p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

ENGLISH REVIEW Journal of English Education
Published by English Language Study Program, Faculty of Teachers Training & Education, Universitas Kuningan, Indonesia in collaboration with AISEE (The Association of Indonesian Scholars of English Education)
AISEE

DEVELOPING MULTIMODAL PRAGMATIC COMPETENCE IN SYNCHRONOUS ONLINE EFL LEARNING: A CONVERSATION ANALYSIS STUDY

Fridolini Fridolini[1]
[1]*English Language and Culture Study Program, Universitas Darma Persada, Jakarta, Indonesia*

Corresponding author's email: fridoliniunsada77@gmail.com

APA Citation: Fridolini, F. (2026). Developing multimodal pragmatic competence in synchronous online EFL learning: A conversation analysis study. *English Review: Journal of English Education, 14*(2), 601-620. https://doi.org/10.25134/englishreview.v14i2.190

Received: 27-01-2026 Accepted: 25-03-2026 Published: 10-06-2026

Abstract: This study investigates the development of multimodal pragmatic competence among 156 advanced Indonesian EFL learners during eight weeks of Zoom-mediated instruction. Drawing on the CA-for-SLA framework, data were collected from video-recorded interactions, chat logs, and stimulated recall interviews. A corpus of 87 request sequences and 64 disagreement sequences was analyzed to trace developmental changes in learners' multimodal pragmatic performance. The findings reveal three developmental phases. During Weeks 1–2, learners relied predominantly on spoken language, with 82% of face-threatening acts performed verbally. In Weeks 3–5, multimodal coordination emerged as verbal-only interactions declined to 46–52%, accompanied by increased use of gestures and chat-based communication. By Weeks 6–8, learners demonstrated advanced multimodal integration, with only 18–23% of interactions remaining verbal-only. At this stage, speech, gesture, chat, and screen-sharing functions were strategically combined to perform requests and disagreements. Platform affordances, particularly private chat and screen sharing, were frequently employed to manage face-threatening situations, especially in breakout-room interactions. These findings demonstrate that multimodal pragmatic competence develops progressively through sustained video-mediated interaction, extending usage-based perspectives on second language development. Pedagogically, the study highlights the value of multimodal awareness-raising and systematic use of digital platform affordances in online language learning.
Keywords: *multimodal pragmatic competence; CA-for-SLA; synchronous online learning; developmental phases; Indonesian EFL learners*.

INTRODUCTION

The rapid growth of synchronous online learning has transformed language education worldwide. Video-conferencing platforms such as Zoom and Google Meet have become integral components of contemporary English as a Foreign Language (EFL) instruction, enabling learners to engage in real-time communication across diverse educational contexts. Unlike traditional face-to-face classrooms, these environments provide multiple communication channels, including spoken interaction, gestures, facial expressions, written chat, emojis, and screen-sharing functions. As a result, successful participation increasingly depends on learners' ability to coordinate multiple semiotic resources simultaneously.

Among the various dimensions of communicative competence, pragmatic competence is particularly important because it enables learners to perform socially appropriate actions and manage interpersonal relationships effectively. In synchronous online interaction, however, pragmatic performance extends beyond spoken language alone. Learners must integrate verbal, embodied, and platform-mediated resources when performing face-threatening speech acts such as requests and disagreements. This highlights the importance of multimodal pragmatic competence, which involves the strategic coordination of multiple communicative resources to accomplish interactional goals.

Although previous studies have examined online language learning, computer-mediated

communication, and multimodal interaction, limited research has explored how learners develop multimodal pragmatic competence over time. Existing studies have primarily focused on interactional patterns or technological affordances rather than the developmental processes through which learners learn to coordinate speech, gesture, gaze, written chat, and platform-specific features during authentic classroom interaction. Consequently, little is known about how multimodal pragmatic competence emerges and develops in synchronous online learning environments.

This gap is particularly relevant in the Indonesian EFL context, where opportunities for authentic English interaction outside the classroom remain relatively limited. Moreover, Indonesian communicative practices are strongly influenced by cultural values emphasizing politeness, respect, and social harmony, which may shape how learners perform face-threatening speech acts in online settings. Despite the growing use of synchronous online learning in Indonesian higher education, empirical investigations of multimodal pragmatic development remain scarce.

To address these gaps, this study investigates how advanced Indonesian EFL learners develop multimodal pragmatic competence during eight weeks of Zoom-mediated classroom interaction. Drawing on Conversation Analysis (CA), the study examines how learners progressively coordinate verbal, embodied, and platform-mediated resources when performing requests and disagreements. The findings are expected to contribute to current understandings of pragmatic development, multimodal interaction, and language learning in technology-mediated environments.

Accordingly, the study addresses the following research questions: (1) How do advanced Indonesian EFL learners develop multimodal pragmatic competence when performing face-threatening speech acts in synchronous online learning environments? (2) How do learners coordinate verbal, embodied, and platform-mediated resources to construct and negotiate pragmatic meaning during video-mediated interaction?

In this study, multimodal pragmatic competence is defined as the ability to strategically coordinate verbal, embodied, and platform-mediated semiotic resources to accomplish socially appropriate pragmatic actions in video-mediated interaction

The increasing digitalization of communication has transformed how language is used, learned, and negotiated across educational settings. Traditional views of communication often positioned spoken and written language as the primary channels of meaning-making. However, contemporary multimodal theories argue that communication is inherently multimodal, involving the coordinated use of multiple semiotic resources, including speech, gesture, gaze, posture, facial expression, images, and digital artifacts. Within computer-assisted language learning (CALL), for instance, semiotic approaches recognize that speech and writing as the main modes of communication are no longer sufficient for contemporary and increasingly digital communication, as multimodal texts are more complex and require more nuanced meaning-making processes (Satar, 2025). Meaning is therefore not produced through language alone but emerges from the interaction among various communicative modes.

This perspective has significant implications for second language (L2) learning. In addition to acquiring linguistic knowledge, learners must develop the ability to interpret and coordinate multiple communicative resources in socially appropriate ways. Consequently, researchers have begun to conceptualize pragmatic competence as a multimodal phenomenon rather than a purely linguistic ability (Taguchi & Youn, 2022).

Recent evidence suggests that pragmatic competence develops most effectively when learners receive systematic instructional support alongside opportunities for meaningful interaction. Ren et al. (2023) demonstrated that pragmatics instruction exerts a strong positive influence on learners' pragmatic development across a wide range of educational settings.

It enables language users to perform speech acts, manage interpersonal relationships, negotiate meaning, and achieve communicative goals effectively. Traditionally, pragmatic competence has been examined through verbal realizations of speech acts such as requests, refusals, apologies, and disagreements. More recent studies, however, emphasize that pragmatic action is frequently accomplished through combinations of verbal and non-verbal resources (Kang et al., 2019).

The notion of multimodal pragmatic competence extends conventional understandings of pragmatics by recognizing that successful communication often depends on the coordinated deployment of speech, gesture, gaze, facial expression, and technological resources. For example, a disagreement may be softened not only

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

through linguistic hedges but also through smiles, nods, gaze behavior, or supportive chat messages. Similarly, requests may be mitigated through embodied conduct that signals politeness, deference, or solidarity.

Recent studies have highlighted the increasing importance of multimodal resources in technology-mediated language learning. Research has shown that learners construct meaning through the coordinated use of speech, gesture, gaze, and digital communication tools during online interaction (Balaman & Sert, 2023; Benson et al., 2023). In video-mediated environments, multimodal competence extends beyond linguistic knowledge to include the strategic deployment of platform-specific affordances that facilitate interactional goals and relationship management.

Research has increasingly demonstrated that multimodal resources play important roles in interpersonal communication. Gesture and gaze contribute to turn management and recipient design, while facial expressions and body orientation influence how utterances are interpreted by interlocutors (Goodwin, 2018). In digital communication environments, additional resources such as emojis, reactions, and written chat messages further expand the repertoire available to language users. Therefore, understanding pragmatic competence in contemporary communication requires attention to how learners coordinate multiple semiotic resources simultaneously.

Despite growing recognition of multimodal communication, research examining the developmental trajectory of multimodal pragmatic competence remains limited. Existing studies have largely focused on the identification of multimodal resources rather than the processes through which learners gradually acquire the ability to use them strategically. This limitation is particularly evident in online language learning contexts, where multimodal interaction constitutes a central component of communicative practice.

The present study is grounded in Conversation Analysis for Second Language Acquisition (CA-for-SLA), an approach that combines the methodological principles of Conversation Analysis (CA) with concerns related to language learning and development (Kasper & Wagner, 2011; Markee, 2015).

Recent developments in CA-for-SLA research have increasingly focused on how second language development emerges through participation in digitally mediated interaction. Studies have demonstrated that learners' interactional repertoires evolve through repeated engagement with multimodal and technology-supported communicative practices (Hauser, 2022; Jakonen et al., 2024).

Conversation Analysis originated within ethnomethodology and seeks to understand how participants organize social interaction through sequentially ordered practices (Sacks et al., 1974). Rather than treating communication as the transmission of information, CA views interaction as a collaborative accomplishment in which participants continuously construct and negotiate meaning. Core analytical concerns include turn-taking organization, sequence structure, repair mechanisms, recipient design, and the accomplishment of social actions.

CA-for-SLA extends these principles to language learning contexts by viewing acquisition as a socially situated process emerging through participation in interaction. From this perspective, learning is observable in the changing ways learners participate in communicative activities rather than solely through gains in linguistic knowledge. Development becomes evident when learners demonstrate increasing abilities to manage interactional contingencies, accomplish social actions, and coordinate available resources effectively (Pekarek Doehler et al., 2018).

One of the major strengths of CA-for-SLA is its ability to capture developmental processes as they occur in naturally occurring interaction. Rather than relying on experimental tasks or elicited responses, researchers examine authentic communicative practices and trace how learners' interactional repertoires evolve over time. This approach is particularly suitable for investigating pragmatic development because pragmatic competence is fundamentally interactional in nature.

Recent CA-for-SLA research has increasingly acknowledged the multimodal dimensions of interaction. Studies have demonstrated that learning is not restricted to verbal language but involves the coordinated use of embodied actions, material artifacts, and technological resources (Balaman & Sert, 2023). These developments suggest that analyses of pragmatic development must move beyond linguistic forms and examine how learners mobilize multiple semiotic resources during interaction.

The concept of interactional competence provides an important theoretical lens for understanding language use in online learning environments. Originally proposed as an alternative to individualistic models of

communicative competence, interactional competence emphasizes participants' ability to jointly construct meaningful social interaction (Hall, 2018).

Interactional competence encompasses a range of abilities, including managing turn-taking, interpreting contextual cues, responding appropriately to interlocutors, and accomplishing social actions within particular interactional settings. Importantly, interactional competence is viewed not as an individual possession but as a collaborative achievement emerging through interaction itself (Young, 2011).

As language learning increasingly occurs in digital environments, scholars have extended the concept of interactional competence to technology-mediated contexts. Online communication requires learners to engage with unique participation frameworks that differ substantially from face-to-face interaction. Participants must manage delayed responses, technological disruptions, screen-mediated visibility, and multiple communication channels simultaneously (Moorhouse et al., 2021).

In video-mediated interaction, learners frequently navigate between spoken language, visual representations, written text, and platform-specific affordances. Consequently, successful participation requires not only linguistic competence but also the ability to coordinate diverse semiotic resources strategically. These demands have led researchers to conceptualize interactional competence in online environments as inherently multimodal and technologically situated.

Recent studies indicate that learners gradually develop new participation practices adapted to online interaction. These include strategic use of chat functions, visual displays of engagement through gaze and gestures, and collaborative management of digital resources. Such findings suggest that interactional competence develops differently across communication environments and must be understood within the specific contexts in which interaction occurs.

The concept of affordances provides a useful framework for understanding how technological environments shapes communicative behavior. Originally proposed by Gibson (1979), affordances refer to action possibilities emerging from the relationship between individuals and their environment. Within educational technology research, affordances are understood as opportunities and constraints that influence how users engage with technological tools.

Video-conferencing platforms such as Zoom provide a range of affordances that fundamentally shape interaction. These include video displays, chat functions, screen sharing, breakout rooms, reaction icons, annotation tools, and private messaging systems. Each affordance introduces new possibilities for communication while simultaneously altering conventional interactional practices (Satar & Wigham, 2023; Balaman, 2023).

Research on video-mediated communication demonstrates that platform affordances influence turn-taking, recipient design, participation structures, and pragmatic action. For example, chat functions allow participants to contribute simultaneously without interrupting ongoing talk. Screen sharing enables collaborative examination of documents and visual materials. Breakout rooms create smaller interactional spaces that often encourage increased participation and experimentation.

Importantly, affordances are not neutral technological features. They become meaningful through users' interactional practices. Participants learn to exploit these affordances strategically to accomplish communicative goals, manage interpersonal relationships, and negotiate social actions. Therefore, platform affordances should be viewed as active mediators of interaction rather than passive technological tools.

The increasing integration of video-conferencing technologies into language education highlights the need to understand how learners appropriate these affordances during language development. Investigating such processes contributes to broader understandings of communication, learning, and participation in digitally mediated environments.

Pragmatic behavior is deeply influenced by cultural norms and sociocultural expectations. In Indonesia, interpersonal communication is strongly shaped by values emphasizing social harmony (rukun), respect (hormat), and avoidance of direct confrontation (Goddard, 2018). These values frequently influence how speakers perform face-threatening speech acts such as requests, refusals, and disagreements.

Previous studies have shown that Indonesian EFL learners often employ indirect strategies, hedging devices, and mitigation techniques when interacting in English (Dwiana & Syahri, 2023). Such practices reflect learners' efforts to maintain positive interpersonal relationships while navigating culturally sensitive communicative situations.

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

Given the increasing prevalence of online education in Indonesia, understanding how learners develop multimodal pragmatic competence in video-mediated settings represents an important area of inquiry. Such research can contribute to both theoretical understandings of pragmatic development and practical efforts to support effective online language learning.

METHOD

This study employs a qualitative dominant longitudinal CA-for-SLA design. While the primary analysis is qualitative and interactional in nature, descriptive quantitative summaries are provided to illustrate developmental trends across the dataset.

The study received ethical approval from the participating institutions prior to data collection. All participants were informed about the objectives of the study, data collection procedures, confidentiality measures, and their rights as participants before providing written informed consent. Participation was entirely voluntary, and students who chose not to participate continued their regular course activities without penalty or academic disadvantage.

The researchers did not serve as instructors in the participating courses. Their role was limited to observing, recording, transcribing, and analysing classroom interaction. To minimize potential pressure and ensure voluntary participation, recruitment and consent procedures were conducted independently of course assessment activities. Participants were informed that their decision to participate, decline participation, or withdraw from the study at any stage would not affect their grades, attendance records, or academic standing. To protect participants' privacy, all names appearing in transcripts and extracts were replaced with pseudonyms, and identifying information was removed during transcription and analysis.

156 advanced EFL students studying English communication classes at four universities—Universitas Indonesia in Jakarta (n = 42), Universitas Padjadjaran in West Java (n = 38), Universitas Trisakti in Jakarta (n = 36), and Universitas Pasundan in West Java (n = 40)—were involved in the study. The range of ages of the participants was 18 to 26 years (M=20.8, SD=1.9), with 62 male and 94 female students. Based on institutional placement assessments given at the start of the semester, all participants achieved either B2 (upper-intermediate, n=89) or C1 (advanced, n=67) levels of competence on the Common European Framework of Reference (CEFR).

Course lecturers who delivered the research during the first week of the term had a number of people on board. Those who chose not to participate in the study were free to continue with their regular coursework without any consequences. Each participant provided informed consent, and each participating institution's research ethics boards granted ethical approval. Although they were informed that their participation would help with comprehending language learning processes in online environments, participants weren't given any money. The four universities had been selected to represent variations in program variety, institutional size, and regional setting. They were also places where teachers were prepared to work together, providing direct access to advanced students.

The choice of advanced-level learners was intentional, as pragmatic competence research is enhanced by participants with adequate linguistic resources to prioritize pragmatic appropriateness over fundamental grammatical accuracy (Derakhshan & Shakki, 2023). Also, advanced learners are more likely to use the complicated speech acts (like requests and disagreements) that this study is studying. Participant demographics and characteristics are presented below.

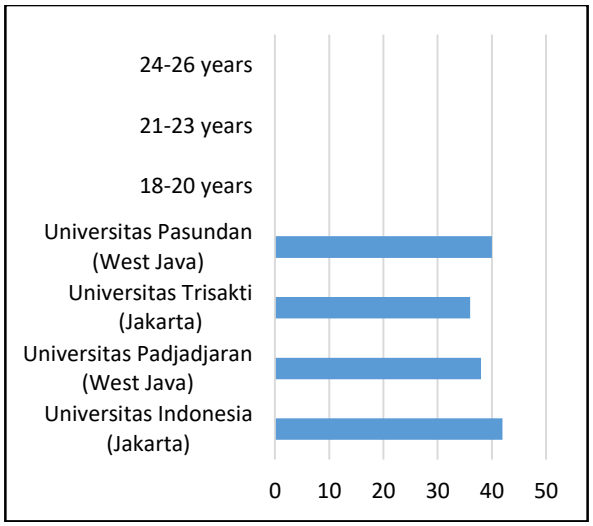


Figure 1. *Participant demographics (university & age)*

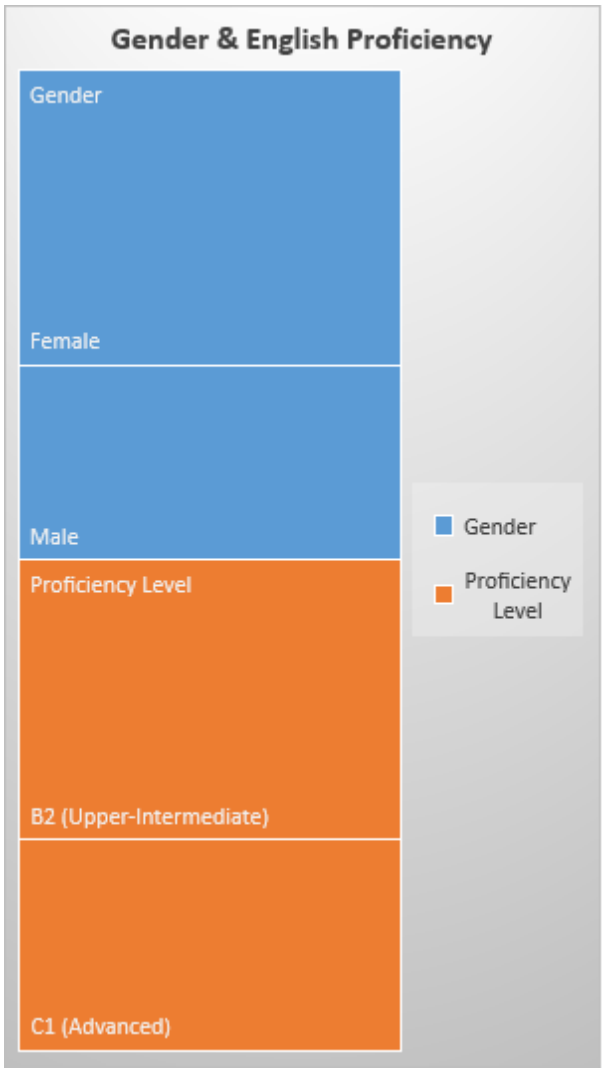


Figure 2. *Gender & English proficiency*

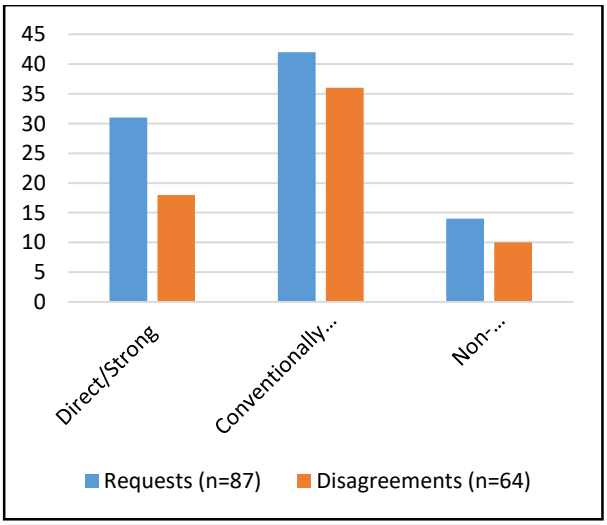


Figure 3 *Pragmatic strategies in speech acts*

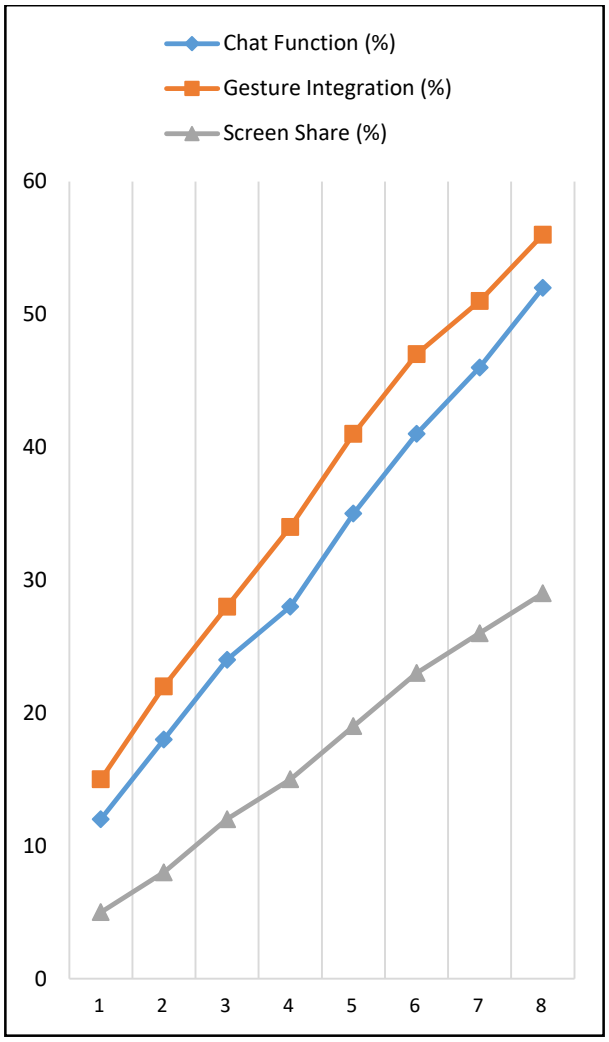


Figure 4. *Platform affordance utilization*

Data collection occurred over eight consecutive weeks during regularly scheduled English communication courses. Data were collected over an eight-week period during the middle phase of the academic semester (Weeks 5–12). This period was selected because students had already become familiar with the course structure, their classmates, and the Zoom platform, allowing interactional practices to emerge more naturally.

The courses met twice weekly for 90-minute sessions conducted via Zoom, totaling 16 sessions per participant. Data collection encompassed three complementary sources:

All Zoom sessions were recorded using the platform's native recording function, which captured both gallery view, displaying all participants simultaneously, and speaker view, highlighting the active speaker. The recordings also included screen-sharing activities and chat interactions, allowing comprehensive documentation of verbal, visual, and platform-mediated communication. This recording procedure enabled the analysis of individual

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

learners' multimodal conduct as well as the broader organization of classroom participation. Recent studies on multimodal online interaction similarly emphasize the importance of capturing multiple interactional layers in video-mediated communication to examine how participants coordinate embodied and technological resources during interaction (Balaman, 2023; Wigham et al., 2020). In total, 187 hours of video data were collected across all participants and classroom sessions.

Recording settings were standardized across all sessions: 1080p video quality, separate audio files for each participant (when technically feasible), and automatic recording of chat logs. Participants were reminded at the beginning of each session that recording was in progress, and they retained the right to request that specific segments be excluded from analysis.

The Zoom chat function served as a significant semiotic resource within participants' multimodal pragmatic practices. Chat records, including timestamps, sender information, and recipient visibility (public or private), were automatically preserved by the platform and exported for analysis following each session. The analysis focused on the interactional functions of chat messages, their coordination with spoken discourse, and their role in the management of face-threatening speech acts. Chat was examined as an integral component of multimodal interaction rather than as a supplementary communication channel.

Chat messages were classified as (a) parallel to verbal interaction (e.g., typing agreement while another person speaks), (b) sequential to verbal interaction (e.g., following up on a verbal request with written clarification), or (c) independent of verbal interaction (e.g., making a written request without verbal accompaniment). This classification facilitated the examination of how textual resources were incorporated into the larger multimodal framework.

Data gathering took place over a period of eight weeks in a row, starting from the third week of the semester when students had finished their orientation for the course and had acquainted themselves with the online learning system. This timing for data collection provided students the opportunity to interact with the course material under consistent teaching conditions, while also allowing for the monitoring of progress throughout the semester. Classes were held twice a week for 90 minutes each, using Zoom, amounting to 16 sessions for each individual. The data collection included three supporting sources.

Stimulated recall interviews were carried out with a carefully chosen group of 32 individuals (around 20% of the overall sample) to understand the viewpoints of participants regarding their practical decisions and the use of various communication This method has been widely employed in EFL teaching research to investigate learners' cognitive processes and strategy deployment during classroom interaction (Zainil & Arsyad, 2021).

This sample was chosen to reflect diversity in gender, university connections, and skill levels, effectively documenting a wide range of interactional encounters and awareness of pragmatics. Every stimulated recall interview was conducted separately through Zoom within two days of the recorded interaction sessions, lasting about 30 to 45 minutes each. This schedule aimed to maintain the accuracy of participants' memories while providing enough opportunity for thoughtful consideration. To prepare and enhance recall accuracy, brief video excerpts (one to two minutes in length) were given to participants approximately 24 hours before the interview. They were informed that the excerpts would serve as prompts for reflection on their communication choices, not for evaluative purposes. During the interviews, participants viewed the same excerpts again with the researcher, focusing particularly on segments involving potentially face-threatening speech acts. These segments served as prompts to elicit participants' reflections on their pragmatic intentions, their selection and coordination of verbal and nonverbal resources, and their awareness of multimodal strategies used to manage interpersonal meaning.

. To facilitate deeper reflection, participants were guided by semi-structured interview questions designed to explore the relationship between pragmatic intention, multimodal resource use, and interactional context. "What were you trying to do at this moment? "Why did you decide to use [gesture/chat/specific phrasing]?"How did you feel about the other person's answer?"

The interviews were conducted in Indonesian to allow participants to articulate nuanced reflections without linguistic limitations, and were later transcribed and translated into English by bilingual research assistants. Back-translation procedures (Brislin, 1970) were used to check the accuracy of the translation.

Data analysis followed established CA-for-SLA procedures adapted for multimodal video-mediated interaction (Sert, 2015; Mondada, 2018). Analysis proceeded through four iterative phases:

The first analytical phase involved systematic identification of face-threatening speech acts within the video corpus. Following established CA work on requests (Willcox-Ficzere, 2025) and disagreements (Buker et al., 2025), the analysis identified sequences where participants: (a) solicited action from interlocutors (requests), or (b) expressed opposition to prior assessments or proposals (disagreements).and the preference organization of agreement/disagreement sequences as a site for face management (Fauziah et al., 2024), the analysis identified sequences where participants: (a) solicited action from interlocutors (requests), or (b) expressed opposition to prior assessments or proposals (disagreements).Request sequences were defined as interactional moments where a participant sought to have an interlocutor perform an action, ranging from direct imperatives ("*Can you share your screen?") to highly indirect formulations ("I'm wondering if anyone has thoughts on this issue"). Disagreement sequences were identified where participants expressed epistemic or evaluative misalignment with prior turns, including both strong disagreements ("I don't think that's accurate") and mitigated disagreements ("That's interesting, but maybe we could also consider…").*

Through systematic viewing of all recorded sessions, that identified 87 request sequences and 64 disagreement sequences meeting our criteria. These sequences were extracted with sufficient surrounding context (typically 30 seconds before and after the target action) to enable analysis of sequential positioning and recipient responses.

Target sequences were transcribed using contemporary Jeffersonian transcription conventions for verbal interaction (Clift et al., 2024) together with multimodal transcription conventions developed by Mondada (2019) for embodied conduct. The transcripts captured: (1) Verbal content with prosodic features (emphasis, intonation, pauses). (2) Temporal alignment of talk and embodied action. (3) Gesture type, timing, and relationship to verbal content. (4) Gaze direction and facial expressions (when visible). (5) Chat messages with timestamps. (6) Platform-specific actions (screen sharing, reactions, hand-raising)

Transcription was conducted using ELAN (EUDICO Linguistic Annotator) software, which enables frame-by-frame analysis and precise temporal alignment of multiple semiotic channels (Hanlon & Keeffe, 2024). Two trained research assistants independently transcribed 20% of sequences, achieving 89% inter-coder agreement on multimodal feature identification. Discrepancies were resolved through discussion and consultation with the research team.

The core analytical phase involved detailed examination of how participants deployed and coordinated multimodal resources to accomplish face-threatening speech acts. Analysis focused on: (1) Sequential positioning: Where in the interaction did the speech act occur, and how did prior turns shape its design? (2) Turn construction: What verbal, prosodic, and grammatical resources were employed? (3) Embodied conduct: How did gesture, facial expression, and body orientation contribute to pragmatic action? (4) Textual resources: How did chat messages function in relation to verbal turns? (5) Recipient response: How did interlocutors orient to the speech act, and what did their responses reveal about its pragmatic force? (6) Repair and modification: Did participants modify or repair their initial attempts, and what multimodal resources were deployed in these adjustments?

This phase produced comprehensive case analyses of exemplary instances illustrating various multimodal configurations and developmental stages. Regular data sessions with the research team made sure that the analysis was thorough and consistent

The final analytical phase involved identifying patterns across individual cases and tracking developmental changes over the eight-week period. Sequences were organized chronologically and compared across weeks to identify shifts in: (1) Frequency and types of multimodal resource deployment. (2) Sophistication of multimodal coordination. (3) Strategic use of platform affordances. (4) Recipient uptake and interactional success.

Quantitative patterns (e.g., percentage of sequences employing multiple modalities) were calculated to complement qualitative microanalysis. Stimulated recall interview data were integrated to triangulate observational findings with participants' reported intentions and awareness.

Multiple measures ensured analytical validity and reliability. First, investigator triangulation was employed, involving three researchers who independently analyzed critical sequences and engaged in discussions to reach interpretive consensus (Denzin, 2012). Second, member checking was conducted through follow-up interviews, during which participants reviewed the analyses and provided feedback on the interpretations, following contemporary approaches to participant validation using video-

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

based and asynchronous methods (Schafer & Phillippi, 2025). Third, detailed records of analytical decisions and supporting evidence were maintained to ensure transparency and confirmability (Lincoln & Guba, 1985).

In addition, deviant case analysis was employed to identify interactional episodes that did not conform to emerging analytical patterns. These cases were examined in detail to challenge preliminary interpretations and refine analytical claims. By systematically considering both typical and atypical instances, the analysis sought to provide a comprehensive account of the observed interactional phenomena and enhance the credibility and robustness of the findings.

RESULTS AND DISCUSSION

Three distinctive stages of development in the multimodal pragmatic competence of Indonesian EFL learners are revealed by analyzing 87 request sequences and 64 disagreement sequences over the course of eight weeks of synchronous online interactions. Growing strategic knowledge of platforms features, increasingly complex coordination of various semiotic fields, and progressive integration of verbal, gestural, and platform-specific textual resources are characteristics of these phases. The present findings are organized based on stage of development, with an emphasis on the distinctions between whole-class and breakout room settings.

Phase 1 (weeks 1-2): Verbal-dominant strategies

In the first phase of participants' online interaction, they rarely employed gestural or textual resources. During this time, 82% of face-threatening speech acts were done mostly with words, with only a little bit of multimodal elaboration.

Requests in phase 1

Early requests typically employed conventionally indirect formulations, reflecting participants' awareness of face-threat but limited deployment of multimodal mitigation strategies. Extract 1 illustrates a typical Phase 1 request sequence from a Week 1 breakout room discussion:

Extract 1: "Can you explain more?" (Week 1, Breakout Room)

01 Rina: so: uh (0.8) about the first question
02 (1.2)
03 Rina: can you explain more about ↑that?
04 (0.6)
05 Dian: oh okay (.) um: (1.4) the first one is about
06 the environmental issue

In this part, Rina makes an indirect request ("can you explain more") without any gestures, a shift in gaze toward the camera, or any text to back it up. The 1.2-second pause at line 02 could mean that Rina is looking for the right words or is unsure of what to say next. However, she doesn't use any multimodal resources to fill this gap or make her request stronger. Her rising intonation on "that" (line 03) is the main pragmatic mitigation because it makes the request seem tentative.

Stimulated recall interviews revealed that participants in Phase 1 were primarily focused on verbal formulation and were not yet consciously attending to multimodal possibilities. As Rina explained:

> "I was just thinking about the right words in English. I didn't think about using chat or gestures. I was worried about my grammar."

This quote exemplifies the verbal-dominant orientation characteristic of Phase 1, where cognitive resources were devoted to linguistic encoding rather than multimodal strategy deployment.

Disagreements in phase 1

Early disagreements similarly relied on verbal mitigation strategies, with minimal gestural or textual elaboration. Indonesian participants' cultural orientation toward maintaining social harmony (rukun) was evident in their preference for indirect disagreement formulations, yet these were not yet enhanced through multimodal resources.

Extract 2: "I think maybe different" (Week 2, Whole-Class Discussion)

01 Budi: I think the main problem is economic
02 (0.4)
03 Sari: mm: (.) I think maybe: (0.6) different?
04 (1.8)
05 Sari: because uh (.) the article mentions about
06 social problem

Sari's disagreement (line 03) employs multiple verbal mitigation devices: the epistemic hedge "I think," the modal "maybe," and the rising intonation transforming the assertion into a question-like format. However, her embodied conduct during this turn consisted only of gazing at her own video feed, with no visible gesture or facial expression variation. She did not use the chat function to soften or elaborate her disagreement.\

The 1.8-second gap at line 04 suggests Budi's misreferred response, yet Sari does not deploy additional multimodal resources during this extended silence to manage the interactional delicacy. Instead, she proceeds with verbal elaboration (lines 05-06), providing an account for her disagreement.

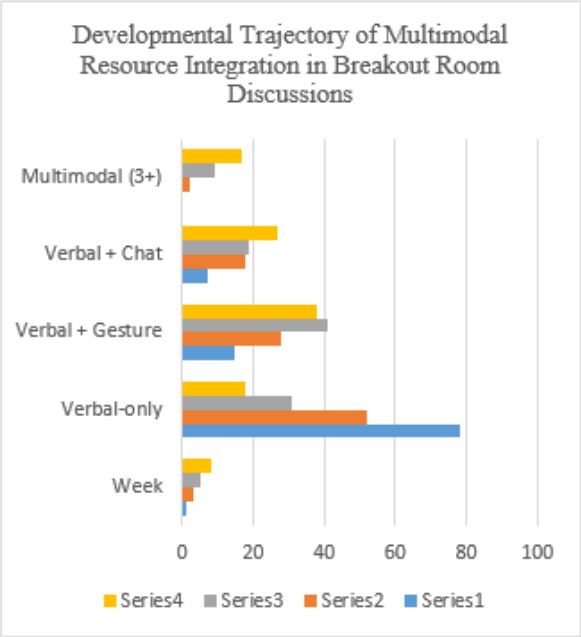


Figure 5. *Developmental trajectory of multimodal resource integration in breakout room discussion*

Phase 2 (Weeks 3-5): Emerging multimodal coordination

By Week 3, participants began integrating gestural and textual resources more consistently into their pragmatic action. This phase was characterized by experimentation with different multimodal configurations and growing awareness of platform affordances as pragmatic tools.

Gesture integration in requests

Participants increasingly employed hand gestures to emphasize requests, project turn completion, or mitigate face-threat. Extract 3 demonstrates this emerging gestural integration:

Extract 3: "Could you share your idea?" (week 4, breakout room)

01 Andi: so: (0.4) about this topic
02 *opens palm toward camera*
03 Andi: could you sha:re your idea?
04 *maintains open palm, slight forward lean*
05(0.8)
0 Fitri: oh yes uh I think (.) the main point is.

Andi's open-palm gesture (line 02) precedes and extends through his verbal request (line 03), creating a multimodal ensemble that softens the face-threat. The gesture, combined with the modal "could" and the possessive "your idea" (acknowledging Fitri's epistemic authority), constructs the request as an invitation rather than an imposition. Andi maintains the gesture through the 0.8-second gap (line 05), displaying continued orientation toward Fitri's response.

Stimulated recall revealed Andi's conscious deployment of this gesture: "I used my hand to show I'm really interested in her opinion. I think it's more polite than just asking with words."

This metacognitive awareness marks a significant shift from Phase 1, where participants reported not consciously considering multimodal options.

Chat function as pragmatic resource

The chat function emerged as a valuable pragmatic resource during Phase 2, particularly for providing simultaneous support or mitigation for verbal actions. Extract 4 illustrates this development:

Extract 4: "I'm not sure I agree" + chat message (week 5, whole-class discussion)

01 Rudi: so we should focus on the economic aspect
02 (0.6)
03 Nina: um::: (0.8) I'm not sure I agree
04 *simultaneously types in chat*
05 [Chat: "But maybe we can consider both economic and social? ☺"]
06 (0.4)
07 Nina: because (.) the question asks about social impact.

Nina's disagreement is accomplished through coordinated verbal and textual action. Her verbal turn (line 03) expresses disagreement using the epistemic hedge "I'm not sure," while simultaneously typing a more elaborated alternative proposal in the chat (line 05). The chat message includes the mitigating "maybe," the inclusive "we," and a smiling emoji all serving to soften the face-threat of disagreement.

This multimodal strategy allows Nina to accomplish two pragmatic goals simultaneously: expressing disagreement verbally while offering a face-saving alternative textually. The emoji adds an affective dimension not easily conveyed through verbal means in video-mediated

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

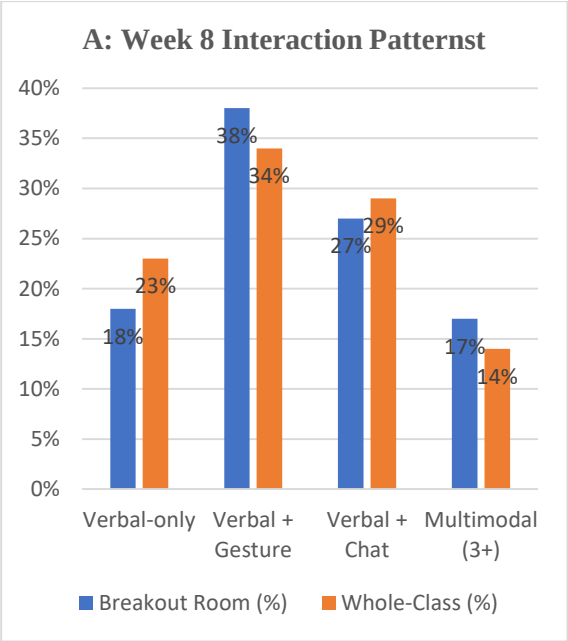


Figure 6. *Interaction patterns*

Interaction, where facial expressions may be obscured by camera quality or screen size.

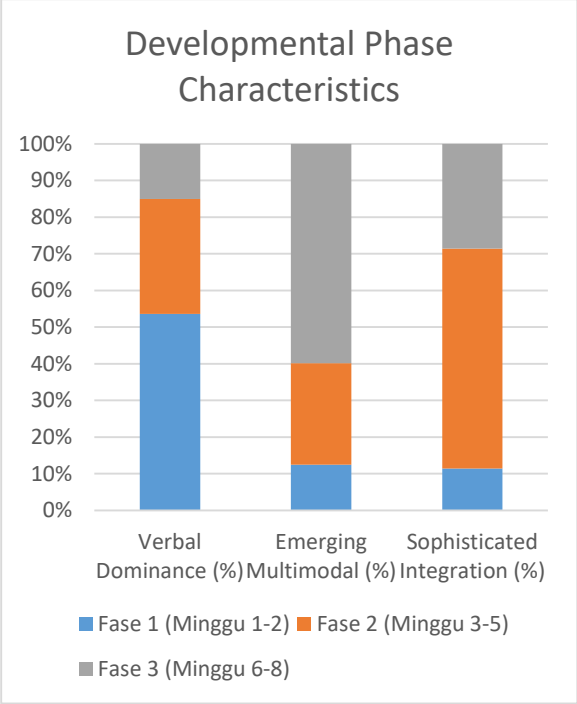


Figure 7. *Comparison of interaction contexts and developmental phases*

Phase 3 (weeks 6-8): sophisticated multimodal integration

By Week 6, participants demonstrated sophisticated coordination of verbal, gestural, and textual resources, with strategic adaptation to interactional context and recipient design. This phase was characterized by fluid movement between modalities and conscious exploitation of platform affordances.

Complex multimodal requests

Late-phase requests often involved intricate coordination of three or more semiotic resources. Extract 5 exemplifies this sophisticated integration:

Extract 5: "Would it be possible…" (week 7, breakout room)

01 Dewi: okay so:: (0.4) we need to organize the presentation
02 *gaze toward Arif's video*
03 Dewi: would it be po:ssible *raises eyebrows, slight head tilt*
04 for you to (.) start with the introduction?
05 *types in private chat to Arif*
06 [Private chat: "You're really good at intros! 👍"]
07(0.6)
08 Arif: *nods, smiles* sure no problem
09 [Chat: "Thanks! I'll prepare it"]

Dewi's request mobilizes multiple semiotic resources in a carefully orchestrated sequence. She establishes joint attention through directed gaze (line 02), employs a highly indirect verbal formulation with the modal "would it be possible" (line 03), produces embodied mitigation through raised eyebrows and head tilt (line 03), and provides simultaneous positive facework through a private chat message complimenting Arif's ability (line 06).

This multimodal ensemble accomplishes several pragmatic functions: the verbal formulation minimizes imposition, the embodied conduct displays deference, and the private chat message provides face-enhancing support. The private nature of the chat message is strategically significant it offers encouragement without putting Arif on the spot publicly.

Arif's response is similarly multimodal, combining verbal acceptance with embodied agreement (nod, smile) and textual confirmation. The efficiency and smoothness of this exchange contrasts sharply with Phase 1 sequences, where requests often occasioned extended gaps and verbal-only responses.

Multimodal disagreement management

Late-phase disagreements demonstrated particularly sophisticated multimodal coordination, as participants managed the dual challenge of expressing epistemic misalignment while maintaining social solidarity. Extract 6 illustrates:

Extract 6: "I see it differently" (week 8, whole-class discussion)

01 Hadi: the data clearly shows the problem is economic
02 (0.4)
03 Lina: *nods slowly* mm hm::: (.) that's one way to see it
04 *begins typing in chat*
05 Lina: but I see it a bit differently
06 *open palm gesture, slight smile*
07 [Chat: "Here's another perspective - what if we consider the social factors mentioned on page 3?"]
08 Lina: if we look at page three *shares screen, highlights text*
09 there's also social factors
10 (0.6)
11 Hadi: *leans forward* oh: right (.) I didn't notice that part.

Lina's disagreement sequence exemplifies sophisticated multimodal integration. She begins with embodied agreement tokens (slow nodding, "mm hm") that display recipiency and partial alignment (line 03), then produces a verbal token agreement "that's one way to see it" before introducing her disagreement (line 05). Her open palm gesture and smile (line 06) provide embodied mitigation, while her chat message (line 07) frames the disagreement as offering "another perspective" rather than contradicting Hadi.

Most strikingly, Lina employs the screen-sharing affordance to provide evidential support for her alternative interpretation (line 08), transforming the disagreement from a subjective opinion clash into a collaborative examination of textual evidence. This strategic use of platform affordances to manage face-threat represents a significant development from earlier phases.

Hadi's response (line 11) demonstrates successful pragmatic action his embodied engagement (leaning forward) and acknowledgment token "oh: right" display genuine uptake rather than defensive resistance. The disagreement is resolved collaboratively rather than competitively, suggesting effective multimodal face management.

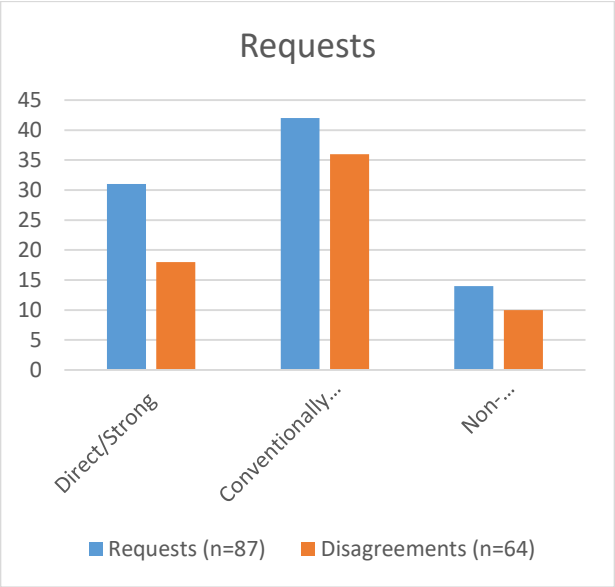


Figure 7. *Distribution of pragmatic strategies in face-threatening speech acts*

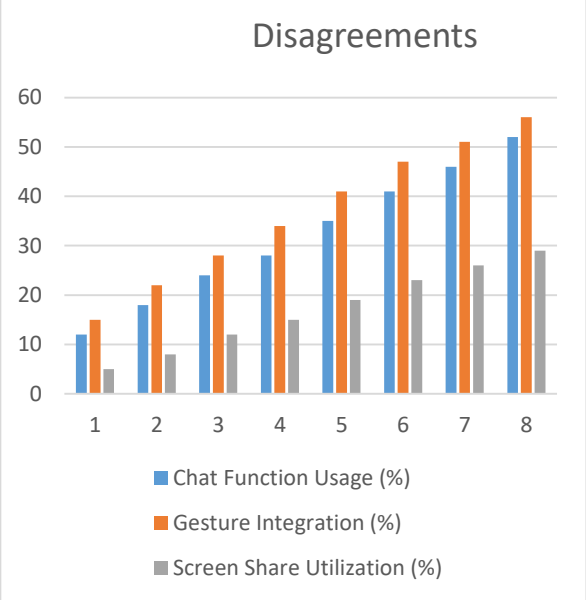


Figure 8. *Distribution of pragmatic strategies in face-threatening speech acts*

Platform affordances as mediational tools

Across all three phases, platform affordances emerged not merely as contextual features but as active mediators of pragmatic strategy deployment. Three affordances proved particularly significant:

Chat function

The chat function served multiple pragmatic purposes across developmental phases. In Phase 1, chat was used primarily for logistical coordination (sharing links, clarifying technical issues). By Phase 2, participants began using chat for simultaneous commentary, backchanneling, and

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

mitigation of face-threatening acts. In Phase 3, strategic deployment of public versus private chat messages enabled sophisticated face management, as illustrated in Extract 5.

Quantitative analysis revealed a steady increase in pragmatically-oriented chat usage: Week 1 (12% of participants), Week 3 (24%), Week 5 (35%), and Week 8 (52%). This trajectory suggests growing recognition of chat as a valuable pragmatic resource rather than merely a technical feature.

Gallery view configuration

Gallery view, which displays all participants simultaneously in a grid, shaped pragmatic action in distinctive ways. Unlike face-to-face interaction where gaze direction clearly indicates addressee, gallery view created ambiguity about recipient design. Participants developed various practices for managing this ambiguity, including: (1) Verbal addressing ("Rina, I want to ask you…") (2) Pointing gestures toward specific video tiles. (3) Private chat messages to clarify addressee. (4) Screen annotation to highlight specific speakers

By Phase 3, participants fluently employed these practices, demonstrating adaptation to the unique participation framework created by gallery view.

Screen sharing

Screen sharing emerged as a powerful resource for collaborative pragmatic action, particularly in disagreement sequences. By sharing screens displaying source texts, data visualizations, or collaborative documents, participants transformed potentially face-threatening disagreements into joint problem-solving activities. This affordance was rarely used in Phase 1 (5% of disagreement sequences) but became common by Phase 3 (29% of disagreement sequences).

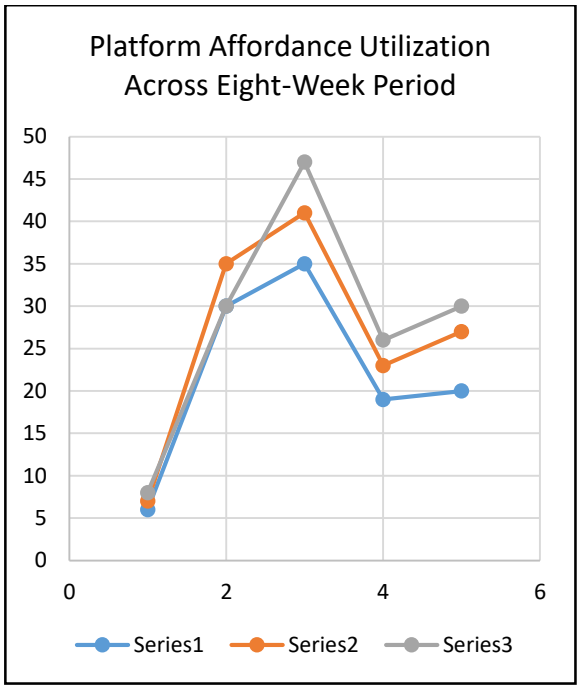


Figure 9. *Platform affordance utilization across eight-week period*

Contextual variation: Breakout rooms vs. whole-class discussions

Multimodal pragmatic strategies varied systematically between breakout room discussions (smaller groups of 4-5 students) and whole-class discussions (entire class with instructor present). Although multimodal integration appeared slightly more frequent in breakout-room interactions than in whole-class discussions, the observed difference was relatively modest. Therefore, the findings should not be interpreted as evidence that breakout rooms inherently foster substantially greater multimodal development. Instead, both interactional settings provided meaningful opportunities for learners to coordinate verbal, embodied, and platform-mediated resources when performing requests and disagreements. Breakout rooms occasionally afforded additional opportunities for the use of private chat and screen-sharing functions; however, the overall patterns suggest that multimodal pragmatic development occurred across both participation frameworks.

Table 1. *Multimodal resource integration by context and developmental phase*

Interaction Mode	Breakout Room (%)	Whole-Class (%)
Week 1 (Phase 1)		
Verbal-only	82	84
Verbal + Gesture	10	8
Verbal + Chat	6	6

Multimodal (3+ resources)	2	2
Week 4 (Phase 2)		
Verbal-only	46	52
Verbal + Gesture	28	24
Verbal + Chat	18	16
Multimodal (3+ resources)	8	8
Week 8 (Phase 3)		
Verbal-only	18	23
Verbal + Gesture	38	34
Verbal + Chat	27	29
Multimodal (3+ resources)	17	14

Table 1 presents the distribution of multimodal resource use across two interactional contexts (breakout rooms and whole-class discussions) at three time points representing Phases 1, 2, and 3 of development. Data were collected from 156 Indonesian EFL learners across eight weeks of synchronous online learning via Zoom. Percentages indicate the proportion of face-threatening speech acts (requests and disagreements) performed using each multimodal configuration. "Verbal-only" refers to turns accomplished solely through spoken language without gesture, chat, or other semiotic resources. "Verbal + Gesture" includes coordinated use of hand gestures, gaze shifts, facial expressions, or head movements. "Verbal + Chat" refers to spoken turns accompanied by simultaneous or sequential text-based messages in the Zoom chat function. "Multimodal (3+ resources)" indicates integration of three or more semiotic channels (e.g., speech + gesture + chat, or speech + gesture + screen sharing). Developmental progression from Phase 1 to Phase 3 is marked by a steady decline in verbal-only exchanges (from 82-84% to 18-23%) and a corresponding increase in coordinated multimodal strategies.

Table 2. *Developmental phases of multimodal pragmatic competence*

Phase	Weeks	Characteristics	Verbal-only	Multimodal Integration
Phase 1	Week 1-2	Verbal-dominant, limited multimodal awareness	82-84%	Minimal (2-8%)
Phase 2	Week 3-5	Emerging coordination, experimentation	46-52%	Moderate (8-24%)
Phase 3	Week 6-8	Sophisticated integration, strategic platform use	18-23%	High (14-38%)

Table 2 summarizes the three-phase developmental trajectory observed among Indonesian EFL learners over eight weeks of synchronous online interaction. Phase 1 (Weeks 1-2) is characterized by verbal-dominant strategies with minimal multimodal elaboration; 82-84% of face-threatening speech acts were performed using spoken language alone. Phase 2 (Weeks 3-5) marks emerging multimodal coordination as learners begin experimenting with gestures and chat functions; verbal-only exchanges decline to 46-52%, while multimodal integration increases. Phase 3 (Weeks 6-8) represents sophisticated integration, with only 18-23% of interactions remaining verbal-only; learners demonstrate strategic deployment of multiple semiotic resources including speech, gesture, chat, and screen sharing. The "Multimodal Integration" column reflects the combined proportion of Verbal+Gesture, Verbal+Chat, and 3+ Resources categories

Individual variation and trajectory diversity

While the three-phase developmental pattern characterized the majority of participants, individual variation was evident. Approximately 15% of participants demonstrated rapid multimodal integration, reaching Phase 3 characteristics by Week 4. These “early adopters” tended to have prior experience with video-conferencing platforms and higher self-reported technological comfort.

Conversely, approximately 12% of participants maintained predominantly verbal-dominant strategies throughout the eight weeks, showing limited multimodal integration even in Phase 3. Stimulated recall interviews suggested that these participants experienced higher cognitive load managing both linguistic formulation and technological navigation, leaving limited attentional resources for conscious multimodal strategy deployment.

These individual differences highlight that developmental trajectories are not uniform, and pedagogical interventions should account for varying rates of multimodal pragmatic development.

Developmental trajectories of multimodal pragmatic competence

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

Our findings reveal that multimodal pragmatic competence in synchronous online language learning develops through progressive integration of semiotic resources rather than as a sudden acquisition. The three-phase trajectory from verbal-dominant strategies through emerging multimodal coordination to sophisticated integration demonstrates that learners require time and practice to recognize, experiment with, and strategically deploy the full range of resources available in video-mediated environments.

This developmental pattern aligns with usage-based approaches to L2 acquisition, which emphasize that competence emerges through repeated participation in interactional practices. Murakami and Ellis (2022) demonstrate that grammatical development is shaped by learners' accumulated exposure to recurring linguistic patterns and contextual associations. Similarly, multimodal pragmatic competence develops through repeated experience coordinating verbal, embodied, and textual resources during pragmatically consequential interaction.

The eight-week timeframe proved sufficient to observe substantial development, challenging claims that pragmatic competence requires extended immersion. While this study does not claim that eight weeks is sufficient to produce native-like pragmatic competence, the findings demonstrate that meaningful developmental changes can occur within this period when learners engage in frequent, authentic interaction involving face-threatening speech acts.

By emphasizing that technology-mediated methods were fundamental elements of successful engagement, these findings further develop the model of interactional competence. While earlier definitions focused primarily on verbal interaction and direct contact, current research shows that proficiency in real-time online environments also necessitates proficient utilization of platform-specific resources, such as private messages and sharing screens. According to this new perspective, interactional competence is contextually built into online environments, where the platform's capabilities shape what constitutes appropriate and significant social engagement.

The role of platform affordances

A central finding of this study is that platform affordances function not merely as contextual features but as active mediators of pragmatic action. The chat function, gallery view, and screen sharing capabilities shaped what pragmatic strategies were available, how face-threat could be managed, and what developmental trajectories were possible.

This finding extends ecological approaches to language learning, which emphasize the inseparability of learner, language, and environment (Zheng et al., 2024). In video-mediated contexts, the technological environment is not a neutral container for interaction but an active participant that affords certain actions while constraining others. Recognizing this has important pedagogical implications: rather than treating platform features as mere tools, instruction should explicitly address how these features can serve pragmatic functions (Gonzalez-Lloret, 2025).

These findings resonate with recent research demonstrating that online learning platforms are not merely delivery mechanisms but interactional environments that shape participation and meaning-making processes. Balaman (2023) and Hampel and Stickler (2023) argue that learners' ability to exploit technological affordances has become an integral component of successful online communication. The present findings further suggest that such affordances contribute directly to the development of multimodal pragmatic competence.

The mediational role of these affordances operates through several interconnected mechanisms. First, they enable semiotic expansion, providing learners with new modes (text, emoji, annotation) for expressing stance and mitigating face-threats. Second, they effect participation restructuring, as visual layouts such as gallery view alter turn-taking and recipient design. Third, they support cognitive distribution, allowing learners to balance verbal expression with textual elaboration. Collectively, these mechanisms demonstrate that platform affordances actively shape the trajectory of pragmatic development rather than simply supporting it.

Face-threatening speech acts in video-mediated interaction

My analysis of requests and disagreements reveals how face-threat management adapts to video-mediated affordances and constraints. While participants employed familiar pragmatic strategies from face-to-face interaction (indirectness, hedging, accounts), they also developed platform-specific practices such as simultaneous chat messages, private messaging for face-saving, and screen sharing to transform disagreements into collaborative inquiry.

The Indonesian cultural context added a distinctive dimension to face-threat management. Participants' strong preference for indirect formulations and their extensive use of mitigation strategies reflect Indonesian cultural values of social harmony (rukun) and respect (hormat) (Goddard, 2018). However, rather than simply transferring L1 pragmatic strategies, participants adapted these cultural orientations to the affordances of the video-mediated environment, creating hybrid practices that honored Indonesian values while exploiting platform-specific resources.

The findings also align with recent investigations of politeness and disagreement management in synchronous online communication. Hou and Yu (2023) demonstrated that participants frequently combine linguistic mitigation with platform-mediated resources to negotiate interpersonal relationships. Similar patterns were observed in the present study, where learners strategically integrated chat functions and screen-sharing tools to reduce face threat while maintaining communicative effectiveness.

Implications for interactional competence theory

The findings extend interactional competence theory (Hall, 2018; Pekarek Doehler & Pochon-Berger, 2015) by demonstrating that competence in video-mediated contexts involves not only traditional linguistic and pragmatic abilities but also technological literacy and multimodal coordination skills. Participants who developed sophisticated multimodal pragmatic competence were those who learned to treat platform affordances as pragmatic resources rather than merely technical features.

This suggests that interactional competence should be reconceptualized as inherently multimodal and technologically situated. Assessment of interactional competence in contemporary language education must account for learners' abilities to coordinate multiple semiotic resources and strategically exploit platform affordances, not merely their verbal pragmatic knowledge.

The growing prevalence of synchronous online learning in higher education highlights the need for pedagogical frameworks that explicitly address multimodal communication. Recent studies suggest that effective online language teaching requires not only linguistic instruction but also the development of learners' digital interactional competence and awareness of platform affordances (Badem, 2025; Hampel & Stickler, 2023).

Additionally, the finding that competence develops differently across contexts (e.g., breakout rooms versus whole-class discussions) reinforces the practice-specific nature of interactional competence. Learners require experience across multiple participation frameworks to develop flexible and context-sensitive pragmatic abilities.

Pedagogical implications

The findings generate several evidence-based recommendations for fostering multimodal pragmatic competence in synchronous online language teaching:

Explicit multimodal awareness-raising

Given that participants in Phase 1 were not consciously attending to multimodal possibilities, instruction should explicitly raise awareness of available semiotic resources. Activities might include: (1) Analyzing video-recorded interactions to identify multimodal strategies. (2) Comparing how the same pragmatic action can be accomplished through different modality configurations. (3) Discussing cultural and contextual factors that influence modality selection. (4) Practicing strategic use of platform affordances for specific pragmatic goals.

Scaffolded multimodal practice

Developmental trajectories suggest that learners benefit from scaffolded opportunities to experiment with multimodal strategies. Rather than expecting immediate sophisticated integration, instruction should: (1) Begin with awareness-raising and single-modality focus (e.g., gesture alone). (2) Progress to two-modality coordination (e.g., verbal + chat). (3) Advance to complex multimodal ensembles (three or more resources). (4) Provide feedback on pragmatic effectiveness rather than prescribing "correct" strategies.

Context-varied practice

The differences between breakout room and whole-class interaction suggest that learners need experience across multiple participation frameworks. Instructional design should: (1) Alternate between small-group and whole-class activities. (2) Explicitly discuss how pragmatic strategies might vary by context. (3) Provide opportunities for reflection on contextual adaptation. (4) Encourage experimentation in lower-stakes contexts (breakout rooms) before whole-class performance.

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

Platform affordance exploration

Rather than treating platform features as self-evident, instruction should explicitly explore their pragmatic potential: (1) Demonstrate how chat can serve pragmatic functions beyond information exchange. (2) Discuss strategic use of public versus private messaging. (3) Model screen sharing as a collaborative problem-solving tool. (4) Encourage creative exploitation of reactions, annotations, and other platform-specific features.

Limitations and future directions

Several limitations should be acknowledged. First, the eight-week timeframe, while sufficient to observe developmental changes, does not capture the long-term stabilization of multimodal pragmatic competence. Future research should examine whether the observed patterns persist and continue to develop over extended periods.

Second, the focus on advanced-level learners limits the generalizability of the findings to other proficiency levels. Intermediate learners may display different developmental trajectories, potentially requiring longer timelines or encountering different challenges in multimodal coordination. Future studies should therefore examine multimodal pragmatic development across a broader range of proficiency levels.

Third, although this study included 156 participants from four universities, all institutions were located in West Java and Jakarta, Indonesia. Cultural and institutional factors specific to this context may have influenced the findings. Comparative research across diverse cultural contexts, educational systems, and technological infrastructures would help determine which findings are context-specific and which may be more broadly generalizable.

Fourth, the analysis focused on two face-threatening speech acts (requests and disagreements). Other speech acts such as compliments, apologies, invitations, and refusals may exhibit different multimodal patterns and developmental trajectories. Future research should expand the range of pragmatic actions examined to provide a more comprehensive understanding of multimodal pragmatic competence development.

` Finally, while stimulated recall interviews were employed to access participants' perspectives, these retrospective accounts may not fully capture in-the-moment cognitive processes. Future research incorporating eye-tracking, think-aloud protocols, or other real-time measures could provide additional insights into multimodal strategy deployment.

CONCLUSION

This study showed that the growing integration of verbal, gestural, and platform-specific textual resources lead to the development of multimodal pragmatic competence in synchronous online language learning. The eight-week developmental trajectory of Indonesian EFL learners from verbal-dominant strategies to emerging multimodal coordination to sophisticated integration shows that in order to fully take advantage of the semiotic resources available in mediated video learning environments, learners require time, practice, and direct awareness-raising.

Platform affordances emerged as active mediators of pragmatic action, shaping how face-threatening speech acts could be performed and how face-threat could be managed. The chat function, gallery view configuration, and screen sharing capabilities functioned not merely as technical features but as pragmatic tools that learners progressively learned to deploy strategically.

The findings extend interactional competence theory to video-mediated contexts, demonstrating that competence in these environments involves not only traditional linguistic and pragmatic abilities but also technological literacy and multimodal coordination skills. The three-phase developmental pattern identified in this study provides a framework for understanding how learners progress from limited multimodal awareness to sophisticated and context-sensitive deployment of semiotic resources.

For language pedagogy, these findings underscore the importance of explicit multimodal awareness-raising, scaffolded practice opportunities, and strategic exploration of platform affordances. Rather than assuming that learners will naturally discover and deploy multimodal resources, instruction should actively cultivate multimodal pragmatic competence through targeted activities and reflective practice.

Understanding how pragmatic competence develops in synchronous online learning environments is becoming more and more important as synchronous online learning continues to play a major role in language teaching worldwide. By providing comprehensive, empirically supported insights into the developmental processes by which students become proficient multimodal communicators in video-mediated engagement, this study improves that knowledge.

Future research should extend these findings by examining longer developmental timelines, diverse

proficiency levels, varied cultural contexts, and expanded ranges of pragmatic actions. Additionally, intervention studies testing the effectiveness of pedagogical approaches designed to foster multimodal pragmatic competence would provide valuable guidance for practitioners. As video-mediated interaction becomes increasingly ubiquitous in professional, academic, and social domains, cultivating learners' abilities to communicate effectively across multiple semiotic channels represents an essential goal for contemporary language education.

REFERENCES

Badem, D. (2025). Sustaining synchronous online language learning in post-pandemic Asian universities: Challenges and opportunities. *RELC Journal*, *56*(1), 45–63.

Balaman, U. (2023). Video-conferencing technology and interactional organization in online language teaching. *Language Learning & Technology*, *27*(1), 1–18.

Balaman, U., & Sert, O. (2023). Embodied participation in video-mediated L2 interaction: A conversation analytic perspective. *Applied Linguistics*, *44*(3), 412–437.

Benson, P., Chik, A., Lim, H., Nakamura, K., Sunakawa, Y., & Watanabe, Y. (2023). Multimodal interaction in video-mediated L2 conversations. *Language Learning & Technology*, *27*(1), 1–21.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Buker, A., Smith, E., Perepelkina, O., & Vinciarelli, A. (2025). Multimodal analysis of disagreement in dyadic conversations: An approach based on emotion recognition. *In Proceedings of the 27th ACM International Conference on Multimodal Interaction (ICMI '25)* (pp. 228–237). ACM. https://dl.acm.org/doi/10.1145/3716553.3750754

Clift, R., Kendrick, K. H., Raymond, C. W., & Robinson, J. D. (2024). Jeffersonian transcription conventions. In J. D. Robinson, R. Clift, K. H. Kendrick, & C. W. Raymond (Eds.), *The Cambridge handbook of methods in conversation analysis* (pp. 974–989). Cambridge University Press. https://doi.org/10.1017/9781108936583.035

Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, *6*(2), 80–88. https://doi.org/10.1177/1558689812437186

Derakhshan, A., & Shakki, F. (2023). *Instructed second language pragmatics for the speech acts of request, apology, and refusal: A meta-analysis*. Springer. https://doi.org/10.1007/978-3-031-37093-9

Dwiana, N. R., & Syahri, I. (2023). Indonesian EFL learners' refusal: Refusal-speech act specific motivation on refusal realization. *Indonesian Journal of EFL and Linguistics*, *9*(1), 145–162. https://doi.org/10.21462/ijefl.v9i1.763

Fauziah, M., Widianto, E., Vrika, R., Cardoso, L. M., & Durgaryan, A. (2024). Politeness strategy found in the third debate of presidential candidates for the 2024 election. *English Franca: Academic Journal of English Language and Education*, *8*(2), 112–125. https://doi.org/10.29240/ef.v8i2.11209

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Goddard, C. (2018). "Cultural keywords" and "cultural scripts": Their theoretical and applied potential. In F. Sharifian (Ed.), *The Routledge handbook of language and culture* (pp. 127–142). Routledge.

Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.

González-Lloret, M. (2025). Technology-enhanced pragmatic instruction. *Applied Pragmatics*, *6*(2), 131–146. https://doi.org/10.1075/ap.00023.gon

Hall, J. K. (2018). From L2 interactional competence to L2 interactional repertoires: Reconceptualising the objects of L2 learning. *Classroom Discourse*, *9*(1), 25–39. https://doi.org/10.1080/19463014.2018.1433050

Hampel, R., & Stickler, U. (2023). Developing online language teaching competence: A framework for practice. *System*, *116*, Article 102997. https://doi.org/10.1016/j.system.2023.102997

Hanlon, G., & Keeffe, A. (2024). A multimodal analysis of roleplays between upper intermediate level learners. In *Multimodal communication in language learning* (pp. 206–221). Routledge. https://doi.org/10.4324/9781003361701-20

Hauser, E. (2022). Conversation analysis and second language pedagogy: Microgenesis in interaction. *Applied Linguistics*, *43*(6), 1089–1110. https://doi.org/10.1093/applin/amac014

Hou, J., & Yu, S. (2023). Politeness strategies in synchronous computer-mediated communication: Managing disagreement in online interaction. *Journal of Pragmatics*, *210*, 45–58. https://doi.org/10.1016/j.pragma.2023.02.004

Jakonen, T., Evnitskaya, N., & Laanemets, A. (2024). Classroom interaction in the digital age: Navigating between physical and virtual spaces. *Language Teaching Research*, *28*(1), 8–26.

Jenks, C. J. (2022). *Video-mediated intercultural conversations*. Routledge.

Kang, T., Arvizu, M. N. G., Chaipuapae, P., & Lesnov, R. O. (2019). Reviews of academic English listening tests for non-native speakers. *International Journal of Listening*, *33*(1), 1–38.

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 14, Issue 2, June 2026

https://doi.org/10.1080/10904018.2016.1185210

Kasper, G., & Wagner, J. (2011). A conversation-analytic approach to second language acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 117–142). Routledge.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, *20*(3), 336–366. https://doi.org/10.1111/josl.1_12177

Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, *51*(1), 85–106. https://doi.org/10.1080/08351813.2018.1413878

Mondada, L. (2019). Transcribing silent actions: A multimodal approach of sequence organization. *Social Interaction: Video-Based Studies of Human Sociality*, *2*(1). https://doi.org/10.7146/si.v2i1.113150

Moorhouse, B. L. (2024). Beginning teaching during COVID-19: Newly qualified Hong Kong teachers' preparedness for online teaching. *Educational Studies*, *50*(4), 499–515.

Murakami, A., & Ellis, N. C. (2022). Effects of input factors in L2 English grammatical morpheme accuracy: Frequency, contingency, and formulaicity. *Language Learning*, *72*(3), 812–852. https://doi.org/10.1111/lang.12509

Pekarek Doehler, S., & Pochon-Berger, E. (2015). The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 233–267). De Gruyter Mouton.

Pekarek Doehler, S., Wagner, J., & González-Martínez, E. (2018). Longitudinal research on second language use and learning: Introduction to the special issue. *The Modern Language Journal*, *102*(S1), 3–18. https://doi.org/10.1111/modl.12448

Ren, W., Li, S., & Lü, X. (2023). A meta-analysis of the effectiveness of second language pragmatics instruction. *Applied Linguistics*, *44*(6), 1010–1029. https://doi.org/10.1093/applin/amac055

Satar, H. M., & Wigham, C. R. (2023). Multimodal instruction-giving practices in synchronous online language teaching. *System*, *115Trace*, Article 103042. https://doi.org/10.1016/j.system.2023.103042

Satar, M. (2025). Semiotic approaches in CALL. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave encyclopedia of computer-assisted language learning*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-51447-0_10-1

Schafer, R., & Phillippi, J. C. (2025). Updating and advancing member-checking methods: Use of video and asynchronous technology to optimize participant engagement. *International Journal of Qualitative Methods*, *24*, 1–12. https://doi.org/10.1177/16094069251315395

Taguchi, N. (2019). Second language acquisition and pragmatics: An overview. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 1–14). Routledge.

Willcox-Ficzere, E. (2025). How do advanced L2 learners demonstrate pragmatic competence by aligning linguistic choices with their interpretation of the social context in oral discourse? *Baltic Journal of English Language, Literature and Culture*, *15*, 143–165. https://doi.org/10.22364/BJELLC.15.2025.10

Zainil, Y., & Arsyad, S. (2021). Teachers' perception of their code-switching practices in English as a foreign language classes: The results of stimulated recall interview and conversation analysis. *SAGE Open*, *11*(2), Article 21582440211013802. https://doi.org/10.1177/21582440211013802

Zheng, D. (2024). The new ecolinguistics: Learning as languaging with digital technologies. *Frontiers of Digital Education*. Advance online publication. https://doi.org/10.1007/s44366-024-0026-7