

ChatGPT assistants in online higher education and student satisfaction: a case study

Asistentes ChatGPT en educación superior en línea y satisfacción del alumnado: un caso de estudio



 Miguel-Ángel Cabeza-Rodríguez - *Universidad Francisco de Vitoria, UFV (Spain)*

ABSTRACT

The perception and satisfaction of students about a virtual assistant based on OpenAI ChatGPT 3.5, integrated in 21 different subjects of the virtual campus of an online university, have been analyzed in this study. Using a mixed methodological approach, information was collected on a sample of 391 students using the validated COMUNICA questionnaire, which included four constructs: Virtual Assistant Efficiency, Learning Impact, Skill Development, and Technical and Accessibility Aspects. The analysis included descriptive statistics, inferential statistical tests, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), complemented by a qualitative analysis of student and teacher comments. The quantitative results highlight that the female gender values the effectiveness of the assistant more than the male gender. The CFA confirmed that the factors can be grouped under a single latent variable: student satisfaction. In addition, the efficiency of the virtual assistant was found to be the most significant factor in the perception of student satisfaction, followed by the impact on learning, skill development and technical aspects. The qualitative analysis revealed mostly positive perceptions, highlighting the usefulness of the assistant in learning, an interest in extending its use to other subjects and suggestions for improvement in the accuracy of answers and functionality. It is concluded that virtual assistants have a positive impact on higher education, optimizing autonomous learning and educational interaction, although technical and design challenges persist that limit their full potential.

Keywords: artificial intelligence; chatbot; ChatGPT; higher education; online teaching; technology.

RESUMEN

En este estudio se han analizado la percepción y satisfacción de los estudiantes sobre un asistente virtual basado en OpenAI ChatGPT 3.5, integrado en 21 asignaturas diferentes del campus virtual de una universidad en línea. Utilizando un enfoque metodológico mixto se recopiló información sobre una muestra de 391 estudiantes mediante el cuestionario validado COMUNICA, que incluyó cuatro constructos: eficiencia del asistente virtual, impacto en el aprendizaje, desarrollo de habilidades, y aspectos técnicos y de accesibilidad. El análisis incluyó estadística descriptiva, pruebas estadísticas inferenciales, análisis factorial exploratorio (AFE) y confirmatorio (AFC), complementado con un análisis cualitativo de comentarios del alumnado y docentes. Los resultados cuantitativos destacan que el género femenino valora más la eficacia del asistente que el masculino. El AFC confirmó que los factores se pueden agrupar bajo una única variable latente: la satisfacción del alumnado. Además, la eficiencia del asistente virtual resultó ser el factor más significativo en la percepción de satisfacción del alumnado, seguido por el impacto en el aprendizaje, desarrollo de habilidades y aspectos técnicos. El análisis cualitativo reveló percepciones mayoritariamente positivas, resaltando la utilidad del asistente en el aprendizaje, un interés en extender su uso a otras asignaturas y sugerencias de mejora en la precisión de las respuestas y la funcionalidad. Se concluye que los asistentes virtuales tienen un impacto positivo en la educación superior, optimizando el aprendizaje autónomo y la interacción educativa, aunque persisten desafíos técnicos y de diseño que limitan su potencial completo.

Palabras clave: inteligencia artificial; chatbot; ChatGPT; educación superior; enseñanza en línea; tecnología.

INTRODUCTION

General context

In recent years, online education has experienced significant growth, driven by technological advances and the increasing demand for flexible and accessible learning modalities (Pokhrel & Chhetri, 2021). The COVID-19 pandemic accelerated the adoption of digital technologies in education, highlighting the need for innovative solutions to maintain the quality and continuity of distance learning through virtual mobility (Ruiz-Corbella & García-Aretio, 2023).

Chatbots are conversational agents, or “virtual assistants”, created by software that use artificial intelligence to simulate human conversations; they have emerged as promising tools in the educational context (Labadze et al., 2023). The ability of these virtual assistants to provide immediate and personalized assistance makes them valuable tools for enhancing the online educational experience.

In this scenario, where there is a clear relationship between self-regulated learning and academic performance (Cheng et al., 2023), virtual assistants have been used to address frequent queries, offer technical support and provide basic academic guidance, helping to alleviate the workload of faculty and administrative staff (Dwivedi et al., 2021; Thottoli et al., 2024).

State of the art

The application of artificial intelligence in education has been extensively discussed (Chen, Chen & Lin, 2020), as well as more specifically in higher education (Wang et al., 2023). The implementation of these virtual assistants in higher education has generated academic interest and debate (Hwang & Chang, 2023; Perez et al., 2020; Peters et al., 2024) because of their potential to improve interaction between students and their educational institutions. These tools can offer real-time support, answer frequently asked questions, provide personalized learning resources, and facilitate communication, especially in online courses where human interaction may be limited (Okonkwo & Ade-Ibijola, 2021).

The implementation of virtual assistants in educational environments has demonstrated multiple benefits according to recent academic literature (Kuhail et al., 2023; Perez et al., 2020). In terms of access and availability, virtual assistants provide educational support without time constraints, allowing students to access resources and assistance at any time. This is particularly valuable for students with unconventional schedules, or in different time zones with respect to the teacher (Winkler & Soellner, 2018).

Personalization of learning is another aspect that has been analyzed (Pérez-Marín, 2021). Virtual assistants can adapt their interaction according to the level of knowledge and learning pace of each student, with results indicating significant improvements in information retention when a personalized approach is used. They also enable greater efficiency in teaching management by reducing the workload of teachers, allowing them to focus on more complex pedagogical tasks (Chen, Xie et al., 2020; Onal & Kulavuz-Onal, 2024).

Educational virtual assistants present a number of challenges in their implementation, ranging from technical to pedagogical and ethical issues (Hwang & Chang, 2023). Among the technical challenges are limited technical functionality to

handle variations in questions, and the need for continuous training to improve their performance (Okonkwo & Ade-Ibijola, 2021)

Another technical difficulty is to mimic human language in an authentic and natural way. Virtual assistants must be able to construct coherent conversations based on previous messages, which is complex; however, since the release of OpenAI ChatGPT (Adiguzel et al., 2023), there has been not only a qualitative but also a quantitative leap in terms of commercialization and popularization of Large Language Models (LLM's) (Han et al., 2021; Tamkin et al., 2021), which are part of deep learning in neural networks (Perrotta & Selwyn, 2020), and this in turn in the field of machine learning, which belongs to artificial intelligence.

The effectiveness of virtual assistants in educational environments (Al-Emran et al., 2024) depends on several key factors: The efficiency of the virtual assistant, their actual impact on the learning process, their contribution to the development of specific skills, and the technical and accessibility challenges they present are determinants of student satisfaction. Understanding how these factors influence the educational experience is crucial to optimize the design and implementation of these tools in online education.

Objective

This study aims to analyze the satisfaction of students in an online university with a virtual assistant implemented on campus through an interface to OpenAI ChatGPT (Adiguzel et al., 2023; Tlili et al., 2023) in 21 different virtual classrooms in 13 different university studies. ChatGPT was chosen because it was considered the LLM with the easiest to develop virtual campus integration and because of its recognition and acceptance by the general public (Zhao et al., 2023).

Through a validated and reliable questionnaire and an inferential statistical analysis, differences by gender, age group or academic center are sought in relation to the four constructs studied: efficiency of the virtual assistant, impact on learning, development of skills, and technical and accessibility aspects.

Relationships of these constructs with student satisfaction are also sought by confirming a proposed factor model.

Additionally, the comments provided by the students are analyzed.

METHODOLOGY

Source

The present research uses a mixed (mostly quantitative) non-experimental, cross-sectional, descriptive and inferential methodology. The sample of students (n=391) used in this study was obtained through non-probabilistic convenience sampling, facilitated by the collaboration of 21 teachers from 5 academic centers of a Spanish online university that provided access to a population of N=3,419 students. A first teacher and researcher provided access to his students as a pilot in the first Applied Teaching Innovation Project (PIDA, Proyecto de Innovación Docente Aplicada) to obtain, through a structured questionnaire, information on the use of a virtual assistant with artificial intelligence integrated in the online campus of his subject. This first PIDA was entitled "Primus: first AI virtual assistant in the PROEDUCA group".

Following the pilot, the academic coordinators of all university departments were contacted to distribute a request for collaboration in a new PIDA entitled “Evaluation of the impact of an artificial intelligence virtual assistant on the learning experience of university students in an online educational environment”. From the teachers who responded, the 20 teachers with the largest number of students were selected and they provided this PIDA with direct access to their classrooms and students.

As for the pedagogical methodology and the means used, in addition to the implementation of the virtual assistant on the campus, no additional management, control or coordination was carried out in the different classrooms, with each teacher maintaining his or her academic freedom.

Instrument

Student questionnaire

As a research instrument to gather information from students, the structured questionnaire COMUNICA (Questionnaire of Opinion on the Management and Use of New Assisted Conversational Interfaces) was designed. This instrument, as shown in Table 1, consists of four items (Hair et al., 2019) for each of its four constructs: Virtual Assistant Efficiency, Learning Impact, Skill Development, and Technical and Accessibility Aspects. Each item includes a 5-value Likert scale, with 1 being equivalent to strongly disagree and 5 to strongly agree. All items are formulated in the affirmative (Haladyna et al., 2002) but, to avoid bias, half of the items (those numbered 3 and 4) were included with an opposite sense to the construct, so these items require scale inversion. An optional comments field was included at the end of the questionnaire. The instrument received a positive evaluation of suitability by the Research Ethics Committee of the online University, with reference UNIR CEI 101/2023.

Table 1

Items and constructs of the COMUNICA questionnaire

Construct	Item id.	Item
Virtual Assistant Efficiency	E1	The AI virtual assistant helped me better understand the concepts of the subject.
	E2	The AI virtual assistant provided information and examples that were relevant and applicable to the subject topics.
	E3	The answers provided by the AI virtual assistant were difficult to comprehend.
	E4	The AI virtual assistant provided inaccurate or irrelevant answers to my queries related to the subject.
Learning Impact	IM1	The AI virtual assistant encouraged me to explore and learn autonomously beyond the contents of the subject.
	IM2	Using the AI virtual assistant helped him to better understand the concepts of the subject.
	IM3	The AI virtual assistant failed to clarify my doubts and questions related to the subject
	IM4	The AI virtual assistant made it difficult for me to understand concepts in the subject.

Construct	Item id.	Item
Skills Development	D1	The use of the AI virtual assistant prompted me to develop and apply critical thinking on the topics of the subject.
	D2	Using the AI virtual assistant helped me improve my research and information-seeking skills.
	D3	Interacting with the AI virtual assistant made it difficult to develop my ability to solve problems related to the subject.
	D4	The AI virtual assistant made it difficult for me to practically apply the knowledge acquired in the subject.
Technical and Accessibility Aspects	A1	Accessing the AI virtual assistant through the campus was simple and enjoyable.
	A2*	When technical issues arose with the AI virtual assistant, they were quickly resolved.
	A3	The AI virtual assistant's interface was complex and difficult to use.
	A4	I encountered technical issues in using the AI virtual assistant.

*Note**: Item A2 was finally eliminated due to reliability optimization criteria.

Source: elaborated by the author.

Within the first PIDA the instrument was validated by 6 experts, and subsequently a pilot was carried out with 66 students of the Physics I course of the Physics Degree of the University in the first term of the academic year 2023/2024, in which the only researcher of that first PIDA was also the teacher. The questionnaire for this pilot obtained responses from 17 students.

At the beginning of the second quarter of the same academic year, the second PIDA was carried out with the collaboration of 20 teachers from 5 of the 6 University centers (see Table 2): the Faculty of Health Sciences, the Faculty of Law, the Faculty of Business and Communication, the Faculty of Education, and the School of Engineering and Technology. The only academic center not present was the Faculty of Social Sciences and Humanities, from which there were no teachers with a sufficient number of students to be selected for the PIDA. A total of 374 students responded to the questionnaire for this second PIDA.

Table 2

Academic centers, studies and subjects involved in the research of the second PIDA

Teacher	Academic Center	Study	Year	Subject
1	Faculty of Education	Bachelor's Degree in Primary Education	3	Knowledge of Social Sciences in Primary Education
2	Faculty of Education	Bachelor's Degree in Primary Education	3	Knowledge of Social Sciences in Primary Education
3	Faculty of Education	Bachelor's Degree in Primary Education	3	Didactics of Physical Education in Primary Education
4	Faculty of Education	Bachelor's Degree in Early Childhood Education Teaching	2	Music education and its didactics
5	Faculty of Education	Bachelor's Degree in Primary Education	3	Morphosyntax and Semantics of the English Language
6	Faculty of Education	Bachelor's Degree in Primary Education	2	Music Education and its Didactics
7	Faculty of Education	Bachelor's Degree in Primary Education	2	Mathematics Teaching in Primary Education
8	Faculty of Education	Bachelor's Degree in Primary Education	1	Information Technologies Applied to Education II
9	Faculty of Education	Bachelor's Degree in Early Childhood Education Teaching	2	Information Technologies Applied to Education I
10	Faculty of Education	Bachelor's Degree in Early Childhood Education Teaching	1	Education and Coexistence Inside and Outside the Classroom
11	Faculty of Education	Master's Degree in Teacher Training for Secondary Education, High School, and Vocational Education	1	Learning and development of the adolescent's personality
12	Faculty of Education	Master's Degree in Teacher Training for Secondary Education, High School, and Vocational Education	1	Learning and development of the adolescent's personality
13	Faculty of Business and Communication	Bachelor's Degree in Finance and Accounting	1	Microeconomics
14	Faculty of Business and Communication	Bachelor's Degree in Business Administration	1	Personal and organizational communication
15	Faculty of Business and Communication	Master in Digital Marketing	1	Mobile Marketing
16	Higher School of Engineering and Technology	Master in Renewable Energies	1	Hydraulic and Geothermal Energy
17	Higher School of Engineering and Technology	Master in Renewable Energies	1	Sea Energy
18	Faculty of Law	Bachelor's Degree in Law	1	Fundamentals of Public Law
19	Faculty of Law	Master's Degree in Access to the Legal Profession and the Attorney General's Office	1	Contentious-administrative procedural practice

Teacher	Academic Center	Study	Year	Subject
20	Faculty of Health Sciences	Master in Clinical Neuropsychology	1	Neuropsychological Rehabilitation

Source: elaborated by the author.

The 20 teachers in the second PIDA were trained by the principal researcher on the virtual assistant, who was also responsible for the configuration of the assistant in each subject. These teachers explained the PIDA to their students, informing them of the limitations of the assistant, including the possibility of biases and errors, and encouraged them to use their critical judgment and to search for confirmation from the teacher, or additional sources, when deemed necessary. At the end of their teaching, they provided their students with the questionnaire designed by the principal researcher, collecting a total sample of 374 responses in the second PIDA, which added to the 17 from the first PIDA constituted 391 total responses.

The data collected from the questionnaire were studied using descriptive statistics. After testing the internal consistency of each construct (Cronbach, 1951), the possible differences due to gender, age groups, and different academic centers in each of the four constructs were analyzed by means of inferential statistics.

After verifying the adequacy of sampling for factor analysis, an *Exploratory Factor Analysis* (EFA) was performed as a preliminary and key step to identify underlying structures, as it allows detecting latent variables reflecting theoretical constructs relevant to the topic of study (Hair et al., 2019). This exploratory approach was necessary due to the initial absence of a clear initial hypothesis on factor structure, which required an exploration of the data before attempting to validate a proposed theoretical model. The tool used to develop the descriptive, inferential and EFA statistical analysis was IBM SPSS 29.0.2.

The EFA results were used to determine the number of dimensions, which provided a solid basis for the subsequent *Confirmatory Factor Analysis* (CFA) (Brown, 2015; Kline, 2023). The tool used for the development of the CFA was e1 R 4.4.2 programming language with the lavaan 0.6-19 library (Rosseel, 2012).

In order to minimize the impact of non-normality in the data (Curran et al., 2003) an *Asymptotic Distribution-Free* (ADF) estimation with a *bootstrapping* of 5000 resamples (Yung & Bentler, 1994) was employed in the AFC. This estimator is also known as *Weighted Least Squares* (WLS) in some tools, such as the R language lavaan library.

The four goodness-of-fit employed in the CFA and their respective cutoff values were: the *Tucker-Lewis index* $TLI > 0.95$ (Hu & Bentler, 1999), the *Standardized Root Mean Square Residual* $SRMR < .1$ (Kline, 2023), the *Comparative Fit Index* $CFI > .95$ (Hu & Bentler, 1999), and the *Root Mean Square Error of Approximation* $RMSEA < .05$ (Browne & Cudeck, 1992).

Finally, the responses to the optional comments field available at the end of the questionnaire were analyzed in order to study the students' opinions in a qualitative way. First, the overall sentiment of the comments was evaluated with machine learning sentiment analysis techniques (Nandwani & Verma, 2021) using the R 4.4.2 programming language and the *SentimentAnalysis* library.

Secondly, to capture in a contextualized way the complexity of the perceptions, from technical aspects to ethical implications, a qualitative methodology with an inductive approach based on open coding and thematic categorization was employed.

The data came from the corpus of 185 student comments, and three phases were followed:

- **Initial coding:** An iterative reading of the comments was carried out to identify emerging concepts, using the *grounded theory* technique (Glaser et al., 1968) to avoid predefined biases.
- **Thematic grouping:** The codes were organized into categories and subcategories through a process of constant comparison, prioritizing recurrence and pedagogical relevance.
- **Cross-validation:** The internal consistency of the categories was verified by triangulation between two researchers, ensuring that each comment was assigned to the most accurate subcategory.

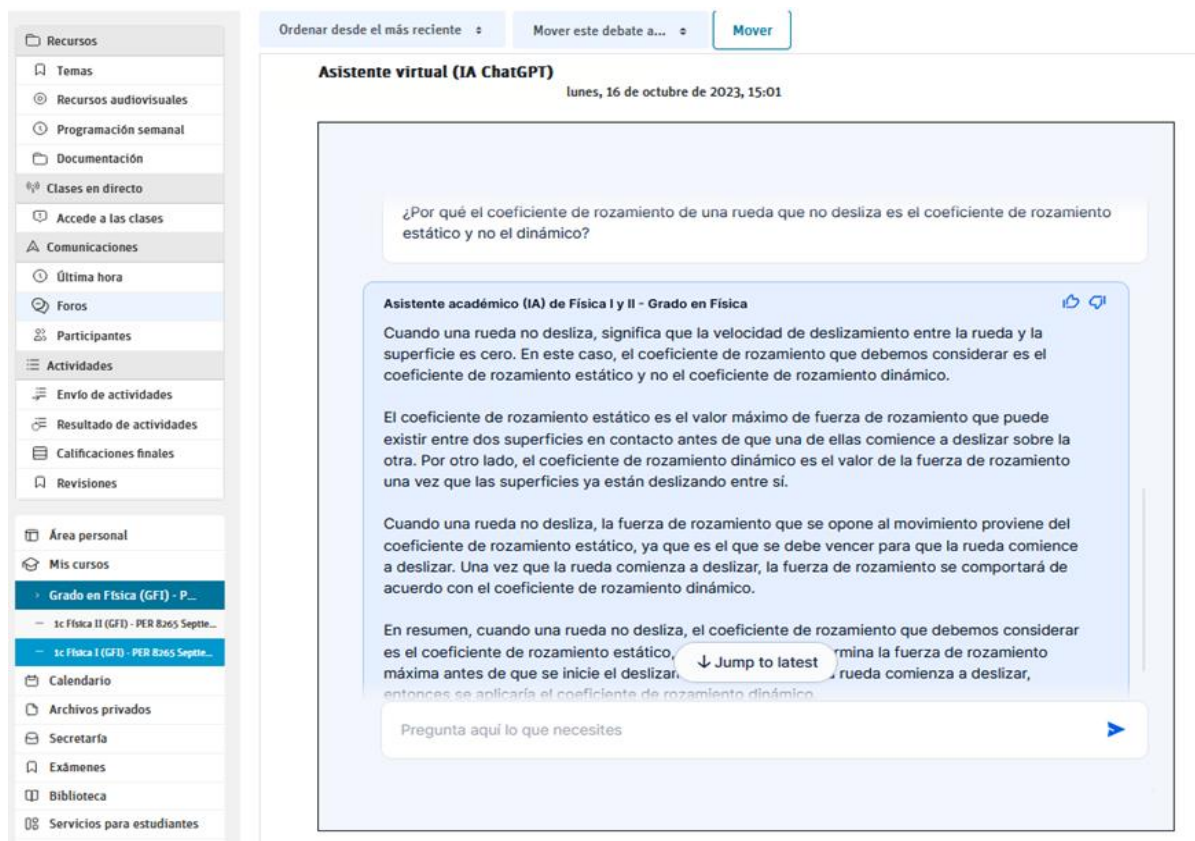
Virtual assistant

Pedagogically, the virtual assistant takes the role of an assistant and coach, with an interaction chosen and initiated by the student (Pérez-Marín, 2021), who must select a link on the campus to display a window (an *iframe* in HTML language) integrated in the subject of the online campus.

The Application Programming Interface (API) of an LLM, namely OpenAI ChatGPT 3.5, was used. The student interface was integrated into the online campus of their respective subjects, see Figure 1. The virtual assistant template for the subjects was adjusted using the LLM API so that the LLM's temperature hyperparameter, which controls their level of creativity, was kept to a minimum, in order to reduce the possibility of hallucinations in student responses (OpenAI, n. d.).

Figure 1

Example of integration of the virtual assistant in the online campus for a subject



Source: elaborated by the author.

Prompt engineering settings (White et al., 2023) were employed to configure the virtual assistant's LLM in its functions towards the specific knowledge domain of the corresponding subject, as well as to minimize the possibility of hallucinations and encourage self-directed learning (Chang et al., 2023). The wording of the text in request engineering is key to the quality of responses (Liu et al., 2024), and in this study the following template was implemented for all 21 virtual assistants:

"You are a virtual assistant for the students of the subject _____ in the undergraduate/master's degree _____.

Your goal is to enrich student learning through accurate, applicable and well-founded answers, while promoting intellectual curiosity and self-directed inquiry.

When responding, be sure to follow the guidelines below:

- 1. Base your answers on established academic principles and theories, citing relevant examples of real-life applications where possible.*
- 2. Prioritize accuracy and academic relevance, avoiding unsubstantiated speculation.*
- 3. Whenever possible, include case studies related to the topic in question, to facilitate understanding and application of the knowledge presented.*
- 4. When providing statistical information or data, clarify the context and source of such information, and be transparent about the reliability and limitations of the data.*

5. *If a question is beyond the scope of your training or is too specific to be answered accurately, guide the user to consult with their teacher.*
6. *For questions involving multiple interpretations or approaches, present a spectrum of academic perspectives, highlighting the diversity of opinions and conclusions.*
7. *Use clear, formal language, appropriate for an academic context, explaining any subject-specific jargon."*

RESULTS

Descriptive statistical analysis

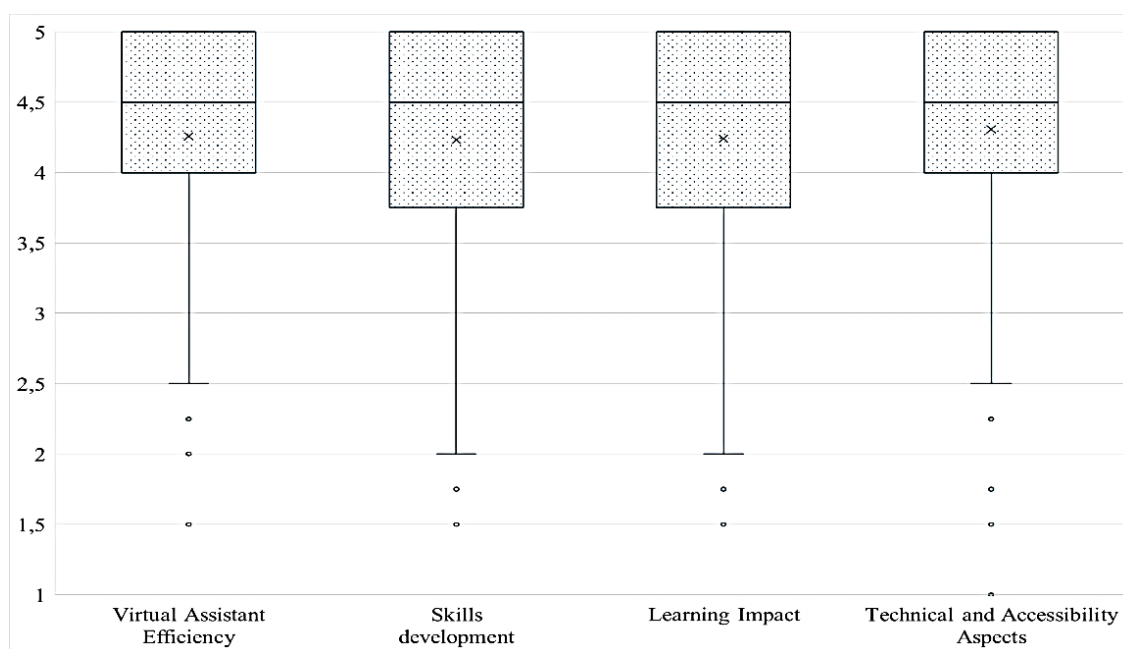
Of the total students enrolled $N=3.419$ in the 21 subjects, 11% responses were obtained, with a sample ($n=391$) in which the gender distribution was 68% female, 32% male, and 0% other genders.

As for the different age generations, the distribution was as follows: born before 1964 (*Baby Boomers*) 1%, between 1965 and 1980 (generation X) 25%, between 1981 and 1996 (generation Y) 52%, and after 1997 (generation Z) 23%.

The student body of the sample was distributed in 5 of the 6 academic centers of the University. In the Faculty of Education 57%, in the Faculty of Business and Communication 20%, in the Faculty of Health Sciences 10%, in the School of Engineering and Technology 7%, and in the Faculty of Law 6%.

Figure 2 shows that there is a clear lack of normality in the distribution of the results. This last aspect is corroborated by a normality test (Anderson & Darling, 1952) on the four constructs: $A_E=17.35$, $A_D=15.08$, $A_I=15.90$, $A_A=18.20$, $p<.001$. A box plot of the questionnaire results for the four constructs shows distributions with a certain similarity; in addition, it can be seen that the four medians are found to be at a value of 4.5.

Figure 2
Box plots of constructs



Source: elaborated by the author.

Inferential statistical analysis

Analysis of constructs according to gender, age groups and academic institution

For the construct of Virtual Assistant Efficiency, there is a higher rating in the female group, $MD_f=4.50$, versus the male group, $MD_m=4.25$, which by means of a median test for independent samples turns out to be significant $p=.07$ ($\alpha=.05$). Furthermore, according to a Mann-Whitney test $U(n_f=271, n_m=120, df=1, p=.03)=14.142.5$ (non-parametric test for comparison of independent distributions) the null hypothesis that both distributions are equal between the two gender categories can be rejected. The effect size turned out to be small (Cohen, 2013; Hattie, 2023) with $\hat{\delta}=.22$ as a measure of effect size that quantifies the difference between the means of two groups in standard deviation units, allowing to assess the practical magnitude of such a difference.

For this same construct, and using Kruskal-Wallis tests (non-parametric comparison of independent distributions), no statistically significant differences were found between age groups $H(n_{BB}=3, n_X=97, n_Y=202, n_Z=89, df=3, p=.42)=2.69$, or between academic centers $H(n_{Education}=224, n_{Business}=77, n_{Health}=40, n_{Law}=22, n_{Engineering}=28, df=4, p=.21)=6.73$.

Regarding the construct Skill development and gender, no statistically significant differences were found between genders according to a Mann-Whitney test $U(n_f=271, n_m=120, p=.89)=13.58$.

For this same construct, and using Kruskal-Wallis tests, no statistically significant differences were found between age groups $H(n_{BB}=3, n_X=97, n_Y=202, n_Z=89, df=3, p=.16)=5.23$, or between academic centers $H(n_{Education}=224, n_{Business}=777, n_{Health}=40, n_{Law}=22, n_{Engineering}=28, df=4, p=.23)=6.42$.

Analyzing the Impact on Learning construct, no statistically significant differences were found between genders according to a Mann-Whitney test $U(n_f=271, n_m=120, p=.28)=15.12$.

For this same construct, and using Kruskal-Wallis tests, no statistically significant differences were found between age groups $H(n_{BB}=3, n_X=97, n_Y=202, n_Z=89, df=3, p=.16)=5.83$, or between academic centers $H(n_{Education}=224, n_{Business}=777, n_{Health}=40, n_{Law}=22, n_{Engineering}=28, df=4, p=.19)=7.59$

Finally, for the Technical Aspects and Accessibility construct, no statistically significant differences were found between genders according to a Mann-Whitney test $U(n_f=271, n_m=120, p=.10)=15.72$.

For this same construct, and using Kruskal-Wallis tests, no statistically significant differences were found between age groups $H(n_{BB}=3, n_X=97, n_Y=202, n_Z=89, df=3, p=.79)=1.52$, or between academic centers $H(n_{Education}=224, n_{Business}=777, n_{Health}=40, n_{Law}=22, n_{Engineering}=28, df=4, p=.07)=9.89$.

Validation, adequacy of sampling and reliability

For the validation of the instrument, the Scale Content Validity Index (S-CVI) was applied to evaluate the relevance and representativeness of the questionnaire items according to the 6 experts, reaching a high average S-CVI/Ave=99.31. In addition, an S-CVI/UA of 95.83% shows that a considerable proportion of items received unanimous agreement from the experts. These index values exceed the commonly accepted levels for confirming content validity (Haynes et al., 1995; Polit et al., 2007).

The adequacy of sampling for factor analysis has been verified from three perspectives. Firstly, a Kaiser-Meyer-Olkin test, obtaining a KMO=.946 which can be considered excellent (Kaiser, 1974). Secondly, a sphericity test (Bartlett, 1954) $\chi^2=4543.25$, $gl=105$, $p<.001$ was passed. And thirdly, an analysis of the correlation matrix was performed, resulting in $r_{12}=.59$.

In this last analysis, the choice of a Promax rotation was consistent with the results of the correlation matrix, where the correlation is moderate, so the use of an oblique rotation is appropriate for this analysis, given that the factors are not orthogonal: The moderate correlation between components 1 and 2 (.59) may reflect a significant relationship between them, and they share approximately 35% of their variance, but they are not perfectly redundant.

The feasibility of the reliability analysis by Cronbach's alpha is supported by two aspects. First, by the existence of an eigenvalue $8.80 > 3$ (Yurdugül, 2008), and secondly by a sample size that exceeds the minimum $141 < n = 391$ (Bonett, 2002). This minimum has been obtained starting from a Cronbach's alpha value in the null hypothesis $CA_0=.7$, an expected value $CA_1=.8$, a power of 90%, a type I error probability $\alpha=.05$, and four items per construct.

Analyzing the reliability of each of the four constructs (see Table 3), it can be seen that they have obtained Cronbach's alphas equal to or higher than .8, indicating a level of internal consistency adequate for basic research (Nunnally, 1978). Additionally, none of them exceeds a Cronbach's alpha of .9, which could indicate an unnecessary level of redundancy between items (Streiner, 2003).

Table 3
Cronbach's alphas of the COMUNICA questionnaire constructs

Construct	Items per construct (initial)	Initial Cronbach's Alpha	Item to remove to improve Alpha	Final Cronbach's Alpha	Items per construct (final)
Virtual Assistant Efficiency	4	.82	None	.82	4
Skills development	4	.82	None	.82	4
Learning Impact	4	.80	None	.80	4
Technical and Accessibility Aspects	4	.79	A2	.80	3

Source: elaborated by the author.

In the case of the Technical and Accessibility Aspects construct, it improves its Cronbach's alpha from $CA_A=.79$ to $CA_A=.80$ when item A2 is removed, so this item is removed from the rest of the analysis and this construct is left with the remaining 3 items. This value of 3 items per construct is sufficient for research (Hair et al., 2019) and compatible with the minimum sample size for 3 items per construct, $159 < n = 391$ (Bonett, 2002).

Exploratory Factor Analysis

In a new exploratory factor analysis of 15 items of the instrument, not having previously established a theoretical framework that can explain correlations between factors, the principal components are studied with a Promax rotation, to obtain a

simple and interpretable factor structure. Of these 15 items analyzed, two eigenvalues greater than 1 (8.80 and 2.09) were obtained, which explain a total accumulated variance of 58.13% and 71.27% respectively.

Table 4
EFA standard matrix

Item	Component	
	1	2
A3	.85	-.04
A4	.83	-.03
D4	.83	.06
I4	.84	.04
E3	.82	-.02
D3	.80	.02
E4	.77	.02
I3	.74	-.02
D1	-.12	.92
E1	.02	.85
I1	-.02	.84
I2	.04	.85
D2	.01	.86
E2	.04	.85
A1	.23	.63

Source: elaborated by the author.

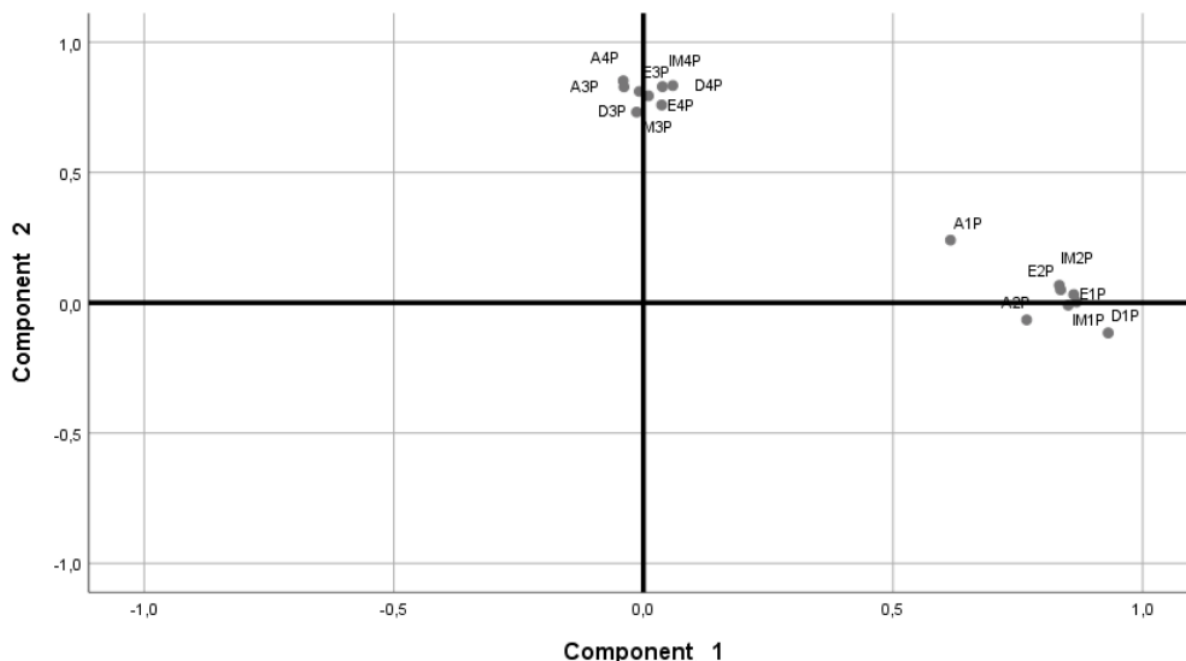
Table 4 shows the standard matrix, obtained in 3 iterations, which adjusts the factor loadings considering the correlation between components by using a Promax oblique rotation with Kaiser normalization. Two components can be clearly seen in it. In the first component the eigenvalue is equal to 8.80. This component contains all 8 items that were originally formulated with a negative association to their respective construct. They are easily recognizable by their numbering in the respective construct, with numerical identifier 3 or 4; their original Likert-scale data required a reversal for data processing.

As for the second component, it has an eigenvalue equal to 2.09. This component contains the 7 items that were formulated with a positive association to their respective construct. They are easily recognizable by their numbering in the construct with numerical identifier 1 or 2.

It is considered that this double component structure, separating questions with positive and negative associations, is nothing more than an effect of the method, that is, it has been produced by the tendency of the items formulated positively or negatively to be grouped into factors separated not by their actual content, but by the format in which they are formulated. Figure 3 shows the rotated space component plot. Responses to negative and positive items tend to load different components, even when both types of items purport to measure the same constructs (Marsh, 1996; Roszkowski

& Soven, 2010). Due to this effect, it is estimated that a factor structure with a mirror reflection of the same component has been obtained, with only one differentiated real dimension that allows explaining a total of 71.27% of the variance.

Figure 3
Components graph in rotated space

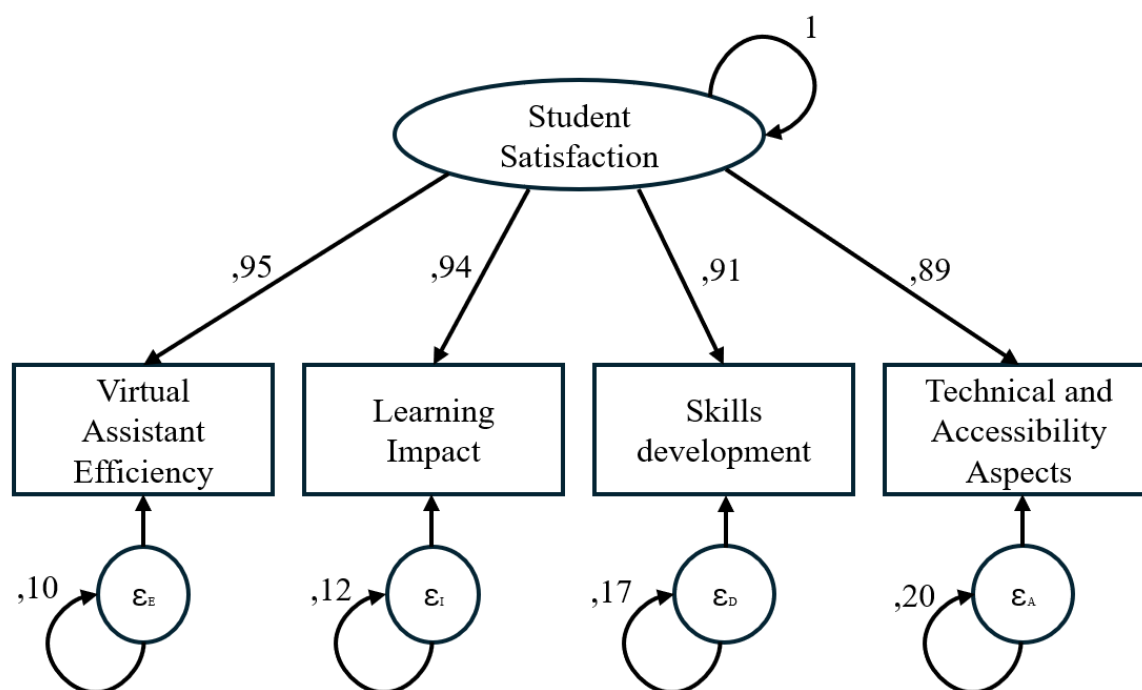


Source: elaborated by the author.

Confirmatory Factor Analysis

Figure 4 shows the model developed in the CFA, in which "Student satisfaction" has been defined as the main latent variable, which is reflected through the four observed variables: Efficiency of the virtual assistant, impact on learning, development of skills, and technical accessibility aspects. Each of these factors has been found to be associated with a factor loading (.95, .94, .91 and .89 respectively), showing the strength of the relationship between the latent variable and the observed indicators. In addition, each indicator includes an associated error term (ϵ_E , ϵ_I , ϵ_D and ϵ_A), with specific error variances (.10, .12, .17 and .20 respectively).

Figure 4
Confirmatory Factor Analysis Model



Source: elaborated by the author.

On the other hand, the minimum sample size to perform this CFA, $30 < n = 391$ (Wolf et al., 2013), with one latent variable and four factors with loadings $\lambda \geq .80$, is exceeded.

As for the goodness-of-fit indices, the Tucker-Lewis index obtained is greater than 0.95 (Hu & Bentler, 1999) with TLI=0.97. The SRMR value obtained=.05. is lower than .09 (Hu & Bentler, 1999) and .1 (Kline, 2023). In the *Comparative Fit Index*, the CFI=.98>.95 (Hu & Bentler, 1999). Finally, an RMSEA index=0.04<0.05 (Browne & Cudeck, 1992). All the indicated indexes confirm the correctness of the fit.

Although the value of $\chi^2/df=1.63 < 2$ with $p=.16 > .05$ may seem correct, it is only stated for information purposes only, as its use as a goodness-of-fit index is inadvisable (Brown, 2015; Wheaton, 1987).

Qualitative analysis

Of the students who participated by answering the questionnaire ($n=391$), 47% ($n_{com}=185$) included their comments and suggestions in an optional field reserved for this purpose at the end of the questionnaire. The sentiments (Nandwani & Verma, 2021) of the 185 comments were analyzed with machine learning using the *SentimentAnalysis* library (Feuerriegel & Pröllochs, 2023) in the R 4.4.2 programming language, using the *SentimentQDAP* index. The result was mostly positive with a *SentimentQDAP index*=66%, where its original output range [-1.1] has been converted to [0.100], with 0% being totally negative, and 100% being totally positive.

Using an inductive approach based on open coding and thematic categorization, the categories and subcategories shown in Table 5, which includes a representative example of each, were obtained. Except for the “Other comments” category that includes acknowledgements without critical evaluation or neutral generic comments,

the three most frequent subcategories and categories were: “Positive comments” of “General perception” (17%), “Academic support” in “Impact on learning” (10%), and requests for “Integration in other areas” such as “Suggestions for improvement” (8%).

Table 5
Categories and subcategories of comments

Category	Subcategory	Number of comments	Percentage	Representative example
1. General perception	1.1. Positive feedback	32	17%	"Excellent tool, congratulations."
	1.2. Negative comments	7	4%	"The chatbot is completely useless."
	1.3. Neutral comments	5	3%	"It's useful, but it doesn't replace the teacher."
2. Functionality and usability	2.1. Ease of use	12	6%	"Intuitive and easy to use."
	2.2. Accuracy of responses	9	5%	"Sometimes it gives erroneous or confusing information."
	2.3. Technical errors	4	2%	"It does not allow you to save a history of sessions."
3. Impact on learning	3.1. Academic support	18	10%	"It clarifies complex scientific concepts."
	3.2. Limitations	8	4%	"It does not solve practical problems, only theoretical ones."
4. Comparison with other tools	4.1. Virtual assistant vs. ChatGPT4	6	3%	"ChatGPT4 gives more complete answers."
	4.2. Competitive advantages	3	2%	"It is adapted to the specific syllabus."
5. Suggestions for improvement	5.1. Extension of functionalities	10	5%	"Include graphics or links to external resources."
	5.2. Training and support	5	3%	"Offer tutorials to use the tool."
	5.3. Integration into other areas	14	8%	"Implement it in all subjects."
6. Educational innovation	6.1. Acceptance of AI	9	5%	"It's the future of education."
	6.2. Resistance or concerns	3	2%	"It encourages laziness in some students."
7. Other comments		40	22%	"Thank you for the tool" (acknowledgements without critical evaluation) or "All good" (neutral generic comments).

Source: elaborated by the author.

DISCUSSION

Analysis of results

Inferential statistical analysis

A statistically significant higher evaluation of the efficiency of the virtual assistant by the female gender was found. Differences in expectations and perceptions between genders may have influenced the evaluation of the efficiency of the virtual assistant, however the result is the opposite of that indicated by some authors (Cai et al., 2017), for whom the male gender shows a more favorable attitude towards the use of technology than the female gender, although with smaller differences.

After finding a single component in the EFA, and proposing student satisfaction as a latent variable, it can be affirmed that all the goodness-of-fit indices of the AFC indicate that the proposed model is acceptable, and therefore there is convergence of the model with the data. The CFA model suggests that “Virtual assistant efficiency” is the most significant indicator of satisfaction, while “Technical and accessibility aspects” have the lowest relationship. This model allows to evaluate the relative contribution of satisfaction to each factor according to the data and the proposed model.

The factor loadings found are within the range .5 and .95 (Bagozzi & Yi, 1989), specifically $.89 < \lambda < .95$ $p < .001$, suggesting that the measured factors are mostly explained by the latent variable, student satisfaction. It is recommended (Hair et al., 2019) that the standardized factor loadings are greater than .7; this implies that the variance explained by the factor (communality) is at least 50% of the total variance of the variable, which in turn means that the uniqueness (error variance) is $\epsilon < .5$, indicating that the observed factors are adequately represented by the latent variable (Hair et al., 2019). These criteria ensure that the latent variable explains a significant proportion of the variance in the observed variables, confirming the validity of the model.

Qualitative analysis

The sentiment analysis using machine learning yielded a mostly positive value (66%). To obtain a more complete picture, the inductive analysis shows that, in terms of the suggestions and opinions expressed by the students in the comments field of the questionnaire, among the three most commented aspects, the usefulness and contribution to learning was found in first place. Most of the students highlight that the AI virtual assistant is a very useful tool that has helped them to understand complex concepts, solve doubts quickly and improve their learning process in the subjects.

The second most commented aspect was the desire for implementation in other subjects. Many students expressed their interest and the need for the assistant to be incorporated into other subjects, considering that its use would be beneficial and would increase the effectiveness of their study in general.

The third most frequent aspect included observed limitations and suggestions for improvement. Some students noted areas of improvement for the virtual assistant, such as improving and accuracy and depth of responses, incorporating clearer examples, better guidance on how to use the tool effectively, and the ability to maintain conversation history, a functionality not present in the implementation conducted.

In some cases, they also expressed that the effectiveness of the virtual assistant was limited because they lacked more guidance on how to use the tool. In addition, some felt that the virtual assistant did not replace the value of critical thinking and interaction with teachers.

In general, the students' comments reflect a mostly positive perception of the virtual assistant as a complement in the educational process, recognizing its benefits, but also pointing out areas where it can be improved to maximize its learning potential. This mostly positive feeling in the comments is reflected in the quantitative analysis of feelings, and in the qualitative analysis of the 185 comments received in the COMUNICA questionnaire.

Although the clearly positive comments (17%) clearly outnumber the negative ones (4%), there were some qualified acceptances because it was perceived as a complement, not a replacement, for human interaction (“it does not replace the teacher”). In addition, 10% of the comments were related to the impact on learning in the academic support provided by the virtual assistant. Finally, the majority of suggestions (8% of all comments) requested the extension of the assistant's existence to all subjects of the students' studies. Overall, the qualitative analysis underscores the need to balance technological advances with student-centered pedagogies, ensuring that AI in education prioritizes academic rigor and accessibility.

Research limitations

This study employs non-probability convenience sampling, which limits the generalizability of the results. For future work, probability sampling and designing strategies to balance the sample by gender would be considered, the latter, although representative of the studies, is unbalanced with 68% women. In addition, a longitudinal analysis would enrich the understanding of the sustained effect of the tool on student learning and satisfaction. For this reason, a line of future research could be a longitudinal analysis to follow-up at various stages, which would certainly increase the robustness of the findings and offer a dynamic vision.

It was not possible to implement functions that would allow a subsequent learning analysis (*Learning Analytics*) (Chang et al., 2023), so in future implementations this feature should be included in order to extract more information about the use of the tool by the students, such as number of sessions, session duration, etc.

During the second term of the academic year 2023/2024, in the execution of the second PIDA with the 20 teachers involved, version 4 of ChatGPT (Peters et al., 2024) was already available, so some students made comparisons and found that in most cases the response with direct access to this version was more thorough than with version 3.5 implemented in the virtual campus for the assistant.

No specific mechanisms were incorporated to control for other external variables (e.g., socioeconomic conditions, previous experience with digital technologies, or other environmental factors) that could influence the perception of the virtual assistant. Future research could explore designs that allow for the control or inclusion of a greater number of covariates, as well as the use of statistical methods (e.g., multivariate analysis or structural modeling) to mitigate the impact of possible external variables.

The 20 teachers of the second PIDA were contacted in unstructured interviews, but no qualitative or quantitative research on their opinions was formally developed due to time constraints. These interviews mainly highlighted the need for greater

customization in the configuration of the virtual assistant to adapt it to the particular characteristics of each subject, in order to improve the accuracy of their responses

For future research, it might be advisable to create a questionnaire as an additional structured instrument to collect the perceptions of teachers involved in the development of the assistant in their classrooms, which will provide valuable information in addition to that of the student body (Kasneci et al., 2023).

Conclusions

This study confirms the relevance and effectiveness of artificial intelligence-based virtual assistants in the field of online higher education, highlighting both their benefits and the challenges of their implementation, not only in a given field of knowledge (Polverini & Gregorcic, 2024; Taani & Alabidi, 2024; Vierhauser et al., 2024; Wardat et al., 2023) but for 21 different university degrees from different centers in an online university.

First, the quantitative findings show that the female gender has rated the efficiency of the virtual assistant more highly. Males and females differ significantly in how they perceive the usefulness and usability of technological tools, and in particular, women tend to be more influenced by social factors and to have different expectations about the usefulness of technology compared to men, which may be the reason for the higher valuation of the efficiency of the virtual assistant by the female gender. While the male gender only considers factors related to productivity, the female gender takes into account various aspects, including productivity as well, when making decisions about technology adoption and use (Venkatesh & Morris, 2000).

The results of the CFA underline that the efficiency of the virtual assistant is the most significant factor in the students' perception of satisfaction. This aspect is closely linked to the virtual assistant's ability to provide accurate and relevant answers, this reinforces its role as a support tool in autonomous learning, which is in agreement with Følstad & Brandtzæg (2017)

Likewise, the results of the confirmatory factor analysis show that the factors associated with the impact on learning, skill development, and technical and accessibility aspects also contribute significantly to the perceived usefulness of the virtual assistant, although to a lesser extent. These dimensions reflect both the pedagogical potential of virtual assistants and the importance of their technical design to ensure an optimal user experience (Bahrini et al., 2023).

However, in addition to analyzing the potential of virtual assistants (Grassini, 2023) this work also novelly identifies key limitations, such as significant differences in perceived assistant efficiency among students of different genders, as well as the lower significant correlation of technical and accessibility aspects with overall satisfaction. This indicates the need to further explore how to optimize these tools to ensure an equitable and accessible experience for all users.

On the other hand, the qualitative analysis is consistent with previous studies highlighting the relationship between artificial intelligence tools and improved autonomous learning (Ait Baha et al., 2024).

Finally, this study contributes to the existing literature by providing empirical evidence on the implementation and impact of virtual assistants in educational contexts (Motlagh et al., 2023). The findings have practical implications for academic institutions, educational technology developers, and policy makers, who will be able to

make more informed decisions on how to integrate these tools effectively and sustainably into teaching and learning processes (Dempere et al., 2023).

Conflict of interest

No potential conflict of interest is declared with respect to the research, authorship and/or publication of this article.

Financing

Both Applied Teaching Innovation Projects associated with this research were funded by the International University of La Rioja during the academic year 2023/2024, receiving the second prize in the VI edition of the “Awards for best practices in the virtual classroom” of the International University of La Rioja on July 10, 2024.

REFERENCES

- Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 15(3), ep429. <https://doi.org/10.30935/cedtech/13152>
- Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, 29(8), 10153-10176. <https://doi.org/10.1007/s10639-023-12166-w>
- Al-Emran, M., AlQudah, A. A., Abbasi, G. A., Al-Sharafi, M. A., & Iranmanesh, M. (2024). Determinants of Using AI-Based Chatbots for Knowledge Sharing: Evidence From PLS-SEM and Fuzzy Sets (fsQCA). *IEEE Transactions on Engineering Management*, 71, 4985-4999. <https://doi.org/10.1109/TEM.2023.3237789>
- Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2), 193-212. <https://doi.org/10.1214/aoms/1177729437>
- Bagozzi, R. P., & Yi, Y. (1989). The Degree of Intention Formation as a Moderator of the Attitude-Behavior Relationship. *Social Psychology Quarterly*, 52(4), 266. <https://doi.org/10.2307/2786991>
- Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, Opportunities, and Threats. *2023 Systems and Information Engineering Design Symposium (SIEDS)*, 274-279. <https://doi.org/10.1109/SIEDS58326.2023.10137850>
- Bartlett, M. S. (1954). A Note on the Multiplying Factors for Various χ^2 Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 16(2), 296-298. <https://doi.org/10.1111/j.2517-6161.1954.tb00174.x>
- Bonett, D. G. (2002). Sample Size Requirements for Testing and Estimating Coefficient Alpha. *Journal of Educational and Behavioral Statistics*, 27(4), 335-340. <https://doi.org/10.3102/10769986027004335>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230-258. <https://doi.org/10.1177/0049124192021002005>
- Cai, Z., Fan, X., & Du, J. (2017). Gender and attitudes toward technology use: A meta-

- analysis. *Computers & Education*, 105, 1-13.
<https://doi.org/10.1016/j.compedu.2016.11.003>
- Chang, D. H., Lin, M. P.-C., Hajian, S., & Wang, Q. Q. (2023). Educational Design Principles of Using AI Chatbot That Supports Self-Regulated Learning in Education: Goal Setting, Feedback, and Personalization. *Sustainability*, 15(17), 12921.
<https://doi.org/10.3390/su151712921>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264-75278.
<https://doi.org/10.1109/ACCESS.2020.2988510>
- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100002.
<https://doi.org/10.1016/j.caeai.2020.100002>
- Cheng, Z., Zhang, Z., Xu, Q., Maeda, Y., & Gu, P. (2023). A meta-analysis addressing the relationship between self-regulated learning strategies and academic performance in online higher education. *Journal of Computing in Higher Education*.
<https://doi.org/10.1007/s12528-023-09390-1>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (0 ed.). Routledge.
<https://doi.org/10.4324/9780203771587>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
<https://doi.org/10.1007/BF02310555>
- Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite Sampling Properties of the Point Estimates and Confidence Intervals of the RMSEA. *Sociological Methods & Research*, 32(2), 208-252.
<https://doi.org/10.1177/0049124103256130>
- Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education*, 8, 1206936.
<https://doi.org/10.3389/feduc.2023.1206936>
- Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., Gupta, B., Lal, B., Misra, S., Prashant, P., Raman, R., Rana, N. P., Sharma, S. K., & Upadhyay, N. (2021). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. *International Journal of Information Management*, 55, 102211.
<https://doi.org/10.1016/j.ijinfomgt.2020.102211>
- Feuerriegel, S., & Pröllochs, N. (2023). *Sentiment Analysis R Package* (Versión 1.3-5) [R]. <https://cran.r-project.org/web/packages/SentimentAnalysis/vignettes/SentimentAnalysis.html>
<https://doi.org/10.1016/j.ijinfomgt.2020.102211>
- Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, 24(4), 38-42.
<https://doi.org/10.1145/3085558>
- Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The Discovery of Grounded Theory; Strategies for Qualitative Research. *Nursing Research*, 17(4), 364.
<https://doi.org/10.1097/00006199-196807000-00014>
- Grassini, S. (2023). Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Education Sciences*, 13(7), 692.
<https://doi.org/10.3390/educsci13070692>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (Eighth edition). Cengage.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-333.
https://doi.org/10.1207/S15324818AME1503_5
- Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., Han, W., Huang, M., Jin, Q., Lan, Y., Liu, Y., Liu, Z., Lu, Z., Qiu, X., Song, R., ... Zhu, J. (2021). Pre-trained models: Past, present and future. *AI Open*, 2, 225-250.

- <https://doi.org/10.1016/j.aiopen.2021.08.002>
- Hattie, J. (2023). *Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement* (1.^a ed.). Routledge. <https://doi.org/10.4324/9781003380542>
- Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological assessment*, 7(3), 238. <https://doi.org/10.1037/1040-3590.7.3.238>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Hwang, G.-J., & Chang, C.-Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7), 4099-4112. <https://doi.org/10.1080/10494820.2021.1952615>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36. <https://doi.org/10.1007/BF02291575>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling* (Fifth edition). Guilford.
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973-1018. <https://doi.org/10.1007/s10639-022-11177-3>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2024). GPT understands, too. *AI Open*, 5, 208-215. <https://doi.org/10.1016/j.aiopen.2023.08.012>
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, 70(4), 810-819. <https://doi.org/10.1037/0022-3514.70.4.810>
- Motlagh, N. Y., Khajavi, M., Sharifi, A., & Ahmadi, M. (2023). *The Impact of Artificial Intelligence on the Evolution of Digital Education: A Comparative Study of OpenAI Text Generation Tools including ChatGPT, Bing Chat, Bard, and Ernie* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2309.02029>
- Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81. <https://doi.org/10.1007/s13278-021-00776-6>
- Nunnally, J. C. (1978). *Psychometric theory* (2d ed). McGraw-Hill.
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033. <https://doi.org/10.1016/j.caeai.2021.100033>
- Onal, S., & Kulavuz-Onal, D. (2024). A Cross-Disciplinary Examination of the Instructional Uses of ChatGPT in Higher Education. *Journal of Educational Technology Systems*, 52(3), 301-324. <https://doi.org/10.1177/00472395231196532>
- OpenAI. (s. f.). *OpenAI API Reference*. Recuperado 25 de noviembre de 2024, de

- <https://platform.openai.com/docs/api-reference/chat/create#chat-create-temperature>
- Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6), 1549-1565. <https://doi.org/10.1002/cae.22326>
- Pérez-Marín, D. (2021). A Review of the Practical Applications of Pedagogic Conversational Agents to Be Used in School and University Classrooms. *Digital*, 1(1), 18-33. <https://doi.org/10.3390/digital1010002>
- Perrotta, C., & Selwyn, N. (2020). Deep learning goes to school: Toward a relational understanding of AI in education. *Learning, Media and Technology*, 45(3), 251-269. <https://doi.org/10.1080/17439884.2020.1686017>
- Peters, M. A., Jackson, L., Papastephanou, M., Jandrić, P., Lazaroiu, G., Evers, C. W., Cope, B., Kalantzis, M., Araya, D., Tesar, M., Mika, C., Chen, L., Wang, C., Sturm, S., Rider, S., & Fuller, S. (2024). AI and the future of humanity: ChatGPT-4, philosophy and education – Critical responses. *Educational Philosophy and Theory*, 56(9), 828-862. <https://doi.org/10.1080/00131857.2023.2213437>
- Pokhrel, S., & Chhetri, R. (2021). A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. *Higher Education for the Future*, 8(1), 133-141. <https://doi.org/10.1177/2347631120983481>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, 30(4), 459-467. <https://doi.org/10.1002/nur.20199>
- Polverini, G., & Gregorcic, B. (2024). How understanding large language models can inform the use of ChatGPT in physics education. *European Journal of Physics*, 45(2), 025701. <https://doi.org/10.1088/1361-6404/ad1420>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113-130. <https://doi.org/10.1080/02602930802618344>
- Ruiz-Corbella, M., & García-Aretio, L. (2023). Virtual mobility in higher education, ¿chance or utopy? *Revista Española de Pedagogía*, 68(246). <https://doi.org/10.22550/2174-0909.3568>
- Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, 80(1), 99-103. <https://doi.org/10.1207/S15327752JPA800118>
- Taani, O., & Alabidi, S. (2024). ChatGPT in education: Benefits and challenges of ChatGPT for mathematics and science teaching practices. *International Journal of Mathematical Education in Science and Technology*, 1-30. <https://doi.org/10.1080/0020739X.2024.2357341>
- Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2102.02503>
- Thottoli, M. M., Alruqaishi, B. H., & Soosaimanickam, A. (2024). Robo academic advisor: Can chatbots and artificial intelligence replace human interaction? *Contemporary Educational Technology*, 16(1), ep485. <https://doi.org/10.30935/cedtech/13948>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Venkatesh, V., & Morris, M. G. (2000).

- Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior. *MIS Quarterly*, 24(1), 115. <https://doi.org/10.2307/3250981>
- Vierhauser, M., Groher, I., Antensteiner, T., & Sauerwein, C. (2024). Towards Integrating Emerging AI Applications in SE Education. *2024 36th International Conference on Software Engineering Education and Training (CSEE&T)*, 1-5. <https://doi.org/10.1109/CSEET62301.2024.10663045>
- Wang, T., Lund, B. D., Marengo, A., Pagano, A., Mannuru, N. R., Teel, Z. A., & Pange, J. (2023). Exploring the Potential Impact of Artificial Intelligence (AI) on International Students in Higher Education: Generative AI, Chatbots, Analytics, and International Student Success. *Applied Sciences*, 13(11), 6716. <https://doi.org/10.3390/app13116716>
- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7), em2286. <https://doi.org/10.29333/ejmste/13272>
- Wheaton, B. (1987). Assessment of Fit in Overidentified Models with Latent Variables. *Sociological Methods & Research*, 16(1), 118-154. <https://doi.org/10.1177/0049124187016001005>
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT* (arXiv:2302.11382). arXiv. <http://arxiv.org/abs/2302.11382>
- Winkler, R., & Soellner, M. (2018). Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. *Academy of Management Proceedings*, 2018(1), 15903. <https://doi.org/10.5465/AMBPP.2018.15903abstract>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, 73(6), 913-934. <https://doi.org/10.1177/0013164413495237>
- Yung, Y., & Bentler, P. M. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 47(1), 63-84. <https://doi.org/10.1111/j.2044-8317.1994.tb01025.x>
- Yurdugül, H. (2008). Minimum sample size for Cronbach's coefficient alpha: A Monte-Carlo study. *Hacettepe Üniversitesi eğitim fakültesi dergisi*, 35(35), 1-9.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., ... Wen, J.-R. (2023). *A Survey of Large Language Models* (Versión 15). arXiv. <https://doi.org/10.48550/ARXIV.2303.18223>

Date of reception: 1 December 2024

Date of acceptance: 12 March 2025

Date of approval for layout: 23 April 2025

Date of publication in OnlineFirst: 30 April 2025

Date of publication: 1 July 2025