

EFL Learner's Perceptions of Multimodality in Online Speaking Class: Exploring Linguistic, Visual, and Gestural Aspects¹

Seftika², Zulfa Sakhiyya³, Januarius Mujiyanto⁴, Abdurrachman Faridi⁵,

Widhiyanto⁶, Universitas Negeri Semarang, Semarang, Jawa Tengah, Indonesia

Abstract

The article discusses the students' perceptions of the benefits and challenges of using multimodality, such as the linguistic, visual, and gestural aspects of speaking. Grounded within the framework of multimodality as a social semiotic approach (Kress et al., 2014), the study employed a qualitative case study method. Forty students of the English Education Department at one of the universities in Lampung, Indonesia participated in the study. The data was derived from twelve semi-structured interviews. To maintain data validity, we invited two validators to review and provide input on the interview protocol and utilized back translation to ensure the instrument's validity and the participants' interview comprehension. Then, the interview data were thematically analyzed. The findings indicated that multimodality (linguistic, visual, and gestural aspects) helps students understand through the various modes combined in oral communication to improve vocabulary, reduce nervousness, make presentations confidently, convey messages orally, and improve speaking skills. Visual elements were considered more challenging to use when speaking, however most students agreed that multimodality helped them practice speaking. In theory, students have never learned about multimodality in learning English speaking. Still, in practice, they had used verbal and non-verbal modes of speaking, especially in online learning, where in the current study they were assigned to produce speaking videos as the output of the speaking course. Their speaking project used technological media, including multimodality elements such as linguistic, visual, and gestural aspects. It is argued that it would be better if students intentionally learned multimodality theories and praxis as these would further assist them in learning to speak in English.

Resumen

El artículo analiza las percepciones de los estudiantes sobre los beneficios y desafíos del uso de la multimodalidad, incluyendo los aspectos lingüísticos, visuales y gestuales del habla. Enmarcado en la multimodalidad como enfoque semiótico social (Kress et al., 2014), el estudio empleó una metodología cualitativa de estudio de caso. Cuarenta estudiantes del Departamento de Educación en Inglés de una universidad en Lampung, Indonesia, participaron en el estudio. Los datos se obtuvieron a partir de doce entrevistas semiestructuradas. Para garantizar la validez de los datos, se invitó a dos revisores a revisar el protocolo de las entrevistas y aportar comentarios. Además, se utilizó la retrotraducción para asegurar la validez del instrumento y la comprensión de las entrevistas por parte de los participantes. Posteriormente, los datos de las entrevistas se analizaron temáticamente. Los resultados indicaron que la multimodalidad (aspectos lingüísticos, visuales y gestuales) ayuda a los estudiantes a comprender a través de la combinación de los diversos modos en la comunicación oral, lo que mejora el vocabulario, reduce el nerviosismo, facilita presentaciones con confianza, transmite mensajes oralmente y perfecciona sus habilidades de expresión oral. Si bien los elementos visuales se consideraron más difíciles de usar al hablar, la mayoría de los estudiantes coincidió en que la multimodalidad les ayudó a practicar la expresión oral. En teoría, los estudiantes nunca han aprendido sobre multimodalidad en el aprendizaje del inglés oral. Sin embargo, en la práctica, han utilizado modos de comunicación verbales y no verbales, especialmente en el aprendizaje en línea, donde en este estudio se les asignó la producción de videos de conversación como parte del curso. Su proyecto de conversación empleó medios tecnológicos, incluyendo elementos multimodales como aspectos lingüísticos, visuales y gestuales. Se argumenta que sería más beneficioso que los estudiantes aprendieran intencionalmente las teorías y prácticas de la multimodalidad, ya que esto les ayudaría a aprender a hablar inglés.

Introduction

The term multimodality emerged in the 1990s. Experts such as Kress and van Leeuwen (1996), van Leeuwen (1999), Lemke (1998), the New London group (1996), and Roth (1994) initiated research in the field of multimodality to construct meanings in social interactions. Since then, this research has expanded to various disciplines.

¹This is a refereed article. Received: 18 January, 2023. Accepted: 24 August, 2023. Published: 3 December, 2025.

²seftika@students.unnes.ac.id, 0000-0001-6438-7505, Correspondent.

³zulfa.sakhiyya@mail.unnes.ac.id, 0000-0003-4183-977X

⁴yanmujiyanto@mail.unnes.ac.id, 0000-0002-2627-9789

⁵pakdur.faridi53@gmail.com, 0000-0002-8438-0340

⁶pakwidhi_english@mail.unnes.ac.id, 0000-0001-7351-8976

Multimodality is defined as using more than one mode of communication for meaning-making (Jewitt et al., 2016; Kress, 2010). With various modes of representation besides language, the preference for multimodality has dominated print and mobile media. These media use various sources (modes) to make meaning, such as language, images, movement, and color (Karatzka, 2020). Furthermore, according to the New London Group (1996), meanings can be communicated through semiotic sources, including linguistics, and visual, audio, spatial, and gestural means.

The rapid development of digital technology requires literacy skills beyond traditional methods making communication to become multimodal through which people communicate using various modes. Furthermore, incorporating multimodal elements in language learning is considered important, especially with the growth of online learning at the university level, as Nugroho and Sakhiyya (2022) stated that the design of online learning for foreign languages needs careful consideration of social, cultural, and individual factors to create safe, effective, and engaging programs for today's students. Online learning has provided opportunities for lecturers and students to become technologically literate; digital literacy has been achieved, and students' multimodal literacy skills have also been improved. Multimodal literacy allows students to produce and analyze information using two or more modes of communication to create meaning (Ross et al., 2020). Students are expected to learn effectively and creatively using multiple modes, semiotic sources, and digital equipment at the higher education levels. Therefore, in learning a language, students must be equipped with multimodality knowledge and trained to use multimodal elements such as linguistics, images, and gestures in making meanings.

Several researchers have explored the use of multimodality in the language teaching field (e.g., Bourelle et al., 2016; Acuña Cárcamo et al., 2016; Choi & Yi, 2016; Dressman, 2020; Franceschi, 2018; Gyogi & Lee, 2022; Huang, 2019; Jiang, 2017; Li, 2020; Marchetti, 2015; Meneses et al., 2018; Sin et al., 2020; Song, 2018), and these have covered various aspects of language components and language skills. For example, the study by Acuña Cárcamo et al. (2016) investigated the language component of multimodality. The study revealed that multimodal instructions can affect vocabulary acquisition. The findings suggest that students in public schools can benefit from multimodality exposure in terms of their vocabulary development at various grade levels.

Similarly, Song (2018) studied the teaching of English speaking and listening in multimodal contexts with undergraduate accounting majors. The findings suggest that, depending on the media, multimodality should be used to present the course materials so students can learn through various methods and improve their ability to master the subject matter of multimodal texts.

These existing studies have suggested the positive impacts of multimodality in language classes. Multimodality allows students to improve their ability to acquire vocabulary, to read, to write, and to better understand the subject matter. The literature also indicates the potential of multimodality in teaching English. Nonetheless, much previous research deals with reading, writing, and English components such as vocabulary. Studies on integrating multimodality into online learning and speaking skills are still limited. Here, Blair and Monske (2015) asserted that, although technology has been widely applied in teaching, multimodality is slower to enter the online arena, and online classes still often use text-based assignments. However, in language learning, previously, the focus of multimodality was still related to students' receptive skills, so there was a need for a movement toward students' productive skills (Magnusson & Godhe, 2019). Moreover, Kress (2010) claimed that many formal environments, such as schools, still use traditional ways to build meanings (making meaning) through writing and paper-based text. In the era of technology and online learning, there needs to be a change in learning, where multimodality can be a solution, especially in learning English.

Considering the potential contributions of multimodality to language learning, this article explores the students' views on their multimodality practice in an online speaking class. Furthermore, as most of the multimodality research in the language field has been related to reading and writing skills and the most frequently studied multimodality aspects were verbal or linguistics and visual, this study focused on students who used linguistics and visual modes and implemented gestural modes in a speaking class. Thus, with these idea in mind, this research explores students' perspectives on multimodal practice using linguistics, visual, and gestural aspects in an online speaking class. The following research question guided the study: How do the participants perceive the benefits and challenges of using the three aspects of multimodality in developing English speaking skills?

Literature Review

Multimodality

In today's communication landscape, meaning is rarely conveyed through language alone. People increasingly draw upon multiple modes—such as visuals, sounds, and gestures—to construct and interpret messages in context.

When combined with two or more semiotic systems, these modes give special meanings to texts. For example, in a video, all elements, including sound, images, colors, layout, and gestures, are seen as a whole to convey a particular message or information. The exchange of information using signs or modes is also known as "semiotic sources" (van Leeuwen, 2005), as actions and artifacts people use in communication, whether they are physiologically produced with our vocal apparatus, with muscles to make facial expressions and gestures, by technological tools such as pens, ink, computer hardware and software, scissors and sewing machines, and others.

The modes of expression in multimodal texts serve several functions and carry various meanings depending on the settings in which they are used. Van Leeuwen (2005) contends semiotic resources have a pre-existing meaning due to their past use and a relevant set of abilities, which will be actualized in a concrete social context and subject to some form of semiotic regime. For example, print and screen-based technologies offer different modes of meaning-making in which people use technology in various ways. These modes of expressions can include words, sounds, still and moving images, and animation and colors (Jewitt, 2005). Furthermore, Kress and van Leeuwen (2001) note that all semiotic meanings people give to an object/text/phenomenon are through the mode used. Therefore, multimodality is an approach in spoken and written communication in which people use various semiotic modes to communicate meanings precisely and gracefully. This study examined three aspects of multimodality (i.e., linguistic, visual, and gesture) used in online speaking classes. Linguistic multimodality includes aspects such as a person's word choices, word order, pronunciation, and grammar of oral and written language. Visual aspects can include colors, layouts, style, size, vectors, and viewpoint, usually in still and moving images. Gestural aspects include hand gestures, facial expressions, body language, and interactions between people (New London Group, 1996).

Using these three learning modes aligns with the multimodal approach theory in communication, as Kress et al. (2014) presented, suggesting using linguistics, visual, and gestural elements for communication in learning. A multimodality approach focuses on using multiple modes to represent learning. This approach is in accordance with the real conditions in the classroom that verbal is not the only way of communicating in the teaching and learning process. Sometimes teachers and students use verbal communication accompanied by visuals, gestures, or both.

Multimodality in language learning

Multimodal studies apply new theoretical frameworks and methodologies to analyze communication that integrates modes outside of verbal language (Kress & van Leeuwen, 1996; Smith, 2011). According to Dressman (2020), multimodality is important for language learning and can help students learn languages more effectively. Therefore, multimodal competence benefits EFL students in an interactive learning process. Furthermore, mediation occurs across all modalities with a multimodal approach, e.g., meaning is created, distributed, interpreted, and reconstructed through various communication sources (Jewitt, 2005; Kress & van Leeuwen, 2001).

In classes with a multimodality approach, the learning materials are presented to students using various modes whose implications can encourage students' attention in learning and improve learning performance (Moreno & Mayer, 2007). Further in the learning process, students engage in multimodality tasks to create meanings. In this case, making meaning refers to continuous activities and processes.

Kress (2010) regards meaning-making as learning and Kalantzis and Cope (2012) define it as making a design which can be understood in a variety of ways. The stages of such a design include design available (e.g., assignments and study resources), design (student work with assignments), and redesign/new designs available. From a multimodal perspective, assignments can encourage students to understand the learning materials better and convey messages. However, the usefulness of assignments also depends on how the assignments are prepared, assessed, and communicated (Magnusson & Godhe, 2019).

Applying multimodality in language learning has various benefits. Metiti Group (2008) stated that students in a learning environment with a multimodal design outperformed students in a learning environment with a traditional one-mode approach. Picciano (2009) also added that a multimodal learning environment

provides a pleasant and challenging environment where students can learn independently and engage with the numerous elements available. In detail, multimodality elements such as linguistics, visual, and gestural modes offer many advantages in learning. For example, Choi and Yi (2016) suggest a visual presentation accompanies linguistic texts to facilitate students' understanding of content knowledge and that multimodality builds students' cognitive and affective engagement with content knowledge. For instance, using visual representations in learning can help students' understanding of linguistic texts. Additionally, gestural aspects, non-verbal aspects, including tone, gaze, facial expressions, body language, and postures, are crucial for good communication for second language students (Busà, 2010). Rance-Roney (2010) suggests that students should be taught how to use body language appropriate to the context, for example, in learning functional language and how gestures are used in conversational strategies.

Therefore, we argue that the multimodality approach potentially offers many advantages in learning English, especially in developing speaking skills. First, these semiotic sources are used together to make meanings. Oral communication accompanied by gestures is helpful for attracting the audience's attention, while pictures are useful as references to clarify the meanings conveyed orally. Second, in multimodal classes, learning materials can be presented through semiotic modes, such as verbals, images, and gestures. Using various multimodal resources can support students' learning as students can become more active and attentive, thus improving their performance. Third, the application of multimodality helps students enrich their vocabulary mastery. It helps improve students' ability regarding language subject contents, allowing students to improve their speaking skills. Fourth, multimodal enables students to be more familiar with multimedia for learning. As Cook and Kirchoff (2017) reported, helping students acquire multimodal literacy skills requires multimodality. Finally, multimodality can facilitate undergraduate students to practice English using various modes of communication, which leads them to acquire multimodal literacy.

Current research on multimodality in ELT

Many researchers have investigated multimodality in English Language teaching. In the context of vocabulary and reading, Acuña Cárcamo et al. (2016) found that multimodality can influence students' vocabulary acquisition. Meneses et al. (2018) also utilized multimodal sources, both verbal and visual, to assess students' reading comprehension skills in a science class revealing that teachers should be innovative in science education and that students can be given several opportunities to learn science and improve their reading comprehension. Huang's (2019) research findings highlighted the power connections generated in texts due to the interaction between the verbal and visual components; notably how the visual mode can rationalize the constructed power relations of the verbal mode. Similarly, Bao's (2017) study suggests that employing multimodality in reading instruction can be effective. The findings also indicated that multimodal instruction is more popular among English majors because it can motivate students to read outside of class and foster their confidence in English reading. In addition, Anari et al. (2019) examined how multimodality affects Iranian EFL learners' reading comprehension and vocabulary retention. The findings showed that multimodality improved L2 reading comprehension and EFL vocabulary recall, and students preferred texts with visuals.

In the writing field, Shin et al. (2020) investigated the writing component of a sixth-grade L2 writers using digital multimodal texts. Using multimedia presentations in the writing activity, the students created *PowerPoint* slides explaining the greenhouse effect and *Glogster*-based multimedia texts. The findings illustrated that the students employed linguistic and visual modes in drawing the writing. Then, Magnusson and Godhe (2019) examined multimodality in the Swiss education curriculum, meaning-making for students to do poetry assignments. The study analyzed the multimodal approach in language education and how the Swedish curriculum supports multimodality in learning. This study used a sociosemiotic, multimodality approach. The students had the task of writing poetry. The poem was viewed from the perspective of multimodality, that is, different modes and media were used. Cook and Kirchoff (2017) investigated the implementation of reading and writing graphic novels as a text for teaching multimodal literacy. The findings showed that multimodal text can help students develop multimodal literacy skills. Collier and Kendrick (2016) examined how multimodal compositional processes can affect learning across different forms of presentation. This study suggests that writing, including multimodal writing, takes time, effort, and practice. Teachers and literacy researchers must evaluate students' responses to teacher instructions and the multimodal environment of classroom composing practices. Using multimodality in a students' writing assignment allows them to determine media and modes that suit their interests and needs in writing practice.

In the speaking field, Marsakawati et al. (2019) studied multimodality aspects, namely evaluative language and gestures, in an EFL class. In their research, Indonesian students used language and gestures to convey interpersonal meanings in oral persuasive presentations. Similarly, Peng (2019) examined the effects of multimodal pedagogy, classroom environment, and willingness to communicate (WTC) in English. The results showed that the participants considered the teacher's audio/video and voice/facial expressions the most satisfying. In this case the visual element of *PowerPoint* slides was less than satisfactory. The teacher's voice/facial expression contributes to the classroom environment, and the teacher's gestures and spatial position create WTC. These findings have direct implications for L2 teaching and can enable L2 teachers to harness the potential of multimodal pedagogy to promote students' WTC and the benefits of their learning. Then, Franceschi (2018) combined a multimodal approach to teach medical English effectively. A Gestalt technique was used to improve students' learning. Gestalt is a technique to encourage learning through hands-on experiences. It promotes spontaneity and leads the individual to act and speak rather than just think or recall something. The students used this technique for speaking practice as they did an exercise in physician-patient communication. This approach facilitates student memorization of vocabulary and typical conversation patterns used in doctor-patient interactions, enhancing students' general communication skills and relationship-building abilities.

Song (2018) examined multimodality while teaching a college English course on seeing, listening, and speaking for accounting students. This research categorized multimodality into two categories, language and non-language. Language studies include 1) pure languages, such as teacher and students' voices; 2) paralinguistic, such as sound, tone, and font layout in textbooks, *PowerPoint* slides, and students' works. The existing literature suggests that using multimodal, namely visual, audio, gesture, and spatial position, creates a comfortable language learning environment and motivates students to communicate to convey messages orally.

Literature has also indicated that the multimodal approach can help improve students' EFL skills, such as reading, listening, speaking, and writing, and improve students' vocabulary mastery. However, only a few studies have considered the speaking skills and experiences of multimodality in online speaking classes at higher education levels. Also, these studies have not particularly examined multimodality from the students' point of view. Therefore, this study intends to fill this gap. Multimodality for higher education students learning to develop speaking skills was investigated in an online speaking class. Then, students' perceptions of using linguistic, visual, and gestural aspects of multimodalities in online speaking classes were examined.

Methodology

Research design and context

This study employed a qualitative case study approach. It was felt to be appropriate because, as Yin (2009) argued, it allows a real-time analysis of a problem and it can be done to determine individual or group perceptions. The current research was carried out in an online speaking class where students could multimodal linguistic, visual, and gestural elements. The goal was to discover students' perspectives regarding the use of these elements and to collect their opinions about the perceived benefits and challenges they encountered.

This study is framed in a social semiotic perspective on multimodality (Kress et al., 2014), based on the premise that linguistic, visual, and gestural elements as modes of representation in the meaning-making processes. Multimodal and social semiotic approaches begin from visuals, gestures, and action and have evolved through their social use into semiotic systems articulated in the same way as language.

The present study focused on undergraduate students in Lampung, Indonesia. This study was conducted in the post-advanced speaking course in which the students practiced speaking English face-to-face through *Google Meet* and their speaking projects. To acquire the data, forty students participated in a semi-structured interview.

Informed consent

Before conducting interviews, written informed consent forms were provided to 54 potential interview participants to seek their consent. Then, they were invited to meet on *Google Meet*. At this meeting, details of the research project were given and research confidentiality was guaranteed. Next, the students were asked to fill out to a *Google* form where they could indicate whether they would be willing to participate in the study. Forty students eventually agreed to participate. Interview schedules were then set up.

Data collection

The data collection process began by designing a semi-structured interview protocol, which two multimodality experts validated to check its validity. Furthermore, back translation was used to ensure the validity, readability, and understandability of the interview protocol. For the convenience of the participants, the semi-structured interview protocol was translated into Indonesian. At the back translation stage, interview drafts were constructed in English by adapting other relevant research instruments. After this, the interview protocol was translated from English into the Indonesian language. Next, colleagues were invited to translate the interview back into English. Then, the English translation was compared with the interview instrument that had been first constructed and differences were reconciled. Finally, participants were given the finalized Indonesian translation. The interview included 12 semi-structured questions. The participants were interviewed individually using *Google Meet*. These sessions were videotaped.

Data analysis

Interview data were analyzed using thematic analysis (Braun & Clarke, 2006) to find the themes contained in the data. To do so, the data was analyzed in the following stages. First, the researchers familiarized themselves with the data by repeatedly listening to the recorded *Google Meet* videos, transcribing the interview data, and reading and re-reading the interview transcripts. Then notes were taken to mark the important points in the interview transcripts. Second, it was determined which data in the interview transcript needed to be coded. After coding the interview transcripts, all the codes were reviewed and evaluated and it was determined which codes were relevant to the study. Then, codes that had the same or similar meanings were combined. The names of these groups were assigned according to the content (codes). Third, groups that had the same meaning were put into a group of tentative themes. Then each tentative theme was compared with other themes to identify similarities or differences. This was also done to find the relationships between one theme and another. Fourth, as the tentative themes were reviewed, each theme's nature and relationship was defined. After the tentative themes had been compiled, they were arranged again based on its significance and relation to the research question and selected major themes. Fifth, a detailed analyses of each chosen theme was written. Finally, the writing stage was considered to be the final data analysis step, since last round of analysis could be done as we presented the chosen themes to the readers.

To further maintain the trustworthiness and rigor of the research findings, member-checking was also performed in which participants were invited to look at the transcripts and comment on the identified major themes.

Results

Students' perceptions of the benefits and challenges of using the three aspects of multimodality in improving their speaking skills were analyzed. The themes that emerged from the data analysis indicated that multimodality was viewed positively by the students as it helped them improve their vocabulary, reduce nervousness, perform presentations more confidently, convey messages orally, and improve their speaking skills.

Improving vocabulary mastery

The participants described visual elements, such as pictures, as beneficial for students when understanding and remembering new vocabulary.

Yes, multimodality is good for speaking. If I speak, indeed, I already know and memorize the vocabulary. The visual element helps me to know the meaning of words. So, remembering the words by looking at the original picture can be more helpful. (Student 15)

The participants maintained that they were able to increase their vocabulary acquisition using multimodality aspects in speaking. Some of them found that they enjoyed this because they got new vocabulary they could use to communicate.

I enjoyed the preparation step on multimodal project because, at this phase, I obtained the new vocabulary in which the lecturer asked us to make a video, and it was given a theme. So, I got many new vocabularies. (Student 14)

Yes. For example, in vocabulary, we are given a different theme in each meeting, so I got new vocabulary from what we were talking about. Then, we can learn how to speak more freely and relax by using gestures. (Student 39)

Based on the students' opinions they felt that if students practiced speaking in English often, they could automatically increase their vocabulary because the speaking topics changed every two meetings. During these two weeks of working on speaking assignments, students had enough time to create content that fit

the topic. They learned vocabulary related to the topic and they practiced speaking independently at home on a variety of topics. The more students practiced speaking independently, the more vocabulary they mastered.

Performing presentations more confidently

Participants reported that multimodality aspects made them more confident in English presentations. Additionally, they pay attention to their pronunciation to convey the message confidently. Speaking with good pronunciation can give them confidence since the speakers are not worried that many people might see the speaking video uploaded to the *YouTube* channel. Furthermore, the grammatical aspect was considered by the speakers to be necessary to convey the message correctly and clearly.

Grammar and pronunciation need to be considered so as not to be embarrassed when many people see the video speaking. (Student 4)

Yes, because if we say something that does not match the pronunciation and has a different meaning. (Student 10)

Multimodality makes me more confident in speaking English and makes me practice speaking more often to hone my English-speaking skills. (Student 6)

Reducing nervousness

Participants explained that using multimodality, combining linguistic aspects with gestures and other aspects, could reduce their feelings of nervousness when speaking in English. As can be seen in the excerpt below, some participants said that the gestural element was one of the crucial elements in speaking. For students, using gestures such as facial expressions, smiles, eye contact, hand, head, and shoulder movements reduced nervousness.

Using gestures lets us be more flexible. We will be too weird and monotonous if we keep talking and quiet. So, body language makes me more relaxed and more accessible. So, I can explain the information better. (Student 30)

Gesture is making [sic] us more comfortable in speaking English. I can express my feelings well and be more relaxed. (Student 39)

They also reported that gestures when used appropriately can affect the speaker's presentation; for example, smiles and hand movements can help students be more relaxed while speaking, making it easier to convey messages to the audience.

Using gestures is automatic. It is convincing the other person you're talking to. When we talk using gestures, the other person will understand better what we are talking about. (Student 17)

Yes. For example, in vocabulary, we are given a different theme in each meeting, so I got new vocabulary from what we were talking about. Then, we can learn how to speak more freely and relax by using gestures. (Student 39)

Conveying messages orally

Participants reported that using multimodal aspects helped them convey oral messages more easily. Some participants said pictures could be used as material or practical examples when speaking in English to clarify what was being expressed.

When making [sic] a speaking video assignment, I used body language because it can clarify what we say. For example, in an assignment, I promoted a product. It is impossible for me to look sad, so I have to be fun, smile and use gestures to explain the products to the audience. (Student 12)

Using vocabulary, picture, and hand movement help me in speaking. For instance, I think the audience will feel more convinced when I speak using gestures. But then, visuals can be used to provide examples in speaking. (Student 20)

Yes, it is beneficial. I can be more flexible in speaking if I have prepared the text. Then, I can prepare the appropriate gesture and visual element so that the audiences understand the information quickly. Therefore, speaking is more exciting and not monotonous. (Student 30)

Images are mainly used as supporting elements when speaking English because they have the following benefits: they can be used as support in speaking so that the audience can understand and believe the information conveyed.

To support the topic of conversation in speaking English, for example, when playing the role of a newscaster, displaying pictures is needed to support the facts of the news being conveyed. So that the audience believes that the news conveyed is not a hoax. So, pictures can be evidence. (Student 1)

Images and colors are used in a video to make it look attractive.... visual elements can beautify the presentation's appearance in English. So, visual elements support aesthetics. Then, pictures can help to convey ideas in speaking. Through pictures, viewers can understand what the speaker is saying. (Student 3)

Visuals are also beneficial for speakers because sometimes they often make mistakes in pronunciation and grammar. However, the listener or interlocutor can still understand the message conveyed by looking at the pictures.

Visual element [sic] makes the audience understand more because sometimes I say something wrong in English, and then it can be explained more using animation and picture. (Student 16)

When I convey something, if there is an expression, if I add body movements, and if I explain the material using a picture, the audience will understand more easily. So, it is beneficial. (Student 15)

The benefit of gestures is students are more fluent in expressing ideas, and the audience can understand the conversation clearly. (Student 7)

Improving speaking skills

The data analysis suggests that the use of multimodalities such as linguistic, visual, and gestural in speaking can improve students' speaking skills, as indicated in the excerpts below.

Linguistic

I got a speaking assignment recorded using video, then looked for references relevant to the speaking theme. In my speaking assignment, I also learned to have good pronunciation, choose the appropriate vocabulary, and use good grammar. So, I got improvements in terms of pronunciation, vocabulary, and grammar. (Student 18)

The multimodality aspect affects my speaking skills, such as in preparing the script, being able to add to my vocabulary, learning grammar again, what should we use in the sentence. So, it helps to communicate verbally. Then, I became more knowledgeable about technology, for example in the editing process I understood how to insert backgrounds, effects, and backgrounds like that. (Student 33)

Definitely, it has a positive impact. For example, grammar, pronunciation, and vocabulary facilitate students to practice English. But, in oral communication, we need to speak using the proper grammar. (Student 40)

Gestures

Multimodality is necessary so that we are communicative in speaking. In speaking English, sometimes I need to remember what vocabulary is right. Using gestures, I can remember the vocabulary that fits the speaking topic. (Student 12)

Visuals

Yes, multimodality is good for speaking. If I speak, indeed, I already know and memorize the vocabulary. The visual element helps me to know the meaning of words. So, remembering the words by looking at the original picture can be more helpful. (Student 15)

At that time, I have a speaking task in the form of news. The news cannot be made up, there must be evidence, so I have to insert the video related to the news I convey. (Student 39)

In addition to benefits, students also have challenges when using multimodality aspects while speaking English. Below are three major themes that illustrate these challenges.

Lack of grammatical ability

The data reveals that students still view grammar as an inhibitor while speaking English. Some students describe grammar as the difficult aspect of speaking and are still confused about using the right grammatical rules.

I am still confused about the grammar. Regarding vocabulary, I can use Google Translate. If there is strange diction [sic], I will try to translate it again. (Student 39)

In my opinion, the most difficult aspect is grammar. The easiest is pronunciation. (Student 40)

Students consider grammar difficult and as it is difficult, they tend to ignore it. They think that without grammar, they can still speak, but without vocabulary, they cannot convey anything.

Inability to choose the right visual

Students also said that the visual aspect was also difficult to implement in speaking. They found it difficult to find visuals that could illustrate the theme.

I mostly used linguistics [sic] and gesture elements. I rarely used images because I got [sic] difficulty obtaining the appropriate pictures. (Student 2)

I have difficulty with the drawing aspect, ma'am. It took much work to find a suitable image. I have never learned how to use proper pictures, animations, and colors in learning. (Student 13)

A lack of adequate editing time

Students also have problems editing, so it is difficult for them when they have to include visual elements.

I have been unable to edit the background when making videos speaking in English. The image editing part is the most difficult of the three elements of multimodality. (Student 1)

It is not easy in terms of the editing process to make a speaking video. I could [sic] improve at computers. So, when I edited the visual elements, I had to search and look at the book first. The parts that make it difficult are the addition of animation, the poster part, and the alternating colors. (Student 40)

Discussion

This study explored the students' perceptions of the benefits and challenges of using multimodality, including linguistic, visual, and gestural elements, in speaking English. The research findings show the benefits were in terms of increasing vocabulary, reducing nervousness, performing presentations confidently, conveying messages orally, and improving speaking skills. In addition, the findings of this research also showed that the students experienced some challenges in using multimodality aspects.

Benefits of using multimodality aspects in online speaking classes

First, students experienced an increase in vocabulary acquisition. This study found that linguistics, visual, and gestural aspects assist students in obtaining new vocabulary and improving their vocabulary acquisition. It is supported by Acuña Cárcamo et al. (2016), who stated that multimodality affects students' vocabulary learning. The students in the current student were enthusiastic about finding new vocabulary independently because, based on their perceptions, it was easier to speak if the content-related vocabulary has been mastered. This is in line with the opinion of previous research that vocabulary size determines students' success in language learning, especially when dealing with speaking activities (Salam et al., 2021).

Second, using multimodal aspects in speaking were found to increase the students' confidence in their oral presentations. The more often students practiced speaking using multimodality, the more confidence they had in speaking. This is supported by Tyas and Widhiyanto's (2020) study that showed that students could be more active and confident when speaking English and could be less afraid of making mistakes. Also by using appropriate pronunciation, the students in the present research enjoyed speaking more, spoke more fluently, and were less embarrassed on giving presentations. Shak et al. (2016) argued that making pronunciation errors can be embarrassing for college students, because they are frequently obliged to give oral presentations in a variety of subjects. Although errors are still found in pronunciation, students' English speech can still be understood. In addition, by having appropriate pronunciation, students can be more confident when speaking, as is supported by Gilakjani (2011) who reported that good pronunciation can enhance students' confidence since they can engage in class discussions and talk more freely to other students. However, grammar still remained a concern for the students in this article, although not as much as vocabulary and pronunciation.

Third, the use of multimodal aspects such as linguistics, visuals, and gestures was found to reduce feelings of nervousness in speaking English. For example, the use of gestures in speaking helped students reduce their feelings of nervousness and resultant stiffness so that oral communication became smoother. For this reason, Rance-Roney (2010) suggests that it is necessary to introduce proper body language explicitly in the context of conversational strategies and the study of functional languages. Thus, using several modes of instruction can foster a positive learning environment in which students are more relaxed in speaking and can encourage students to continue learning independently as they observe English speakers in videos.

Fourth, multimodality aspects can help students convey oral messages well. For example, the gestural aspect is a natural mode students use in verbal communication. Consciously or subconsciously, gestures have been used in speaking. The forms of gestures that students often use are body language, hand movements, facial expressions, and eye contact. Gestures are an essential role in oral communication. Facial expressions help the speaker convey meaning. Facial expressions and hand gestures in speaking help students convey ideas. The speaker's gestures can reflect their thinking. When the speaker is unable to verbalize his thoughts, gestures can help him communicate his point, this way the audience can comprehend the communicating partner (Goldin-Meadow & Alibali, 2013).

Students also used visual of many kinds, including colorful pictures, animations, and backgrounds. The benefits of used in oral presentations attracted, motivated, and kept the learners' attention so they could

understand the information more easily. Besides, pictures could be used as references to clarify the meaning conveyed orally. Mujiyanto (2017) said providing visual examples to help people understand verbal passages was essential. With multimodality, non-verbal elements such as pictures can help students communicate verbally. For example, in an English presentation, displaying pictures related to the topic can help the speaker convey the message, and listeners who may not be fluent in English can understand the information better when they see the picture displayed. This statement is in agreements with Choi and Yi (2016) who said that when students have limited language skills or they do not have fluency in the target language yet, visuals can help them understand oral messages.

Fifth, this study also found that multimodal elements such as linguistic, visual, and gestural elements were effective in the students' speaking practice. One of the most popular visual practices involved the creation of videos. These can help students increase their confidence in speaking and can carry over to face-to-face interactions. This finding aligns with Busà's (2015) that reported that oral learning activities such as making videos benefit EFL students in integrating verbal and non-verbal communication skills. In the videos the three aspects of multimodality complement each other. For example, in making a video speaking assignment that contains elements of multimodality, students will have improved their speaking whether they realize it or not and they have become accustomed to using elements of multimodality and multimedia. The linguistic aspect helps students improve their language skills, and the visual and gestural aspects help them communicate meaning non-verbally so that messages conveyed orally can be accepted and understood by the audience. As Elmiana (2019) stated, by using visual elements, students can learn effectively and help students practice understanding the meanings of a text's linguistic elements. This finding aligns with Busà (2010) who says that in learning, students who are exposed to various multimedia sources such as pictures and *YouTube* videos can engage in metalinguistic reflection on how verbal and non-verbal communication is used in different settings to gain two advantages at once: gaining knowledge about non-verbal communication, and also practicing oral communication.

The contribution of non-verbal elements, such as visual and gestural, in oral communication, is significant because speaking requires not only language proficiency but also knowledge of how to use appropriate language in social interactions such as correct intonation, voice quality, and gestures which accompany face-to-face communication (Camiciottoli & Campoy-Cubillo, 2018). In addition, Kress et al. (2014) suggests that language, visual images, and gestures are resources that can be represented as modes of communication.

The challenges of using multimodality

In using the multimodality aspects, students did experience challenges. In relation to language, students had difficulties in terms of grammar. However, the visual mode was considered to be the most difficult element for some students because they needed more time to think about and prepare visual objects to support the speaking content of their video presentations. Students had to be skilled at inserting and editing images, combining colors, creating backgrounds and animations in. Besides the selected visual aspects had to support the speaking topic. Choi and Yi (2016) stated that visual aspects were important to help students understand texts. Then, every visual element used in speaking must have a role in conveying meaning since it can help overcome communication problems that words can fail to do. Neuroscience research has also revealed that significant improvements in learning can be achieved through visual and verbal multimodal learning information (Metiti Group, 2008).

Other challenges that students faced in this study were that students were not aware of the theory of multimodality. At the beginning of the interview, they admitted that they did not know about multimodality, even though they had applied elements of multimodality, such as verbal (linguistic aspect), visual, and gestural elements in the past. Because students did not know the concept of multimodality, they used it to construct video speaking without clear guidelines and only based on intuitiveness. Students who felt insecure about presenting their speaking project had difficulties choosing images that matched the content of their speaking skills and had problems creating visual elements in the speaking project. Some students were not confident with their work and thought their visual elements construction was not as good as others. Prain and Waldrip (2006) said that students who understood the theory and could create a connection among various elements of multimodality had a more sophisticated conceptual understanding than those who did not. Because of the importance of multimodality in learning speaking skills, especially for adult learners, it is highly recommended to raise awareness and practice the concept of multimodality in the teaching and learning process. This way, students would be able to use and integrate various modes into their practice.

Conclusion

This study revealed that multimodality elements, such as linguistic, visual, and gestural modes, contributed to students' oral communication. However, a few of the students in this study had difficulties in applying multimodal elements in speaking, namely obstacles in terms of grammar, a lack of knowledge about multimodality, and working with visual aspect. Nevertheless, most of them agreed that using multiple speaking modes could help them communicate. Multimodality helped students make themselves understood in oral communication by improving vocabulary, reducing nervousness, giving presentations confidently, conveying messages orally, and improving their speaking skills in general. The current research has provided a real example of the role of multimodality elements in language learning, especially related to speaking skills. This learning environment is suitable for application in the current digital communication technology era. The use of various semiotic modes can help students understand learning well through meaning-making done by students so that they can learn independently, challenge, innovate, and become technology and information proficient. It is, therefore, essential for educators and students to know multimodality both in theory and practice.

In addition, the use and research of multimodal elements in online speaking classes with EFL learners can contribute to the current literature. First, this study can be used as a reference for learning speaking skills based on a multimodality approach. Second, this study recommends that students become aware and practice the concept of multimodality in the language learning process. Third, it suggests a multimodal approach as an innovation in learning to speak because students can determine their multimodal representation based on their learning needs. Lastly, this research has provided new insights into students' perspectives regarding the benefits of multimodality in learning English in today's digital era.

References

- Acuña Cárcamo, M. M., Concha Cartes, R. A., Estrada Velásquez, N. E., & Díaz Larenas, C. H. (2016). The impact of multimodal instruction on the acquisition of vocabulary. *Trabalhos Em Linguística Aplicada*, 55(1), 129-154. <https://doi.org/10.1590/010318134842170942>
- Anari, N. N., Aboo saeedi, A. A. R., & Shariati, M. (2019). The effects of multimodality on reading comprehension and vocabulary retention among Iranian EFL learners. *Iranian Journal of English for Academic Purposes (IJEAP)*, 8(4), 86-101. <https://dor.isc.ac/dor/20.1001.1.24763187.2019.8.4.6.2>
- Bao, X. (2017). Application of multimodality to teaching reading. *English Language and Literature Studies*, 7(3), 78-84. <https://doi.org/10.5539/ells.v7n3p78>
- Blair, K. L., & Monske, E. A. (2009). Developing digital literacies and professional identities: The benefits of e-portfolios in graduate education. *Journal of Literacy and Technology*, 10(1), 40-68.
- Bourelle, A., Bourelle, T., Knutson, A. V., & Spong, S. (2016). Sites of multimodal literacy: Comparing student learning in online and face-to-face environments. *Computers and Composition*, 39, 55-70. <https://doi.org/10.1016/j.compcom.2015.11.003>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706QP0630A>
- Busà, M. G. (2010). Sounding natural: Improving oral presentation skills. *Language Value*, 2(1), 51-67.
- Busà, M. G. (2015). Teaching learners to communicate effectively in the L2: Integrating body language in the students' syllabus. *Lingue e Linguaggi*, 15, 83-98. <https://doi.org/10.1285/i22390359v15p83>
- Choi, J., & Yi, Y. (2016). Teachers' integration of multimodality into classroom practices for English language learners. *TESOL Journal*, 7(2), 304-327. <https://doi.org/10.1002/tesj.204>
- Cook, M. P., & Kirchoff, J. S. J. (2017). Teaching multimodal literacy through reading and writing graphic novels. *Language and Literacy*, 19(4), 76-95. <https://doi.org/10.20360/G2P38R>
- Collier, D. R., & Kendrick, M. (2016). I wish I was a puppy: A multimodal view of writing process assessment. *Pedagogies: An International Journal*, 11(2), 168-188. <https://doi.org/10.1080/1554480X.2016.1169187>
- Camiciottoli, B. C., & Campoy-Cubillo, M. C. (2018). Introduction: The nexus of multimodality, multimodal literacy, and English language teaching in research and practice in higher education settings. *System*, 77, 1-9. <https://doi.org/10.1016/j.system.2018.03.005>
- Dressman, M. (2020). Multimodality and language learning. In M. Dressman & R. W. Sadler, (Eds.), *The handbook of informal language learning* (pp. 39-55). Wiley.
- Elmiana, D. S. (2019). Pedagogical representation of visual images in EFL textbooks: A multimodal perspective. *Pedagogy, Culture & Society*, 27(4), 613-628. <https://doi.org/10.1080/14681366.2019.1569550>
- Franceschi, D. (2018). Physician-patient communication: An integrated multimodal approach for teaching medical English. *System*, 77, 91-102. <https://doi.org/10.1016/j.system.2018.02.011>
- Gyogi, E., & Lee, V. (2022). Multimodality in translation: A look into EFL and JSL classroom. *Language Awareness*, 31(2), 232-249. <https://doi.org/10.1080/09658416.2021.2023557>
- Gilakjani, A. P. (2011). A study on the situation of pronunciation instruction in ESL/EFL classrooms. *Journal of Studies in Education*, 1(1). <https://doi.org/10.5296/JSE.V1I1.924>

- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257-283. <https://doi.org/10.1146/annurev-psych-113011-143802>
- Huang, S.-Y. (2019). A critical multimodal framework for reading and analyzing pedagogical materials. *English Teaching: Practice & Critique*, 18(1), 52-69. <https://doi.org/10.1108/ETPC-08-2018-0078>
- Jiang, L. (2017). The affordances of digital multimodal composing for EFL learning. *ELT Journal*, 71(4), 413-422. <https://doi.org/10.1093/elt/ccw098>
- Jewitt, C. (2005). Multimodality, "Reading" and "Writing" for the 21st century. *Discourse: Studies in the Cultural Politics of Education*, 26(3), 315-331. <https://doi.org/10.1080/015966300500200011>
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Routledge.
- Kalantzis, M., & Cope, B. (2012). *Literacies*. Cambridge University Press.
- Karatza, S. (2020). Multimodal literacy and language testing: Visual and intersemiotic literacy indicators of reading comprehension texts. *Journal of Visual Literacy*, 39(3-4), 220-255. <https://doi.org/10.1080/1051144X.2020.1826222>
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Routledge.
- Kress, G. R., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2014). *Multimodal teaching and learning: The rhetorics of the science classroom*. Bloomsbury.
- Lemke, J. L. (1998). *Multiplying meaning: Visual and verbal semiotics in scientific text*. Routledge.
- Li, M. (2020). Multimodal pedagogy in TESOL teacher education: Students' perspectives. *System*, 94. <https://doi.org/10.1016/j.system.2020.102337>
- Magnusson, P., & Godhe, A.-L. (2019). Multimodality in language education – Implications for teaching. *Designs for Learning*, 11(1), 127-137. <https://doi.org/10.16993/df1.127>
- Marchetti, L., & Cullen, P. (2015). A multimodal approach in the classroom for creative learning and teaching. *Psychological and Creative Approaches to Language Teaching*, 5(1), 39-51. <https://journals.muni.cz/casalc-review/article/view/20569/16487>
- Marsakawati, N. P. E., Mujiyanto, J., Agustien, H. I. R., & Astuti, P. (2019). An Indonesian EFL learner's use of evaluative language and gestures in a spoken persuasive presentation: A case study. *The Journal of Asia TEFL*, 16(1). <http://dx.doi.org/10.18823/asiatefl.2019.16.1.30.401>
- Meneses, A., Escobar, J.-P., & Véliz, S. (2018) The effects of multimodal texts on science reading comprehension in Chilean fifth-graders: Text scaffolding and comprehension skills. *International Journal of Science Education*, 40(18). 2226-2244. <https://doi.org/10.1080/09500693.2018.1527472>
- Metiti Group (2008). *Multimodal learning through media: What the research says*. Cisco Systems. https://curriculumredesign.org/wp-content/uploads/Multimodal_learning_through_media.pdf
- Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychological Review*, 19(3), 309-326. <https://doi.org/10.1007/s10648-007-9047-2>
- Mujiyanto, Y. (2017). The dependence of verbal passages on visual representation in meaning-making. *International Seminar PRASASTI III: Current Research in Linguistics*, 884-890. <https://doi.org/10.20961/pras.v0i0.1708>
- Nugroho, K. Y. & Sakhiyya, Z. (2022). Photovoice: Young children online English language learning, parents' voices and its implication to educational policy and provision. *Indonesian Journal of Applied Linguistics*, 11(3), 612-622. <https://doi.org/10.17509/ijal.v11i3.35083>
- Peng, J.-E. (2019). The roles of multimodal pedagogic effects and classroom environment in willingness to communicate in English. *System*, 82, 161-173. <https://doi.org/10.1016/j.system.2019.04.006>
- Prain, V., & Waldrip, B. (2006). An exploratory study of teachers' and students' use of multi-modal representations of concepts in primary science. *International Journal of Science Education*, 28(15), 1843-1866. <https://doi.org/10.1080/09500690600718294>
- Picciano, A. G. (2009). Blending with purpose: The multimodal model. *Journal of Asynchronous Learning Networks*, 13(1), 7-18. <https://doi.org/10.24059/OLJ.V13I1.1673>
- Rance-Roney, J. A. (2010). Reconceptualizing interactional groups: Grouping schemes for maximizing language learning. *English Teaching Forum*, 48(1), 20-26. <https://eric.ed.gov/?id=EJ914885>
- Ross, J., Curwood, J. S., & Bell, A. (2020). A multimodal assessment framework for higher education. *E-Learning and Digital Media*, 17(4), 290-306. <https://doi.org/10.1177/2042753020927201>
- Roth, W.-M. (1994). Thinking with hands, eyes, and signs: Multimodal science talk in a grade 6/7 unit on simple machines. *Interactive Learning Environments*, 4(2), 170-187. <https://doi.org/10.1080/1049482940040204>
- Salam, N., Ubaidillah, M. F., & Putri, A. N. (2021). Indonesian ESP students' willingness to communicate in English: Focusing on situated factors. *The Journal of ASIA TEFL*, 18(1), 336-344. <http://dx.doi.org/10.18823/asiatefl.2021.18.1.25.336>
- Shak, P., Lee, C. S., & Stephen, J. (2016). Pronunciation problems: A case study on English pronunciation errors of low proficient students. *International Journal of Language Education and Applied Linguistics (IJLEAL)*, 4, 25-35. <https://doi.org/10.15282/IJLEAL.V4.483>
- Shin, D.-s., Cimasko, T., & Yi, Y. (2020). Development of metalanguage for multimodal composing: A case study of an L2 writer's design of multimedia texts. *Journal of Second Language Writing*, 47. <https://doi.org/10.1016/j.jslw.2020.100714>

- Song, J. (2018). A case study on multimodality teaching in a college English listening and speaking course. *International Journal on Studies in English Language and Literature*, 6(3), 1-4. <https://doi.org/10.20431/2347-3134.0603001>
- The New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60-93. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Tyas, G. A., & Widhiyanto, W. (2020). A study on interpersonal meaning negotiated teacher-students classroom interaction in vocational high school. *ELT Forum: Journal of English Language Teaching*, 9(2), 193-202. <https://doi.org/10.15294/elt.v9i2.37221>
- van Leeuwen, T. (1999). *Speech, music, sound*. Bloomsbury.
- van Leeuwen, T. (2005). *Introducing social semiotics: An introductory textbook*. Routledge.
- Wang, H. (2018). *A case study of English language learners' multimodal compositions and identity representations* [Unpublished doctoral dissertations], Georgia State University. <https://doi.org/10.57709/12049324>
- Yin, R.K. (2009). *Case study research: Design and method* (4th ed.). Sage.

Appendix

Semi-structured interview

Multimodality in an online speaking class

No.	Questions	Answer
1	You have recently completed a video speaking assignment. Is this the first time that you have created a digital video?	
2	When you were given the assignment about the creation of the video, can you tell me your first reaction towards such an assignment?	
3	Did you understand what was expected from you?	
4	What are the moments you enjoyed the most while doing the video speaking?	
5	Tell us a bit about any difficult /frustrating moments (if any) you encountered during the process of this video assignment.	
6	Did you have any challenges or difficulties working on the multimodality element?	
7	What mode(s) did you use the most frequently and why?	
8	Why did you choose to use (or not use) linguistics elements in your multimodal composition?	
9	Why did you choose to use (or not use) visual elements in your multimodal composition?	
10	Why did you choose to use (or not use) gestural elements in your multimodal composition?	
11	Tell us about your multimodal composing process in your speaking project.	
12	Can multimodal aspects help you with your speaking in the classroom? If so, how? If not, why not?	

Adapted from Wang (2018)