Avances en Interacción Humano-Computadora, 2025, Volumen 10, No. 1, pp. 52-59
DOI: 10.47756/aihc.y10i1.201

RESEARCH PAPER

Designing a Generative AI–Powered Virtual Tutor: Studying Interaction Modality and Avatar Presence on the User Experience

Diseñando una Tutora Virtual con IA Generativa: Estudio de los Efectos de la Modalidad y la Presencia de Avatar en la Experiencia de Usuario

Ramón J. Carabeo [1], Luis A. Castro [1,*]

Published: 30 November 2025

Abstract
Advances in AI-driven educational technologies have increased interest in virtual tutoring systems, yet empirical evidence in Latin America, particularly in Mexico, remains limited. This study presents the design and preliminary evaluation of a generative AI–powered virtual tutor prototype. The study examines the effect of user interaction modality with the tutor (voice vs. text) and the presence of a visual avatar, considering metrics such as user satisfaction, perceived usefulness, trust in the system, and behavioral engagement. A 2×2 *between-subjects* experimental design was conducted with 22 undergraduate engineering students. The results revealed significant behavioral differences: participants in the voice condition exhibited a longer user active time (UAT), whereas those in the text condition produced a higher number of messages. The presence of the avatar did not yield statistically significant effects on any of the variables. These findings suggest that the interaction modality influences user engagement dynamics without altering users' overall perceptions, which may have practical implications for the adaptive design of AI-driven virtual tutors in higher education and contribute to understanding how students interact with generative AI tools in academic environments.

Keywords:
Virtual tutoring; Generative artificial intelligence; Human–computer interaction; Perceived usefulness; Trust; Behavioral engagement.

Resumen
En los últimos años, los tutores virtuales basados en IA han despertado creciente interés en la educación superior; sin embargo, aún existe poca evidencia empírica en América Latina y, en particular, en México. Este trabajo describe el diseño y la evaluación preliminar de un prototipo de un tutor virtual basado en Inteligencia Artificial generativa. En este trabajo se estudia el efecto de la modalidad de interacción del usuario con el tutor (voz vs. texto) y de la presencia de un avatar visual, considerando métricas como satisfacción, percepción de utilidad, confianza en el sistema y comportamiento interactivo. Se utilizó un diseño experimental 2×2 *between-subjects* con 22 estudiantes universitarios de ingeniería. Los resultados muestran diferencias conductuales. Los participantes en la condición de voz mostraron mayor tiempo activo, mientras que quienes usaron texto enviaron más mensajes. La presencia del avatar no generó efectos estadísticamente significativos en ninguna variable. Estos hallazgos sugieren que la modalidad de interacción influye en la dinámica de interacción, aunque no modifica la percepción general de la plataforma, lo cual puede tener implicaciones prácticas para el diseño adaptativo de tutores virtuales educativos basados en inteligencia artificial.

Palabras clave:
Tutoría virtual; Inteligencia artificial generativa; Interacción humano-computadora; Percepción de utilidad; Confianza; Comportamiento interactivo.

Carabeo, R. J.[1], Castro, L. A.[1,*]
Dept. of Computing and Design
Sonora Institute of Technology (ITSON)
Ciudad Obregon, Mexico
Email: ramon.carabeo228497@potros.itson.edu.mx, luis.castro@acm.org

* Corresponding author

1 Introduction

In recent years, virtual tutors and educational chatbots have gained increasing prominence in higher education and self-directed learning [1, 2, 13]. Multiple studies have highlighted their ability to provide personalized responses and rapid access to content [2, 13, 15]. Artificial intelligence (AI) driven technologies enable not only immediate feedback but also increased availability of academic support and a perceived sense of support during the learning process [13, 14].

Despite increasing global interest in AI-based tutoring technologies [1, 14, 15], empirical research in Latin America remains scarce. In the Mexican higher education context, the adoption and evaluation of virtual tutoring systems powered by generative AI are still emerging areas of study. Understanding how local students perceive and interact with these systems is therefore essential for assessing their potential impact and generalizability.

The presentation of a digital system directly shapes how users perceive and interpret it. According to numerous studies in the field of Human-Computer Interaction (HCI), and drawing on the Media Equation theory, people tend to interact with computers as if they were social agents [4]. This social dimension of interaction has motivated the development of theoretical frameworks and instruments that systematically assess technology acceptance and user experience. Among these, the Technology Acceptance Model (TAM) stands out, emphasizing perceived usefulness and ease of use [5], as well as the System Usability Scale (SUS), widely used to measure perceived usability in interactive interfaces [6].

Within the design factors of a virtual tutor, the interaction modality (i.e., voice vs. text) and avatar display have been identified as variables that may affect user satisfaction, trust, and behavioral engagement [7, 8]. Recent studies have reported mixed findings: some highlight that voice and avatars enhance social presence and may foster users' perceived motivation [9, 10], whereas reviews and studies in higher education reveal inconclusive effects on perceived usefulness and learning outcomes [15, 20]. However, in the Latin American context, empirical research on AI-based virtual tutors remains limited [12]. This gap highlights the need to explore how students in local higher education contexts perceive such technologies. Moreover, it remains unclear whether avatar presence or interaction modality (voice or text) affects the overall user experience.

This study contributes empirical evidence on how interaction modality (voice vs. text) and avatar presence influence behavioral engagement in AI-powered tutoring systems within a Mexican engineering education context. While prior research has explored these factors in general HCI environments, empirical studies grounded in Latin American higher education remain scarce. By examining these design elements in a real academic setting, this work provides contextualized insights that support the development and evaluation of AI-driven tutoring tools tailored to local needs.

2 Related Work

Virtual tutors and educational chatbots have gained relevance in higher education by offering personalization, immediate feedback, and greater accessibility, thereby fostering students' self-regulation and engagement [1, 2, 13]. These systems enable learning to be tailored to individual needs, enhancing motivation and perceived usefulness [15, 17]. However, challenges remain in designing natural interactions and establishing clear metrics to evaluate their impact on learning [16, 19]. Despite growing international interest, empirical research in Latin American contexts is limited, with few studies addressing sociocultural specificities [30]. In Mexico, evidence on AI applications in higher education remains scarce [31], underscoring the need for studies exploring how students perceive these technologies within local educational contexts. This study addresses this gap by evaluating a virtual tutor powered by generative artificial intelligence within a Mexican higher education context, exploring design factors that could enhance user experience among students.

Design factors of virtual tutors, such as interaction modality (voice vs. text) and the presence of visual avatars, play a crucial role in shaping user experience [7, 9, 10, 22]. Voice-based interactions typically enhance conversational naturalness and fluency, resulting in greater behavioral engagement and satisfaction compared with text-based interactions [7, 22]. Meanwhile, visual avatars can enhance trust and social presence, fostering positive attitudes toward the system, particularly in educational settings such as learning management systems or simulations [9, 10, 24]. Nevertheless, the effects of these factors on personalized tutoring remain underexplored, and results regarding satisfaction and perceived usefulness are mixed, reflecting the heterogeneity reported in prior reviews [15, 20]. While some studies report increases in motivation and positive perceptions, others find no significant differences or even indicate frustration with less natural responses, possibly due to differences in agent realism, synthesized voice quality, or the type of academic task used in experiments [18, 25, 26].

Interactive behavior, measured through metrics such as User Active Time (UAT) and Turns Per Minute (TPM), reflects the level of behavioral engagement during interaction [27]. Although no consensus exists regarding their operationalization, these metrics help capture the dynamics of system use [28, 29]. In this work, we examine the effects of interaction modality and the presence of a female avatar selected to align with perceptions of empathy, accessibility, and support in educational environments [9, 10, 24] on satisfaction, perceived usefulness, trust, and behavioral engagement among university students, contributing to a deeper understanding of these design factors in educational contexts.

3 Method

3.1 Participants

Undergraduate students were recruited for the study through direct messages sent by the program coordinator. Participation was voluntary. Students were invited through an online form that also collected demographic information (e.g., age, gender, semester) and previous experience with chatbots. All participants provided informed consent prior to participating in the study.

To balance the groups, participants were randomly assigned to one of the four experimental conditions, which combined interaction modality (voice or text) and avatar presence (with or without avatar).

3.2 Variables

The independent variables in this study were:

- **Interaction modality of the tutor:** voice vs. text.
- **Visual avatar presence:** with avatar vs. without avatar.

The dependent variables were as follows:

- **User satisfaction**: Reflects the general degree of enjoyment with the interaction experience [24]. It was measured using a 7-point Likert scale (1 = Strongly disagree; 7 = Strongly agree).
- **Perceived usefulness:** Assesses the user's perception of the system's usefulness. The items were adapted from the Technology Acceptance Model (TAM) [5] through direct translation into Spanish and slight linguistic adjustments to

ensure clarity for Mexican engineering students. A 7-point Likert scale was used (1 = Strongly disagree; 7 = Strongly agree).

- **User Active Time (UAT):** This metric represents the sum of interaction intervals between consecutive messages sent by the participant, considering only intervals less than or equal to 90 seconds to exclude long inactivity periods [27, 28]. In other words, UAT reflects the moments when the student is genuinely active during the conversation. Formally: UAT = Σ(intervals ≤ 90 seconds).
- **Turns Per Minute (TPM):** Corresponds to the number of turns (messages sent) divided by the total session duration in minutes. The timestamp between the first and last message, including pauses, was taken into consideration. To avoid overestimation, messages within three-second intervals were merged into a single conversational turn [28, 29].
- **Trust in the system:** Measures the user's trust in automated systems. The items from Jian et al. [21] were translated into Spanish and reformulated minimally to maintain their original meaning while improving readability for the target population. It was assessed using a 7-point Likert scale (1 = Strongly disagree; 7 = Strongly agree).

3.3 Hypotheses

The hypotheses guiding this study are the following for Interaction Modality (H1a, H1b1, and H1b2), and for Avatar Presence (H2a and H2b):

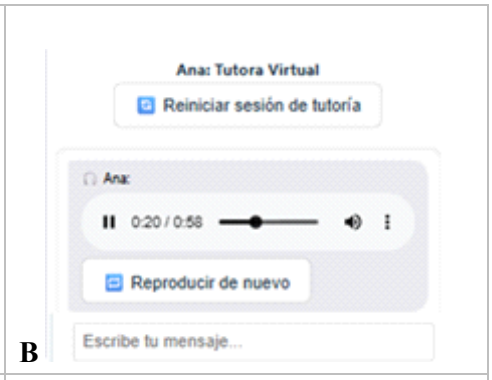
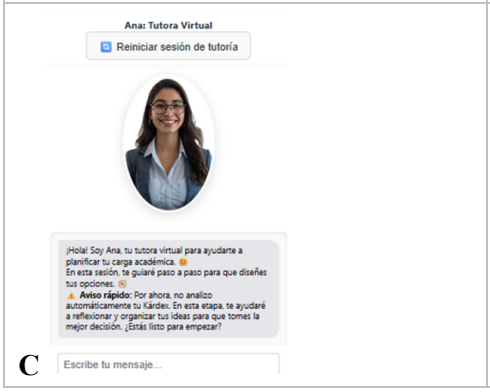
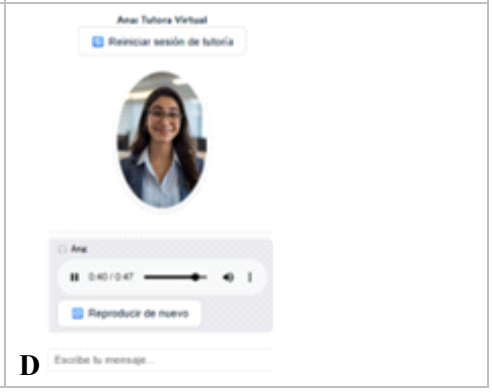
- **H1a**: The level of user satisfaction will be higher among students who interact with the tutor through voice compared with those who interact through text only.
- **H1b1**: User Active Time (UAT) will be higher among students who interact with the tutor through voice compared with those who interact through text only.
- **H1b2**: Turns Per Minute (TPM) will be higher among students who interact with the tutor through voice compared with those who interact through text only.
- **H2a**: The level of perceived usefulness will be higher among students who interact with a virtual tutor that displays a visual avatar compared with those who interact with one without an avatar.
- **H2b**: The level of trust will be higher among students who interact with a virtual tutor that displays a visual avatar compared with those who interact with one without an avatar.

3.4 Materials and Instruments

A web-based prototype of the chat module was developed to simulate interaction with a virtual tutor named Ana, powered by generative artificial intelligence. User input remained text-based in all conditions, while the tutor's interaction modality varied according to the assigned condition (i.e., voice vs. text).

All interaction with the virtual tutor, including text, audio responses, and interface elements, was entirely in Spanish. In the voice conditions, the tutor's responses were synthesized using a Spanish Text-to-Speech (TTS) engine with a natural female voice.[1]

Table 1. Screenshots of the different versions of the tutors used across the four conditions

	Text	Audio
No Avatar		
With Avatar		

[1] https://cloud.google.com/text-to-speech

The tutoring scenario consisted of a predefined script comprising six conversational blocks: introduction, reflection on the previous academic term, setting educational goals, course planning, validation, and closing. This structure ensured that all participants progressed through the same sequence of topics, allowing comparable interaction patterns across conditions. The script was designed to promote self-reflection and academic decision-making while maintaining consistency in the complexity and length of the tutor's prompts.

Although generative AI powered the system, the tutoring flow did not rely entirely on open-ended generation. Instead, a structured base prompt was provided to the Gemini 2.5 Flash model, specifying the tutor's role, tone, behavioral rules, and the sequence of conversational blocks to follow. This prompt included the complete script that defined what the tutor should address in each stage (introduction, academic reflection, goal setting, course planning, validation, and closing), ensuring that all participants experienced the same thematic progression. The AI model generated natural-sounding responses within each block while remaining constrained to the predefined structure, which helped maintain consistency across participants and prevented unintended deviations in content or order.

Satisfaction, perceived usefulness, and trust were measured through a questionnaire implemented in Google Forms. Behavioral

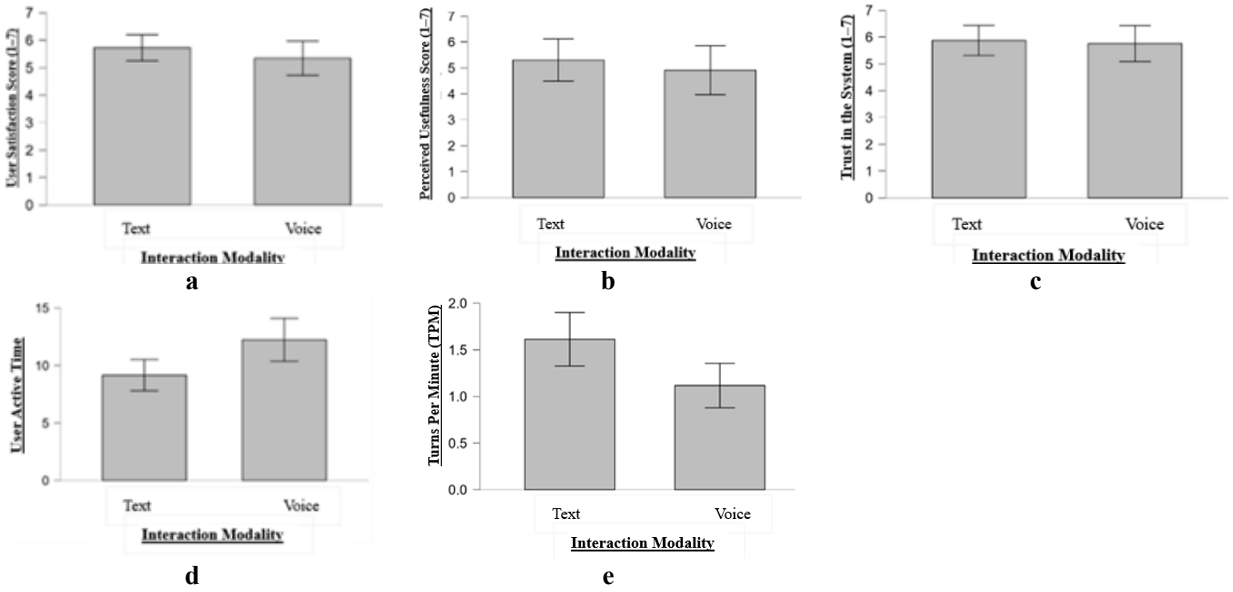


Figure 1. Text vs. Voice : Mean values with 95% confidence intervals related to a) user satisfaction, b) perceived usefulness, c) trust in the system, d) User Active Time (UAT), and e) Turns Per Minute (TPM).

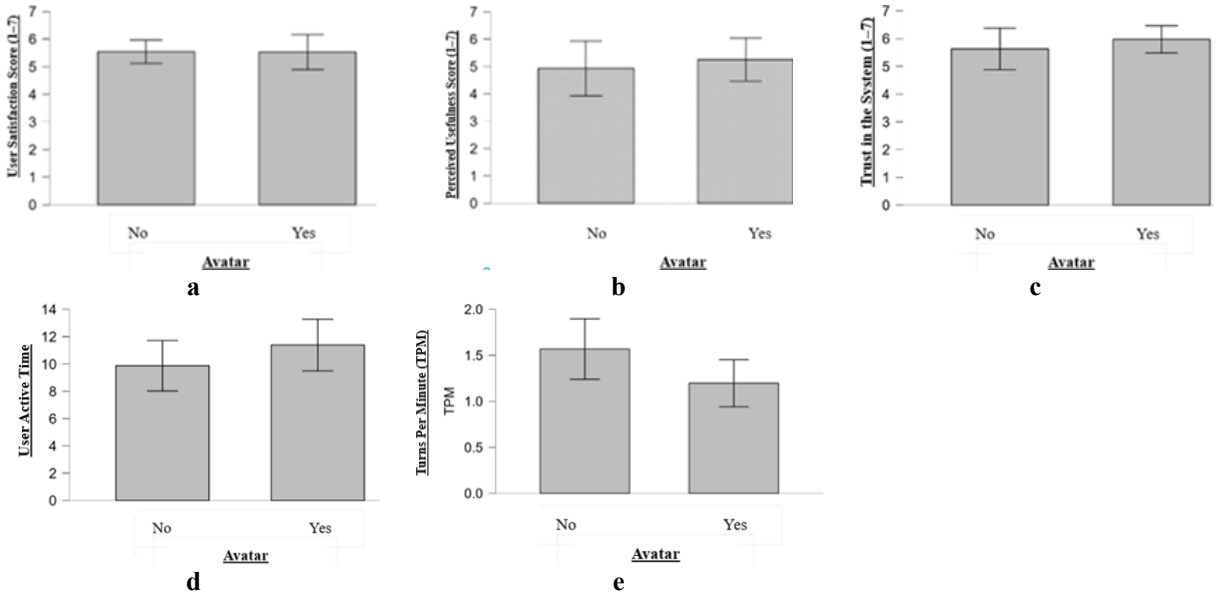


Figure 2. 4.2 Avatar Presence: Mean values of with 95% confidence intervals related to a) user satisfaction, b) perceived usefulness, c) trust in the system, d) User Active Time (UAT), and e) Turns Per Minute (TPM)

engagement was computed from the timestamp logs of each message. For each participant, the following metrics were calculated:

- **User Active Time (UAT):** The sum of interaction intervals between consecutive messages by the human participant, counting as active time only intervals less than or equal to 90 seconds, to exclude long inactivity pauses. In other words, this metric reflects only the moments in which the student is actually active during the conversation.
- **Turns Per Minute (TPM):** The number of turns divided by the total session duration (in minutes), treating messages separated by 3 seconds or less as a single turn.

In Table 1, we present a screenshot of the four different conditions used in the study. In the Text–Avatar condition, the avatar was presented as a static image, as it was not coherent to display facial animation without audio output.

In the Voice–Avatar condition, the avatar video consisted of a 6-second clip that looped during each response turn, simulating visual activity without synchronization to the audio.

3.5 Data Analysis

Independent-samples t-tests were conducted to evaluate the effects of interaction modality (Text vs. Voice) and avatar presence (with vs. without avatar) on the dependent variables: satisfaction, perceived usefulness, trust, User Active Time (UAT), and Turns Per Minute (TPM). Interaction effects were not included, as the analysis focused on the main effects of each factor in this preliminary study.

3.6 Procedure

Each participant interacted individually with the Ana virtual tutor in a session lasting approximately 12 to 20 minutes, following the predefined script described in Section 3.4. Sessions were grouped into three temporal phases: (a) Introduction and orientation (2–3 min), (b) Guided interaction (8–14 min), and (c) Evaluation (2–3 min). The platform automatically recorded interaction metrics such as User Active Time (UAT) and Turns Per Minute (TPM).

To ensure consistency across participants, all sessions followed the same structured prompt and conversational flow, regardless of condition. Instructions were standardized, and the study was conducted through a web-based platform, ensuring that all participants interacted with the same underlying application logic and progression of conversational blocks. A small pilot test was conducted prior to data collection to verify the stability of the AI-generated responses and to refine the structured prompt. These measures provided comparable conditions across the four experimental groups.

4 Results

A total of 22 undergraduate students (8 women and 14 men) participated in the experiment (M = 20.2 years; SD = 2.3; range = 17–26 years). Most participants (86.4%, n = 19) reported frequently using chatbots, while two participants (9.1%) occasionally, and one participant (4.5%) used them rarely. Participants were randomly assigned to one of the four experimental conditions of the 2×2 design: text without an avatar (n = 5), text with an avatar (n = 6), voice without an avatar (n = 5), and voice with an avatar (n = 6).

4.1 Text vs. Voice

Table 2 presents the descriptive results and t-test comparisons for the Interaction Modality factor (Text vs. Voice). No statistically significant differences were found in perceived satisfaction, usefulness, or trust. Descriptively, participants in the Text condition reported slightly higher mean scores on these three perceptual variables compared with those in the Voice condition. As shown in Figure 1a, the mean values of User Satisfaction were slightly higher for the text condition compared with the voice condition. Figure 1b displays the mean perceived usefulness, while Figure 1c shows the mean trust scores across both modalities.

Figure 1d illustrates the differences in User Active Time (UAT) across modalities, and Figure 1e shows the comparison of Turns Per Minute (TPM).

In contrast, participants in the Text condition showed significantly higher Turns Per Minute (TPM) (M = 1.61) compared with the Voice condition (M = 1.12), $t(20) = 2.97$, $p = .008$, $d = 1.26$. These results indicate that although subjective perceptions remained similar across conditions, interaction modality influenced the dynamics of participation.

4.2 Avatar Presence

Table 3 shows the results for the Avatar presence. As can be seen, none of the variables shows any statistical difference. Regarding behavioral variables, User Active Time (UAT) was higher in the With Avatar condition (M = 11.39) than in the Without Avatar condition (M = 9.87), although the difference was not statistically significant ($p = .222$). As shown in Figure 2a, user satisfaction remained stable across avatar conditions. Figure 2b presents perceived usefulness, and Figure 2c illustrates trust levels in both conditions. Differences in behavioral engagement are shown in Figure 2d for UAT and Figure 2e for TPM. In contrast, Turns Per Minute (TPM) showed a marginally significant difference, with a higher message production rate in the Without Avatar group (M = 1.57) compared with the With Avatar group (M = 1.20), $t(20) = 2.02$, $p = .057$, $d = 0.87$.

5 Discussion

The objective of this study was to examine whether interaction modality (voice vs. text) and avatar presence (with vs. without avatar) influenced user satisfaction, perceived usefulness, trust, and behavioral engagement when interacting with a virtual tutor. Regarding the proposed hypotheses, the results of the independent samples t-tests did not reveal statistically significant differences in satisfaction, usefulness, or trust; therefore, none of the perceptual hypotheses were confirmed. These results do not support hypothesis H1a, as no differences were found in satisfaction across modalities.

However, the analysis of behavioral variables revealed significant differences. Participants in the Voice condition exhibited greater User Active Time (UAT; Mean = 12.24, SD = 2.78), which aligns with previous findings reporting higher naturalness and conversational fluency in voice-based interactions compared with text-based ones [22]. These differences, also visible in Figures 1d and Figure 1e, suggest that interaction modality alters the rhythm and temporal structure of engagement rather than users' subjective evaluation of the system. This effect could also be related to the additional time users spend listening to and processing auditory responses. In contrast, those who interacted in the Text condition showed a significantly higher rate of Turns Per Minute (TPM; Mean = 1.61, SD = 0.43) compared with Voice (Mean = 1.12, SD = 0.35; $p = .008$), contradicting hypothesis H1b2, which expected faster engagement in the voice condition. A plausible explanation, extending beyond previous modality research [23], is that written interaction allows for quicker, sequential responses compared to listening. Taken together, these findings suggest that interaction modality influences the dynamics

Table 2. Independent Samples t-Test Results for Interaction Modality (Text vs. Voice)

Variable	Text Mean (SD) N=11	Voice Mean (SD) N=11	t(df)	p	d	95% CI d
User Satisfaction	5.73 (0.70)	5.35 (0.93)	1.09(20)	.288	0.47	–0.39, 1.31
Perceived Usefulness	5.30 (1.20)	4.91 (1.40)	0.71(20)	.488	0.30	–0.54, 1.14
Trust in the System	5.87 (0.84)	5.76 (1.00)	0.29(20)	.773	0.13	–0.71, 0.96
User Active Time (UATUAT)	**9.16 (2.02)**	**12.24 (2.78)**	**-2.97(20)**	**0.008**	**-1.27**	**–2.18,-0.33**
Turns Per Minute (TPM)	**1.61 (0.43)**	**1.12 (0.35)**	**2.97(20)**	**0.008**	**1.26**	**0.33,2.17**

Table 3. Independent Samples t-Test Results for Avatar Presence (With vs. Without Avatar)

Variable	Without Avatar Mean (SD) N=10	With Avatar Mean (SD) N=12	t(df)	p	d	95% CI d
User Satisfaction	5.54 (0.59)	5.53 (1.01)	0.02(20)	.985	0.01	–0.83, 0.85
Perceived Usefulness	4.93 (1.41)	5.25 (1.23)	–0.56(20)	.579	–0.24	–1.08, 0.60
Trust in the System	5.63 (1.05)	5.97 (0.78)	–0.89(20)	.385	–0.38	–1.22, 0.47
User Active Time (UAT, min)	9.87 (2.58)	11.39 (2.98)	-1.26(20)	.222	-0.54	–1.39, 0.32
Turns Per Minute (TPM)	1.57 (0.46)	1.2(0.4)	2.02(20)	.057	0.87	-0.03,1.74

of engagement, although it does not necessarily affect users' subjective evaluation of the platform.

Regarding avatar presence, no significant effects were found for any of the dependent variables. However, a marginal difference emerged in TPM ($p = .057$), where the group without an avatar produced more messages than the group with an avatar. As illustrated in Figure 2e, the absence of an avatar was associated with slightly higher TPM values. The slight reduction in interaction rate in the avatar condition may suggest a mild visual distraction effect; however, this interpretation remains speculative given the simplicity of the avatar design (a static image or short looping animation without audio synchronization). More controlled studies with higher-fidelity avatars are needed to confirm whether such effects persist with more expressive or realistic agents. Although not statistically significant, these findings indicate that including an avatar may be associated with a slight reduction in interactive activity, especially when combined with the text modality. The lack of effects related to avatar presence does not support hypothesis H2b.

These findings can be interpreted in light of prior work on interaction modality in HCI. Voice-based interactions often promote a slower conversational rhythm due to the time required to listen to system responses [22], which aligns with the higher User Active Time observed in this study. In contrast, text-based interaction enables faster turn-taking, resulting in a higher number of messages per minute. This interpretation is consistent with research showing that voice interfaces increase cognitive processing time while maintaining conversational naturalness [22, 23].

Regarding avatar presence, the absence of significant effects on perceptual measures aligns with earlier mixed evidence on social presence in educational agents [7, 15, 20]. Given the simplicity of the visual design used in this prototype, it is plausible that the avatar did not provide sufficient cues to meaningfully influence satisfaction, usefulness, or trust [8, 9]. The slight decrease in message frequency in the avatar condition may suggest a mild cognitive or visual load effect, although this interpretation remains tentative [9].

Taken together, these results highlight the importance of analyzing not only subjective evaluations but also the temporal dynamics of interaction, particularly in early-stage prototypes of AI-driven tutoring systems.

These interpretations are further supported by recent empirical work. For example, recent studies [8, 9] have shown that while avatars and synthetic voices can enrich the interaction experience, their effects largely depend on the specific design, task, and context. Other studies have also reported that some students perceive greater social presence and motivation when interacting with AI-generated avatars in learning management systems. However, such findings are primarily based on qualitative evidence rather than statistical testing [10]. These findings suggest that perceived benefits may not necessarily translate into measurable behavioral outcomes.

Overall, the results indicate that the Ana platform provides a consistent experience for students, with neither interaction modality nor avatar presence substantially altering users' perceptions. From a practical perspective, this implies that the tool can be flexibly implemented in different formats without negatively affecting user satisfaction or trust.

From a Latin American perspective, these findings offer valuable insight into how AI-based tutoring tools may perform in contexts where such technologies are still emerging, suggesting that basic implementations can achieve acceptable user experiences without requiring highly sophisticated avatar designs or complex interaction modalities.

5.1 Limitations and Future Work

This study presents several limitations. First, the small sample size ($N = 22$) and gender imbalance limit the generalizability of the findings. A second limitation lies in the relatively short interaction duration (15–30 minutes), which may not have been sufficient for the interaction modality or avatar presence to produce perceptible effects on students' experience.

Finally, the technical implementation was basic: the avatar was displayed as a static image in the Text–Avatar condition and as a short animated clip in the Voice–Avatar condition, accompanied by

a Text-to-Speech (TTS)-generated voice. The avatar design was intentionally kept simple due to the exploratory nature of this study. As the primary focus was on interaction modality rather than avatar realism, a minimal visual agent (static image or short looping animation) allowed the interface to remain consistent while avoiding unintended confounds. Additionally, implementing real-time facial synchronization with lip movement matched to the Spanish text-to-speech output would have required technical and resource demands beyond the scope of this initial prototype. These design decisions prioritized experimental control and feasibility, although they may have reduced the potential impact of the avatar on user experience and limited the generalizability of the findings regarding the avatar's effects. These design characteristics likely attenuated the observed effects, particularly the marginal difference in TPM ($p = .057$) for the avatar condition, as well as the unexpected result of higher TPM in Text compared to Voice [22, 23].

For future research, it will be necessary to work with larger and more gender-balanced samples, as well as to include more varied interaction scenarios that more closely reflect authentic academic contexts. It is also recommended to test avatars with a higher level of animation and more expressive voices, which may increase the perception of social presence and, consequently, produce more evident differences between conditions. Specifically, future studies should investigate why avatar presence slightly reduces TPM, possibly due to visual distraction [9], and why text interactions generated higher TPM than voice, contrary to expectations (H1b2). Finally, integrating qualitative methods (e.g., interviews, focus groups, observations) would allow for a deeper understanding of how students interpret and evaluate their experiences with AI-based virtual tutors.

6 Conclusion

This study evaluated the impact of interaction modality (voice vs. text) and avatar presence (with vs. without avatar) on the experience of university students interacting with Ana, a virtual tutor developed using generative AI technology. Based on a 2×2 between-subjects experimental design and independent samples t-tests, no statistically significant differences were found in satisfaction, perceived usefulness, or trust.

However, the behavioral variables revealed significant effects: students in the Voice condition exhibited greater User Active Time (UAT). In contrast, those in the Text condition showed a higher number of Turns Per Minute (TPM). Avatar presence produced a marginal reduction in TPM, suggesting a possible distractive effect that warrants further investigation. These findings indicate that although subjective evaluations of the platform remained stable across conditions, the dynamics of interaction varied depending on the modality used.

From a practical perspective, the results suggest that the platform can be flexibly implemented without compromising users' perceptions, and that academic content and functionality are key elements for user acceptance. Nevertheless, the observed differences in interaction metrics highlight relevant avenues for future research, particularly when incorporating more advanced technologies such as animated avatars or expressive voices.

In conclusion, this study offers initial empirical evidence on the use of AI-based virtual tutors in a higher education context, contributing to the international discussion on the effectiveness of such technologies in learning environments. The findings underscore the need for continued research on how design variables such as interaction modality and avatar presence influence the student experience, particularly in authentic academic scenarios.

Given the limited availability of empirical studies on AI-powered tutoring systems in Mexico, these findings offer initial evidence of how engineering students interact with generative AI tools in local higher education contexts. Highlighting this regional perspective is essential for understanding the extent to which design decisions and user responses may generalize or require adaptation when applied to different sociocultural settings.

7 Acknowledgements

The authors acknowledge the use of ChatGPT (OpenAI) to assist in the revision of English grammar in this manuscript.

8 References

[1] Schei, O. M., Møgelvang, A., & Ludvigsen, K. Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies, Education Sciences, vol. 14, no 8, p. 922, Aug. 2024, https://doi.org/10.3390/educsci14080922

[2] Belda-Medina, J., & Kokošková, V. Integrating chatbots in education: insights from the Chatbot-Human Interaction Satisfaction Model (CHISM), *International Journal of Educational Technology in Higher Education*, vol. 20, n.º 1, p. 62, Dec. 2023, https://doi.org/10.1186/s41239-023-00432-3

[3] Fink, M. C., Robinson, S. A., Ertl, B. AI-based avatars are changing the way we learn and teach: benefits and challenges, *Front. Educ.*, vol. 9, Jul. 2024, https://doi.org/10.3389/feduc.2024.1416307

[4] Nass, C., & Reeves, B. (2000). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publications.

[5] Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, vol. 13, no. 3, pp. 319-340, 1989, https://doi.org/10.2307/249008

[6] Brooke, J. (1996). *SUS: A "quick and dirty" usability scale*. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189–194). CRC Press.

[7] Zhang, Y., Lucas, M., Bem-haja, P., & Pedro, L. AI versus human-generated voices and avatars: rethinking user engagement and cognitive load, *Educ Inf Technol*, jun. 2025, https://doi.org/10.1007/s10639-025-13654-x

[8] Ma, N., Khynevych, R., Hao, Y., & Wang, Y. Effect of anthropomorphism and perceived intelligence in chatbot avatars of visual design on user experience: accounting for perceived empathy and trust, *Front. Comput. Sci.*, vol. 7, may 2025, https://doi.org/10.3389/fcomp.2025.1531976.

[9] Tan, S. F. Perceptions of students on artificial intelligence-generated content avatar utilization in learning management system, *Asian Association of Open Universities Journal*, vol. 19, no. 2, pp. 170-185, Sep. 2024, https://doi.org/10.1108/AAOUJ-12-2023-0142

[10] Islam, M. Z., & Wang, G. Avatars in the educational metaverse, Visual Computing for Industry, Biomedicine, and Art, vol. 8, no. 1, p. 15, jun. 2025, https://doi.org/10.1186/s42492-025-00196-9.

[11] Zaky, Y. A. M., & Gameil, A. A. Exploring the Use of Avatars in the Sustainable Edu-Metaverse for an Alternative

Assessment: Impact on Tolerance, *Sustainability*, vol. 16, n.º 15, p. 6604, ene. 2024, https://doi.org/10.3390/su16156604.

[12] Hilliger, I., Ortiz-Rojas, M., Pesántez-Cabrera, P., Scheihing, E., Tsai, Y. S., Muñoz-Merino, P. J., ... & Pérez-Sanagustín, M. Towards learning analytics adoption: A mixed methods study of data-related practices and policies in Latin American universities, *British Journal of Educational Technology*, vol. 51, no. 4, pp. 915-937, 2020, https://doi.org/10.1111/bjet.12933.

[13] Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. Educational AI Chatbots for Content and Language Integrated Learning, *Applied Sciences*, vol. 12, no. 7, p. 3239, Jan. 2022, https://doi.org/10.3390/app12073239

[14] Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. Interacting with educational chatbots: A systematic review, Educ Inf Technol, vol. 28, n.o 1, pp. 973-1018, Jan. 2023, https://doi.org/10.1007/s10639-022-11177-3.

[15] Labadze, L., Grigolia, M., & Machaidze, L. Role of AI chatbots in education: systematic literature review, International Journal of Educational Technology in Higher Education, vol. 20, no. 1, p. 56, oct. 2023, https://doi.org/10.1186/s41239-023-00426-1

[16] Davar, N. F., Dewan, M. A. A., Zhang, X. AI Chatbots in Education: Challenges and Opportunities, Information, vol. 16, n.o 3, p. 235, mar. 2025, https://doi.org/10.3390/info16030235.

[17] Zou, M., & Huang, L. To use or not to use? Understanding doctoral students' acceptance of ChatGPT in writing through technology acceptance model, *Front. Psychol.*, vol. 14, oct. 2023, https://doi.org/10.3389/fpsyg.2023.1259531.

[18] Yetişensoy, O., & Karaduman, H. The effect of AI-powered chatbots in social studies education, Educ Inf Technol, vol. 29, n.o 13, pp. 17035-17069, Sep. 2024, https://doi.org/10.1007/s10639-024-12485-6.

[19] Kooli, C. Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions, Sustainability, vol. 15, no. 7, p. 5614, Jan. 2023, https://doi.org/10.3390/su15075614.

[20] Pontual Falcão, T., Ferreira Mello, R., Lins Rodrigues, R. Applications of learning analytics in Latin America., *British Journal of Educational Technology*, vol. 51, no. 4, p. 871, Jul. 2020, https://doi.org/10.1111/bjet.12978.

[21] Jian, J. Y., Bisantz, A. M., Drury, C. G. Foundations for an Empirically Determined Scale of Trust in Automated Systems, *International Journal of Cognitive Ergonomics*, vol. 4, no. 1, pp. 53-71, mar. 2000, https://doi.org/10.1207/S15327566IJCE0401_04.

[22] Reicherts, L., Rogers, Y., Capra, L., Wood, E., Duong, T. D., & Sebire, N. It's Good to Talk: A Comparison of Using Voice Versus Screen-Based Interactions for Agent-Assisted Tasks, *ACM Trans. Comput.-Hum. Interact.*, vol. 29, n.º 3, p. 25:1-25:41, ene. 2022, https://doi.org/10.1145/348422.

[23] Rzepka, C., Berger, B., & Hess, T. Voice Assistant vs. Chatbot – Examining the Fit Between Conversational Agents' Interaction Modalities and Information Search Tasks, *Inf Syst Front*, vol. 24, no. 3, pp. 839-856, jun. 2022, https://doi.org/10.1007/s10796-021-10226-5

[24] Chae, S. W., Lee, K. C., Seo, Y. W. Exploring the Effect of Avatar Trust on Learners' Perceived Participation Intentions in an e-Learning Environment, *International Journal of Human–Computer Interaction*, vol. 32, no. 5, pp. 373-393, May 2016, https://doi.org/10.1080/10447318.2016.1150643.

[25] Baake, J., Schmitt, J., Metag, J. Balancing realism and trust: AI avatars, In science communication, *JCOM*, vol. 24, n.º 2, p. A03, abr. 2025, https://doi.org/10.22323/2.24020203.

[26] Canales, R., Roble, D., & Neff, M. The Impact of Avatar Stylization on Trust. Proc. IEEE VR 2024, 418–428. https://doi.org/10.1109/VR58804.2024.00063

[27] Doherty, K., & Doherty, G. Engagement in HCI: Conception, Theory and Measurement, ACM Comput. Surv., vol. 51, n.o 5, p. 99:1-99:39, nov. 2018, https://doi.org/10.1145/3234149.

[28] Karimah, S. N., & Hasegawa, S. Automatic engagement estimation in smart education/learning settings: a systematic review of engagement definitions, datasets, and methods, *Smart Learning Environments*, vol. 9, n.º 1, p. 31, Nov. 2022, https://doi.org/10.1186/s40561-022-00212-y.

[29] Ray, A. E., Greene, K., Pristavec, T., Hecht, M. L., Miller-Day, M., & Banerjee, S. C. Exploring indicators of engagement in online learning as applied to adolescent health prevention: a pilot study of REAL media, *Education Tech Research Dev*, vol. 68, n.º 6, pp. 3143-3163, dic. 2020, https://doi.org/10.1007/s11423-020-09813-1.

[30] Rivadeneira, L., Bellido de Luna, D., Fernández, C. (2025). Exploring the role of ChatGPT in higher education institutions: Where does Latin America stand? Digital Government: Research and Practice. https://doi.org/10.1145/3689370

[31] Salas-Pilco, S. Z., & Yang, Y. Artificial intelligence applications in Latin American higher education: a systematic review, International Journal of Educational Technology in Higher Education, vol. 19, art. 21, 2022, https://doi.org/10.1186/s41239-022-00326-w.