

## Generative AI Mediation in EFL Oral and Written Production: Colombian Student and Teacher Perceptions

Dionelio Jesús Moreno Villalobos<sup>1\*</sup>

<sup>1</sup>Universidad EAN, Bogotá, Colombia

\*Corresponding author's email: Djmorenov@universidadean.edu.co

 <https://orcid.org/0009-0006-8046-1123>

 <https://doi.org/10.54855/ijaile.26321>

® Copyright (c) 2026 Dionelio Jesús Moreno Villalobos

Received: 01/05/2025

Revision: 01/08/2026

Accepted: 11/08/2026

Online: 13/08/2026

### ABSTRACT

**Keywords:** generative artificial intelligence; EFL production; teacher mediation; learner agency; AI literacy

Generative artificial intelligence (GenAI) is increasingly used to support EFL writing and speaking, yet research has focused more heavily on written products than on the combined processes of oral and written production, authorship, and teacher mediation. This qualitative study explored how students and teachers in a Colombian university language program perceived GenAI-supported production. Participants were seven students who completed an open-ended semi-structured questionnaire and five teachers who participated in semi-structured interviews. Data were examined through iterative thematic coding, analytic memoing, and comparison across participant groups. Four patterns emerged: GenAI as a rehearsal and revision partner; tension between polished output and language ownership; teacher mediation shifting toward critical language awareness; and the need for explicit ethical and assessment guidance. The study highlights that GenAI is pedagogically valuable when transparent, process-oriented, and dialogic routines preserve learner agency, human feedback, and responsible authorship.

### Introduction

Generative artificial intelligence has changed the ordinary ecology of language learning. Students no longer approach writing and speaking tasks with only textbooks, dictionaries, classmates, and teacher feedback. They can now ask a chatbot to outline a paragraph, reformulate a sentence, explain an error, produce examples at a given proficiency level, simulate an interview, or prepare possible answers before an oral assessment. Recent studies have shown that GenAI tools can provide language examples, immediate feedback, conversational practice, and writing support, but they also create risks related to overreliance,

uneven feedback quality, inappropriate output, privacy, authorship, and assessment validity (Barrot, 2023; Kohnke et al., 2023; Warschauer et al., 2023; Yang & Li, 2024).

This development is especially relevant in Colombian university EFL contexts. Regional evidence indicates that AI tools are entering higher education and EFL practice rapidly, while their educational effects remain shaped by local conditions, institutional cultures, and unevenly developed guidance (George-Reyes et al., 2025; Gortaire-Díaz et al., 2025; UNESCO, 2023). Within the program examined here, English is linked to academic, professional, and international mobility, and students may use GenAI for planning, translation, drafting, rehearsal, feedback, and revision before teachers and programs have established shared criteria for acceptable assistance, transparency, authorship, and assessment. This mismatch creates the pedagogical gap examined in the study.

For EFL teachers, therefore, the central question is not simply whether students use GenAI. The more pedagogically important question is what happens to learning when a tool participates in the production of language that will later be assessed as evidence of student competence. Oral and written production are not merely products; they are mediated processes through which learners develop communicative competence, strategic control, voice, audience awareness, interactional responsiveness, and responsibility for meaning (Canale & Swain, 1980; Hymes, 1972; Lantolf & Thorne, 2006). When GenAI intervenes in those processes, it may support learning, but it may also obscure who made particular linguistic, rhetorical, or pragmatic decisions (Godwin-Jones, 2024; Warschauer et al., 2023). A polished paragraph or fluent oral script may therefore represent genuine learning, partial assistance, or substitution of learner agency.

The research gap addressed here is threefold. First, reviews and empirical studies show that AI-assisted writing and automated feedback dominate the literature, while speaking and the relationship between oral and written preparation remain less developed (Barrot, 2023; Crompton et al., 2024; Yan, 2024). Second, studies often emphasize affordances, attitudes, or product improvement, leaving authorship, language ownership, and teacher mediation less directly examined as interrelated processes (Kohnke et al., 2023; Pham & Le, 2024; Warschauer et al., 2023). Third, cross-national evidence has begun to include Latin American participants, but locally situated EFL studies from the region remain comparatively limited (George-Reyes et al., 2025; Gortaire-Díaz et al., 2025). The present study addresses these gaps by examining student and teacher perceptions of GenAI across connected oral and written production practices in a Colombian university program.

This article reports a qualitative study situated in a Colombian university language program. It does not ask whether GenAI should be accepted or prohibited. Instead, it asks how students and teachers perceive GenAI as a mediational resource in EFL oral and written production and what classroom routines may support transparent, responsible, and pedagogically meaningful use. The argument developed here is that GenAI becomes educationally meaningful only when its use is made visible, discussed, and connected to process-oriented assessment and teacher mediation. Without such mediation, the tool may improve the appearance of student output while weakening reflection, authorship, spontaneous performance, and human feedback.

## Literature Review

### *GenAI as a Mediational Resource in Language Learning*

This study is grounded in a sociocultural understanding of learning, where tools, signs, interaction, and social activity mediate development (Lantolf & Thorne, 2006; Vygotsky, 1978). From this perspective, GenAI should not be understood as a neutral accessory added to pre-existing language tasks. Tools change the activity in which they participate. A learner who asks a chatbot to compare two ways of expressing an idea is not simply receiving a correction; the learner is entering a mediated interaction in which noticing, choice, and self-regulation may be reorganized.

This mediational framing helps avoid two extremes. The first is technological enthusiasm, which treats GenAI as an automatic path to better learning because it generates immediate language. The second is technological suspicion, which treats every use of GenAI as a shortcut or threat. A mediational perspective asks a more analytical question: What kind of learning activity does the tool make possible, and under what pedagogical conditions does that activity strengthen or weaken learner agency? The same AI-generated suggestion can serve as a resource for reflection if students compare, evaluate, and explain it, or it can become a substitute for learning if they accept it without understanding.

Recent work on distributed agency in AI-supported language education similarly suggests that learning is shaped by the interaction among learners, teachers, tools, tasks, policies, and assessment practices (Godwin-Jones, 2024). This is particularly important in EFL contexts, where learners may overvalue fluent output because it resembles expert language, even when they cannot fully explain the choices that produced it. GenAI, therefore, functions as a mediational resource with uncertain pedagogical effects: it can scaffold production, but it can also narrow learning if used primarily to avoid difficulty.

### *Writing, Automated Feedback, and the Risk of Surface Improvement*

Writing has been one of the most visible areas of GenAI use in language education. Students can use chatbots to generate outlines, correct grammar, propose titles, simplify language, expand vocabulary, improve coherence, and rewrite paragraphs. Research on ChatGPT and L2 writing has reported both possibilities and concerns. Barrot (2023) discusses the potential of ChatGPT for second language writing while warning about overreliance, uneven feedback quality, and the possibility that students may submit polished texts without developing stronger independent control. Escalante et al. (2023), Guo and Wang (2024), and Yan (2024) similarly show that AI-generated feedback can be useful, but only when learners actively process it, and teachers help them interpret the suggestions. More recent empirical work sharpens this distinction. In a double-blind, paired-comparison study, Tsai et al. (2024) found that ChatGPT-assisted revisions raised EFL essay scores, with the largest gains for initially weaker writers, yet cautioned that such revisions do not reveal learners' own writing competence and therefore call for new forms of writing assessment. In a multiple-case study, Yan and Zhang (2024) likewise showed that L2 writers' engagement with ChatGPT feedback varied with proficiency and strategy use, and that learners valued the tool partly because it reduced the discomfort of receiving critical feedback from humans. Read together, these studies indicate that AI-supported writing can improve products and lower affective barriers while leaving open whether independent competence actually develops—the precise tension this study examines in the EFL production process.

This concern is also reflected in recent IJAILE research on AI-supported academic writing. Tran, Le, and Tran (2025) reported that AI tools can support academic writing development

by improving skills and providing assistance, while also creating challenges related to over-reliance and limited support for complex writing problems. Their findings are relevant to the present study because they show that the value of AI support lies not only in the improved final product but also in whether learners understand and appropriate the revisions made.

The distinction between surface improvement and learning is crucial. When a text becomes more fluent after AI support, teachers may find it harder to evaluate the student's actual writing development. The final product may be cleaner, but it may not show stronger planning, organization, editing ability, or ownership of language choices. For this reason, the present study treats AI-supported writing not as a detection problem but as a mediation problem: how students use the tool, what they learn from it, and what they can still explain without it become central pedagogical questions.

### *Oral Production, Rehearsal, and Communicative Agency*

A systematic review of artificial intelligence in English language teaching and learning has mapped affordances across speaking, writing, reading, pedagogy, and self-regulation, alongside recurring challenges such as technological breakdowns, limited tool capabilities, and the standardization of learner language (Crompton et al., 2024). Within this body of work, however, writing remains considerably better documented than speaking. Although writing has received more attention, GenAI also matters for oral production. Learners can use chatbots to prepare answers, create role-play scripts, translate possible expressions, generate vocabulary lists, simulate interviews, and rehearse before speaking in class. As voice-based interfaces become more common, learners may increasingly use AI not only for written preparation but also for spoken interaction. These practices may reduce anxiety and provide additional practice, especially for students who have limited opportunities to speak English outside the classroom.

Recent empirical studies on AI-mediated speaking support this point while also revealing its limits. Nguyen and Le (2025) found that the AI conversational agent, Call Annie, increased students' confidence and willingness to communicate in English-language lessons, although concerns about chatbot-mediated interaction remained. In a Latin American EFL setting, Gortaire-Díaz et al. (2025) likewise found that AI tools were perceived as accessible and helpful for reducing error-related anxiety in classroom language practice. These findings suggest that AI-mediated rehearsal can have affective value, but they do not resolve whether confidence developed with AI transfers to spontaneous human interaction.

Communicative competence provides a useful lens for this concern. Canale and Swain (1980) and Hymes (1972) remind us that language production involves more than formal accuracy. Learners need to make choices appropriate to context, purpose, audience, and interaction. In oral production, they must also manage hesitation, repair, turn-taking, clarification, and unpredictability. GenAI can help students prepare language, but it cannot guarantee that learners can use that language flexibly. Teachers therefore need to design oral tasks that allow AI-supported preparation while still requiring personal response, interaction, and justification.

### *Voice, Authorship, and Language Ownership*

The concept of voice is central to this study because EFL production is not simply a matter of producing correct language. Learners gradually develop ways of positioning themselves, expressing attitudes, making choices, and sounding like legitimate participants in a second or foreign language. GenAI complicates this process because it can generate language that is more fluent, more academic, or more idiomatic than the learner's current repertoire. The result may be impressive, but it may also feel detached from the learner's own intention.

Authorship in AI-supported language production should not be understood as a binary category. A learner may remain meaningfully involved when using AI for planning, prompting, selecting, rejecting, revising, and explaining. However, authorship becomes weaker when the learner cannot account for the decisions that shaped the final version. In this sense, authorship is not only a legal or academic integrity category; it is also a pedagogical category related to language ownership and learner agency.

Pham and Le (2024) found that students perceived ChatGPT as useful for language learning, engagement, and knowledge development, while also warning that over-reliance may limit critical thinking and creativity. This ambivalent pattern is important for the present study because it suggests that GenAI can support learner voice only when students are guided to adapt, question, and appropriate the tool's output rather than treat it as authoritative language.

### *Digital Literacy, Ethics, and Pedagogical Transparency*

Digital literacy in the GenAI era involves more than technical ability. This understanding extends the multiliteracies tradition, which reframes literacy as a plural, socially situated practice shaped by changing technologies and communicative contexts (New London Group, 1996). Students may know how to use a chatbot, but they may not know how to evaluate the quality of its response, identify hallucinated information, recognize inappropriate register, avoid overdependence, or document assistance ethically. In language education, these concerns are intensified because the object of learning is language itself. If a tool generates the language that students are expected to develop, then digital literacy and language pedagogy become inseparable.

Research on AI literacy emphasizes the need for users to understand, evaluate, and critically engage with AI systems (Ng et al., 2021). UNESCO (2023) likewise calls for human-centered, transparent, and responsible use of generative AI in education. For EFL programs, these principles imply that AI policies should not remain abstract institutional statements. They need to be translated into classroom-level agreements and assessment practices that clarify which forms of assistance are permitted, which require disclosure, and which undermine the learning objective.

A Latin American perspective on higher education further complicates this issue. George-Reyes et al. (2025) show that students across Latin America and Europe perceive ChatGPT as useful but complex in its educational impact, reinforcing the need to examine local institutional cultures rather than assume that global technologies are adopted in identical ways. In Colombian EFL classrooms, responsible GenAI use therefore requires attention to access, teacher workload, assessment expectations, and students' developing academic and professional language trajectories.

### *Research Questions*

This study was guided by the following main research question:

How do students and teachers perceive the role of GenAI in EFL oral and written production, particularly in relation to revision, rehearsal, authorship, teacher mediation, assessment, and responsible use?

The analysis also considered three supporting questions:

- How do students describe their use of GenAI when preparing written and oral production tasks?

- How do teachers interpret the pedagogical, ethical, and assessment-related implications of GenAI use?
- What classroom routines may support transparent and responsible GenAI use in EFL production?

## Methods

### *Research Design and Context*

The inquiry follows a qualitative exploratory design with a phenomenological orientation. This orientation was chosen because the study seeks to understand how participants experience and give meaning to GenAI in relation to language learning, rather than to measure the effect of a specific intervention. The study does not claim that GenAI improves or harms EFL production in general. Instead, it examines how students and teachers interpret the tool in relation to learning, assessment, authorship, and pedagogy.

The study is situated in a private Colombian university language program where learners complete communicative tasks, written assignments, oral presentations, and classroom interactions in English. The Colombian EFL context is relevant because students often link English learning to academic performance, professional mobility, and intercultural participation, while teachers must respond to institutional assessment demands and rapidly evolving student practices. GenAI enters this environment not as an isolated innovation but as part of a broader transformation in how students access language support.

### *Participants and Sampling*

A purposeful, non-probabilistic sampling strategy guided participant selection, consistent with the study's qualitative and phenomenological orientation. The focal corpus included 12 information-rich participants: seven students from an English-language and culture course in a modern languages program, aged 17 to 23, and five teachers affiliated with the same program, aged 30 to 55. Participants were selected because they had direct experience with GenAI tools in oral and written production and were willing to participate.

The small sample size is acknowledged as a limitation in breadth, but it is methodologically appropriate for this exploratory qualitative study. The aim was not statistical representativeness, but analytic depth: to identify how participants make sense of GenAI use, where tensions emerge, and how student and teacher perspectives converge or diverge. In this sense, the findings are intended as analytically transferable insights rather than generalizable claims about all Colombian EFL learners or teachers.

Table 1.

## Participant Profile for the Article Corpus

| Participant group | n  | Age range | Primary data source                                    | Role in analysis                                                         |
|-------------------|----|-----------|--------------------------------------------------------|--------------------------------------------------------------------------|
| Students          | 7  | 17-23     | Open-ended semi-structured questionnaire through Forms | Accounts of use, benefits, concerns, ownership, and transparency         |
| Teachers          | 5  | 30-55     | Semi-structured in-depth interviews                    | Interpretations of mediation, assessment, feedback, and ethical guidance |
| Total             | 12 | 17-55     | Qualitative perception data                            | Focal corpus for thematic analysis                                       |

*Data Collection*

Data collection followed two clearly differentiated phases. First, the seven student participants completed an open-ended semi-structured questionnaire through Forms. The student questionnaire asked about uses, benefits, concerns, and experiences with AI-supported oral and written production. It included prompts about planning, vocabulary, grammar correction, speaking rehearsal, confidence, dependence, teacher guidance, disclosure, and perceived risks.

Second, the five teacher participants completed semi-structured interviews. These interviews did not recruit a smaller subgroup from the student questionnaire; rather, they constituted the teacher component of the corpus. The interviews explored GenAI in relation to feedback, authorship, dependence, learner confidence, teacher mediation, assessment, and ethical responsibility. The semi-structured protocol included guiding questions such as: How do you perceive students' use of GenAI in oral and written production? What forms of AI assistance appear pedagogically acceptable or problematic? How does GenAI affect feedback, authorship, and assessment? What kind of teacher guidance is needed for responsible use?

Students, therefore, provided open-ended questionnaire accounts, while teachers provided interview data. The two sources were compared to identify convergences and contrasts between learners' experiences and teachers' interpretations. Documentary materials provided contextual support, but participant perception data remained the primary empirical basis of the article. To make the evidence traceable while protecting anonymity, the teacher interview excerpt is identified with a T code, whereas student questionnaire excerpts are identified sequentially as SR1, SR2, and so forth. The SR labels distinguish the quoted response excerpts used in the analysis rather than asserting participant-level links beyond the anonymized evidence presented here.

*Data Analysis and Trustworthiness*

The analysis used iterative thematic coding supported by analytic memos. Initially, the researcher read the data several times to identify recurrent meanings related to revision, rehearsal, voice, authorship, autonomy, feedback, mediation, dependence, assessment, and responsible use. Preliminary codes were then grouped into broader interpretive categories. Following Braun and Clarke's (2006, 2021) approach to thematic analysis, the process

involved moving between coded segments, emerging themes, participant accounts, and the overall research question.

Trustworthiness was strengthened through four procedures. First, the analysis compared student and teacher perspectives as a form of source triangulation. Second, dated analytic memos documented decisions about coding and theme formation and recorded the boundary maintained between descriptive findings and interpretive or pedagogical claims. Third, disconfirming and ambivalent cases were retained in the findings rather than removed; for example, some students described GenAI as useful support, while others warned that it could produce language beyond their real level. Fourth, an audit trail was maintained through dated coding notes, successive versions of the thematic tables, and a documented record of analytic decisions.

Because the study is qualitative and interpretive, reliability was not treated as inter-rater statistical agreement. Instead, methodological rigor was pursued through transparency of procedures, reflexivity, thick description of context, attention to divergent cases, and clear linkage between participant excerpts and interpretive claims (Lincoln & Guba, 1985). The researcher's position as a language educator provided contextual sensitivity but also required reflexive caution so that the analysis did not simply confirm prior pedagogical expectations.

## Results/Findings

The analysis points to four interrelated patterns that help explain how GenAI is experienced in EFL production tasks. The findings below are organized to foreground empirical evidence, with interpretive and pedagogical claims developed in the Discussion and Pedagogical Implications sections. Each pattern is supported by student and teacher accounts, including divergent views and tensions within the corpus.

Across student accounts, GenAI most frequently appeared as a resource for organizing ideas, expanding vocabulary, correcting language, preparing for oral tasks, and reducing uncertainty before performance. Teachers, however, interpreted these benefits alongside concerns about authorship, over-polished output, and the loss of human feedback. This contrast is important: students tended to emphasize usefulness and confidence, while teachers tended to emphasize mediation, assessment, and the need for transparent criteria.

**Table 2. Evidence-Based Thematic Structure**

| Theme                                     | Evidence from participants                                                                                        | Tension identified in the data                                                           | Analytical meaning                                                                         |
|-------------------------------------------|-------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| GenAI as rehearsal and revision partner   | Students described using AI to organize ideas, find vocabulary, rehearse answers, and receive immediate feedback. | Support lowered anxiety but could also reduce productive struggle if accepted passively. | GenAI functioned as a low-stakes preparation space whose value depended on reflective use. |
| Polished output versus language ownership | Some students valued clearer language; others noted that AI could produce language beyond their real level.       | A better-looking text or script did not always indicate stronger independent control.    | Authorship became a process of selecting, adapting, rejecting, and explaining AI output.   |

|                                                  |                                                                                                                               |                                                                                                        |                                                                                                             |
|--------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Teacher mediation as critical language awareness | Teachers emphasized that AI cannot replace the affective and formative dimension of human feedback.                           | AI feedback was fast, but human mediation was needed to interpret appropriateness, level, and purpose. | Teacher work shifted from correction alone toward judgement, transparency, and critical language awareness. |
| Need for ethical and assessment guidance         | Participants expressed uncertainty about disclosure, acceptable support, and how AI-supported performance should be assessed. | Unclear rules encouraged private judgement and possible concealment.                                   | Responsible use required explicit classroom agreements and process evidence.                                |

### *GenAI as a Low-Stakes Rehearsal and Revision Partner*

The first pattern concerns the way students experience GenAI as a low-stakes partner for rehearsal and revision. Students described using GenAI at moments when production felt difficult to begin. In writing, they used it to organize ideas, improve vocabulary, compare possible phrasings, and revise grammar. In speaking, they used it to prepare answers, anticipate vocabulary, rehearse ideas, and reduce anxiety before participating in class or completing an oral task.

One student summarized this supportive role as follows:

"GenAI has helped me organize my ideas. I feel more confident when structuring a sentence, and I have learned vocabulary because it offers synonyms and examples for speaking in an interview. The feedback is immediate; sometimes it is not as personalized as teacher feedback, but it guides me a lot." (SR1)

A second response described a more selective use:

"I normally try to use it as a model without finishing - a draft or brainstorm from which I can take various ideas and build a new one." (SR2)

These accounts show that students did not necessarily experience AI as a final author; several framed it as a starting point for planning, vocabulary expansion, and initial organization.

The same pattern appeared in oral rehearsal. One student reported using AI as a base for a presentation and then adapting it:

"I take the answer as a base for how to begin the presentation; after that I keep building it, I tell the AI how I want it to be and then I do the final adaptation, in which I change words, structures, and so on, depending on what I want to achieve." (SR3)

This account suggests an iterative relationship between learner intention and AI-generated support rather than simple copy-and-paste use.

However, the pattern was not entirely positive. Students recognized that rehearsal support may lead to dependence when the generated version replaces the learner's own effort. This tension is important because GenAI appears to lower the threshold for participation, but the data do not show that this support automatically becomes deeper learning. It becomes pedagogically meaningful only when students adapt, explain, and retain control over the language they use.

### *Polished Output Does Not Always Mean Language Ownership*

The second pattern concerns the tension between polished output and language ownership. Students appreciated clearer phrasing, more precise vocabulary, and improved organization, yet some also recognized that AI-supported production could exceed their real level of control. One student stated:

"In my case, there were times when the AI said things I would not know how to formulate on my own, demonstrating a level I did not really have and would not know how to demonstrate if I had to do it without AI." (SR4)

This account illustrates a level-displacement problem, in which AI-supported output can outpace the learner's independent competence.

Another student expressed a similar risk in relation to oral presentations:

"The main risk is that AI suggests vocabulary or grammatical structures that are too advanced or complex, that we have not yet seen in class. This makes the text lose its naturalness and the student fails to memorize or defend the oral presentation well." (SR5)

This quotation is important because it links advanced vocabulary, loss of naturalness, and inability to defend the performance. In other words, the problem is not only ethical but also communicative and pedagogical.

A divergent view also appeared in the data. One student argued that AI does not necessarily create a false level because the learner still has to perform:

"Structurally, maybe the AI is better than the student; but in expressing ideas, body language, and pronunciation, the student will always have a better level, and likewise in improvising on the spot." (SR6)

This disagreement suggests that AI-supported production should not be reduced to plagiarism or dependence. It may displace language ownership in some cases, but it may also support students when they remain able to adapt, pronounce, explain, and respond spontaneously.

These accounts show that authorship in AI-supported language production is negotiated through multiple decisions: what the student wants to say, how the student prompts the tool, which suggestions are accepted and which are rejected, how the final text or script is revised, and whether the student can explain the final choices. The evidence, therefore, supports a process-based understanding of ownership rather than a binary distinction between AI use and no AI use.

### *Teacher Mediation Shifts from Correction to Critical Language Awareness*

The third pattern concerns the transformation of teacher mediation. Teacher accounts did not describe GenAI as making pedagogy irrelevant. Instead, they suggested that teachers must help learners interpret, evaluate, and contextualize AI support. One teacher summarized this concern as follows:

"There is more confidence for speaking in simulated contexts, but GenAI lacks the affective and formative dimension of human feedback." (T1)

This comment captures a central concern among teachers: AI may provide rapid language support, but it does not replace the relational, affective, and formative dimensions of teacher feedback.

Student accounts also confirmed the value of guidance. One student explained:

"It helped me that the teacher always made us see AI not as a problem but as a tool. Likewise, the teacher gave us ideas about how to ask the AI correctly - not so that it would do the work for us, but so that we could receive orientation in case of confusion." (SR7)

This excerpt shows that teacher mediation can reposition AI use from a hidden shortcut to a guided learning resource.

Another student described teacher-guided oral practice with a conversational interface:

"My English teacher recommended I use the ChatGPT avatar to practise as if I were having a real conversation or doing an exposition. This tool helped me improve my pronunciation, corrected mistakes, and suggested expressions that sounded more natural in English." (SR8)

This account suggests that teacher mediation can shape the purpose of AI use, especially when the tool is framed as rehearsal and feedback rather than as a substitute for performance.

The difference between student and teacher perspectives is therefore not a simple opposition. Students tended to value immediacy, confidence, and vocabulary support. Teachers tended to emphasize the need for human feedback, appropriateness, and responsibility. Together, these perspectives suggest that GenAI shifts teacher mediation from correction alone toward critical language awareness: helping learners decide whether AI-generated language is appropriate, understandable, level-consistent, and genuinely theirs.

#### Ethical and Assessment Guidance Is a Pedagogical Need

The fourth pattern concerns the need for explicit ethical and assessment guidance. Students did not speak about disclosure in a uniform way. Some viewed it as necessary; others treated it as dependent on the teacher or on how much of the AI output was used. One response linked disclosure to teacher reactions:

"It depends on the teacher, since some take it badly and would make an issue of it." (SR9)

A contrasting response framed unacknowledged use as acceptable when AI served only as inspiration:

"It is not necessary, as long as it is not exactly the same, but rather you take what it gave you as inspiration." (SR10)

These responses suggest that unclear rules lead students to make private decisions about what counts as legitimate use.

Other students articulated a stronger position on transparency. One stated:

"Yes, I consider it important to inform the teacher when artificial intelligence is used to prepare an oral activity. This promotes transparency and lets the teacher understand how the work was done. Moreover, using AI as support does not take merit away from learning when it is used responsibly." (SR11)

One respondent explained that disclosure helps "show what we learned or whether we have doubts about something the AI explained" (SR12).

These accounts link transparency not only to integrity but also to formative feedback.

Assessment tensions also arose when students considered whether AI could make a performance appear stronger than the learner's actual level:

One respondent warned that memorizing AI-generated language could lead to an "excellent" presentation without understanding what was being said (SR13).

Another observed that AI may introduce more technical vocabulary when asked to produce a script (SR14).

Together, these accounts support the claim that AI-mediated production creates a visibility problem: teachers may see the product but not the process by which it was produced.

The Colombian university setting matters here because students and teachers are negotiating the use of GenAI in a context where institutional expectations, course-level practices, and everyday student uses are still being aligned. The data do not support a simplistic ban, but neither do they support uncritical adoption. They instead point to the need for explicit agreements regarding acceptable support, disclosure, process evidence, and spontaneous demonstration of communicative control.

## Discussion

The findings show that GenAI is best understood as a pedagogically unstable mediational resource. It can support EFL production by providing vocabulary, examples, rehearsal, feedback, and confidence, but it can also create dependence, level displacement, loss of ownership, and uncertainty about assessment. This interpretation extends previous research by showing that the central issue is not only whether AI improves output but also whether students remain able to explain, adapt, and perform in the language that AI helped them prepare.

The writing-related findings are similar to Barrot (2023), Guo and Wang (2024), Yan (2024), and Tran et al. (2025) in that they confirm the usefulness of AI feedback and revision support while also warning against overreliance. The difference is that the present study foregrounds language ownership as an interpretive and pedagogical issue. In this Colombian EFL corpus, students did not simply ask whether AI made writing better; they also questioned whether AI-generated language still represented their own level and voice. This adds a learner-ownership dimension to discussions often centered on feedback quality or writing improvement.

The oral-production findings also converge with Nguyen and Le (2025) and Gortaire-Díaz et al. (2025), who report that AI-supported practice may increase confidence and reduce anxiety. However, the present study qualifies that claim by showing that confidence may be rehearsal-bound. A student may feel safer after practicing with AI, but that confidence may not transfer to spontaneous interaction if the learner memorizes a script or uses language that cannot be defended. Thus, AI-mediated speaking should be evaluated not only through perceived confidence but also through spontaneous follow-up, explanation, and adaptation.

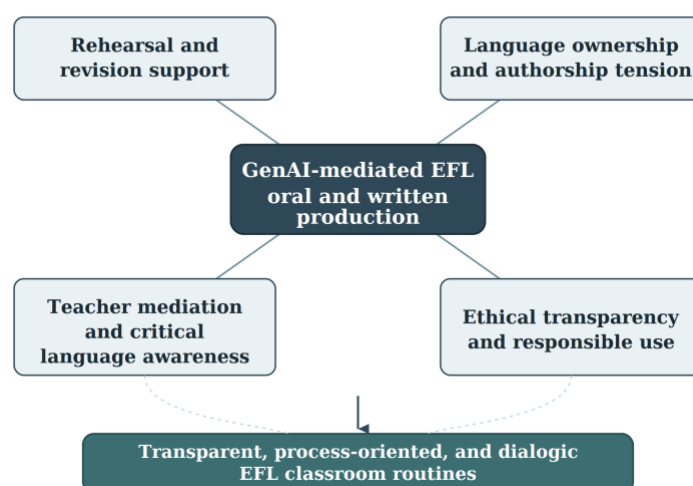
The findings further extend the work of Pham and Le (2024) and Kohnke et al. (2023) by showing that critical AI literacy is not an abstract competence. It appears in small classroom decisions: choosing vocabulary the student can actually use, rejecting language that sounds too advanced, disclosing AI assistance, or asking the teacher how to use AI without outsourcing the task. In this sense, teacher mediation does not restrict autonomy; it creates the conditions for critical autonomy. Without such mediation, students may be operationally independent but pedagogically dependent.

A counterargument must also be considered. GenAI may expand access to language support for students who lack opportunities for practice outside the classroom. It can provide immediate input, reduce embarrassment, and offer explanations at any time. From this perspective, overly restrictive AI use may reduce equitable access to support. Nevertheless, the negative evidence in this study suggests that access alone is insufficient. If AI-generated language is too advanced, if students cannot defend their performance, or if disclosure rules are unclear, the tool may produce a more fluent product while making learning less visible. The pedagogical task is therefore not to choose between access and restriction, but to design mediated access.

The study's unique contribution is its integrated view of oral and written production, teacher mediation, and authorship from a Colombian university EFL context. Rather than treating GenAI solely as a writing assistant, the findings show that students move between writing and speaking: they draft, translate, rehearse, script, revise, and prepare oral responses with the same tool. This connection means that AI-aware pedagogy must address productive skills together and design assessment practices that distinguish preparation from performance.

Figure 1.

Conceptual Model of Responsible GenAI Mediation in EFL Oral and Written Production



*Note.* The figure is read from the four surrounding dimensions inward to the central practice and downward to the outcome. The four dimensions correspond to the interpretive patterns reported in the findings.

The model synthesizes the four interpretive patterns identified in the findings. It places GenAI-mediated EFL oral and written production at the center, surrounded by the dimensions through which participants experienced the tool: rehearsal and revision support; the tension between polished output and language ownership; the repositioning of teacher mediation toward critical language awareness; and the need for ethical transparency and responsible use. The downward arrow indicates that these dimensions do not function in isolation. GenAI becomes pedagogically valuable only when its use is converted into transparent, process-oriented, and dialogic classroom routines through teacher guidance, learner reflection, and ethical transparency. In the absence of such mediation, the same tool may yield more polished output while leaving learning less visible.

### *Pedagogical Implications for EFL Programs*

The implications below are derived from the findings rather than presented as general advice detached from the data. Because students reported both support and risk, EFL programs should avoid categorical responses that either ban GenAI completely or allow it without guidance.

First, language programs should teach the use of AI as a decision-making process. The student who described using AI as a draft or brainstorm illustrates a productive pattern: the learner begins with AI support but builds a personal version. Teachers can reinforce this by asking students to identify what they accepted, changed, rejected, and learned.

Second, the assessment should include evidence of process. The level-displacement accounts show that a polished final product may not represent independent competence. Process-oriented assessment can include initial drafts, AI-assisted revisions, oral defenses, spontaneous follow-up questions, short reflection notes, and teacher-student conferences.

Third, transparency should be normalized rather than treated only as a disciplinary matter. Several students linked disclosure to fear of teacher reaction, while others linked it to responsible learning. This means that students need clear, task-specific statements explaining when AI use is allowed, when it must be declared, and which parts of the task must be produced independently.

Fourth, teacher education should focus on AI-aware task design. Teachers need examples of prompts, rubrics, disclosure statements, revision activities, speaking tasks, and feedback routines. The findings suggest that teachers remain central because they help learners evaluate whether AI-generated language is appropriate for level, purpose, audience, and communicative context.

### *A Classroom Protocol for Responsible GenAI Use*

The classroom protocol proposed here is grounded directly in the study's findings: students valued GenAI as support, warned against overdependence, and asked for clearer guidance about transparency and assessment. It is offered as a transferable framework rather than a prescriptive sequence, to be adapted to local proficiency levels, task types, and institutional norms.

Before submitting an AI-supported written or oral task, learners can complete four brief prompts: (a) I used GenAI to help me with...; (b) the most useful suggestion was...; (c) I accepted, changed, or rejected the suggestion because...; and (d) this is what I can now explain or perform without the tool. These prompts make the process visible without turning the assignment into surveillance.

For writing tasks, students can submit an initial draft, an AI-assisted revision, and a short commentary explaining changes. For speaking tasks, students can use GenAI during preparation but must answer at least one unscripted follow-up question, respond to a peer, or explain a vocabulary choice. For feedback tasks, students can compare teacher feedback and AI feedback, identify where they agree or differ, and justify which suggestion they used.

The protocol should be adapted to the proficiency level. Beginners may use simple disclosure frames such as "I used AI for vocabulary" or "I changed these words." More advanced students can write fuller reflective commentaries. The goal is not to punish support, but to protect the distinction between assistance, appropriation, and substitution.

### *Limitations and Future Directions*

This study has limitations that should be stated explicitly. First, the sample is small ( $n = 12$ ), which limits the breadth of perspectives represented. The findings should therefore be read as exploratory and analytically transferable rather than statistically generalizable. Second, the study is situated in a single Colombian university language program; institutional culture, teacher practices, student access to technology, and assessment norms may differ in other contexts.

Third, the study relies primarily on self-reported perceptions through student questionnaire responses and teacher interviews. These data are valuable for understanding meaning-making, but they do not directly measure learning outcomes or demonstrate actual improvement in oral or written production. Fourth, the study did not include classroom observation, AI interaction logs, prompt histories, written drafts, or recorded oral performances. As a result, the analysis cannot fully compare participants' reported behavior with their actual behavior during AI-supported production.

Future research should incorporate larger, more diverse participant groups, compare institutions, and include classroom observations, AI use logs, writing samples, prompt records, and oral performance data. Longitudinal designs would be especially valuable for examining whether AI-supported rehearsal and revision lead to durable language development or mainly improve immediate task performance. Future studies should also investigate Latin American EFL contexts more systematically, since current international evidence remains concentrated in other regions.

### **Conclusion**

GenAI is already reshaping how learners approach EFL writing and speaking. The question for teachers is not whether the technology exists, but how language programs will respond pedagogically. The findings discussed in this article suggest that students and teachers are not simply divided between enthusiasm and resistance. They are negotiating a complex terrain in which GenAI can support revision, rehearsal, vocabulary growth, and confidence while also raising concerns about authorship, dependence, level displacement, assessment, and the loss of human feedback.

The unique contribution of this study is to show that GenAI-mediated EFL production should be understood as a matter of mediated ownership. Students may use AI responsibly when they select, adapt, explain, and disclose support. They may lose ownership when they accept language they cannot understand, defend, pronounce, or adapt. Teachers, therefore, remain essential as mediators of critical language awareness, process visibility, and ethical judgment.

For EFL programs, this means treating GenAI as a curricular issue rather than merely a disciplinary problem. Students need to learn how to use AI support without surrendering ownership of their language. Teachers need institutional support to design tasks that make the process visible. Programs need shared expectations that distinguish responsible assistance from substitution. If these conditions are present, GenAI may help learners develop greater awareness of language choices. If they are absent, the tool may simply produce better-looking texts and scripts while leaving learning less visible.

### *Compliance with Ethical Standards*

Ethical considerations: The study followed institutional ethical principles for research involving human participants. Participants were informed about the purpose of the study, the

voluntary nature of their participation, and the confidential treatment of their responses. The use of open-ended questionnaires and interviews was aligned with the qualitative purpose of understanding perceptions rather than evaluating participant performance.

**Informed consent:** Informed consent was obtained from participants before data collection. Participants were informed that their contributions would be anonymized and used for academic research purposes.

**Conflict of interest:** The author declares no conflict of interest.

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Data availability:** Due to the qualitative nature of the study and the need to protect participant confidentiality, the raw data are not publicly available. Anonymized excerpts may be reported in publications when they do not compromise participant identity.

#### *Declaration of the Use of AI Technologies in the Writing Process*

During the preparation of this manuscript, OpenAI was used only for language editing, clarity enhancement, organization of ideas, and editorial refinement. The author was responsible for the study design, theoretical framing, data interpretation, argument development, reference verification, and final approval of the manuscript. No AI tool was used to generate research data, conduct data analysis, fabricate references, or replace scholarly judgment. No AI tool is listed as an author.

## References

- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47. <https://doi.org/10.1093/applin/1.1.1>
- 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>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
- George-Reyes, C. E., Avello-Martínez, R., & Buenestado-Fernández, M. (2025). Perceptions of ChatGPT and the complexity of its impact among higher education students: Evidence across ten countries of Latin America and Europe. *Educational Process: International Journal*, 15, Article e2025171. <https://doi.org/10.22521/edupij.2025.15.171>

- Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5-30. <https://doi.org/10.64152/10125/73570>
- Gortaire-Díaz, D., Almache-Granda, G., Real-Roby, R., & Mora-Herrera, E. (2025). Bridging the gap: How AI tools support self-directed EFL skill development inside the classroom. *Revista InveCom*, 6(2), Article e602012.
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Penguin.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60-92. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Nguyen, L. A. D., & Le, T. T. P. (2025). Exploring the effects of an AI chatbot on emotional engagement in English speaking lessons: Insights from Call Annie. *International Journal of AI in Language Education*, 2(2), 79-99. <https://doi.org/10.54855/ijaile.25225>
- Pham, V. P. H., & Le, A. Q. (2024). ChatGPT in language learning: Perspectives from Vietnamese students in Vietnam and the USA. *International Journal of Language Instruction*, 3(2), 59-72. <https://doi.org/10.54855/ijli.24325>
- Tran, H. N., Le, T. T. N., & Tran, V. B. U. (2025). AI tools in learning academic writing: Benefits and challenges for MA students in the English language studies at the Industrial University of Ho Chi Minh City. *International Journal of AI in Language Education*, 2(1), 74-91. <https://doi.org/10.54855/ijaile.25215>
- Tsai, C.-Y., Lin, Y.-T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, 29(17), 22427-22445. <https://doi.org/10.1007/s10639-024-12722-y>
- UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or

- foreign language. *Journal of Second Language Writing*, 62, Article 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
- Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), 1-19. <https://doi.org/10.64152/10125/73597>
- Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11, Article 1086. <https://doi.org/10.1057/s41599-024-03543-y>
- Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, 124, Article 103351. <https://doi.org/10.1016/j.system.2024.103351>

## Biodata

Dionelio Jesús Moreno Villalobos is a university language professor and doctoral researcher in Education based in Bogotá, Colombia. His research focuses on generative artificial intelligence in language education, multilingualism, digital literacies, learner autonomy, and authentic assessment. He has experience in university language teaching, academic research, translation, and scholarly publishing.