Research Club Scientia Vol. 4, No 4
Frecuencia: Octubre - Diciembre
ISSN: 3103 - 1315
DOI: https://doi.org/10.65415/2aj29975

# *Fostering Speaking Skills Through AI-Moderated Debate: Exploring Artificial Intelligence in EFL Instruction*

## *Fomento de las habilidades de expresión oral mediante debates moderados por inteligencia artificial: exploración de la inteligencia artificial en la enseñanza del inglés como lengua extranjera*

## *Sviluppo delle abilità di produzione orale attraverso dibattiti moderati dall'intelligenza artificiale: esplorazione dell'intelligenza artificiale nell'insegnamento dell'inglese come lingua straniera*

**Mauricio José Sánchez Galindo**[1]
**Email:** msanchezg40@unemi.edu.ec
**ORCID:** https://orcid.org/0009-0004-2818-6430

**Vanessa Monserrat Vélez Santana**[2]
**Email:** vvelezs3@unemi.edu.ec
**ORCID:** https://orcid.org/0009-0008-0637-6830

**Ruben Enrique Hidalgo Mosqueda**[3]
**Email:** rhidalgom5@unemi.edu.ec
**ORCID:** https://orcid.org/0009-0004-0076-0939

**Jorge Francisco Zambrano Pachay**[4]
**Email:** jzambranop10@unemi.edu.ec
**ORCID:** https://orcid.org/0000-0001-9456-2765

**Correspondencia:** Mauricio José Sánchez Galindo[1] **Email:** msanchezg40@unemi.edu.ec

**Artículo de investigación**

* **Recibido:** 15-09-2026 * **Aceptado:** 23-11-2026 * **Publicado:** 25-11-2026

1. Facultad de Posgrados, Universidad Estatal de Milagro, Ecuador.
2. Facultad de Posgrados, Universidad Estatal de Milagro, Ecuador.
3. Facultad de Posgrados, Universidad Estatal de Milagro, Ecuador.
4. Facultad de Posgrados, Universidad Estatal de Milagro, Ecuador.

**Abstract**

This study examined the influence of AI-moderated debates on the development of speaking skills in English as a Foreign Language (EFL) instruction. A mixed-methods, quasi-experimental design was implemented at a private educational institution in Machala, Ecuador. Twenty second-year high school students were divided equally into experimental and control groups. Over one month, the experimental group attended four debate sessions moderated by ChatGPT, which organized speaking turns, generated follow-up questions, and provided feedback based on a six-criterion analytic rubric. Speaking performance was assessed through pre-tests and post-tests, while six English teachers completed a questionnaire exploring their perceptions of the strategy. The experimental group's overall rubric score increased from 14 to 22 out of 30, compared with an increase from 13 to 18 in the control group. The most noticeable improvements involved fluency, coherence, vocabulary, and argumentation, whereas interaction and responsiveness remained unchanged in the experimental group. Teachers generally expressed favorable views, although some raised concerns about the feasibility of AI moderation in classroom settings. These findings suggest that AI-moderated debates may support speaking development by combining structured oral interaction with targeted feedback. Nevertheless, the limited sample size, short intervention period, and methodological constraints prevent definitive claims about their superiority over conventional instruction.

**Keywords:** artificial intelligence, speaking skill, AI-moderated debates, English as a foreign language, language teaching.

**Resumen**

El presente estudio examinó la influencia de los debates moderados por inteligencia artificial en el desarrollo de la expresión oral en inglés como lengua extranjera (EFL). Se empleó un enfoque mixto con un diseño cuasiexperimental en una institución educativa privada de Machala, Ecuador. Participaron 20 estudiantes de segundo de bachillerato, distribuidos en un grupo experimental y otro de control, con diez integrantes cada uno. Durante un mes, el grupo experimental participó en cuatro sesiones de debate moderadas por ChatGPT, que organizó las intervenciones, formuló preguntas de seguimiento y proporcionó retroalimentación basada en una rúbrica analítica de seis criterios. El desempeño oral se evaluó mediante pruebas iniciales y finales, mientras que seis docentes de inglés respondieron un cuestionario sobre sus percepciones de la estrategia. Los resultados mostraron que el grupo experimental aumentó su puntuación de 14 a 22 puntos sobre 30, frente al incremento de 13 a 18 puntos registrado en el grupo de control. Las mayores mejoras se observaron en fluidez, coherencia, vocabulario y argumentación, mientras que la interacción y capacidad de respuesta permanecieron sin cambios en el grupo experimental. Las percepciones docentes fueron mayoritariamente favorables, aunque surgieron reservas sobre la viabilidad de la estrategia. Los hallazgos sugieren que los debates moderados

por inteligencia artificial pueden favorecer la práctica oral estructurada, aunque las limitaciones metodológicas impiden establecer su superioridad frente a la enseñanza convencional.

**Palabras clave:** inteligencia artificial, expresión oral, debates moderados, inglés como lengua extranjera, enseñanza de idiomas.

**Riassunto**

Il presente studio ha esaminato l'influenza dei dibattiti moderati dall'intelligenza artificiale sullo sviluppo delle competenze di produzione orale nell'insegnamento dell'inglese come lingua straniera (EFL). È stato adottato un approccio metodologico misto, con un disegno quasi-sperimentale, presso un istituto scolastico privato di Machala, in Ecuador. Hanno partecipato 20 studenti del secondo anno di scuola secondaria superiore, suddivisi equamente in un gruppo sperimentale e uno di controllo. Nell'arco di un mese, il gruppo sperimentale ha svolto quattro sessioni di dibattito moderate da ChatGPT, utilizzato per organizzare i turni di parola, formulare domande di approfondimento e fornire riscontri basati su una rubrica analitica composta da sei criteri. Le competenze orali sono state valutate attraverso prove iniziali e finali, mentre sei docenti di inglese hanno compilato un questionario sulle proprie percezioni della strategia. Il punteggio del gruppo sperimentale è aumentato da 14 a 22 punti su 30, rispetto all'incremento da 13 a 18 punti registrato nel gruppo di controllo. I miglioramenti più evidenti hanno riguardato fluidità, coerenza, lessico e argomentazione, mentre le capacità di interazione e risposta sono rimaste invariate nel gruppo sperimentale. Le valutazioni dei docenti sono risultate prevalentemente positive, pur evidenziando alcune perplessità sulla fattibilità dell'approccio. I risultati suggeriscono che i dibattiti moderati dall'intelligenza artificiale possano favorire lo sviluppo delle competenze orali, sebbene i limiti metodologici non consentano di affermare la superiorità rispetto all'insegnamento tradizionale.

**Parole chiave:** intelligenza artificiale, produzione orale, dibattiti moderati dall'IA, inglese come lingua straniera, didattica delle lingue.

## Introduction

Developing speaking proficiency remains one of the most demanding goals in English as a Foreign Language instruction. Oral communication requires learners to integrate not only fluency, pronunciation, but also grammatical accuracy, lexical resources, coherence, as well as interactional competence while responding to others in real time. In several EFL settings, however, opportunities for sustained oral communication are limited, which may restrict students' confidence and willingness to speak. Recent research consequently emphasizes the need for instructional practices that increase meaningful oral production and expose learners to communicative situations in which English is used purposefully rather than reproduced mechanically.

Artificial intelligence has broadened the possibilities for creating such environments. AI-supported conversational systems can generate prompts, sustain interaction, adapt questions, and provide learners with repeated opportunities to use English orally. As it was stated by Fathi et al. (2024) AI-mediated interaction improved EFL learners' speaking performance and willingness to communicate, including dimensions such as fluency, coherence, vocabulary, grammatical accuracy, and pronunciation. Similarly, Tai and Chen (2024) mentioned positive effects of generative AI chatbot interaction on speaking development. Shafiee Rad (2024) also portrayed that an AI-based speaking application contributed to improvements in L2 speaking proficiency and willingness to communicate. Collectively, these findings suggest that AI can provide additional spaces for communicative practice when it is incorporated into clearly structured learning activities.

The affective dimension of speaking is equally relevant. Learners may hesitate to participate because of anxiety, fear of making mistakes, or low perceptions of their communicative competence. Recent studies highlight that AI-supported interaction may help address some of these barriers. Research on AI speaking assistants has identified increases in enjoyment and willingness to communicate alongside reductions in foreign-language anxiety. In the same way, studies comparing conversational generative AI tools have examined their effects on speaking anxiety, self-perceived communicative competence, and oral performance. Zhang et al. (2024) have uncovered that interaction with an AI chatbot could strengthen learners' willingness to communicate, although their results also showed that the pedagogical effects of different AI systems are not necessarily identical.

Nevertheless, the availability of AI does not ensure language development. Its effectiveness depends on pedagogical design, the nature of the interaction, as well as the role assigned to the technology. Ericsson and Johansson (2023) found that conversational AI can provide sustained speaking opportunities and support learners' cognitive, emotional, and social engagement. Zhai and Wibowo (2023) similarly highlighted the relevance of AI dialogue systems for developing interactional competence among university EFL learners. More recently, Du and Daniel (2024), in a systematic review of AI-powered chatbots for EFL speaking practice,

showed growing evidence of their potential while also identifying the need for stronger pedagogical frameworks. Lai and Lee (2024) concluded that conversational AI research in English language teaching has expanded rapidly, although much of the existing evidence remains concentrated in Asian EFL contexts.

Debate offers an appropriate pedagogical structure for integrating these technological affordances. Debate requires learners to present not only viewpoints, construct arguments, respond to opposing positions, request clarification, defend claims, but also react spontaneously. Pham (2024) explained that debate-based instruction contributed to improvements in fluency, vocabulary use, critical thinking, and confidence among EFL learners. Similar findings were reported by Ban et al. (2023), who linked English debate participation with the development of speaking skills among university learners. More recent research conducted by Oktaviani et al. (2025) mentioned gains in speaking performance, confidence, and classroom interaction. These studies support the use of debate as a communicative technique that combines language production with higher-order reasoning.

Within this framework, AI can assume a more specific role as a debate moderator. Instead of generating arguments for students, the system can introduce the debate motion, establish rules, organize speaking turns, control time, ask follow-up questions, request clarification, and maintain the progression of the interaction. Such a configuration preserves learner-to-learner communication while assigning AI a regulatory and supportive function. Current research has largely examined AI as a conversational partner, tutor, feedback provider, or speaking assistant; comparatively less attention has been given to AI as a moderator of structured EFL debates. This gap is relevant because recent systematic reviews have called for more rigorous, context-sensitive investigations of conversational AI and broader applications beyond conventional chatbot interaction (Du & Daniel, 2024; Lai & Lee, 2024; Zhai & Wibowo, 2023).

Accordingly, the problem addressed in this study concerns the limited empirical evidence regarding how AI-moderated debates influence high school EFL students' speaking performance and perceptions of oral interaction. The general objective is to examine the influence of AI-moderated debates on high school EFL students' speaking performance and their perceptions of this instructional strategy through a mixed-methods approach. Quantitative evidence will be obtained through a speaking pre-test and post-test applied to high school students, whereas qualitative evidence will be gathered through a survey with a likert scale administered to university and high school English instructors. By combining measurable speaking outcomes with participant and teacher perspectives, the study seeks to determine not only whether changes occur but also how AI moderation is perceived pedagogically. To conclude, the contribution of the study lies in positioning AI as an interactional moderator rather than as a substitute for human communication.

## Methodology

### Research Design

The study adopted a mixed-methods, quasi-experimental design to examine the influence of AI-moderated debates on the speaking performance of second-year high school students. The quantitative component was based on a pre-test/post-test comparison between an experimental group and a control group, together with a Likert-scale teacher survey. The qualitative component was an optional open-ended comment included in the teacher survey, which was used to contextualize the numerical response patterns when comments were provided. This design was appropriate since the study sought to determine measurable changes in oral performance while also considering educators' interpretations of the pedagogical acceptability of the intervention.

Recent research has shown that AI-mediated interaction can support EFL speaking development and willingness to communicate when it is mixed in structured oral tasks (Fathi et al., 2024). Similar experimental work has also reported positive speaking outcomes when generative AI is incorporated into guided language-learning interactions (Tai & Chen, 2024).

### Setting and Participants

The intervention was conducted in a private educational institution located in Machala, in the coastal region of Ecuador. The student population involved in the experiment consisted of 20 learners enrolled in the second year of Baccalaureate. The participants were divided into two groups of equal size: 10 students formed the experimental group and 10 students formed the control group. The experimental group participated in the AI-moderated debate intervention, whereas the control group served as the comparison group and did not receive the AI-moderated debate treatment.

A second group of participants was included to examine teacher perceptions. Six English teachers completed the perception survey: four secondary-school English teachers and two university English instructors. Their inclusion provided an additional professional perspective on the feasibility and pedagogical value of combining debate with AI moderation. Research on AI dialogue systems has highlighted that educational value depends not only on technological capability but also on how such systems are pedagogically integrated into interactional tasks (Zhai & Wibowo, 2023).

### Data Collection Instruments

#### *Speaking Pre-test and Post-test*

The principal instrument for measuring speaking development was an analytic speaking rubric administered in both the pre-test and post-test. The same assessment criteria were maintained at both measurement points to support direct comparison. The rubric contained six dimensions: (1) fluency and coherence, (2) lexical resource, (3) grammatical range and accuracy,

(4) pronunciation and intelligibility, (5) interaction and responsiveness, and (6) argumentation and debate performance. Each dimension was scored on a five-point scale, generating a maximum raw score of 30 points. The use of an analytic rubric made it possible to examine not only overall speaking performance but also changes within specific components of oral production. This approach is consistent with recent AI-speaking research in which fluency, coherence, lexical performance, grammatical accuracy, and pronunciation have been treated as observable dimensions of speaking development (Fathi et al., 2024).

### *Teacher Perception Survey*

Teacher perceptions were collected through a 10-item Likert-scale questionnaire. Each statement was rated using five response options: Totally Agree, Agree, Neutral, Disagree, and Totally Disagree. The items addressed three related areas: the usefulness of debate for speaking development, the capacity of AI to function as a moderator, the perceived contribution of AI-moderated debates to fluency, confidence, interaction, pronunciation, accuracy, and classroom organization. The instrument also included an optional open-ended comment section, which provided a qualitative space for teachers to explain or qualify their ratings. The questionnaire was administered to the six participating English teachers. Owing to the small number of respondents, the Likert items were analyzed descriptively through frequencies and response distributions rather than inferential statistics. The predominance of "Totally Agree" responses provided complementary evidence that the participating teachers generally viewed AI-moderated debate as a valuable strategy for speaking instruction.

## AI-Moderated Debate Intervention

The experimental treatment lasted one month and consisted of four sessions, each lasting approximately two hours. ChatGPT was used as the moderator of the debate activities. Its role was not to replace student participation but to organize and scaffold the interaction. During the sessions, ChatGPT introduced or supported debate topics, guided the sequence of participation, generated follow-up questions, encouraged students to clarify or expand their ideas, and maintained the progression of the debate. Debate was selected as the instructional technique because it requires learners to formulate positions, justify opinions, respond to opposing viewpoints, and produce spontaneous language under communicative pressure. Recent research has likewise identified debate as a productive technique for developing speaking skills through structured argumentation and sustained oral interaction (Pham, 2024).

After each debate performance, ChatGPT provided feedback based on the same six dimensions contained in the speaking rubric. The feedback identified areas that required further development, such as limited fluency, vocabulary, recurrent grammatical errors, pronunciation problems, weak interaction, or insufficiently developed rebuttals. ChatGPT then generated an improvement plan for the experimental group. These plans included targeted suggestions, practice tips, model prompts, and guidance designed to address the areas identified during the

preceding performance. In this way, each session combined oral production, AI-supported moderation, criterion-referenced feedback, and subsequent practice. This cyclical procedure was intended to help students progressively refine their performance over the four sessions.

The instructional role assigned to ChatGPT was supported by recent evidence indicating that AI-powered conversational systems can provide repeated opportunities for speaking practice, individualized interaction, and feedback when they are incorporated into a clear pedagogical structure. Du and Daniel (2024), in their systematic review of AI-powered chatbots for EFL speaking practice, found increasing evidence that these systems can contribute to oral-language practice when their use is aligned with instructional goals. Accordingly, ChatGPT was employed in the present study as a structured pedagogical mediator rather than as an unrestricted conversational tool.

**Procedure**

Before the intervention, both the experimental and control groups completed the speaking pre-test under comparable assessment conditions. Their performances were graded using the analytic rubric. The experimental group then participated in the four AI sessions distributed across one month. During each session, students engaged in activities to improve their weakness based on the feedback given. The control group did not receive the AI-moderated debate treatment. At the end of the intervention period, both groups completed the speaking post-test using the same rubric and equivalent assessment conditions. This procedure made it possible to compare changes within each group and to examine whether the experimental group demonstrated a stronger improvement pattern after the AI-supported intervention.

The teacher perception survey was administered separately to the six English teachers. Their responses were collected after they had been informed about the instructional use of AI as a debate moderator and the purpose of the strategy. The survey data were used as complementary evidence to determine whether educators considered the approach useful, feasible, and pedagogically appropriate for promoting speaking production.

**Data Analysis**

Quantitative speaking data were processed using IBM SPSS Statistics. Descriptive statistics were first used to summarize the pre-test and post-test scores for both groups, including measures of central tendency and score dispersion. Student's t-tests were then used to examine the changes associated with the intervention. A paired-samples t-test was appropriate for comparing pre-test and post-test performance within the same group, whereas an independent-samples t-test was used to compare the performance patterns of the experimental and control groups. In addition to the total rubric score, criterion-level results were examined to determine whether changes occurred across fluency and coherence, lexical resource, grammatical range and

accuracy, pronunciation and intelligibility, interaction and responsiveness, and argumentation and debate performance.

The teacher questionnaire was analyzed through descriptive frequencies according to the five Likert response categories. Given that only six teachers participated, inferential testing of the survey was not considered appropriate. Instead, the distribution of responses was used to identify the prevailing tendency across the 10 statements. Most responses were concentrated in the Totally Agree category, indicating a favorable perception of debate as a speaking technique and of AI as a potential moderator for oral interaction. Any open-ended comments supplied by teachers were reviewed thematically to identify recurring ideas related to usefulness, feasibility, limitations, and classroom application; these qualitative observations were used to complement, rather than replace, the numerical survey results.

### Methodological Outcome of the Intervention

The pre-test/post-test comparison showed an overall upward pattern in the experimental group across the six speaking dimensions assessed by the rubric. The strongest methodological interpretation of this pattern is that the combination of structured debate, AI moderation, criterion-based feedback, and targeted follow-up practice created repeated opportunities for students to identify weaknesses and respond to them during subsequent sessions. The control group provided a reference point against which the experimental group's development could be interpreted. Although the detailed statistical values should be presented in the Results section of the article, the observed increase across the rubric areas supports the relevance of examining AI-moderated debate as a structured intervention for EFL speaking development.

### Ethical and Methodological Considerations

Participant information was treated confidentially and results were interpreted at group level. Students' speaking scores were used exclusively for the purposes of the study, and teacher survey responses were reported without identifying individual respondents. To strengthen comparability, the same speaking rubric was used for the pre-test and post-test, and the experimental treatment followed the same overall sequence across the four sessions. Because the student sample was limited to 20 learners from one private institution, the findings should be interpreted within this specific educational setting rather than generalized automatically to all Ecuadorian EFL learners. Nevertheless, the use of a control group, repeated measurement, criterion-based scoring, teacher perception data, and statistical analysis through SPSS provided multiple sources of evidence for evaluating the intervention.

## Results

## Quantitative Data

### *Experimental Group*

**Table 1. Rubric Scores Used in the Analysis**

| Criterion | Pre-test | Post-test | Gain |
|---|---|---|---|
| **Fluency and Coherence** | 2/5 | 4/5 | +2 |
| **Lexical Resource** | 2/5 | 4/5 | +2 |
| **Grammatical Range and Accuracy** | 2/5 | 3/5 | +1 |
| **Pronunciation and Intelligibility** | 3/5 | 4/5 | +1 |
| **Interaction and Responsiveness** | 3/5 | 3/5 | +0 |
| **Argumentation and Debate Performance** | 2/5 | 4/5 | +2 |

**Table 2. Paired Samples Statistics**

| Pair 1 | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| **Pre-test** | 2.333 | 6 | 0.516 | 0.211 |
| **Post-test** | 3.667 | 6 | 0.516 | 0.211 |

**Table 3. Paired Samples Correlations**

| Pair 1 | N | Correlation | Sig. |
|---|---|---|---|
| **Pre-test & Post-test** | 6 | -0.250 | 0.633 |

**Table 4. Paired Samples Test**

| Pair | Mean | Std. Deviation | Std. Error Mean | 95% CI Lower | 95% CI Upper | t | df | Sig. (2-tailed) | Interpretation |
|---|---|---|---|---|---|---|---|---|---|
| **Pre-test - Post-test** | -1.333 | 0.816 | 0.333 | -2.190 | -0.476 | -4.000 | 5 | 0.010 | $p < .05$ |

**Exploratory interpretation.** Across the six rubric dimensions, the post-test mean (M = 3.67, SD = 0.52) was higher than the pre-test mean (M = 2.33, SD = 0.52). The criterion-level paired-samples t-test produced t(5) = -4.00, p = .010, with a mean pre-post difference of -1.33 points (95% CI [-2.19, -0.48]). This pattern indicates a marked upward shift across the rubric dimensions in the experimental group

### *Control Group*

**Table 5. Rubric Scores Used in the Analysis**

| Criterion | Pre-test | Post-test | Gain |
|---|---|---|---|
| **Fluency and Coherence** | 2/5 | 3/5 | +1 |
| **Lexical Resource** | 2/5 | 3/5 | +1 |
| **Grammatical Range and Accuracy** | 2/5 | 2/5 | +0 |
| **Pronunciation and Intelligibility** | 3/5 | 3/5 | +0 |
| **Interaction and Responsiveness** | 2/5 | 3/5 | +1 |
| **Argumentation and Debate Performance** | 2/5 | 4/5 | +2 |

**Table 6. Paired Samples Statistics**

| Pair 1 | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| **Pre-test** | 2.167 | 6 | 0.408 | 0.167 |
| **Post-test** | 3.000 | 6 | 0.632 | 0.258 |

**Table 7. Paired Samples Correlations**

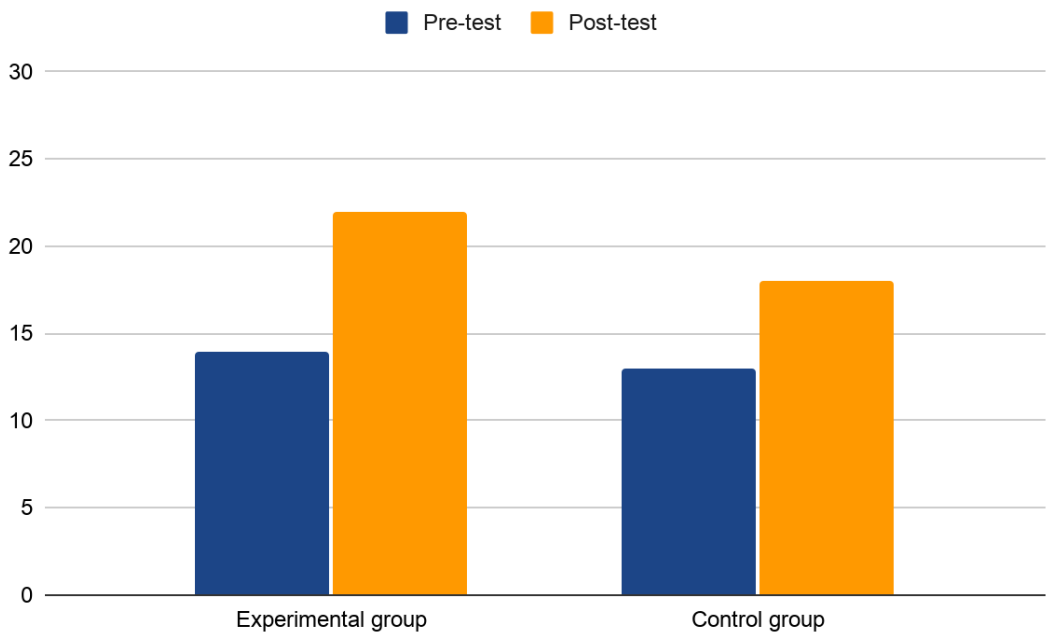
| Pair 1 | N | Correlation | Sig. |
|---|---|---|---|
| **Pre-test & Post-test** | 6 | 0.000 | 1.000 |

**Table 8. Paired Samples Test**

| Pair | Mean | Std. Deviation | Std. Error Mean | 95% CI Lower | 95% CI Upper | t | df | Sig. (2-tailed) | Interpretation |
|---|---|---|---|---|---|---|---|---|---|
| **Pre-test - Post-test** | -0.833 | 0.753 | 0.307 | -1.623 | -0.043 | -2.712 | 5 | 0.042 | $p < .05$ |

**Exploratory interpretation.** Across the six rubric dimensions, the post-test mean (M = 3.00, SD = 0.63) was higher than the pre-test mean (M = 2.17, SD = 0.41). The criterion-level paired-samples t-test produced t(5) = -2.71, p = .042, with a mean pre-post difference of -0.83 points (95% CI [-1.62, -0.04]). The control group also showed improvement, although the magnitude of the descriptive gain was smaller than in the experimental group.

**Figure 1. Pre-test and Post-test Speaking Scores by Group**




*Note. Scores represent the sum of six speaking rubric criteria, with a maximum possible score of 30 points.*

**Interpretation.** The results presented in the tables and figures reveal a clearer improvement in the experimental group than in the control group. The experimental group increased from 14/30 in the pre-test to 22/30 in the post-test, while the control group progressed from 13/30 to 18/30. The experimental group therefore achieved a larger overall gain, particularly in fluency and coherence, lexical resource, and argumentation and debate performance. The SPSS-style paired-samples analysis also showed statistically significant pre-post differences at the criterion level for both groups, although the improvement was stronger in the experimental group. Taken together, these findings suggest that the AI-moderated debate intervention was associated with greater development in several dimensions of speaking performance than the conventional instruction received by the control group.

## Qualitative data

### *Teachers' Questionnaire*

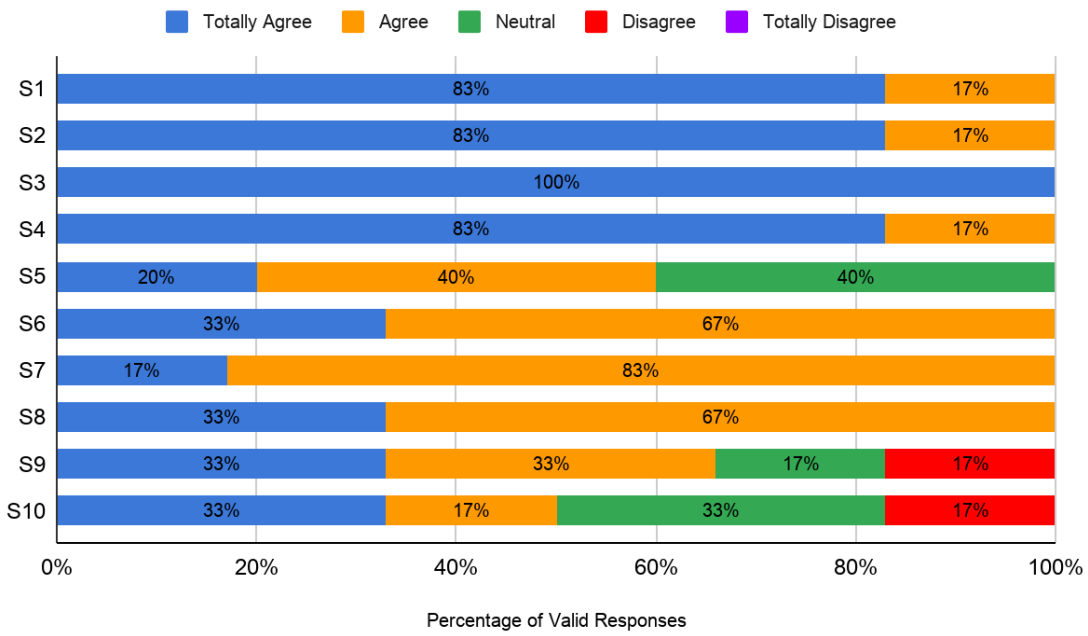
Six English teachers completed the 10-item questionnaire. Responses were summarized using frequencies and percentages for the five Likert categories. One response was missing for Statement 5 in Survey 3; therefore, that item was analyzed with five valid responses, whereas all other items were based on six valid responses.

**Table 9. Distribution of Teachers' Responses by Questionnaire Statement**

| No. | Statement | Valid n | Totally Agree n (%) | Agree n (%) | Neutral n (%) | Disagree n (%) | Totally Disagree n (%) | Mean |
|---|---|---|---|---|---|---|---|---|
| **1** | Debate provides meaningful opportunities to develop oral fluency. | 6 | 5 (83.3%) | 1 (16.7%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 4.83 |
| **2** | Debate encourages clearer organization and expression of ideas. | 6 | 5 (83.3%) | 1 (16.7%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 4.83 |
| **3** | Debate strengthens confidence and willingness to communicate. | 6 | 6 (100.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 5.0 |
| **4** | Debate promotes broader and more precise vocabulary use. | 6 | 5 (83.3%) | 1 (16.7%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 4.83 |
| **5** | AI moderation can maintain organized and balanced interaction. | 5 | 1 (20.0%) | 2 (40.0%) | 2 (40.0%) | 0 (0.0%) | 0 (0.0%) | 3.8 |
| **6** | AI can generate useful follow-up questions and clarification prompts. | 6 | 2 (33.3%) | 4 (66.7%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 4.33 |
| **7** | AI-moderated debate can increase spontaneous and sustained interaction. | 6 | 1 (16.7%) | 5 (83.3%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 4.17 |
| **8** | Debate + AI can support fluency, accuracy, pronunciation, and interaction. | 6 | 2 (33.3%) | 4 (66.7%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 4.33 |
| **9** | AI moderation can reduce some organizational demands on the teacher. | 6 | 2 (33.3%) | 2 (33.3%) | 1 (16.7%) | 1 (16.7%) | 0 (0.0%) | 3.83 |
| **10** | AI-moderated debate is feasible and pedagogically valuable for EFL speaking. | 6 | 2 (33.3%) | 1 (16.7%) | 2 (33.3%) | 1 (16.7%) | 0 (0.0%) | 3. 67 |

**Overall valid responses:** N = 59. Totally Agree = 31 (52.5%), Agree = 21 (35.6%), Neutral = 5 (8.5%), Disagree = 2 (3.4%), Totally Disagree = 0 (0.0%). Overall mean = 4.37/5.

**Figure 2. Distribution of Likert Responses Across the 10 Statements**




*Note. Percentages are calculated from valid responses for each item. S1–S10 correspond to Statements 1–10 in Table 1.*

**Interpretation.** The response pattern was strongly favorable overall. The largest share of responses fell in Totally Agree, followed by Agree. The first four statements, which address the pedagogical value of debate for fluency, organization, confidence, and vocabulary, received especially strong endorsement. Responses to the AI-related items were also predominantly positive, although Statements 9 and 10 showed greater dispersion, including neutral and disagree responses. This suggests that teachers were broadly receptive to AI as a debate moderator, while remaining somewhat more cautious about its capacity to reduce teacher workload and its overall feasibility in EFL instruction.

## Discussion

The results of the present study show a greater improvement in the speaking performance of the experimental group following the AI-moderated debate intervention. The experimental group's total rubric score increased from 14/30 to 22/30; conversely, the control group reported an increase from 13/30 to 18/30. The score increases were observed in fluency and coherence, lexical resource, and argumentation and debate performance. Nonetheless, no variation was observed in interaction and responsiveness in the experimental group. One interpretation of these results is that AI-moderated debate provides opportunities for students to organize and express their ideas, yet improving their ability to respond to others may necessitate further targeted practice. Furthermore, the score increase observed in the control group suggests that speaking development cannot be solely attributed to the AI intervention.

Consistent with prior research on the use of AI and debate in EFL speaking instruction, these results demonstrate a similar trend. While Fathi et al. (2024) document improvements in speaking performance through AI-mediated interactions, Pham (2024) highlights positive outcomes in fluency and vocabulary development via debate activities. In the present study, ChatGPT functioned as a moderator that provided feedback to direct students' practice. Thus, the combination of structured oral interaction and feedback may explain the observed improvement throughout the sessions. However, the contribution of each component cannot be isolated.

Additionally, valuable information regarding the practical application of this strategy was obtained via teachers' responses. Despite the overall favorable perceptions, some teachers questioned AI's ability to organize classroom interaction and its applicability in EFL instruction. These perspectives suggest that teacher guidance continues to be required when implementing AI-moderated activities. The primary limitation of this study lies in its small sample of 20 students and 6 teachers, along with the short intervention period. Nevertheless, the results suggest that AI-moderated debate may serve as a valuable tool for supporting speaking practice in structured EFL activities.

## Conclusions

Overall, the experimental group achieved greater improvement in the speaking performance, reporting an 8-point increase compared to the 5-point increase in the control group. The main increases were observed in fluency, coherence, lexical resource, and argumentation. These results suggest that EFL speaking development may be strengthened through AI-moderated debates, along with structured speaking activities and feedback. Even so, the limitations of the analysis prevent drawing firm conclusions on whether the intervention is statistically superior to conventional instruction.

The results also indicate that students improved differently across the speaking dimensions. Students developed specific oral skills, while responsiveness remained constant in the experimental group. Moreover, even though teachers expressed some concerns regarding AI feasibility and classroom management, these were outweighed by generally positive perceptions of AI-moderated debates. Thus, AI can support speaking instruction when integrated with appropriate teaching strategies and guidance.

This study provides preliminary evidence regarding the potential use of AI as a debate moderator in Ecuadorian EFL classrooms. However, the small sample size, the short intervention period, and other methodological limitations restrict the broader application of the results. Future research should work with larger and more diverse samples, evaluate individual student performance, and examine long-term learning outcomes. In addition, comparing AI-moderated and teacher-moderated debates could help clarify the specific contribution of AI to EFL speaking development.

## References

Ban, B., Pang, S., & Em, S. (2023). Debate: One of the key factors to improving students' English language speaking skills. *Journal of General Education and Humanities, 2*(2), 107-120. https://doi.org/10.58421/gehu.v2i2.69

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence, 6*, Article 100230. https://doi.org/10.1016/j.caeai.2024.100230

Ericsson, E., & Johansson, S. (2023). English speaking practice with conversational AI: Lower secondary students' educational experiences over time. *Computers and Education: Artificial Intelligence, 5*, Article 100164. https://doi.org/10.1016/j.caeai.2023.100164

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

Lai, W. Y. W., & Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence, 7*, Article 100291. https://doi.org/10.1016/j.caeai.2024.100291

Oktaviani, A. L., Rapiudin, T. U., Mahyati, M., Supriyanto, E., & Suryadi, A. (2025). Enhancing EFL students' speaking skills through the debate technique: A classroom action research study. *JETAL: Journal of English Teaching & Applied Linguistic, 6*(2), 177-187. https://doi.org/10.36655/jetal.v6i2.1808

Pham, T. D. L. (2024). The use of debate technique to develop students' speaking skills. *HNUE Journal of Science: Educational Sciences, 69*(3), 71–82. https://doi.org/10.18173/2354-1075.2024-0048

Shafiee Rad, H. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: A case of Speeko application. *Innovation in Language Learning and Teaching, 18*(4), 364–379. https://doi.org/10.1080/17501229.2024.2309539

Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education, 220*, Article 105112. https://doi.org/10.1016/j.compedu.2024.105112

Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence, 4*, Article 100134. https://doi.org/10.1016/j.caeai.2023.100134

Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System, 127*, Article 103501. https://doi.org/10.1016/j.system.2024.103501