Volume 2 Issue 1 Year 2025 Pages 22-41
e–ISSN 3090-6245 | DOI: 10.70152
https://journal.akademimerdeka.com/ojs/index.php/matcha/index

# Real-Time AI Transcription as an Interactional Resource in Academic EFL Classrooms: An Exploratory Mixed-Methods Study

Irma Ratna Ningsih[1*], Ken Paul M. Espinosa[2]

[1] Institut Seni Budaya Indonesia Bandung, Indonesia
[2] Baliuag University, Philippines

[*]Corresponding author's email: irma.ratna@isbi.ac.id
DOI: https://doi.org/10.70152/matcha.v2i1.331

**Abstract:** This study examines how students use real-time AI transcription during speaking activities and how they perceive its role in supporting spoken interaction in academic EFL classrooms. A total of 40 students participated in classroom activities over six weeks. Data were collected through classroom observations, semi-structured interviews with five eleven selected participants, and a self-report questionnaire. The qualitative findings reveal that students employ AI transcription in self-directed and interaction-sensitive ways, such as monitoring speech accuracy, managing communication breakdowns, supporting turn-taking, and reducing anxiety during real-time interaction. These practices indicate that AI transcription functions not merely as a corrective aid but as an interactional resource embedded within ongoing classroom discourse. Quantitative results further show generally positive student perceptions regarding the usefulness of real-time AI transcription for enhancing clarity, confidence, and engagement in spoken interaction.. This study highlights how AI transcription can function as a supportive classroom resource when used flexibly by learners.

**Keywords:** artificial intelligence, EFL classrooms, exploratory mixed methods, real-time transcription, spoken interaction

## INTRODUCTION

Spoken interaction plays a central role in academic learning, serving not only as a medium for language development but also as a key resource for knowledge construction, participation, and meaning-making in classroom settings. In English as a Foreign Language (EFL) context, however, engaging in spoken interaction often poses significant challenges for learners (Yessenbekova, 2024; Zhang, 2023). These challenges include limited linguistic resources (Kulsum et al., 2025), concerns about pronunciation and intelligibility (Suratin & Sribayak, 2025), fear of making errors (Attia & Algazo, 2025), and anxiety related to public performance (Quvanch et al., 2024). As a result, many students participate minimally in classroom discussions, group work, and oral tasks, despite recognizing the importance of spoken interaction for academic success.

In recent years, advances in artificial intelligence (AI) have introduced new technological tools that may help address some of these challenges. Among these developments, real-time AI transcription, also referred to as speech-to-text technology, has become increasingly accessible through mobile devices, web-based platforms, and integrated classroom applications (Broeren et al., 2025). Unlike traditional language learning technologies that focus on post-task feedback or scripted practice, real-time AI transcription converts spoken language into written text instantaneously, allowing speakers to see a visual representation of their speech as it unfolds (Harihar et al., 2025). This immediacy opens new possibilities for supporting spoken interaction during live classroom communication rather than after it has concluded.

Within EFL classrooms, real-time AI transcription has the potential to function as an interactional support rather than a mere assessment tool. By displaying spoken output in written form, the technology may help learners monitor intelligibility, notice mismatches between intended and produced language, and maintain participation in ongoing interaction (Shadiev et al., 2024). In academic settings where spoken interaction often involves explaining ideas, responding to peers, and negotiating meaning, such support may be particularly valuable (Sok & Shin, 2025). However, the integration of real-time AI transcription into classroom interaction also raises important questions about how students actually use the technology and how they perceive its role in supporting spoken communication.

Existing research on AI-assisted language learning has largely focused on outcomes such as pronunciation improvement, speaking accuracy, or fluency gains, often within controlled or task-specific environments (Darmawansah et al., 2025; Qiao & Zhao, 2023). While these studies provide valuable insights into the potential effectiveness of AI tools, they offer limited understanding of how learners engage with such technologies in authentic classroom interaction. In particular, there remains a lack of empirical research examining how students use real-time AI transcription during spontaneous or semi-spontaneous spoken interaction in academic EFL classrooms. Understanding how students use AI transcription during real-time speaking activities is crucial (Adiguzel et al., 2023), as the pedagogical value of AI tools depends not only on their features but on how learners use them in classroom interaction.

This study focuses on how real-time AI transcription is used within spoken interaction in academic classrooms, particularly in supporting students during ongoing communication. It involves turn-taking, repair, negotiation of meaning, and alignment with interlocutors, all of which unfold in real time (Chen & Ye, 2022; Kunitz et al., 2022). Introducing AI transcription into this interactional space may shape how learners manage speech production, respond to others, and sustain participation (Pituxcoosuvarn et al., 2025; Xiao, 2025). For example, students may use the transcription to check if their speech is understood, improve their pronunciation while speaking, or help them continue their ideas when they feel stuck. At the same time, students may choose not to use the technology in certain moments, suggesting selective and strategic engagement rather than uniform adoption.

Another important dimension concerns learners' perceptions of real-time AI transcription. Students' beliefs about the usefulness, ease of use, and relevance of AI tools can strongly influence whether and how they engage with them during learning activities (Boubker, 2024; Siu & White, 2025; Zhang et al., 2023). In speaking contexts,

perceptions related to confidence, anxiety reduction, and communicative support are particularly salient. If learners perceive AI transcription as supportive and non-threatening, they may be more willing to take risks in spoken interaction (Huang et al., 2024). Conversely, if the technology is perceived as distracting, unreliable, or evaluative, students may avoid using it altogether (Pérez-Juárez et al., 2023). Exploring learners' perceptions is therefore essential for understanding the pedagogical potential and limitations of AI transcription in spoken interaction.

However, existing research has not yet examined how students actually use AI tools during real-time interaction while also considering their perceptions of such use. Most studies focus either on learning outcomes or general attitudes, rather than what students do during interaction (Karataş & Yüce, 2024; Ng et al., 2024). To address this gap, the present study investigates how students use real-time AI transcription in classroom speaking activities and how they perceive its role in supporting interaction. The study also focuses on naturally occurring classroom activities, where the use of AI transcription is optional rather than mandated. By examining self-directed use, the study seeks to understand how learners themselves make sense of and appropriate the technology within communicative tasks. This perspective foregrounds learners' experiences and practices, offering insights that are directly relevant to classroom teaching and technology integration. The study is guided by two research questions:

1. How do EFL students use real-time AI transcription to support spoken interaction in academic classroom settings?
2. How do EFL students perceive the role of real-time AI transcription in facilitating spoken interaction during classroom activities?

The significance of this study lies in its contribution to research on AI-assisted speaking and classroom interaction. First, it extends existing literature by shifting attention from learning outcomes to interactional use, highlighting how AI tools are embedded in everyday classroom practices. Second, it provides empirical evidence on real-time AI transcription, a technology that is increasingly available but still under-researched in authentic EFL interactional contexts. Third, the study offers pedagogical insights for teachers and curriculum designers by illustrating how AI transcription can support spoken interaction without replacing human communication. Finally, by adopting an exploratory mixed-methods approach, the study contributes methodologically to research on emerging educational technologies, demonstrating how qualitative and quantitative data can be integrated to capture both practice and perception.

## LITERATURE REVIEW

### Analytical Lens

This study is informed by an analytical lens that integrates perspectives from interaction-oriented approaches to second language acquisition and research on technology-mediated communication. Spoken interaction is viewed not merely as a site for language practice, but as a socially situated process through which learners negotiate meaning, manage participation, and construct understanding in real time (Mifka-Profozic, 2023). From this perspective, learning opportunities emerge through interactional processes such as turn-taking, repair, clarification, and alignment with interlocutors (Eguchi et al., 2025; Nanbu & Greer, 2023). Any technological tool introduced into this space, including real-time AI

transcription, becomes part of the interactional ecology and may influence how these processes unfold.

Within interaction-oriented frameworks, visibility of language plays a crucial role in supporting learners' noticing of linguistic form and meaning. Traditionally, such visibility has been achieved through teacher feedback, peer correction, or post-task reflection (Sun et al., 2023; Tsikou & Papadopoulou, 2025). Real-time AI transcription introduces a different form of visibility by rendering spoken language into text instantaneously. This transformation externalizes speech, allowing learners to see their oral production as it occurs (Bashori et al., 2024; Jiang et al., 2023; W. Qiao & Yijun, 2023). From an analytical standpoint, this visual mediation may shape how learners monitor intelligibility, initiate self-repair, and sustain participation during interaction without interrupting communicative flow.

Another important aspect of the analytical lens is the notion of learner-directed technology use. Rather than treating digital tools as fixed instructional interventions, recent research emphasizes how learners appropriate technologies in ways that align with their goals, preferences, and contextual constraints (Gumbheer et al., 2022). In classroom interaction, such appropriation is often selective and strategic. Learners may choose to engage with a tool at particular moments, for specific purposes, or not at all (Pratiwi & Waluyo, 2023; Schmid et al., 2023). This study therefore approaches real-time AI transcription as a resource that learners may draw upon flexibly to support spoken interaction, rather than as a deterministic factor shaping behaviour in uniform ways.

The analytical lens also foregrounds affective dimensions of spoken interaction. Anxiety, confidence, and willingness to communicate are widely recognized as influential factors in EFL speaking contexts (Lin et al., 2025; Solhi, 2024). Technologies that provide private, immediate feedback may help learners manage affective barriers by reducing fear of negative evaluation and increasing a sense of control over speech production (Liu et al., 2025; W. Sun, 2023). From this perspective, real-time AI transcription can be understood not only as a linguistic support, but also as an affective scaffold that shapes learners' engagement in interaction.

Finally, this study adopts an exploratory orientation toward AI in language education (Belda-Medina & Calvo-Ferrer, 2022). Given the rapid development of AI tools and their uneven integration into classrooms, it is premature to impose rigid evaluative frameworks focused solely on effectiveness or outcomes. Instead, an exploratory analytical lens prioritizes understanding how technologies are used, interpreted, and experienced by learners in authentic contexts. This orientation aligns with the study's mixed-methods design, which seeks to capture both interactional practices and learner perceptions without making causal claims.

## Previous Studies

Research on technology-supported speaking in EFL contexts has a long history, encompassing tools such as audio recording, computer-mediated communication, mobile applications, and automated feedback systems. Early studies highlighted the potential of technology to provide additional speaking opportunities, reduce anxiety, and promote learner autonomy (Bashori et al., 2024; Butarbutar et al., 2023). More recent work has focused specifically on AI-driven tools, including automated speech recognition and

speech-to-text systems (Benlaghrissi & Ouahidi, 2024; Hirai & Kovalyova, 2024), which offer new forms of feedback and interactional support.

Studies examining automated speech recognition have often emphasized pronunciation training and accuracy development. Many of these studies are conducted in controlled settings, where learners read scripted texts or complete isolated speaking tasks (Bashori et al., 2024a; Dillon & Wells, 2023). Findings generally suggest that exposure to recognition feedback can raise learners' awareness of pronunciation issues and encourage repeated practice (Inceoglu et al., 2024). However, such studies typically focus on individual performance rather than interaction, offering limited insight into how speech recognition functions during live classroom communication.

A smaller body of research has begun to explore speech-to-text tools in more communicative contexts. These studies suggest that real-time transcription can support comprehension, assist with vocabulary recall, and provide visual cues that help learners track spoken discourse (X. Chen et al., 2023). In collaborative tasks, transcription has been shown to facilitate shared understanding and support participation, particularly for learners who struggle with listening or speaking simultaneously (Zhao et al., 2024). Nevertheless, many of these studies focus on comprehension or accessibility rather than on spoken interaction as a dynamic, reciprocal process.

Research on learner perceptions of AI-assisted speaking tools consistently reports generally positive attitudes, particularly regarding ease of use and perceived usefulness (Wang & Wang, 2025; Zou et al., 2023). Learners often value immediate feedback and the ability to practice independently. At the same time, concerns have been raised about over-reliance on technology, inaccuracies in recognition output, and potential distractions during communication (Chang et al., 2023). These mixed perceptions highlight the importance of examining not only attitudes, but also how learners navigate the affordances and limitations of AI tools in practice.

Importantly, few studies have combined observational data with learner perceptions to examine real-time use of AI transcription in academic speaking contexts. Many rely solely on questionnaires or post-task reflections, which may not fully capture how tools are actually used during interaction (Crompton et al., 2024; Fathi et al., 2024). Others focus on experimental comparisons, which can obscure the complexity of classroom practices (Muqaibal et al., 2026; Zakian et al., 2022). This gap points to the need for research designs that integrate qualitative and quantitative data to explore both use and perception.

In addition, much of the existing literature treats AI tools as instructional interventions introduced by teachers or researchers. Less attention has been paid to self-directed use, where learners independently decide when and how to engage with technology (Fitriati & Williyan, 2025). Understanding self-directed use is particularly important in academic settings, where students are expected to manage their own participation and communication strategies.

While previous studies demonstrate the potential of AI-powered speech recognition and transcription to support speaking and comprehension, they leave several questions unanswered. There is limited empirical evidence on how real-time AI transcription is embedded in spoken interaction during authentic classroom activities, how learners appropriate the tool to support participation, and how observed practices align with

learners' perceptions. By addressing these gaps, the present study builds on and extends existing research, offering a more comprehensive account of real-time AI transcription as a resource for spoken interaction in EFL classrooms.

## METHODS

### Research Design

This study adopts an exploratory mixed-methods design, which is suitable for investigating emerging classroom practices that are not yet well established (Hirose & Creswell, 2023). This approach allows the study to capture both students' actual use of AI transcription during interaction and their perceptions of its usefulness. An exploratory mixed-methods approach is appropriate when examining an emerging phenomenon for which existing theoretical frameworks and empirical evidence remain limited. Rather than testing hypotheses or establishing causal relationships, the design aims to document practices, identify patterns, and develop an empirically grounded understanding of how learners engage with new technologies in authentic contexts.

In this study, qualitative and quantitative data were collected during the same instructional period, with qualitative inquiry serving as the primary strand. Classroom observations and semi-structured interviews were used to explore how students integrated real-time AI transcription into spoken interaction and how they made sense of its use during academic tasks. Quantitative data were collected through a self-report questionnaire to capture broader trends in students' perceptions of the technology. The quantitative component was not intended to explain or predict behavior but to complement qualitative findings by providing an overview of perceived usefulness and engagement.

The exploratory nature of the design is reflected in three key features. First, the study focused on naturally occurring classroom practices, with no experimental manipulation or prescribed use of the AI tool. Second, the use of real-time AI transcription was self-directed, allowing students to decide when and how to engage with the technology. Third, integration of qualitative and quantitative data occurred at the interpretive stage, where insights from both strands were synthesized to address the research questions. This design allows for a rich, contextualized understanding of technology use while avoiding overgeneralized claims about effectiveness.

### Participants

The participants were undergraduate university students enrolled in English courses that emphasized spoken interaction in academic settings. These courses included activities such as group discussions, collaborative problem-solving tasks, short presentations, and peer-led exchanges. Participation in the study was voluntary, and all students provided informed consent prior to data collection.

Participants represented diverse academic disciplines and a range of English proficiency levels, reflecting the heterogeneity typical of tertiary-level EFL classrooms. No specific proficiency threshold was imposed, as the study aimed to capture varied interactional practices and perceptions rather than to compare proficiency-based outcomes. Inclusion criteria focused on students' regular participation in classroom speaking activities where real-time AI transcription tools were available for optional use.

To minimize disruption to regular classroom practices, the tool was briefly introduced without explicit instruction on when or how to use it. Students were not required to use the tool and could choose whether or not to engage with it during speaking activities. This approach aimed to approximate natural classroom conditions while acknowledging the presence of a new technological resource.

Ethical considerations were addressed throughout the study. Participants were informed of the study's purpose, the types of data to be collected, and their right to withdraw at any time without penalty. Anonymity was maintained by assigning pseudonyms and removing identifying information from all transcripts, observation notes, and questionnaire responses. The study did not interfere with instructional routines or assessment practices.

### Data Collection

Data were collected from multiple sources to support methodological triangulation and to capture both observable practices and subjective perceptions. The primary qualitative data sources were classroom observations and semi-structured interviews, while quantitative data were obtained through a self-report questionnaire.

Classroom observations were conducted during scheduled speaking-focused sessions over a defined instructional period. The observations focused on students' use of real-time AI transcription during spoken interaction, particularly in group and pair-based activities. Detailed field notes were taken to document when and how students activated transcription, the interactional purposes it served, and how students responded to the transcription output while speaking. Attention was also paid to interactional features such as turn-taking, repair, clarification, and peer response, as well as contextual factors including task type and participation structure.

Following the observation phase, Semi-structured interviews were conducted with eleven students selected based on their varying levels of AI use (frequent, occasional, and non-users). This purposive sampling allowed for diverse perspectives. The interviews aimed to gain deeper insights into students' experiences, decision-making processes, and perceptions related to the use of real-time AI transcription. Open-ended questions addressed topics such as reasons for using or not using the tool, perceived benefits and challenges, influence on confidence and participation, and reflections on its role in academic interaction. Interviews were conducted in Bahasa Indonesia (participants' first language) to encourage detailed and reflective responses. All interviews were audio-recorded with consent and transcribed for analysis.

Quantitative data were collected through a questionnaire designed to capture students' perceptions of real-time AI transcription. The questionnaire was adapted from the Technology Acceptance Model (TAM), focusing on perceived usefulness and ease of use. It consisted of six items measured on a 5-point Likert scale. The questionnaire included six items designed to capture key aspects of perceived usefulness and ease of use, consistent with prior exploratory studies. Responses were measured on a five-point Likert scale ranging from strongly disagree to strongly agree. The questionnaire was administered after students had used the tool across six weeks, ensuring that they had adequate experience to form informed opinions.

### Data Analysis

Qualitative and quantitative data were analysed separately before being integrated at the interpretive stage. For the qualitative data, an inductive thematic analysis approach was employed (De Paoli, 2024). Observation notes and interview transcripts were read repeatedly to achieve familiarity with the data. Initial codes were generated to capture recurring patterns related to interactional use, purposes of transcription, and students' experiences during spoken interaction.

These codes were then examined and grouped into broader themes that addressed the first research question. The analysis emphasized how real-time AI transcription was embedded in interactional practices rather than treating use as a binary variable, meaning students were classified as either users (used the tool at least once) or non-users (did not use it at all). Constant comparison was used to refine themes across data sources and participants, ensuring that interpretations were grounded in multiple instances of evidence. Analytical memos were written throughout the process to document emerging insights and to enhance reflexivity.

The quantitative questionnaire data were analysed using descriptive statistics (Fraenkel et al., 2023). Means, standard deviations, and frequency distributions were calculated for each item to identify overall trends in students' perceptions. The analysis focused on describing patterns rather than testing hypotheses or conducting inferential comparisons. This descriptive focus is consistent with the exploratory aim of the study and avoids overstating the implications of the quantitative data.

Integration of the qualitative and quantitative findings occurred during the interpretation phase. Results from both strands were compared to identify areas of convergence and divergence between observed practices and reported perceptions. The qualitative findings provided contextualized explanations of how students used real-time AI transcription during spoken interaction, while the quantitative results offered a broader view of perceived usefulness and engagement. Together, these integrated insights enabled a comprehensive response to the research questions and contributed to a nuanced understanding of real-time AI transcription as a resource for spoken interaction in EFL classrooms.

## FINDINGS

This section presents the findings of the study in relation to the two research questions. Findings for RQ1 are based on qualitative data from classroom observations and semi-structured interviews, while findings for RQ2 draw on quantitative data from a self-report questionnaire. The two strands are presented separately to reflect their distinct analytical purposes.

### Students' Use of Real-Time AI Transcription in Spoken Interaction

The qualitative analysis revealed that students used real-time AI transcription in diverse and purposeful ways to support spoken interaction during academic classroom activities. Rather than employing the tool continuously, students engaged with it selectively, depending on task demands, interactional challenges, and individual needs. Three major themes emerged from the analysis: (1) monitoring intelligibility during interaction, (2)

supporting lexical and structural formulation, and (3) managing affective challenges in spoken participation.

**Table 1.**
Thematic Overview of RQ1 Findings

| Theme | Description | Data Sources |
|---|---|---|
| Monitoring intelligibility | Using transcription output to check clarity and comprehensibility of speech | Observations, Interviews |
| Supporting formulation | Using transcription to assist word choice and sentence construction | Observations, Interviews |
| Managing affective challenges | Using transcription to reduce anxiety and increase confidence | Interviews |

*Monitoring Intelligibility During Interaction*

One prominent pattern observed across classroom activities was students' use of real-time AI transcription to monitor whether their spoken output was intelligible. During group discussions and pair work, students occasionally glanced at the transcription display while speaking, particularly when expressing complex ideas or using unfamiliar vocabulary. When transcription output did not align with their intended meaning, students often repeated or reformulated their utterances. An interview excerpt and classroom observation illustrate this practice:

> *"When I saw the words were different from what I meant, I tried to say it again more clearly. It helped me check if my pronunciation was okay." (Participant 3)*

> *Student A: "I think the government should… uh… reduce the tax for small business." (looks at transcription screen)*
> *Student A: "small businesses… yes… reduce taxes for small businesses."*
> *Student B: "Yes, I agree."*
> *(Classroom Interaction)*

Rather than treating transcription errors as definitive judgments, students interpreted them as signals that their speech might be unclear to listeners. This practice suggests that transcription functioned as an interactional cue, prompting self-repair without interrupting communication. Observations indicated that such monitoring occurred briefly and did not dominate interaction, allowing students to remain engaged with their peers.

*Supporting Lexical and Structural Formulation*

Another theme concerned the use of transcription to support lexical retrieval and sentence formulation during spoken interaction. Students sometimes paused briefly to look at the transcribed text, using it as a visual reference to continue their ideas. This was particularly evident during academic discussions that required explanation, justification, or elaboration. One participant explained:

> *"Sometimes I know the idea, but I'm not sure about the sentence. Seeing the text helps me continue speaking." (Participant 7)*

In these instances, transcription served as a cognitive scaffold, helping students maintain coherence and fluency. Importantly, students did not rely on the tool to generate language for them. Instead, they used it to confirm or adjust language they had already produced. This indicates that real-time AI transcription supported language formulation without replacing active language production.

*Managing Affective Challenges in Spoken Participation*

The third theme highlights the affective dimension of transcription use. Many students reported that real-time transcription helped reduce anxiety associated with speaking in English, especially in academic settings where accuracy and clarity were perceived as important. The ability to privately monitor speech provided reassurance and increased willingness to participate. As one participant noted:

> *"I feel less nervous because I can check myself first. If it looks okay, I feel more confident to continue speaking." (Participant 11)*

This affective support was particularly salient for students who described themselves as less confident speakers. Observational data showed that these students were more likely to engage with transcription during preparation phases and initial turns in interaction. Overall, transcription appeared to function as an emotional buffer that encouraged participation rather than inhibiting it.

## Students' Perceptions of Real-Time AI Transcription

Quantitative findings from the questionnaire provide an overview of students' perceptions of real-time AI transcription as a support for spoken interaction. Descriptive statistics indicate generally positive perceptions, with some variation across items.

**Table 2.**
Descriptive Statistics of Questionnaire Results

| Item | Mean | Median | SD |
|---|---|---|---|
| Helps me speak more clearly | 4.12 | 4 | 0.71 |
| Supports my confidence in speaking | 4.05 | 4 | 0.76 |
| Helps me notice pronunciation issues | 4.18 | 4 | 0.68 |
| Helps me organize ideas while speaking | 3.94 | 4 | 0.82 |
| Is easy to use during class | 4.26 | 4 | 0.64 |
| Overall usefulness for spoken interaction | 4.15 | 4 | 0.69 |

Responses indicate that students generally agreed that real-time AI transcription supported clarity, pronunciation awareness, and confidence. The highest mean score was related to ease of use, suggesting that technical accessibility was not a major barrier during classroom interaction. Slightly lower, though still positive, scores were observed for items related to idea organization, indicating that the perceived cognitive benefits varied among students.

**Table 3.**
Frequency Distribution of Overall Usefulness

| Response Category | Percentage |
|---|---|
| Strongly agree | 38% |
| Agree | 44% |
| Neutral | 14% |
| Disagree | 4% |
| Strongly disagree | 0% |

The majority of students expressed agreement or strong agreement regarding the overall usefulness of AI transcription for spoken interaction. Neutral responses suggest that some students perceived limited benefit, often aligning with interview comments indicating that confident speakers felt less need for technological support.

The quantitative data suggest that students largely perceived real-time AI transcription as a supportive resource rather than a disruptive presence. These perceptions align with qualitative findings showing selective and strategic use. Students who reported higher perceived usefulness often described using transcription to monitor intelligibility and manage anxiety, while those with neutral perceptions tended to view it as optional rather than essential. Importantly, no items indicated strong disagreement, suggesting that negative perceptions were minimal. This pattern supports the interpretation that real-time AI transcription is broadly acceptable to students in academic speaking contexts, even if its perceived value varies depending on individual needs and preferences.

**Table 4.**
Integration of Qualitative and Quantitative Findings

| Theme | Observed Practice | Questionnaire Support |
|---|---|---|
| Intelligibility | Students checked transcription | High clarity scores |
| Formulation | Used to continue ideas | Moderate organization score |
| Confidence | Reduced hesitation | High confidence score |

The findings reveal that real-time AI transcription was integrated into spoken interaction in flexible and learner-directed ways (Table 4). Qualitative data demonstrate how students used the tool to monitor intelligibility, support language formulation, and manage affective challenges. Quantitative data complement these insights by showing generally positive perceptions of usefulness, confidence support, and ease of use. The convergence of observed practices and reported perceptions provides a comprehensive picture of real-time AI transcription as a supportive resource for spoken interaction in EFL classrooms, setting the stage for deeper interpretation in the discussion section.

## DISCUSSION

This study explored how EFL students use real-time AI transcription to support spoken interaction in academic classroom settings and how they perceive its role during interaction. Rather than focusing on effectiveness or learning outcomes, the study adopted an exploratory mixed-methods approach to understand situated practices and learner perspectives. This discussion interprets the findings by connecting them to existing scholarship on spoken interaction, technology-mediated communication, and AI-assisted language learning, while highlighting their pedagogical and conceptual significance.

The findings indicate that real-time AI transcription functions as an interactional resource rather than a corrective or evaluative mechanism. Students did not engage with the tool continuously, nor did they treat it as an authoritative source of linguistic correctness. Instead, transcription was used selectively to support participation, intelligibility, and continuity in spoken interaction. This pattern aligns with interaction-oriented perspectives that view classroom communication as a dynamic process shaped by participants' moment-by-moment decisions (Chen & Ye, 2022; Kunitz et al., 2022; Mifka-Profozic, 2023). The use of transcription as a prompt for self-repair and reformulation indicates that AI-mediated feedback can be embedded within ongoing interaction without interrupting communicative flow. The immediacy of visual transcription cues enabled learners to make real-time adjustments to their speech while remaining actively engaged with their interlocutors. This contrasts with more conventional forms of feedback, which tend to be delayed or teacher-led (Sun et al., 2023), and highlights how transcription supports immediate responsiveness within interaction. In this sense, the technology did not function as an external corrective tool but became part of the interactional ecology, enabling learners to sustain participation during academic speaking tasks (Eguchi et al., 2025; Xiao, 2025).

A central interpretive insight from the study concerns students' pragmatic orientation toward intelligibility. Students evaluated their spoken output primarily in terms of whether it could be successfully transcribed, interpreting mismatches as potential indicators of unclear pronunciation or phrasing. This reflects a communicative focus on being understood rather than achieving linguistic perfection, an orientation that has been widely recognized as central to spoken interaction in academic and communicative EFL contexts (Kunitz et al., 2022; Sok & Shin, 2025). Such an orientation is particularly relevant in academic settings, where clarity and meaning-making are prioritized over formal accuracy. Engagement with transcription errors further illustrates how learners negotiated the limitations of AI tools. Rather than rejecting the technology when errors occurred, students often responded by reformulating their speech or adjusting pronunciation, echoing previous findings that learners can productively work with imperfect AI feedback when it is interpreted critically and contextually (Chang et al., 2023; Hirai & Kovalyova, 2024). The findings therefore contribute to a more nuanced understanding of AI accuracy, showing that usefulness does not depend solely on technical precision but also on how learners make sense of feedback during interaction (Qiao & Yijun, 2023).

The study also highlights the role of real-time AI transcription as a source of both cognitive and affective scaffolding during spoken interaction. Cognitively, transcription supported lexical retrieval and idea formulation, particularly during tasks that required extended explanation or argumentation. By externalizing spoken output in written form,

the tool appeared to reduce processing demands and help learners maintain coherence while speaking, a benefit similarly observed in previous research on speech-to-text and transcription-supported interaction (Chen et al., 2023; Jiang et al., 2023). Affectively, transcription contributed to reduced anxiety and increased confidence for many students. The ability to privately monitor speech provided reassurance, especially in situations where students feared negative evaluation, a key factor influencing willingness to communicate in EFL classrooms (Lin et al., 2025; Solhi, 2024). Rather than inhibiting spontaneity, this support appeared to encourage participation by lowering emotional barriers to speaking, consistent with research emphasizing the affective affordances of AI-assisted speaking tools (Huang et al., 2024; Liu et al., 2025).

One of the strengths of this study lies in the alignment between observed practices and reported perceptions. Students' positive perceptions of usefulness, ease of use, and confidence support correspond closely with qualitative evidence of purposeful engagement during interaction. This convergence suggests that learners' attitudes toward real-time AI transcription were grounded in actual classroom experience rather than abstract expectations about technology, supporting concerns raised in prior research about the limits of perception-only studies (Ng et al., 2024; Crompton et al., 2024). However, most students reported generally positive perceptions, with only minor variation across items. This reinforces the view that real-time AI transcription should not be conceptualized as a universal solution but as a flexible resource whose value emerges through use in specific contexts (Pratiwi & Waluyo, 2023; Schmid et al., 2023).

From a pedagogical perspective, the findings suggest that real-time AI transcription can be meaningfully incorporated into academic speaking activities without undermining interaction. Allowing optional, self-directed use enables learners to engage with the tool according to their needs while preserving the primacy of human communication. This aligns with research emphasizing learner-directed appropriation of digital tools rather than prescriptive implementation (Gumbheer et al., 2022; Fitriati & Williyan, 2025). The study also highlights the importance of pedagogical framing, as students' interpretations of transcription output shape how the technology is used. More broadly, the findings contribute to ongoing discussions about AI integration in language education by emphasizing practice over prescription (Belda-Medina & Calvo-Ferrer, 2022; Siu & White, 2025). Rather than positioning AI transcription as an instructional intervention with predefined outcomes, the study illustrates how its pedagogical value emerges through learners' interactional practices and perceptions.

Conceptually, this study advances understanding of AI-assisted spoken interaction by foregrounding use, experience, and interpretation. It shifts attention away from outcome-driven evaluations toward a more situated view of technology use in classrooms. By documenting how learners integrate real-time AI transcription into interaction, the study contributes to a growing body of research that views AI tools as socially embedded resources rather than deterministic technologies (Eguchi et al., 2025; Pituxcoosuvarn et al., 2025). The discussion underscores that the significance of real-time AI transcription lies not in its technical sophistication alone but in how learners appropriate it to support spoken interaction. These insights contribute to more grounded and human-centered perspectives on AI in EFL education and provide a foundation for continued inquiry into technology-mediated interaction. While AI transcription provided support, it may also increase cognitive load, as students need to divide attention between speaking and reading

the screen. In some cases, over-reliance on transcription may reduce spontaneous speaking.

## CONCLUSION

This study explored how EFL students use real-time AI transcription to support spoken interaction in academic classroom settings and how they perceive its role during oral activities. Adopting an exploratory mixed-methods approach, the findings demonstrate that students employ AI transcription in diverse, self-directed ways, including monitoring accuracy, supporting turn-taking, managing breakdowns in communication, and enhancing confidence during spoken interaction. Rather than functioning merely as a corrective or assessment-oriented tool, real-time AI transcription emerged as an interactional resource that students strategically integrated into ongoing communication. Students' perceptions further indicate that the technology is valued for its immediacy, visibility, and supportive role in reducing cognitive and affective demands during speaking tasks. Together, these findings contribute to a more nuanced understanding of AI-supported spoken interaction by foregrounding students' situated practices and experiences within authentic classroom contexts.

The study offers important pedagogical and theoretical implications for EFL instruction in academic settings. It highlights the potential of real-time AI transcription to mediate spoken interaction in ways that align with communicative and interaction-focused approaches to language teaching. By emphasizing exploratory use rather than instructional imposition, the study suggests that AI tools may be most effective when learners are given agency to appropriate them according to their communicative needs. However, this study is limited by its small sample size and short observation period. Future research should examine long-term use and different proficiency levels. Practically, teachers can use AI transcription as an optional support tool rather than a required component of instruction. For future research, longitudinal studies are needed to examine how students' use of real-time AI transcription evolves over time and how sustained engagement shapes interactional competence. Further research may also investigate how such tools influence peer interaction patterns, classroom discourse dynamics, and teacher roles across different instructional contexts. Comparative studies across proficiency levels, task types, or AI transcription systems would also deepen understanding of the conditions under which real-time transcription most effectively supports spoken interaction in EFL classrooms.

## REFERENCES

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, *15*(3), ep429. https://doi.org/10.30935/cedtech/13152

Attia, S., & Algazo, M. (2025). Foreign language anxiety in EFL classrooms: Teachers' perceptions, challenges, and strategies for mitigation. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1614353

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024a). I can speak : Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, *18*(5), 443–461. https://doi.org/10.1080/17501229.2024.2315101

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024b). 'Look, I can speak correctly': Learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning*, *37*(5–6), 1335–1363. https://doi.org/10.1080/09588221.2022.2080230

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 8427. https://doi.org/10.3390/app12178427

Benlaghrissi, H., & Ouahidi, L. M. (2024). The impact of mobile-assisted project-based learning on developing EFL students' speaking skills. *Smart Learning Environments*, *11*(1), 18. https://doi.org/10.1186/s40561-024-00303-y

Boubker, O. (2024). From chatting to self-educating: Can AI tools boost student learning outcomes? *Expert Systems with Applications*, *238*, 121820. https://doi.org/10.1016/j.eswa.2023.121820

Broeren, M., Gu, Y., Pitt, M., & Tompkins, V. (2025). A script and tutorial for using Rev AI's automatic speech transcription. *Infant and Child Development*, *34*(2). https://doi.org/10.1002/icd.70007

Butarbutar, R., Ruing, F. H., Basri, N., Tuharea, V. U., & Radja Leba, S. M. (2023). Unpacking online collaborative learning in teaching EFL speaking: Insights from three rural area case studies. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2023.6165

Chang, D. H., Lin, M. P.-C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, *15*(17), 12921. https://doi.org/10.3390/su151712921

Chen, M. H., & Ye, S. X. (2022). Extending repair in peer interaction: A conversation analytic study. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.926842

Chen, X., Li, S., Liu, S., Fowler, R., & Wang, X. (2023). MeetScript: Designing transcript-based interactions to support active participation in group video meetings. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW2), 1–32. https://doi.org/10.1145/3610196

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, *55*(6), 2503–2529. https://doi.org/10.1111/bjet.13460

Darmawansah, D., Hwang, G.-J., Lin, C.-J., & Febiyani, F. (2025). An artificial intelligence-supported GFCA learning model to enhance L2 students' role-play performance, English speaking and interaction mindset. *Educational Technology Research and Development*, *73*(3), 1451–1479. https://doi.org/10.1007/s11423-025-10453-6

De Paoli, S. (2024). Performing an inductive thematic analysis of semi-structured interviews with a large language model: An exploration and provocation on the limits of the approach. *Social Science Computer Review*, *42*(4), 997–1019. https://doi.org/10.1177/08944393231220483

Dillon, T., & Wells, D. (2023). Effects of pronunciation training using automatic speech recognition on pronunciation accuracy of Korean English language learners. *English Teaching*, *78*(1), 3–23. https://doi.org/10.15858/engtea.78.1.202303.3

Eguchi, M., Takizawa, K., Saeki, M., Kurata, F., Suzuki, S., Matsuyama, Y., & Sawaki, Y. (2025). Human- versus artificial intelligence-delivered roleplay tasks for assessing interactional competence: An applied conversation analytic study. *TESOL Quarterly*, *59*(S1). https://doi.org/10.1002/tesq.70028

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Fitriati, S. W., & Williyan, A. (2025). AI-enhanced self-regulated learning: EFL learners ' prioritization and utilization in presentation skills development. *Journal of Pedagogical Research*, *9*(2), 22–37. https://doi.org/https://doi.org/10.33902/JPR.202530647

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2023). *How to design and evaluate research in education*. McGraw-Hill Higher Education.

Gumbheer, C. P., Khedo, K. K., & Bungaleea, A. (2022). Personalized and adaptive context-aware mobile learning: Review, challenges and future directions. *Education and Information Technologies*, *27*(6), 7491–7517. https://doi.org/10.1007/s10639-022-10942-8

Harihar, D., Shrivastava, V., Talele, P., & Jahagirdar, A. (2025). Voice-based user interface for hands-free data entry and automation at workplaces. *MethodsX*, *15*, 103596. https://doi.org/10.1016/j.mex.2025.103596

Hirai, A., & Kovalyova, A. (2024). Speech-to-text applications' accuracy in English language learners' speech transcription. *Language Learning & Technology*, *28*(1), 1–21. https://doi.org/10.64152/10125/73555

Hirose, M., & Creswell, J. W. (2023). Applying core quality criteria of mixed methods research to an empirical study. *Journal of Mixed Methods Research*, *17*(1), 12–28. https://doi.org/10.1177/15586898221086346

Huang, F., Wang, Y., & Zhang, H. (2024). Modelling generative AI acceptance, perceived teachers' enthusiasm and self-efficacy to English as a foreign language learners' well-being in the digital era. *European Journal of Education*, *59*(4). https://doi.org/10.1111/ejed.12770

Inceoglu, S., Chen, W.-H., & Lim, H. (2024). Monitoring student behavior in autonomous automatic speech recognition-based pronunciation practice. *System*,

*124*, 103387. https://doi.org/10.1016/j.system.2024.103387

Jiang, M. Y., Jong, M. S., Lau, W. W., Chai, C., & Wu, N. (2023). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. *Journal of Computer Assisted Learning*, *39*(1), 125–140. https://doi.org/10.1111/jcal.12732

Karataş, F., & Yüce, E. (2024). AI and the future of teaching: Preservice teachers' reflections on the use of artificial intelligence in open and distributed learning. *The International Review of Research in Open and Distributed Learning*, *25*(3), 304–325. https://doi.org/10.19173/irrodl.v25i3.7785

Kulsum, E. M., Jamjami, A. Y., Uyun, A. S., & Sri Widianingsih, A. W. (2025). A study on the difficulties in English dpeaking for EFL students. *Journal on Education*, *7*(2), 10648–10659. https://doi.org/10.31004/joe.v7i2.8109

Kunitz, S., Berggren, J., Haglind, M., & Löfquist, A. (2022). Getting students to talk: A practice-based study on the design and implementation of problem-solving tasks in the EFL classroom. *Languages*, *7*(2), 75. https://doi.org/10.3390/languages7020075

Lin, Y., Wang, B., Zhang, G., & Su, Y. (2025). The relations among foreign language anxiety, academic buoyancy and willingness to communicate in EFL classroom. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1634054

Liu, C.-C., Hwang, G.-J., Yu, P., Tu, Y.-F., & Wang, Y. (2025). Effects of an automated corrective feedback-based peer assessment approach on students' learning achievement, motivation, and self-regulated learning conceptions in foreign language pronunciation. *Educational Technology Research and Development*, *73*(4), 2403–2424. https://doi.org/10.1007/s11423-025-10484-z

Mifka-Profozic, N. (2023). Interactive alignment in L2 learning: The link between social interaction and psycholinguistic phenomena. *Education Sciences*, *13*(8), 792. https://doi.org/10.3390/educsci13080792

Muqaibal, M. H., Kasprowicz, R., & Tissot, C. (2026). Evaluating the impact of spaced practice using computer-assisted language learning (CALL) on vocabulary learning in the classroom. *Language Teaching Research*, *30*(1), 365–402. https://doi.org/10.1177/13621688221146146

Nanbu, Z., & Greer, T. (2023). Creating obstacles to progressivity: Task expansion in second language role-plays. *TESOL Quarterly*, *57*(4), 1364–1400. https://doi.org/10.1002/tesq.3197

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, *55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Pérez-Juárez, M. Á., González-Ortega, D., & Aguiar-Pérez, J. M. (2023). Digital

distractions from the point of view of higher education students. *Sustainability*, *15*(7), 6044. https://doi.org/10.3390/su15076044

Pituxcoosuvarn, M., Tanimura, M., Murakami, Y., & White, J. S. (2025). Enhancing EFL speaking skills with AI-powered word guessing: A comparison of human and AI partners. *Information*, *16*(6), 427. https://doi.org/10.3390/info16060427

Pratiwi, D. I., & Waluyo, B. (2023). Autonomous learning and the use of digital technologies in online English classrooms in higher education. *Contemporary Educational Technology*, *15*(2), ep423. https://doi.org/10.30935/cedtech/13094

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1255594

Qiao, W., & Yijun, C. (2023). The role of live transcripts in synchronous online L2 classrooms: Learning outcomes and learner perceptions. *Education and Information Technologies*, *28*(11), 14783–14804. https://doi.org/10.1007/s10639-023-11784-8

Quvanch, Z., Qasemi, A. S., & Na, K. S. (2024). Analyzing levels, factors and coping strategies of speaking anxiety among EFL undergraduates in Afghanistan. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2413225

Schmid, R., Pauli, C., & Petko, D. (2023). Examining the use of digital technology in schools with a school-wide approach to personalized learning. *Educational Technology Research and Development*, *71*(2), 367–390. https://doi.org/10.1007/s11423-022-10167-z

Shadiev, R., Feng, Y., Zhussupova, R., & Huang, Y. (2024). Intercultural competence development through a tele-collaborative project supported by speech-enabled corrective feedback technology. *Journal of Computer Assisted Learning*, *40*(2), 697–714. https://doi.org/10.1111/jcal.12906

Siu, B., & White, J. (2025). When ease becomes a barrier: What influences student intentions to use Generative AI in higher education? *British Educational Research Journal*, *51*(6), 2893–2919. https://doi.org/10.1002/berj.4206

Sok, S., & Shin, H. W. (2025). Do interactions with ChatGPT influence L2 learners' oral speaking ability, summarization ability, and perceptions of generative AI tasks? *TESOL Quarterly*, *59*(S1). https://doi.org/10.1002/tesq.70001

Solhi, M. (2024). The impact of EFL learners' negative emotional orientations on (un)willingness to communicate in in-person and online L2 learning contexts. *Journal of Psycholinguistic Research*, *53*(2), 26. https://doi.org/10.1007/s10936-024-10071-y

Sun, Q., Chen, F., & Yin, S. (2023). The role and features of peer assessment feedback in college English writing. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1070618

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1210187

Suratin, N., & Sribayak, V. (2025). Factors contributing to speaking anxiety and anxiety reduction techniques in Thai adult EFL learners. *LEARN Journal: Language Education and Acquisition Research Network*, *18*(1), 294–319. https://doi.org/10.70730/MYTS3987

Tsikou, M., & Papadopoulou, M. (2025). Video-stimulated recall interviews for multimodal awareness raising in communication among adolescent emergent multilinguals. *Frontiers in Communication*, *10*. https://doi.org/10.3389/fcomm.2025.1432271

Wang, W., & Wang, W. (2025). College students' behavioural intentions of AI-assisted language learning: Based on the technology acceptance model. *Journal of Computer Assisted Learning*, *41*(4). https://doi.org/10.1111/jcal.70075

Xiao, Y. (2025). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: A randomized controlled trial. *Humanities and Social Sciences Communications*, *12*(1), 425. https://doi.org/10.1057/s41599-025-04672-8

Yessenbekova, K. (2024). Understanding the dynamics of English-speaking challenges for students in Kazakhstan. *Asian Education and Development Studies*, *13*(4), 373–385. https://doi.org/10.1108/AEDS-04-2024-0083

Zakian, M., Xodabande, I., Valizadeh, M., & Yousefvand, M. (2022). Out-of-the-classroom learning of English vocabulary by EFL learners: Investigating the effectiveness of mobile assisted learning with digital flashcards. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 16. https://doi.org/10.1186/s40862-022-00143-8

Zhang, C. (2023). The effects of emotional intelligence on students' foreign language speaking: A narrative exploration in China's universities. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2023.6296

Zhang, C., Schießl, J., Plößl, L., Hofmann, F., & Gläser-Zikuda, M. (2023). Acceptance of artificial intelligence among pre-service teachers: A multigroup analysis. *International Journal of Educational Technology in Higher Education*, *20*(1), 49. https://doi.org/10.1186/s41239-023-00420-7

Zhao, L., Gašević, D., Swiecki, Z., Li, Y., Lin, J., Sha, L., Yan, L., Alfredo, R., Li, X., & Martinez-Maldonado, R. (2024). Towards automated transcribing and coding of embodied teamwork communication through multimodal learning analytics. *British Journal of Educational Technology*, *55*(4), 1673–1702. https://doi.org/10.1111/bjet.13476

Zou, B., Lyu, Q., Han, Y., Li, Z., & Zhang, W. (2023). Exploring students' acceptance of an artificial intelligence speech evaluation program for EFL speaking practice: an application of the Integrated Model of Technology Acceptance. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2023.2278608