Innovation in english language teaching: Chatbots based on artificial intelligence to improve oral expression skills

Innovación en la enseñanza del inglés: Chatbots basados en inteligencia artificial para mejorar las habilidades de expresión oral

https://doi.org/10.5281/zenodo.19616664

AUTORES:

Pablo Luis Vásconez Mera
Universidad Técnica de Babahoyo
ORCID: 0000-0001-9035-7166
pvasconezm@utb.edu.ec

Erika Paola García León
Universidad Técnica de Babahoyo
ORCID: 0000-0001-8203-6434
egarcia@utb.edu.ec

María Grazzia González Quinto
Universidad Técnica de Babahoyo
ORCID: 0009-0003-3593-5383
mgonzalezq@utb.edu.ec

Nelly Victoria Ley Leiva
Universidad Técnica de Babahoyo
ORCID: 000-0003-2296-7354
nley@utb.edu.ec

DIRECCIÓN PARA CORRESPONDENCIA: pvasconezm@utb.edu.ec

Fecha de recepción: 23 / 10 / 2025

Fecha de aceptación: 12 / 12 / 2025

ABSTRACT

The study analyzes the impact of the use of artificial intelligence-based chatbots on the development of oral expression skills in third-level English students at the Technical University of Babahoyo. Four variables were evaluated: fluency, lexical variety, pronunciation, and communicative interaction. A quantitative approach was used with a quasi-experimental pretest-posttest design with a control group. A total of 160 students participated, divided into an experimental group and a control group. The independent variables included interaction with the chatbot, the functionalities used, and frequency of use, while the moderating variables included motivation, digital skills, technological experiences, and internet access. The instruments were subjected to validity and reliability processes using Cronbach's alpha, McDonald's omega, and confirmatory factor analysis, which allowed the reliability of the applied rubric to be verified. Statistical analysis was performed using SPSS, JASP, and R. The results showed significant improvements in the experimental group, with increases of between 11 and 14 points in the variables evaluated. Likewise, t-tests for related and independent samples and ANOVA analyses demonstrated high effect sizes, particularly in fluency. The CFA model showed excellent fit indices, validating the unidimensional structure of the rubric. Multiple regression models confirmed that the independent variables did not influence the initial scores, ensuring equivalence between the groups. Taken together, the findings support the systematic use of chatbots as an effective pedagogical intervention for strengthening oral expression in English.

Keywords: *Autonomous learning, Automated feedback, Educational technologies, Oral proficiency, Psychometric validation.*

RESUMEN

El estudio analiza el impacto del uso de chatbots basados en inteligencia artificial en el desarrollo de habilidades de expresión oral en estudiantes del tercer nivel de inglés de la Universidad Técnica de Babahoyo. Se evaluaron cuatro variables: fluidez, variedad léxica, pronunciación e interacción comunicativa. Se empleó un enfoque cuantitativo con un diseño cuasiexperimental pretest–postest con grupo control. Participaron 160 estudiantes divididos en grupo experimental y grupo control. Las variables independientes incluyeron

la interacción con el chatbot, las funcionalidades empleadas y la frecuencia de uso, mientras que las variables moderadoras abarcaron la motivación, las competencias digitales, las experiencias tecnológicas y el acceso a internet. Los instrumentos fueron sometidos a procesos de validez y fiabilidad mediante Alfa de Cronbach, Omega de McDonald y análisis factorial confirmatorio, lo que permitió comprobar la fiabilidad de la rúbrica aplicada. El análisis estadístico se realizó con SPSS, JASP y R. Los resultados evidenciaron mejoras significativas en el grupo experimental, con incrementos entre 11 y 14 puntos en las variables evaluadas. Asimismo, las pruebas t para muestras relacionadas e independientes y los análisis ANOVA demostraron tamaños de efecto elevados, particularmente en fluidez. El modelo de AFC mostró excelentes índices de ajuste, lo que validó la estructura unidimensional de la rúbrica. Los modelos de regresión múltiple confirmaron que las variables independientes no influyeron en los puntajes iniciales, garantizando la equivalencia entre los grupos. En conjunto, los hallazgos respaldan que el uso sistemático de chatbots constituye una intervención pedagógica eficaz para fortalecer la expresión oral en inglés.

Palabras clave: *Aprendizaje autónomo, Competencia oral, Retroalimentación automatizada, Tecnologías educativas, Validación psicométrica.*

INTRODUCTION

Globally, innovation in language teaching has been promoted by technological advances, particularly with the integration of artificial intelligence (AI) and chatbots. These types of tools allow students to interact with the language dynamically, overcoming barriers such as the lack of oral practice in real, everyday contexts. The use of emerging technologies offers opportunities to improve English language learning, providing personalized and accessible experiences that allow students to develop language skills more effectively and efficiently (Li et al., 2025).

In Latin America, the use of technological platforms in English language teaching has increased significantly in recent years, although there are still major challenges. Teaching capacity and educational infrastructure are important areas for ensuring the successful implementation of digital tools. In other words, AI and chatbots are beginning to be

adopted in different institutions, with the potential to transform the way English is taught, especially with regard to oral expression, one of the most sought-after skills in the labor market (Rivadeneira et al., 2025).

In Ecuador, despite all the efforts made by the Ministry of Education to improve English language proficiency, students face difficulties due to a lack of teacher training and technological infrastructure. The integration of AI and chatbots into the classrooms of educational institutions in Ecuador represents an innovative solution to enhance English language learning, especially with regard to oral expression. These types of technologies allow students to practice in a simulated and flexible environment, improving their confidence and fluency when communicating in the language (Buele et al., 2025).

At the Technical University of Babahoyo (UTB), improving students' language skills has been a priority within the educational strategy. Through the Language Center (CENID), it seeks to incorporate technological tools that optimize the English language teaching process. The implementation of AI and chatbots will allow students to practice oral expression in a continuous and adaptive manner, thereby strengthening their ability to communicate in academic settings and preparing them for challenges in the professional sphere.

The overall objective of this article is to analyze how artificial intelligence and chatbots improve the English-speaking skills of students at the Technical University of Babahoyo. This study investigates how to evaluate the effectiveness of the use of AI and chatbots in the English language teaching process and their integration into the university's pedagogical strategies. This research is justified by the need to innovate in language learning, providing the necessary tools to facilitate access to quality education and respond to the demands of a globalized world.

Theoretical framework

The evolution of language teaching in the digital age

The language teaching model has undergone a remarkable transformation with the emergence of digital technologies. From the use of teaching materials such as books and

audio recordings to the incorporation of interactive platforms, the use of technology facilitates new forms of learning. Today, digital tools such as chatbots and AI-based applications support a personalized approach, adapting to students' needs and focusing on the continuous practice of oral expression skills (Seyma & Savaş, 2024).

With the advancement of technology, techniques have adapted. Instead of being a one-way process, language teaching is becoming more dynamic and interactive. The implementation of AI in classrooms will allow students to practice autonomously through conversation simulations and instant feedback. In this way, English language teaching becomes accessible, aligned, and flexible with the demands of a globalized society, where linguistic competence is essential (Handley, 2024).

En el contexto de transformación digital, la evolución de la enseñanza de idiomas no solo implica la incorporación profunda de los roles pedagógicos y de las dinámicas de aprendizaje. La inteligencia artificial ha permitido transitar hacia modelos centrado en el estudiante, donde la retroalimentación inmediata, la adaptabilidad y la personalización se convierte en ejes fundamentales del proceso formativo. Asimismo, las plataformas conversacionales favorecen entornos de practica continua que reducen la ansiedad comunicativa y amplían las oportunidades de exposición al idioma. De este modo, la enseñanza del idioma inglés en la era digital se orienta hacia experiencias más autónomas, flexibles y contextualizadas, coherentes con las demandas comunicativas de una sociedad interconectada y tecnológicamente mediada (Wah, 2025; Zhang & Umeanowai, 2025).

Theories and approaches to teaching English as a foreign language (EFL)

The process of teaching English as a foreign language is authentically based on approaches that promote interaction and effective communication. Traditional methods such as the grammar-based and direct approaches are gradually being replaced by more learner-centered models. This approach, for example, emphasizes communicative skills and oral practice, which opens the door to technological integrations such as chatbots, which facilitate interaction in realistic contexts (Klímová & Ibna Seraj, 2023).

On the other hand, contemporary pedagogical theories, such as constructivism and task-based learning, emphasize the importance of active approaches where students take the lead in their own learning process. These types of theories favor the use of digital tools that can

promote continuous practice and instant feedback, key elements used by chatbots and AI. Through these types of technologies, students have the opportunity to interact with the language in simulated situations, optimizing their oral expression skills (Katsarou et al., 2023).

In line with these approaches, the integration of artificial intelligence-based technologies can be understood as a natural extension of communicative and constructivist models in EFL teaching. By prioritizing meaningful interaction and authentic task completion, chatbots enable the operationalization of student-centered learning principles through conversational simulations that promote meaning negotiation, frequent oral production, and self-regulation. Furthermore, by providing immediate and adaptive feedback, these tools strengthen situated learning and promote the progressive construction of linguistic knowledge, aligning with the fundamentals of task-based learning and contemporary socioconstructivist theories (Lyu et al., 2025; Qiao & Zhao, 2023).

Artificial intelligence in education

AI is emerging as a powerful tool in the field of education, revolutionizing pedagogy and learning methods. In the context of language teaching, it addresses the needs and learning pace of students. Through different adaptive learning systems, many platforms based on the use of AI identify weaknesses in students and offer specific exercises to improve and optimize the educational process (Bobocea et al., 2024).

Furthermore, artificial intelligence is not only used to personalize content, but can also facilitate access to various resources. Students can access learning materials anywhere and anytime, thus breaking down the physical and temporal barriers of traditional education. In English language teaching, this type of technology is used to develop applications that encourage continuous language practice, especially in the areas of listening comprehension and oral expression, facilitating constant and autonomous learning (Cabrera et al., 2025).

Chatbots in English language teaching

Chatbots are gaining popularity in education due to their ability to simulate realistic dialogues and provide instant feedback. In English language teaching, these chatbots are effective tools for improving speaking skills, allowing students to interact in real time without the pressure of a dialogue with a native speaker. This system offers a safe and non-

judgmental environment for students to make mistakes in their practice, thus facilitating English language learning (Vo & Nguyen, 2024).

In addition, chatbots can be used to customize interactions according to students' skill levels, adapting their responses and offering specific exercises that address various areas of difficulty. This technology can be complemented with various advances in AI, such as voice recognition and automatic correction, allowing students to receive useful and effective feedback. Chatbots have the ability to replicate various English conversation scenarios, increasing student motivation and improving their fluency and confidence in oral expression (Deep et al., 2025).

En este sentido, el potencial pedagógico de los chatbots trasciende la simple simulación conversacional, al constituirse en mediadores tecnológicos que amplían el tiempo de exposición y práctica del idioma más allá del aula tradicional. Al integrar sistemas de reconocimiento de voz y análisis automático del discurso, estas herramientas no solo favorecen la corrección fonética y la precisión léxica, sino que también promueven procesos metacognitivos al permitir que el estudiante reflexione sobre su propio desempeño. Asimismo, la posibilidad de adaptar el nivel de dificultad y los escenarios comunicativos fortalece la diferenciación pedagógica, garantizando experiencias de aprendizaje más inclusivas, personalizadas y orientadas al desarrollo progresivo de la competencia comunicativa integral (Shaalan, 2025).

The importance of oral expression in learning English

Oral expression is a fundamental skill in learning English. It involves not only the ability to produce sounds and words, but also to interact effectively in various everyday situations. Oral fluency is essential for students to participate in real interactions, whether in academic, social, or work environments. However, stability can be difficult to develop without constant practice and the opportunity to interact with native speakers, making it necessary to use technologies that can facilitate practice (Taeza, 2025).

Digital tools such as chatbots and various AI-based platforms can offer students the opportunity to practice their oral expression on an ongoing basis, regardless of time and space barriers. These types of systems allow students to practice dialogues in different controlled environments and receive instant feedback, which will help improve their

confidence and fluency. In this way, the integration of different technologies into English language learning can facilitate the acquisition of various oral skills that are essential for mastering the language in different real-life contexts (Wei et al., 2025).

METHODOLOGY

The following research has a quantitative approach with a quasi-experimental design for pretest and posttest plus a control group, with the aim of analyzing the effect of using AI-based chatbots for the development of oral expression skills in the English language. The population consists of third-level students at the Language Center of the Technical University of Babahoyo, with an intentional sample divided into a control group and an experimental group.

The independent variables were student-chatbot interaction, specific functionalities used, and frequency of use. The dependent variables included oral fluency, lexical variety, pronunciation, intangibility, and communicative interaction, evaluated using an analytical rubric aligned with the CEFR. The moderating variables chosen were motivation to learn English, digital skills, previous experience with technological tools, and access to the internet or devices. The moderating variables were measured using a Likert-type questionnaire. Both instruments underwent content validity and reliability analysis.

The procedure used was to administer an oral pretest and the moderator questionnaire to both groups; then, the experimental group developed scheduled sessions with chatbots at different periods, while the control group continued with traditional activities. At the end, the oral posttest was administered, maintaining the same conditions as the initial assessment.

Data tabulation was performed in Microsoft Excel and analyzed in SPSS v25, JASP 0.95.4.0, and R Studio. Descriptive statistics and normality tests were applied, as well as internal reliability analysis (Cronbach's Alpha and McDonald's Omega). The structure of

the instruments was examined using Confirmatory Factor Analysis (CFA). Pre/post-test t-tests, ANOVA, and multiple regression were used to measure differences and effects of the intervention. Were used to analyze the relationships between independent, moderating, and dependent variables, thus ensuring methodological rigor through standardized fit indices.

RESULTS

Descriptive statistics

Descriptive statistics show that the moderating variables present favorable average levels for the use of educational technologies. Motivation to learn registers an average of 3.22 (SD = 0.76), while digital skills and technological experience show values of 2.95 (SD = 0.64) and 2.73 (SD = 0.67), respectively. Similarly, internet access reached an average of 3.64 (SD = 0.69), suggesting adequate conditions for the intervention. Regarding the use of the chatbot, there is marked variability, especially in interaction (M = 23.72; SD = 22.40), reflecting marked differences between the groups. The functionalities used (M = 2.84; SD = 1.33) and the weekly frequency of use (M = 2.84; SD = 2.00) confirm that exposure to the chatbot was intense in the experimental group.

In relation to language skills, the comparison of pretest and posttest shows substantial increases in all indicators of oral expression. Fluency increased from 60.22 to 72.22 points, while lexical variety increased from 62.39 to 74.83, demonstrating a significant expansion of linguistic repertoire. Similarly, pronunciation improved from 61.27 to 72.30 points, and communicative interaction increased from 56.82 to 70.40, representing the most significant increase. Overall, the preliminary results suggest a positive effect of chatbot-based interaction, as the post-test averages far exceed those of the pre-test. This indicates a general strengthening of oral expression skills in English (Table 1).

Table 1

Descriptive statistics for the study variables

Descriptive statistics					
	N	Minimum	Maximum	Mean	Desv. Desviation
Motivation	160	1,24	5,00	3,2276	0,76044
Digital skils	160	1,00	5,00	2,9521	0,64719
Technological Experience	160	1,09	4,24	2,7318	0,67694
Internet Access Devices	160	1,65	5,00	3,6458	0,69871
Chatbot Interaction	160	-1	59	23,72	22,409
Features Used	160	1,00	5,00	2,8441	1,33947
Weekly Frequency of Use	160	0,00	6,46	2,8409	2,00634
Pre-test Fluency	160	40,53	77,89	60,2251	6,45244
Post-test Fluency	160	49,74	93,35	72,2051	9,08269
Pre-test Lexical Variety	160	46,9	77,9	62,397	5,9893
Posttest Lexical Variety	160	50,22	100,00	74,8366	9,37442
Pre-test Pronunciation	160	40,49	77,60	61,2751	6,85011
Posttest Pronunciation	160	53,04	93,88	72,3034	9,09435
Pre-test Communicative Interaction	160	31,39	77,95	56,8271	7,48359
Posttest Communicative Interaction	160	34,68	98,73	70,4009	10,07298
Valid N (per list)	160				

Note: The table shows the minimum, maximum, mean, and standard deviation values for the moderating, independent, and dependent variables included in the study.

Normality tests

The results obtained by the Kolmogórov-Smirnov and Shapiro-Wilk normality tests show that, for most of the variables analyzed, the significance values ($p > 0.05$) allow us to assume that there is a normal distribution in both the control and experimental groups. In particular, the results for fluency, lexical variety, and pronunciation in the pretest and posttest stages meet the normality criterion in both tests, indicating that the data show a symmetrical distribution without significant deviations. This is especially evident in the Shapiro-Wilk tests, where all variables far exceed the significance threshold, with the exception of a few isolated values for communicative interaction.

Complementarily, although two specific cases with $p > 0.05$ are observed (post-test lexical variety in Kolmogorov-Smirnov for the experimental group and post-test communicative interaction in Shapiro-Wilk for the control group), these values represent exceptions and do not alter the general trend of normality of the data set. Therefore, considering the sample size ($n = 80$ per group) and the robustness of parametric tests against slight deviations, it is concluded that the data adequately meet the assumptions of normality. Consequently, parametric tests should be applied in subsequent inferential analyses, such as t-tests for related and independent samples, ANOVA, and structural equation models (Table 2).

Table 2

Results of Kolmogorov–Smirnov and Shapiro–Wilk normality tests for study variablesNormality tests

Group		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	gl	Sig.	Statistic	gl	Sig.
Pretest Fluency	Control	0,057	80	,200*	0,994	80	0,978
	Experimental	0,069	80	,200*	0,982	80	0,322
Posttest Fluency	Control	0,058	80	,200*	0,990	80	0,764
	Experimental	0,096	80	0,067	0,983	80	0,380
Pre-test Lexical Variety	Control	0,062	80	,200*	0,992	80	0,911
	Experimental	0,091	80	0,159	0,975	80	0,111
Posttest Lexical Variety	Control	0,044	80	,200*	0,995	80	0,985
	Experimental	0,104	80	0,033	0,982	80	0,316
Preest Pronunciation	Control	0,056	80	,200*	0,983	80	0,344

	Experimental	0,057	80	,200*	0,988	80	0,636
Posttest	Control	0,059	80	,200*	0,983	80	0,356
Pronunciation	Experimental	0,062	80	,200*	0,987	80	0,570
Pre-test	Control	0,103	80	0,034	0,971	80	0,061
Communicative Interaction	Experimental	0,066	80	,200*	0,987	80	0,606
Posttest	Control	0,078	80	,200*	0,968	80	0,043
Communicative Interaction	Experimental	0,078	80	,200*	0,990	80	0,769

*. This is a lower limit of true significance.

a. Lilliefors significance correction

Note: P values greater than 0.05 indicate normality in the distribution of data. In general, the variables comply with this assumption, allowing the application of parametric analyses in intergroup and intraindividual comparisons.

Internal reliability using Cronbach's alpha and McDonald's omega tests

The reliability results show that the rubric and questionnaire have an acceptable overall internal consistency index (α = 0.653) considering the eight items included. Although this value is slightly below the recommended threshold of 0.70 for adequate reliability, it remains within the expected range for instruments that assess different dimensions of the same complex skill, as is the case with oral expression. Similarly, when analyzing the corrected item-total correlations, it can be seen that several post-test items (fluency, lexical variety, pronunciation, and communicative interaction) have values above 0.48, which shows that there is a better fit to the construct compared to the pre-test values. Similarly, the Cronbach Alpha if Item Deleted column shows that the elimination of none of the items significantly increases the overall coefficient, suggesting that all elements contribute, to a greater or lesser extent, to the overall structure of the instrument. Overall, the results support the preliminary internal validity of the rubric, although they also point to the advisability of continuing to refine some indicators to strengthen overall consistency (Tables 3 and 4).

Table 3

Instrument reliability statistics (Cronbach's Alpha)

Reliability Statistics	
Cronbach's Alpha	N of Items
0,653	8

Tabla 4

Item-total analysis of the oral expression rubric

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Pretest Fluency	470,2449	1142,015	0,147	0,663
Postest Fluidez	458,2649	892,153	0,503	0,575
Posttest Fluency	468,0733	1138,954	0,181	0,655
Postest Lexical Variety	455,6334	874,764	0,514	0,570
Pre-test Lexical Variety	469,1949	1121,973	0,172	0,659
Postest Pronunciacion	458,1666	870,628	0,549	0,561
Posttest Lexical Variety	473,6429	1118,908	0,146	0,667
Postest Interaccion Comunicativa	460,0691	862,550	0,480	0,580

Note: Corrected item–total correlations indicate the degree of association of each item with the entire scale. Values above 0.30 are considered acceptable and reflect an adequate contribution to the construct being evaluated.

The McDonald's Omega coefficient obtained (ω = 0.725) indicates adequate internal consistency of the instrument, exceeding the threshold of 0.70 recommended for educational studies. The reduced standard error (SE = 0.029) and the 95% confidence interval (0.660–0.777) show an accurate and stable estimate, which reinforces the reliability of the construct evaluated. Taken together, these results confirm that the rubric developed has a coherent structure and an optimal level of reliability under a factorial approach (Table 5).

Table 5

McDonald's Omega coefficient and its 95% confidence interval

Frequentist Scale Reliability Statistics				
			95% CI	
Coefficient	Estimate	Std. Error	Lower	Upper
Coefficient ω	0.725	0.029	0.660	0.777

Note: ω values greater than 0.70 are considered indicators of adequate reliability in educational and psychological instruments.

Confirmatory factor analysis (CFA)

The results of the confirmatory factor analysis show that the unidimensional model proposed for evaluating oral expression is an excellent fit, meeting and far exceeding the internationally recommended criteria for structural validation of educational instruments. In fact, the goodness-of-fit indices reached optimal values, with an CFI = 1.000 and TLI = 1.012, both of which are above the threshold of 0.90. This shows that the factorial model adequately reproduces the observed covariance matrix. Similarly, the model's error indices remained within excellent parameters, as the SRMR was 0.019 and the RMSEA = 0.000, well above the standards that establish values equal to or less than 0.08 as acceptable. These results, taken together, confirm the relevance of the theoretical model and its statistical stability under the estimation method implemented in lavaan.

Similarly, the standardized factor loadings show that the four indicators included—fluency, lexical variety, pronunciation, and communicative interaction—contribute significantly and consistently to the latent construct of oral expression. These loadings ranged from 0.562 to 0.707, values that reflect a solid relationship between each item and the underlying factor, as well as a coherent internal structure within the instrument. This consistency is reinforced by the correlation matrix between indicators, whose positive coefficients indicate conceptual convergence between the dimensions evaluated. In summary, the AFC findings empirically validate the unidimensional structure of the rubric, demonstrating that the set of indicators reliably and consistently measures oral expression skills after the intervention.

T-test for related and independent samples

T-test for related samples (pre-test and post-test within each group)

The results of the t-test analysis for related samples show significant differences between pre-test and post-test scores in all dimensions of oral expression evaluated. In each pair compared, the mean difference is negative (between -11.03 and -13.57), indicating a significant increase in post-test scores compared to pre-test scores. These results are accompanied by moderate standard deviations and low standard errors, confirming the stability of the estimates. Similarly, the 95% confidence intervals do not include the value

zero, reinforcing the existence of real and systematic improvements in fluency, lexical variety, pronunciation, and communicative interaction after the intervention.

Additionally, the t-values obtained are high (between -21.49 and -24.71) and highly significant ($p < 0.001$), showing that the improvements found are not due to chance but rather respond to the effect of using chatbots as a pedagogical tool. The magnitude and consistency of these differences suggest that the intervention had a positive impact on the development of oral production skills in the English language. Overall, the analyses confirm that students made significant progress in all dimensions evaluated, providing solid evidence in favor of the effectiveness of the conventional artificial intelligence-mediated learning approach (Table 6).

Table 6

Results of the t-test for related samples in oral expression skills

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest Fluency - Postest Fluency	-11,98000	6,48682	0,51283	-12,99283	-10,96717	-23,361	159	0,000
Pair 2	Pretest Lexical Variety - Postest Lexical Variety	-12,43994	7,10338	0,56157	-13,54904	-11,33084	-22,152	159	0,000
Pair 3	Pretest Pronuntiation - Postest Pronuntiation	-11,02831	6,48855	0,51297	-12,04142	-10,01521	-21,499	159	0,000
Pair 4	Pretest Communicative Interaction - Postest Communicative Interaction	-13,57375	6,94740	0,54924	-14,65850	-12,48900	-24,714	159	0,000

Note: Negative differences indicate an increase in post-test scores. P-values less than 0.05 indicate statistically significant improvements between the two measurements.

Independent samples t-test (control vs. experimental)

The results of the independent samples t-test indicate that there were statistically significant differences between the experimental group and the control group in all oral expression skills assessed in the post-test. First, Levene's tests confirm that there are no differences in variance between the groups ($p > 0.05$ in all dimensions), which allows us to assume equal variances and use the t-test values corresponding to this condition. In this sense, the t-values obtained are high and negative (between -8.959 and -12.848), which shows that the scores of the experimental group were significantly higher than those of the control group. Additionally, the observed mean differences, ranging from -10.52 to -13.35 points, reflect a substantial effect of the chatbot-based intervention on oral performance.

Similarly, the 95% confidence intervals do not include the zero value and remain within narrow ranges, confirming the magnitude and consistency of the differences between the two groups. The fact that all tests have p-values less than 0.001 is conclusive evidence that the improvement observed in fluency, lexical variety, pronunciation, and communicative interaction is due to the pedagogical intervention and not to chance. Overall, the results reinforce the effectiveness of chatbot-mediated learning and empirically support the fact that students exposed to this technology achieved higher levels of oral expression compared to those who followed traditional methods (Table 7).

Table 7

Results of the t-test for independent samples in the oral expression post-test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Postets Fluidez	Equal variances assumed	0,06	0,80	-10,98	158	0,00	-11,91	1,09	-14,05	-9,77
	Equal variances not assumed			-10,98	157,948	0,00	-11,91	1,09	-14,05	-9,77
Postest Variedad Lexica	Equal variances assumed	0,02	0,89	-12,85	158	0,00	-13,36	1,04	-15,41	-11,31
	Equal variances not assumed			-12,85	157,891	0,00	-13,36	1,04	-15,41	-11,31
Post_Pronunciacion	Equal variances assumed	1,46	0,23	-8,96	158	0,00	-10,52	1,17	-12,84	-8,20
	Equal variances not assumed			-8,96	153,851	0,00	-10,52	1,17	-12,84	-8,20
Post_Interaccion_Comunicativa	Equal variances assumed	1,01	0,32	-8,96	158	0,00	-11,66	1,30	-14,23	-9,09
	Equal variances not assumed			-8,96	157,106	0,00	-11,66	1,30	-14,23	-9,09

Note: Non-significant Levene's tests ($p > 0.05$) indicate homogeneous variances between groups. P values lower than 0.001 reflect significant differences in post-test scores between the experimental group and the control group.

Analysis ANOVA

The results obtained from the implementation of ANOVA intra-subject effect analyses reveal statistically significant differences between the experimental group and the control

group in the post-test fluency score. The group effect is highly significant (F (1,158) = 120.485, p < 0.001), indicating that the chatbot-based intervention produced substantially greater improvements in oral fluency compared to the control group. Furthermore, the effect size (partial $\eta^2 = 0.433$) is considered large, demonstrating that approximately 43.3% of the variability in fluency levels can be explained by belonging to one of the groups, evidencing a notable impact of the intervention. The overall model is also significant (p = < 0.001) and has a high coefficient of determination (R2 = 0.433), reflecting a high degree of fit and considerable explanatory power. Taken together, these results confirm that the use of chatbots generated superior and consistent improvements in oral fluency compared to traditional methods (Table 8).

Table 8

ANOVA of between-subject effects for the posttest of fluency between the experimental and control groups

Tests of Between-Subjects Effects						
Dependent Variable:						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5674,877[a]	1	5674,877	120,485	0,000	0,433
Intercept	834172,812	1	834172,812	17710,537	0,000	0,991
Grupo	5674,877	1	5674,877	120,485	0,000	0,433
Error	7441,858	158	47,100			
Total	847289,547	160				
Corrected Total	13116,735	159				

a. R Squared = ,433 (Adjusted R Squared = ,429)

Note: A p-value < 0.05 indicates significant differences between groups. The effect size (partial η^2) reflects the proportion of variance explained by the intervention.

Multiple regression analysis

The results of the multiple regression model indicate that none of the independent variables included (interaction with the chatbot, features used, frequency of use, motivation, digital skills, technological experience, and internet access) have a significant effect on the dependent variable of pretest fluency. This is evident in the significance values ($p > 0.05$ in all cases), as well as in the standardized Beta coefficients, which are low and close to zero, indicating minimal predictive power. Likewise, the t-statistics do not reach relevant levels, confirming that the independent variables significantly explain the initial variations in fluency before the intervention. This pattern was expected, since the students had not yet interacted with the chatbot in the pretest, which reaffirms the internal validity of the experimental design by showing that there were no differences attributable to technological predictors (Table 9).

Table 9

Multiple regression coefficients for predicting fluency in the pretest

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	62,243	5,217		11,931	0,000
	Chabot Interaction	-0,042	0,082	-0,147	-0,516	0,606
	Features_Used	-0,041	1,166	-0,008	-0,035	0,972
	Frequency_of_Weekly_Use_Uso_Semanal	0,509	0,817	0,158	0,623	0,534
	Motivation	0,054	0,705	0,006	0,076	0,939
	Digital_Skills	-0,014	0,821	-0,001	-0,017	0,986
	Technological_Experience	-0,185	0,784	-0,019	-0,235	0,814

Internet_Access_Devices	-0,542	0,759	-0,059	-0,714	0,476

a. Dependent Variable: Pre_Fluidez

Note: No independent variable showed a significant effect ($p > 0.05$), indicating an absence of predictive capacity regarding fluency prior to the intervention.

DISCUSSION

The results obtained in this research demonstrate that the integration of AI-based chatbots produced significant improvements in all dimensions of oral expression (fluency, lexical variety, pronunciation, and communicative interaction), with an increase of 11–14 points on average, and effects consistent with the trend reported in recent research. Ericsson et al. (2023) found that a conversational dialogue system integrated into regular classes consistently optimized the fluency and accuracy of high school students by offering frequent oral practice and individualized feedback. Similarly, Ying et al. found that research studies of university students show significantly favorable results in oral competence and communicative confidence when systems are used as a complement to face-to-face instruction, thanks to their interactive nature and the possibility of practicing in different low-anxiety environments.

Regarding the overall impact of AI on English language learning, recent reviews show that the use of intelligent tools benefits personalization, pronunciation improvement, and active student participation, especially when integrated into clear and sustained pedagogical designs. For Xiao et al (2024), the results of this research confirm this type of orientation: the sample shows average levels of motivation and sufficient digital skills to be able to take advantage of technology, and the systematic use of chatbots is associated with greater improvements in the experimental group versus the control group, as evidenced by t-tests, ANOVA, and high effect sizes. However, in contrast to various studies that report very high reliability indices within their instruments, this research obtained a moderate Cronbach's alpha (0.753), offset by an adequate McDonald's Omega (0.725) and an AFC with excellent fit indices (CFI = 1.000; RMSEA = 0.000).

The results show differential effects between the dimensions of oral expression, with greater increases observed in fluency and communicative interaction, suggesting that chatbots have a particular impact on skills related to dialogic practice and spontaneity. This finding coincides with Ye et al. (2022) y Alenezi & Alenezi (2025), who report significant improvements in confidence and oral participation through conversational AI. In contrast, although pronunciation improved, their progress was not differentially superior, unlike what was indicated by Egas et al. (2024). Likewise, motivation and digital skills acted as contextual facilitators, in line with Pituxcoosuvarn et al. (2025).

CONCLUSIONS

The results obtained in this study demonstrate that the integration of AI-based chatbots is an effective pedagogical strategy for improving oral expression skills in university students. Inferential analyses justified highly significant statistical differences between the experimental group and the control group in all the different dimensions evaluated (fluency, lexical variety, pronunciation, and communicative interaction) with higher increases and large effects in favor of the intervention group. Likewise, confirmatory factor analysis showed an excellent fit of the one-dimensional model of the rubric, validating the consistency of the "oral expression" construct and supporting the relevance of its use in subsequent studies. Similarly, McDonald's Omega coefficient confirmed adequate internal consistency of the instrument, complementing the moderate reliability observed in Cronbach's alpha.

On the other hand, the results of multiple regression and normality analysis support the internal validity of the experimental design, as the independent variables did not show predictive power in the pretest, indicating that both groups started under similar conditions before the intervention. In summary, the empirical evidence obtained supports that the systematic use of chatbots promotes significant development in English oral production, especially when combined with guided pedagogical practices and adequate minimum technological conditions. This confirms that conversational AI can play an important role in

the language teaching process, enhancing both linguistic performance and student participation.

REFERENCES

Alenezi, A., & Alenezi, A. (2025). Evaluating the Effectiveness of Chatbot-Assisted Learning in Enhancing English Conversational Skills Among Secondary School Students. *Education Sciences*, *15*(9). https://doi.org/10.3390/educsci15091136

Bobocea, A., Bologa, R., Bătăgan, L., & Posedaru, B. Ștefan. (2024). Architecture to Transform Classic Academic Courses into Adaptive Learning Flows with Artificial Intelligence. *Amfiteatru Economic*, *26*(65), 363–380. https://doi.org/10.24818/EA/2024/65/363

Buele, J., Sabando-García, Á. R., Sabando-García, B. J., & Yánez-Rueda, H. (2025). Ethical Use of Generative Artificial Intelligence Among Ecuadorian University Students. *Sustainability (Switzerland)*, *17*(10). https://doi.org/10.3390/su17104435

Cabrera, P., Ochoa-Cueva, C., & Castillo-Cuesta, L. (2025). Enhancing EFL Higher Education through Fliki Videos: An Artificial Intelligence Implementation Approach. *World Journal of English Language*, *15*(1), 424–433. https://doi.org/10.5430/wjel.v15n1p424

Deep, P. Das, Martirosyan, N., Ghosh, N., & Rahaman, M. S. (2025). ChatGPT in ESL Higher Education: Enhancing Writing, Engagement, and Learning Outcomes. In *Information (Switzerland)* (Vol. 16, Number 4). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/info16040316

Egas Villafuerte, V. P., Maldonado Palacios, G. M., Narváez Sarango, J. S., Vera Rojas, M. D. P., & Carrazco Colcha, F. H. (2024). The impact of online learning platform in the English teaching -learning process. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, *5*(1). https://doi.org/10.56712/latam.v5i1.1809

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

Handley, Z. (2024). Has artificial intelligence rendered language teaching obsolete? In *Modern Language Journal* (Vol. 108, Number 2, pp. 548–555). John Wiley and Sons Inc. https://doi.org/10.1111/modl.12929

Katsarou, E., Wild, F., Sougari, A. M., & Chatzipanagiotou, P. (2023). A Systematic Review of Voice-based Intelligent Virtual Agents in EFL Education. *International Journal of Emerging Technologies in Learning*, *18*(10), 65–85. https://doi.org/10.3991/ijet.v18i10.37723

Klímová, B., & Ibna Seraj, P. M. (2023). The use of chatbots in university EFL settings: Research trends and pedagogical implications. In *Frontiers in Psychology* (Vol. 14). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2023.1131506

Li, Y., Zhou, X., Yin, H. B., & Chiu, T. K. F. (2025). Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review. In *Smart Learning Environments* (Vol. 12, Number 1). Springer. https://doi.org/10.1186/s40561-025-00379-0

Lyu, B., Lai, C., & Guo, J. (2025). Effectiveness of Chatbots in Improving Language Learning: A Meta-Analysis of Comparative Studies. *International Journal of Applied Linguistics (United Kingdom)*, *35*(2), 834–851. https://doi.org/10.1111/ijal.12668

Pituxcoosuvarn, M., Tanimura, M., Murakami, Y., & White, J. S. (2025). Enhancing EFL Speaking Skills with AI-Powered Word Guessing: A Comparison of Human and AI Partners. *Information (Switzerland)*, *16*(6). https://doi.org/10.3390/info16060427

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

Rivadeneira, L., de Luna, D. B., & Fernandez, C. (2025). Exploring the role of ChatGPT in higher education institutions: Where does Latin America stand? *Digital Government: Research and Practice*, *6*(2). https://doi.org/10.1145/3689370

Seyma, F. K., & Savaş, P. (2024). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL*. https://doi.org/10.1017/S0958344024000168

Shaalan, I. E. N. A. W. (2025). The Impact of Artificial Intelligence and Machine Learning on Linguistic Accuracy, Fluency, and Self-Direction Among Advanced EFL Students. *Theory and Practice in Language Studies*, *15*(6), 1987–1995. https://doi.org/10.17507/tpls.1506.27

Taeza, J. (2025). The role of AI-powered chatbots in enhancing second language acquisition: An empirical investigation of conversational AI assistants. *Edelweiss Applied Science and Technology*, *9*(3), 2616–2629. https://doi.org/10.55214/25768484.v9i3.5853

Vo, C., & Nguyen, A. (2024). Generative Artificial Intelligence and ChatGPT in Language Learning: EFL Students' Perceptions of Technology Acceptance. In *Journal of University Teaching and Learning Practice* (Number 6).

Wah, J. N. K. (2025). Artificial Intelligence in Language Learning: A Systematic Review of Personalization and Learner Engagement. *Forum for Linguistic Studies*, *7*(9). https://doi.org/10.30564/fls.v7i9.10336

Wei, W., Zhao, A., & Ma, H. (2025). Understanding How AI Chatbots Influence EFL Learners' Oral English Learning Motivation and Outcomes: Evidence From Chinese Learners. *IEEE Access*, *13*, 56699–56716. https://doi.org/10.1109/ACCESS.2025.3554545

Xiao, F., Zhao, P., Sha, H., Yang, D., & Warschauer, M. (2024). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*, *4*(2), 300–325. https://doi.org/10.1515/jccall-2022-0032

Ye, Y., Deng, J., Liang, Q., & Liu, X. (2022). Using a Smartphone-Based Chatbot in EFL Learners' Oral Tasks. *International Journal of Mobile and Blended Learning*, *14*(1). https://doi.org/10.4018/IJMBL.299405

Ying, G., Binti, N., Abdul Nauruk, Suyan, Z., & Yiming, C. (2024). Social responsibility of small and medium enterprises in Vietnam through digital transformation and application of artificial intelligence. *LatIA*, *2*. https://doi.org/10.62486/latia202423

Zhang, X., & Umeanowai, K. O. (2025). Exploring the transformative influence of artificial intelligence in EFL context: A comprehensive bibliometric analysis. *Education and Information Technologies*, *30*(3), 3183–3198. https://doi.org/10.1007/s10639-024-12937-z

Taeza, J. (2025). The role of AI-powered chatbots in enhancing second language acquisition: An empirical investigation of conversational AI assistants. *Edelweiss Applied Science and Technology*, *9*(3), 2616–2629. https://doi.org/10.55214/25768484.v9i3.5853

Vo, C., & Nguyen, A. (2024). Generative Artificial Intelligence and ChatGPT in Language Learning: EFL Students' Perceptions of Technology Acceptance. In *Journal of University Teaching and Learning Practice* (Issue 6).

Wei, W., Zhao, A., & Ma, H. (2025). Understanding How AI Chatbots Influence EFL Learners' Oral English Learning Motivation and Outcomes: Evidence from Chinese Learners. *IEEE Access*, *13*, 56699–56716. https://doi.org/10.1109/ACCESS.2025.3554545

Xiao, F., Zhao, P., Sha, H., Yang, D., & Warschauer, M. (2024). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*, *4*(2), 300–325. https://doi.org/10.1515/jccall-2022-0032