (2023), Kress and Selander (2021), and Mayer (2021), who emphasized the motivational and self-efficacy effects of learner-centered multimodal design.

Limitations

Although the study revealed consistent patterns, several limitations should be acknowledged. The participant pool was limited to three Iraqi universities, which constrains generalizability. Individual cognitive factors such as spatial ability and working memory were not systematically measured. While facial expressions and body movements emerged as dominant nonverbal factors, artifacts such as props and visual symbols were rated as less effective, despite research by Bambaeeroo and Shokrpour (2017), which demonstrated that teacher appearance and use of physical space can significantly influence learner attention and academic engagement. This discrepancy may indicate that such static cues operate contextually and require more nuanced measurement. Furthermore, the finding that turn-taking, rather than visuals, was the strongest predictor of communicative success introduces a thematic ambiguity. It remains unclear whether visual support facilitates turn-taking or whether the two operate independently. Doumont (2002) noted that visuals convey intuitive meaning, but their interaction with

social communication structures may require more nuanced investigation.

CONCLUSION

This study provides empirical evidence that multimodal visual methodologies significantly enhance audio-visual comprehension, as well as verbal and nonverbal communication in EFL contexts. The integration of quantitative and qualitative data demonstrates that multimodal input supports language learning through a combination of cognitive, emotional, and social mechanisms. Statistically significant gains across multiple testing phases suggest that multimodal tools strengthen the connection between linguistic and sensory processing, contributing to improved retention and communicative performance. Students' interview responses further confirm the affective value of multimodal instruction, emphasizing reduced anxiety, increased motivation, and greater engagement.

The study contributes to the field by offering a triangulated analysis that moves beyond additive models of learning. Multimodal input was shown to not only supplement traditional instruction but to transform the way learners internalize and apply language. These results align with dual-coding and sociocultural theories, illustrating how visual and nonverbal cues facilitate both comprehension and interactional competence. Importantly, the study highlights the relevance of turn-taking, gestural expression, and contextual cues in bridging receptive and productive skills.

Future studies should focus on isolating the specific mechanisms behind multimodal learning, including cognitive load distribution, emotional engagement, and interactive structure. Experimental designs with randomized control groups would help clarify causal relationships, while cross-cultural comparisons could shed light on the influence of sociolinguistic context. The role of multimodal instruction in virtual and AI-enhanced learning environments also warrants further attention. Adaptive platforms that adjust visual input to learner profiles may offer promising avenues for increasing the efficiency and personalization of EFL instruction.

AI DISCLOSURE STATEMENT

The authors confirm that AI-assisted tools, including deepseek and ChatGPT, were utilized to enhance the manuscript's language clarity and coherence. These tools were employed primarily for linguistic refinement and to assist in addressing reviewer feedback - particularly in aligning critical theory with the study's scope. However, all conceptual arguments, analytical frameworks, and interpretative insights were independently developed by the authors and rigorously reviewed to uphold scholarly originality and academic integrity.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Ibrahim Hassan Ali: Conceptualization; data curation; formal analysis; funding acquisition; inquiry; methodology; administration of the project; resources; writing - original draft; review and editing.

Istabraq Tariq Jawaad Alazzawi: Supervision; investigation; re-sources; validation; review and editing (literature review and discussion sections).

REFERENCES

Abdikarimova, M., Tashieva, N., & Abdullaeva, Z. (2021). Developing students' verbal communication skills and speech etiquette in English language teaching. *Open Journal of Modern Linguistics*, *11*(1), 83-89. http://dx.doi.org/10.4236/ojml.2021.111007

AbdulGhafoor, C. & Challob, A.I. (2021). The Influence of Using Online Discussion on Developing EFL Students' Grammar Knowledge. *Korean Journal of English Language and Linguistics, 21*: 837-855. https://doi.org/10.15738/kjell.21..202109.837

Abdullah, A. H., Soh, H. M., Mokhtar, M., Hamzah, M. H., Ashari, Z. M., Ali, D. F., & Abd Rahman, S. N. S. (2020). Does the use of smart board increase students' higher order thinking skills (HOTS)? *IEEE Access, 9*, 1833-1854. http://dx.doi.org/10.1109/ACCESS.2020.3042832

Abduraximova, M. (2024). Multimodal discourse analysis in English: Integrating verbal and non-verbal modes. *New Uzbekistan Journal of Academic Research*, *1*(7), 62-65. https://doi.org/10.5281/zenodo.11242460

Adhitya, N., & Valiansyah, V. (2024). The impacts of native speaker teachers' nonverbal communication in EFL classrooms. *The Journal of English Language Teaching, Literature, and Applied Linguistics*, *5*(2), 95–108. https://doi.org/10.37742/jela.v5i2.110

AL Nofali, H. A., & Gasim, A. (2024). Adopting verbal and non-verbal communication strategies in oral discussion by Omani university students. *International Journal for Multidisciplinary Research*. https://doi.org/10.36948/ijfmr.2024.v06i05.28823

Arbab, I. A. E. (2020). *Effect of using audio-visual materials on students' language achievements (A case study of secondary schools' students at Eastern Gazeera)* [Unpublished doctoral dissertation]. Sudan University of Science & Technology.

Archvadze, E. (2023). Exploring multimodality-social-semiotic analysis of communication in TEFL. *Language and Culture*. http://dx.doi.org/10.52340/lac.2023.08.05

Bairstow, D., & Lavaur, J. M. (2012). Audiovisual information processing by monolinguals and bilinguals: effects of intralingual and interlingual subtitle. *Audiovisual Translation and Media Accessibility at the Crossroads*, *36,* 273–293. http://dx.doi.org/10.1163/9789401207812_016

Baldry, A., Thibault, P. J., Coccetta, F., Kantz, D., & Taibi, D. (2020). Multimodal Ecological Literacy: Animal and human interactions in the animal rescue genre. *Multiliteracy advances and multimodal challenges in ELT environments* (pp.155-218). Forum.

Bambaeeroo, F., & Shokrpour, N. (2017). The impact of the teachers' non-verbal communication on success in teaching. *Journal of Advances in Medical Education & Professionalism*, *5*(2), 51-59.

Batty, A. O. (2015). A comparison of video-and audio-mediated listening tests with many facet Rasch modelling and differential distractor functioning. Language Testing*, 32*(1), 3–20. http://dx.doi.org/10.1177/0265532214531254

Becker, S. R., & Sturm, J. L. (2017). Effects of audiovisual media on L2 listening comprehension: A preliminary study in French. *Calico Journal*, *34*(2), 147-177. http://dx.doi.org/10.1558/cj.26754

Benetti, S., Ferrari, A., & Pavani, F. (2023). Multimodal processing in face-to-face interactions: A bridging link between psycholinguistics and sensory neuroscience. *Frontiers in Human Neuroscience*, *17*, 1108354. http://dx.doi.org/10.3389/fnhum.2023.1108354

Bezemer, J. & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. Routledge. https://doi.org/10.4324/9781315687537

Bilfaqih, Y., & Qomarudin, M. N. (2017). Multimodal analysis (vol. 1). Dee Publish. https://doi.org/10.23887/jet.v7i1.54336

Bouchey, B., Castek, J., & Thygeson, J. (2021). Multimodal learning. In J. Ryoo, & K. Winkelmann (Eds.), Innovative Learning Environments in STEM Higher Education. SpringerBriefs in Statistics. Springer. https://doi.org/10.1007/978-3-030-58948-6_3

Bromberek-Dyzman, K., Jankowiak, K., & Chełminiak, P. (2021). Modality matters: Testing bilingual irony comprehension in the textual, auditory, and audio-visual modality. *Journal of Pragmatics*, *180*, 219-231. http://dx.doi.org/10.1016/j.pragma.2021.05.007

Burgoon, J. K. (2003). Nonverbal communication skills. *Handbook of Communication and Social Interaction Skills/Lawrence Erlbaum Associates, Inc*.

Campbell, R. (2007). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1001–1010. https://doi.org/10.1098/rstb.2007.2155

Chen, G., & Fu, X. (2003). Effects of multimodal information on learning performance and judgment of learning. *Journal of Educational Computing Research*, *29*(3), 349-362. http://dx.doi.org/10.2190/J54F-B24D-THN7-H9PH

Cheng, J. (2024). Application of ELAN in human-machine mutual translation of English for the news. *Journal of Electrical Systems*, *20*(6s), 2343-2351. http://dx.doi.org/10.52783/jes.3218

Chiriac, R. (2025). The importance of audio-visual materials in teaching ESL vocabulary at primary school level. *LiBRI. Linguistic and Literary Broad Research and Innovation*, *9*(1), 18-25. https://orcid.org/0009-0002-3541-5153

Coffin, C. & Donohue, J. (2014). *A language as social semiotic-based approach to teaching and learning in higher education* (Language Learning Monograph Series). Wiley-Blackwell. https://doi.org/10.1111/lang.2014.64.issue-s

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education.

Creswell, J. W. (2017). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Edam.

Creswell, J. W. (2024). My 35 years in mixed methods research. *Journal of Mixed Methods Research*, 18(3), 203-215. http://dx.doi.org/10.1177/15586898241253892

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2018). *The sage handbook of qualitative research* (5th ed.). Sage publications, Inc.

Doumont, J. L. (2002). Verbal versus visual: A word is worth a thousand pictures, too. *Technical communication*, *49*(2), 219-224.

Ed-Dali, R. (2024). Enhancing EFL learning through multimodal integration: The role of visual and auditory features in Moroccan textbooks. *Journal of World Englishes and Educational Practices*, *6*(3). https://doi.org/10.32996/jweep.2024.6.3.5

Ercan, S. A., Asenbaum, H., Curato, N., & Mendonça, R. F. (2022). *Research methods in deliberative democracy* (p. 529). Oxford University Press. http://dx.doi.org/10.1093/oso/9780192848925.001.0001

Espino, R. M., Castro, F. Z., Jiménez, M. D. C. R., Suárez, J. L. M., & Marrero, A. C. (2006). Modulated frequency systems in students with cochlear implant. *Auditio, 3*(2), 32-36.

Fàbregues, S., Younas, A., Escalante-Barrios, E. L., Molina-Azorin, J. F., & Vázquez-Miraz, P. (2024). Toward a framework for appraising the quality of integration in mixed methods research. *Journal of Mixed Methods Research*. http://dx.doi.org/15586898241257555.

Farías, M., Obilinovic, K., Orrego, R., & Gregersen, T. (2014). Evaluating types and combinations of multimodal presentations in the retention and transfer of concrete vocabulary in EFL learning. *Revista signos*, *47*(84), 21-39.

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive Science*, *37*(7), 1356-1367. https://doi.org/10.1111/cogs.12048

Fitria, T. N. (2024). Teaching IELTS speaking skills: How is the students' preparation for taking the test? *Journal of English Education Program*, *5*(2). http://dx.doi.org/10.26418/jeep.v5i2.75380

Franzen, M. D. (2013). *Reliability and validity in neuropsychological assessment*. Springer Science & Business Media. http://dx.doi.org/10.1007/978-1-4757-3224-5

Ghoushchi, S., Yazdani, H., Dowlatabadi, H., & Ahmadian, M. (2021). A multimodal discourse analysis of pictures in ELT textbooks: Modes of communication in focus. *Jordan Journal of Modern Languages and Literatures, 13*(4), 623-644. http://dx.doi.org/10.47012/jjmll.13.4.2

Gilakjani, A. P., Ismail, H. N., & Ahmadi, S. M. (2011). The effect of multimodal learning models on language teaching and learning. *Theory & Practice in Language Studies, 1*(10). http://dx.doi.org/10.4304/tpls.1.10.1321-1327

Hahl, K., Lehtovuori, K., & Pietarila, M. (2025). Video research in language classrooms: activities and target language use in early language learning. *Scandinavian Journal of Educational Research*, *69*(2), 363-375.

Halliday, M. A. K., and Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford University Press.

Haneef, M., Faisal, M. A., Alvi, A. K., & Zulfiqar, M. (2014). The role of non-verbal communication in teaching practice. *Science International*, *26*(1), 513-517.

Harutyunyan, N. (2023). Decoding multimodal texts of media discourse. *Armenian Folia Anglistika*, *19*(1), 83-94. https://doi.org/10.46991/AFA/2023.19.1.083%20

Heimbürger, A. (2013). Context meets culture. In *Modeling and Using Context: 8th International and Interdisciplinary Conference* (pp. 143-156). Springer Berlin Heidelberg.

Herman, Murni, S. M., Sibarani, B., & Saragih, A. (2019). Structures of representational meta-functions of the "Cheng Beng" ceremony in pematangsiantar: A multimodal analysis. *International Journal of Innovation, Creativity and Change*, *8*(4), 34–46. https://doi.org/10.25134/erjee.v10i2.6244

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning, 33*(5–6), 567–589. https://doi.org/10.1080/09588221.2019.1577898

Huang, J. (2006). English abilities for academic listening: How confident are Chinese students? *College Student Journal*, *40*(1).

Jewitt, C., & Kress, G. (Eds.) (2003). *Multimodal Literacy*. Peter Lang. http://dx.doi.org/10.3102/0091732X07310586

Kartika, D., Siahaan, S., Herman, H., Rumapea, E. L., & Silalahi, T. F. (2023). Implementation of Audio-Visual Teaching Media in Improving Students' Listening Comprehension: A Case on Teaching Method. *Journal of English Language and Education*, *8*(2), 86-96. https://doi.org/10.31004/jele.v8i2.428

Kehe, D., & Kehe, P. D. (1994). *Conversation Strategies: Pair and group activities for developing commupnicative competence* (2nd ed.). Pro Lingua Associates.

Kehe, D., Kehe, P. D., & Toos, A. (1998). *Discussion strategies: Beyond everyday conversation*. Pro Lingua Associates.

Kenzhegaliev, S. A., Aitzhan, D. K., & Zhanuzakhova, K. K. (2023). Communicative potential of the means of non-verbal communication. *Bulletin of the Karaganda University. Philology Series*, *112*(4), 117-123. https://doi.org/10.31489/2023Ph4/117-123

Khadimally, S. (2016). Visual communications and learning. *US-China Education Review*, *6*(8), 466-479. http://dx.doi.org/10.17265/2161-623X/2016.08.002

Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G. R., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.

Kress, G. R., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design*. Routledge. http://dx.doi.org/10.4324/9781003099857

Kress, G., & Selander, S. (2021). *Multimodal teaching and learning: The rhetorics of the classroom*. Bloomsbury.

Lennartsson, B. (2010, April). Visualization in education-support for the cognitive processes in understanding and learning. In *International Conference on Computer Supported Education* (vol. 2, pp. 375-379). SCITEPRESS.

Lesnov, R. O. (2017). Using videos in ESL listening achievement tests: effects on difficulty. *Eurasian Journal of Applied Linguistics, 3*, 67–91. http://dx.doi.org/10.32601/ejal.461034

Lin, L. F. (2016). The impact of video-based materials on Chinese speaking learners' English text comprehension. *English Language Teaching, 9,* 1–13. http://dx.doi.org/10.5539/elt.v9n10p1

Love, N. (2019). The Semiotics of JR Firth. In T. Sebeok & J. Umiker-Sebeok (Eds.), *The Semiotic Web 1990: Recent Developments in Theory and History*. De Gruyter Mouton.

Madella, P., Wharton, T., & Romero-Trillo, J. (2023). *Non-verbal communication and context: multi-modality in interaction*. Cambridge University Press.

Maguire, L. L. (2005). Literature review: Faculty participation in online distance education: Barriers and motivators. *Online Journal of Distance Learning Administration*, *8*(1).

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511816819

Mayer, R. E. (2009*). Multimedia learning*. Cambridge University Press.

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E. (2021). *Multimedia learning*. Cambridge University Press.

McDuffie, A. (2021). Verbal communication. *Encyclopedia of Autism Spectrum Disorders*, 5029-5029. http://dx.doi.org/10.1007/978-1-4419-1698-3_25

Metsämuuronen, J. (2013). Effect of repeated testing to the development of Biblical Hebrew language proficiency. *Journal of Educational and Psychological Development, 3*(1). http://dx.doi.org/10.5539/jedp.v3n1p10

Mikhael, J. N., Rafiqa, S., & Siaahan, I. (2022). The effect of teachers verbal and non-verbal communication on students motivation in learning English at Smak Frater Don Bosco Tarakan. *Borneo Journal of English Language Education*, *4*(2). http://dx.doi.org/10.35334/bjele.v4i2.3136

Miller, S. M. (2007). English teacher learning for new times: Digital video composing as multimodal literacy practice. *English Education, 40*(1), 64–83.

Mohsen, M. A. (2016). The use of help options in multimedia listening environments to aid language learning: a review. *British Journal of Educational Technology, 47*, 1232–1242. http://dx.doi.org/10.1111/bjet.12305

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19, 309-326.

Namaziandost, E., Nasri, M., & Akbari, S. (2019). The impact of teaching listening comprehension by audio and video aids on the intermediate EFL learners listening proficiencies. *Language, Literature and Culture*, *2*(3), 121-128.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66,* 60–92. http://dx.doi.org/10.17763/haer.66.1.17370n67v22j160u

Ngongo, M. (2021). The investigation of modality and adjunct in spoken text of proposing a girl using Waijewa language based on Halliday's systemic functional linguistic approach. *English Review: Journal of English Education*, *10*(1), 223–234. https://doi.org/0.25134/erjee.v10i1.5382

Norte Fernández-Pacheco, N. (2018). The impact of multimodal ensembles on audio-visual comprehension: Implementing vodcasts in EFL contexts. *Multimodal Communication*, *7*(2). http://dx.doi.org/10.1515/mc-2018-0002

Nuzzaci, A. (2019). A picture is worth a thousand words: visual thinking between creative thinking and critical thinking in the teaching-learning processes. *img journal*, (1), 234-253. https://doi.org/10.6092/issn.2724-2463/11071

Olelewe, C. J., Dong, C., Abdullahi, M., & Nwangwu, C. E. (2023). Effects of using a video-clip instructional strategy on students' performance in a computer networking course. *Technology, Pedagogy and Education, 32*(3), 351-365. http://dx.doi.org/10.1080/1475939X.2023.2201931

Oluwatayo, J. A. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research, 2*. https://doi.org/10.5901/jesr.2012.v2n2.391

Pardo-Ballester, C. (2016). Using video in web-based listening tests. *Journal of New Approaches in Educational Research, 5,* 91–98. http://dx.doi.org/10.7821/naer.2016.7.170

Pintado, B. R., & Fajardo, T. (2021). Learning idioms through the multimodal approach. *Learning*, *12*(24). http://dx.doi.org/10.7176/JEP/12-24-03

Pratolo, B. W. (2019). Integrating body language into classroom interaction: The key to achieving effective English language teaching. *Humanities & Social Sciences Reviews*, *7*(3), 121–129. https://doi.org/10.18510/hssr.2019.7319

Proudfoot, K. (2023). Inductive/deductive hybrid thematic analysis in mixed methods research. *Journal of Mixed Methods Research*, *17*(3), 308-326. http://dx.doi.org/10.1177/15586898221126816

Purwaningtyas, T. (2020). Didactic symbol of visual images in EFL textbook: Multi-modal critical discourse analysis. *Pedagogy: Journal of English Language Teaching*, *8*(1), 51-63. http://dx.doi.org/10.32332/pedagogy.v8i1.1959

Putri, Y. S., Setyaningsih, E., & Putra, K. A. (2024). Embracing multimodality to teach EFL ыtudents: english course teachers' perspectives. *International Journal of Educational Research and Social Sciences*, *5*(4), 731–735. https://doi.org/10.51601/ijersc.v5i4.847

Quintana, S. M. & Maxwell, S. E. (1994). A monte Carlo comparison of seven an adjustment procedure in repeated measures designs with small sample sizes. *Journal of Educational and Behavioural Statistics, 19*, 57–71. http://dx.doi.org/10.3102/10769986019001057

Ramirez, D., and Alonso, I. (2007). Using digital stories to improve listening comprehension with Spanish young learners of English. *Language Learning & Technology, 11(1)*, 87–101. http://goo.gl/U5z8S0 (2016-05-19).

Ravid, R. (2024). *Practical statistics for educators*. Rowman & Littlefield.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge University Press.

Rojas, J. H. C. (2024). The effect of flipped classroom audiovisual content on students' talking time and speaking skills in an adult EFL class. *Gist: Education and Learning Research Journal*, *28*(28). https://doi.org/10.26817/16925777.1609

Rosas, E. (2010). *100+ body language tips*. Knesix Institute.

Ruswandi, R., & Arief, M. (2024). Inspiring through interaction: The impact of teachers' verbal and non-verbal communication in EFL classes. *Voices of English Language Education Society*, *8*(2). http://dx.doi.org/10.29408/veles.v8i2.26821

Salamanti, E., Park, D., Ali, N., & Brown, S. (2023). The efficacy of collaborative and multimodal learning strategies in enhancing English language proficiency among ESL/EFL Learners: a quantitative analysis. *Research Studies in English Language Teaching and Learning*, *1*(2), 78-89. https://doi.org/10.62583/rseltl.v1i2.11

Shaojie, T., Samad, A. A., & Ismail, L. (2022). Systematic literature review on audio-visual multimodal input in listening comprehension. *Frontiers in Psychology*, *13*. http://dx.doi.org/10.3389/fpsyg.2022.980133

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language learning*, *55*(4), 661-699. https://psycnet.apa.org/doi/10.1111/j.0023-8333.2005.00320.x

Surguladze, S. A., Calvert, G. A., Brammer, M. J., Campbell, R., Bullmore, E. T., Giampietro, V., et al. (2001). Audio–visual speech perception in schizophrenia: An fMRI study. *Psychiatry Research: Neuroimaging, 106*, 1–14. http://dx.doi.org/10.1016/S0925-4927(00)00081-0

Sutiyatno, S. (2018). The effect of teacher's verbal communication and non-verbal communication on students' English achievement. *Journal of language teaching and research*, *9*(2), 430-437. http://dx.doi.org/10.17507/JLTR.0902.28

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Szawerna, M. (2023). Theoretical and analytical explorations of multimodality. *Anglica Wratislaviensia*, *61*(1), 9-11. https://dx-.doi.org/10.19195/0301-7966.61.1.1

Tang, S. F., & Logonnathan, L. (2016). *Assessment for learning within and beyond the classroom*. Springer Verlag. http://dx.doi.org/10.1007/978-981-10-0908-2

Tussupova, G. K., Temirbekova, A. T., & Kerimbayeva, K. K. (2016). Verbal and non-verbal means of communication in teaching foreign. *New Trends and Issues Proceedings on Humanities and Social Sciences*. 2(5), pp 205-209. http://dx.doi.org/10.18844/PROSOC.V2I5.1123

Van Leeuwen, T. (2020). Multimodality and multimodal research. *The SAGE handbook of visual research methods* (2nd ed.) (pp. 465-483). SAGE.↑

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

Victoria, M. (2021). The verbal and the visual in language learning and teaching: Insights from the 'Selfie Project'. *The Language Learning Journal*, *49*(1), 93-104. http://dx.doi.org/10.1080/09571736.2018.1484797

Vojdanoska M., Cranney, J., & Newell, B. R. (2009). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*(8), 1183–1195. http://dx.doi.org/10.1002/acp.1630

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing, 27*, 493–513. http://dx.doi.org/10.1177/0265532209355668

Wagner, J., Lingenfelser, F., Bee, N., & André, E. (2011). Social Signal Interpretation (SSI) - A framework for real-time sensing of affective and social signals. *KI-Kuenstliche Intelligenz*, *25*, 251-256. http://dx.doi.org/10.1007/s13218-011-0115-x

Wahyuni, A. (2018). The power of verbal and nonverbal communication in learning. In *1st International Conference on Intellectuals' Global Responsibility* (pp. 80-83). Atlantis Press. http://dx.doi.org/10.2991/icigr-17.2018.19

Wang, C., Yan, J., & Liu, B. (2014). An empirical study on washback effects of the internet-based college English test band 4 in China. *English Language Teaching, 7,* 26–53. http://dx.doi.org/10.5539/elt.v7n6p26

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*(1), 10-16. http://dx.doi.org/10.1111/sjop.12093

Yeh, Y. C. (2022). Student satisfaction with audio-visual flipped classroom learning: a mixed-methods study. *International Journal of Environmental Research and Public Health*, *19*(3), 1053.↑ https://doi.org/10.3390/ijerph19031053

Yi, Q., Dong, Z. Y., & Qiao, H. (2024). Enhancing EFL listening and speaking skills: strategies and practice for implementing multimedia and multi-modal approaches. *Journal of Humanities, Arts and Social Science*, *8*(9), 2211–2217. https://doi.org/10.26855/jhass.2024.09.032

Yi, Y. (2014). Possibilities and challenges of multimodal literacies in the learning and teaching of second and world languages*. Language and Linguistics Compass, 8*(4), 158–169. http://dx.doi.org/10.1111/lnc3.12076

Yıldırım, A. and Şimşek, H. (2006). *Qualitative research methods in social sciences*. Seçkin. https://doi.org/10.4236/vp.2023.92003

Zhang, R., & Zou, D. (2022). A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning. *Computer Assisted Language Learning*, *35*(9), 2790-2816.↑ http://dx.doi.org/10.1080/09588221.2021.1896555

Zhyrun, I. (2016). Culture through comparison: Creating audio-visual listening materials for a CLIL course. *American Journal of Content and Language Integrated Learning, 9*(2)*,* 345–373. http://dx.doi.org/10.5294/laclil.2016.9.2.5.