Journal of Infrastructure, Policy and Development 2024, 8(10), 6381.
https://doi.org/10.24294/jipd.v8i10.6381

Article

Unveiling the core constructs: A statistical approach to evaluating user experience with Chatbots in higher education (A case study from a university in Ecuador)

Jomar Elizabeth Guzman Seraquive[1], Patricio Álvarez-Muñoz[1], Kerly Palacios-Zamora[1], Dennis Alfredo Peralta Gamboa[1], Marco Faytong-Haro[1,2,*]

[1] Universidad Estatal de Milagro, Milagro 091050, Guayas, Ecuador
[2] Research Department, Ecuadorian Development Research Lab, Daule 090656, Guayas, Ecuador
*** Corresponding author:** Marco Faytong-Haro, mfaytongh@unemi.edu.ec

CITATION

Guzman Seraquive JE, Álvarez-Muñoz P, Palacios-Zamora K, et al. (2024). Unveiling the core constructs: A statistical approach to evaluating user experience with Chatbots in higher education (A case study from a university in Ecuador). Journal of Infrastructure, Policy and Development. 8(10): 6381. https://doi.org/10.24294/jipd.v8i10.6381

ARTICLE INFO

Received: 14 May 2024
Accepted: 7 June 2024
Available online: 30 September 2024

Abstract: Introduction: Chatbots are increasingly utilized in education, offering real-time, personalized communication. While research has explored technical aspects of chatbots, user experience remains under-investigated. This study examines a model for evaluating user experience and satisfaction with chatbots in higher education. **Methodology:** A four-factor model (information quality, system quality, chatbot experience, user satisfaction) was proposed based on prior research. An alternative two-factor model emerged through exploratory factor analysis, focusing on "Chatbot Response Quality" and "User Experience and Satisfaction with the Chatbot." Surveys were distributed to students and faculty at a university in Ecuador to collect data. Confirmatory factor analysis validated both models. **Results:** The two-factor model explained a significantly greater proportion of the data's variance (55.2%) compared to the four-factor model (46.4%). **Conclusion:** This study suggests that a simpler model focusing on chatbot response quality and user experience is more effective for evaluating chatbots in education. Future research can explore methods to optimize these factors and improve the learning experience for students.

Keywords: factor analysis; chatbot; higher education; virtual assistants; artificial intelligence; Turing test; user experience; satisfaction; response quality; algorithms; e-learning; simulation; multivalent statistics

1. Introduction

Virtual assistants, commonly known as chatbots, have emerged as powerful tools that simulate human conversations through algorithms and artificial intelligence techniques. Initially developed for messaging and customer support services on platforms such as Facebook, Telegram, and WhatsApp, these tools have found valuable applications in education, particularly in their ability to provide real-time, personalized communication (Adamopoulou and Moussiades, 2020). This adaptability makes them particularly useful in digital learning environments, where engagement and interaction can otherwise be challenging to sustain.

The inception of chatbots dates back to 1950 when Alan Turing first proposed the concept, questioning the ability of machines to simulate human thought. This foundational idea led to the development of the Turing test, designed to assess a machine's capability to exhibit intelligent behavior indistinguishable from that of a human (Fox, 2014). Since then, the application of chatbots has evolved significantly, with modern implementations far surpassing the original expectations set by early models like Microsoft's Clippy, which demonstrated practical albeit limited assistance capabilities (Liu et al., 2022).

Recent studies have shown that chatbots can significantly impact student engagement and learning outcomes by providing personalized support and facilitating interaction. For instance, Baskara (2023) highlighted the role of chatbots in flipped learning environments, where they enhance engagement and motivation by offering personalized feedback and supporting collaborative learning activities. Similarly, a meta-analysis by Deng and Yu (2023) demonstrated that chatbots could improve learning outcomes across various educational contexts, particularly in terms of explicit reasoning, learning achievement, and knowledge retention.

Additionally, case studies have shown that chatbots can reduce cognitive load for students by providing timely and accurate information, thus allowing learners to focus more on critical thinking and problem-solving tasks. For example, Pérez et al. (2020) reviewed the development and application of chatbots in education and found that these tools could function effectively as both service assistants and educational agents. They emphasized the importance of chatbots in enhancing user experience by providing seamless, real-time assistance, which is crucial for maintaining student engagement in digital learning environments.

Furthermore, the integration of AI-based educational tools, such as chatbots, can personalize feedback systems for learners, significantly impacting their educational experience. Cao et al. (2023) explored the use of multi-role chatbots in computer science education, demonstrating their potential to enhance engagement and motivation through personalized interactions and support. This approach aligns with the principles of Self-Determination Theory, which highlights the importance of competence, autonomy, and relatedness in fostering effective learning experiences.

Despite these promising findings, the effectiveness of chatbots in enhancing user satisfaction and the quality of interactions in educational settings has not been extensively studied.Most research to date has focused on their technical capabilities and application without thoroughly evaluating the user experience within an educational framework (Chung et al., 2020; Sakulwichitsintu, 2023). This gap in the literature underscores the need for more focused research on the impact of chatbots on user satisfaction and educational outcomes, particularly how these tools can be optimized to support student engagement and learning.

To address this gap, this paper's goal is to examine a model for evaluating user experience and satisfaction with chatbots in higher education. The present study employs a simplified model inspired by Chung et al.'s framework but adapted for the educational context. This model considers two primary factors: the quality of the chatbot's responses and the overall user experience, which are posited to be central to enhancing student engagement and satisfaction. This research collected data through surveys distributed among students and faculty at a major university, employing the Consumer Acceptance of Technology model to quantify aspects such as system quality and information quality, which are hypothesized to significantly impact user satisfaction (Ait Baha et al., 2023; Zarouali et al., 2018).

The methodology adopted in this study involved an exploratory and confirmatory factor analysis to validate the proposed model, with a significant portion of the survey designed to capture detailed feedback on the users' interaction with the chatbot. The findings are expected to contribute valuable insights into the design and implementation of chatbot technologies in educational settings, potentially guiding future applications to better meet the needs of students and educators (Peyton and Unnikrishnan, 2023; Smutny and Schreiberova, 2020).

While chatbots have experienced considerable growth within the business sector, primarily driven by their potential to enhance customer service, their application in education has not been as extensively explored. Education, similar to a commercial service, treats students as potential customers, suggesting that customer service strategies from the business realm can be effectively adapted for educational use. Przegalinska et al. (2019) highlight the critical role of quality service in maintaining competitiveness, underscoring its importance for both the provider and the consumer in any sector.

Informed by the study conducted by Zarouali et al. (2018), this research utilizes the Consumer Acceptance of Technology model, which distinguishes between utilitarian elements—namely, the quality of information and system quality—and a hedonic element, which relates to the user experience of the chatbot. This model serves as a foundation for examining both the cognitive and affective impacts on user satisfaction, enabling a structured investigation into how these elements influence educational outcomes, as detailed in **Table 1** and visualized in **Figure 1**.

Table 1. Description of constructs.

Element	Description
Quality of Information	The presentation of the information enables the receiver to understand and interpret. It encompasses accuracy, timeliness, completeness, and format. Users must perceive the chatbot as able to accurately understand their concern and provide an appropriate response.
System Quality	Related to technical aspects such as usability, reliability, availability, adaptability, and timeliness. Difficulties in use negatively influence user satisfaction. Reliability and adaptability are crucial for adjusting to changes.
Experience with Chatbot	Highlights the importance of the user's emotions during online communication via chatbots, expressing enjoyment and the extent to which the user feels that the information provided allows them to act freely and be in control of their actions.
Customer/User Satisfaction	Customer/User Satisfaction refers to the overall contentment of users with the chatbot's performance. It encompasses their satisfaction with the chatbot, the perceived quality of the chatbot's job, the extent to which the chatbot met their expectations, and their happiness with the chatbot. These elements collectively measure how well the chatbot fulfills user needs and expectations, contributing to a positive user experience.

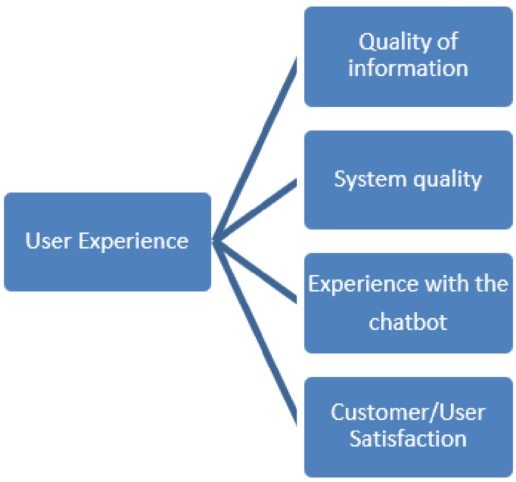

Figure 1. Research model.

To evaluate user experience and satisfaction with chatbots in higher education, this study employed a two-stage analysis. First, exploratory factor analysis was used to identify underlying structures within the data. This analysis informed the development of two competing models: a four-factor model based on prior research in the business sector (Information Quality, System Quality, Chatbot Experience, Customer/User Satisfaction) and a two-factor model derived from the exploration (Chatbot Response Quality and Chatbot Experience and User Satisfaction). Each model was then evaluated using a combination of variance explained and fit indices (Chi-square, degrees of freedom, SRMR, RMSEA, TLI, CFI) to determine which model better represents the data's structure in this educational context. Ultimately, this research aims to bridge the knowledge gap by providing a comprehensive evaluation of chatbot effectiveness in higher education, focusing on user satisfaction and response quality. By detailing the methodologies used, the questions addressed, and the significant findings, this paper seeks to highlight the potential of simplified chatbot models to enhance the educational experience, thus contributing to the ongoing discussion regarding the integration of advanced technologies in learning environments.

2. Methods

2.1. Survey design and participants

This study employed a quantitative approach, utilizing surveys distributed via Google Forms. Approval for the study was granted by the Kennedy Clinic's Institutional Review Board (IRB) in Guayaquil, under the code HCK-CEISH-2022-006. Prior to completing the online questionnaire, informed consent was obtained from participants. The target demographic comprised students and employees of a public higher education institution in Milagro, Guayas province, Ecuador. Data collection took place from April to May 2023.

Participants were asked about their experiences with the UNEMI university chatbot, a tool accessible to all. The chatbot was designed to address inquiries regarding various university-related topics. They were asked via email to fill out the survey. It served as a platform for users to seek information and assistance, enhancing their overall experience with university services. Malhotra et al. (2020) emphasize that questionnaires are particularly effective for gathering large data sets, provided they are designed with clear and concise questions. The questionnaire used in this study included multiple item measures for evaluating constructs such as information quality, system quality, and user experience.

2.2. Sampling

The target population of the study consisted of 40,000 individuals, including students and employees of the university. We used simple random sampling to ensure a representative sample. The sample size calculation was based on a population size of 40,000, a *Z*-score for a 95% confidence level, an estimated population proportion of 0.5 (for maximum variability), and a margin of error of 5%. Although the calculation suggested a sample size of approximately 400 individuals, we decided to send the questionnaire to 1000 participants to account for potential non-responses.

Using Excel, we employed the = RAND () function to assign a random number to each individual in the population. These random numbers were then sorted in ascending order, and the top 1000 individuals were selected to receive the survey. Out of the 1000 individuals who were sent the questionnaire, we received responses from 695 participants. This approach ensured that every member of the population had an equal chance of being included in the study, resulting in a robust sample size that allowed for reliable and precise measurement of the constructs under investigation. This approach ensured that every member of the population had an equal chance of being included in the study, resulting in a robust sample size that allowed for reliable and precise measurement of the constructs under investigation. The sample consisted of 695 responses, with women comprising 50.2% and employees constituting 77.2% of the participants.

2.3. Questionnaire development

The assessment of customer satisfaction was influenced by the methodology outlined in Chung et al. (2020), focusing on four essential elements. It is important to note that certain items from previous questionnaires were omitted in this study to refine the focus, particularly those that addressed the quality of the relationship with the brand, thereby concentrating on the pivotal aspects of chatbot interactions within an educational setting.

Table 2. Variables of the instrument dimensions.

Variables of the Instrument Dimensions	Constructs
Quality of Quality of Information	IQ1: The chatbot provided me with the information required
Quality of Quality of Information	IQ2: The chatbot provided answers to queries as expected
Quality of Quality of Information	IQ3: Chatbot provided enough information
Quality of Quality of Information	IQ4: The information provided by the chatbot (brand) was helpful with respect to my questions or problems
System Quality	SQ1: I found it easy to become proficient in chatbot use (brand)
System Quality	SQ2: I think the chatbot is easy to use
System Quality	SQ3: Using a chatbot requires minimal mental effort
System Quality	SQ4: The chatbot answered quickly
System Quality	SQ5: This chatbot is reliable
Experience with the Chatbot	EWC1: Enjoyed using the chatbot
Experience with the Chatbot	EWC2: The chatbot experience was interesting
Experience with the Chatbot	EWC3: I am happy with the experience of using chatbots
Customer/User Satisfaction	CS1: I am satisfied with the chatbot
Customer/User Satisfaction	CS2: Chatbot did a good job
Customer/User Satisfaction	CS3: The chatbot did what I expected it to do
Customer/User Satisfaction	CS4: I am happy with the chatbot

Note: In the instructions it was stated that all these questions corresponded specifically to the University Chatbot.

The selection of survey items was guided by our research objectives, aiming to evaluate the impact of chatbots on user satisfaction and educational outcomes.

Specifically, the questions were designed to measure four key constructs: Quality of Information, System Quality, Experience with the Chatbot, and Customer/User Satisfaction (See **Table 2**).

Quality of Information: This dimension assesses how effectively the chatbot provides the required information. Items IQ1 to IQ4 were chosen to evaluate various aspects of information quality, such as accuracy, completeness, and relevance. These items are crucial for understanding how well the chatbot meets the informational needs of users, which directly impacts their satisfaction and perception of the tool's utility.

System Quality: This construct focuses on the technical performance of the chatbot. Items SQ1 to SQ5 were selected to measure the ease of use, reliability, and response time of the chatbot. These attributes are essential for ensuring a seamless user experience and minimizing frustration, thereby enhancing overall satisfaction.

Experience with the Chatbot: The questions under this dimension (EWC1 to EWC3) aim to capture users' subjective experiences and enjoyment while interacting with the chatbot. These items help us understand the emotional and experiential aspects of using the chatbot, which are important for fostering positive user attitudes and engagement.

Customer/User Satisfaction: This construct directly measures the users' overall satisfaction with the chatbot. Items CS1 to CS4 were included to gauge users' general satisfaction, the chatbot's performance in meeting expectations, and their happiness with the interactions. These questions are pivotal for assessing the success of the chatbot in achieving its intended outcomes.

The items were evaluated using Likert-type scales ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Additionally, the questionnaire included a section to gather demographic information such as gender and the type of relationship respondents had with the higher education institution (HEI). To ensure the instrument was culturally appropriate, a bilingual researcher initially translated the questionnaire into the local language. Minor adaptations were then made to better fit the specific context of the HEI, and these changes were reviewed by an expert in educational research.

2.4. Statistical methods

Data analysis was conducted using R/R-Studio software, aiming to delineate a structure based on the correlations among various dimensions of the questionnaire. This analysis started by confirming whether the data suited the theoretical framework proposed, using techniques such as exploratory factor analysis, which is notable for its ability to handle variable interdependence effectively (Martínez and Sepúlveda, 2012).

Prior to performing exploratory factor analysis, it was crucial to validate the suitability of the data through Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. Bartlett's test checks if the correlation matrix is an identity matrix, suggesting that the variables do not overlap in the sample (López-Aguado and Gutiérrez-Provecho, 2019), while the KMO test assesses the proportion of variance among variables that might be common variance. High KMO values (close to 1.0) generally indicate that a factor analysis may be useful with the given data. As well, reliability test results demonstrate that the instrument possesses excellent reliability and internal consistency. With a very high Cronbach's Alpha of 0.99 in both its raw and standardized forms, the instrument shows robust internal coherence. Additional indicators, such as the G6 (smc) value of 1 and the average inter-item correlations of 0.92, further confirm the strong correlation among the items. The signal-to-noise ratio of 187 suggests exceptionally high reliability, while the standard error of Alpha, at 0.0003, indicates a very precise estimation of internal consistency. These findings, along with the mean item scores of 3.8 and their standard deviation of 0.93, highlight the instrument's suitability for reliably measuring its intended construct, providing precise and consistent data.

Once the data passed these preliminary tests, the factor analysis proceeded with an appropriate rotation to maximize the clarity of the results by simplifying the loading structures. This rotation redistributes the variance of the original items into factors, thereby enhancing interpretability (Zhang and Preacher, 2015). The Promax rotation was used, as the factors were expected to be correlated with each other. Promax rotation facilitates a more interpretable factor structure in contexts where the factors are not orthogonal. Significant factor loadings were considered to be those greater than 0.4. To determine the number of factors to retain, the criterion of resulted in the retention of two main factors.

Subsequently, a confirmatory factor analysis (CFA) was conducted to test the reliability and validity of the measurement model, utilizing data that reflect the characteristics of the overall population. The adequacy of the model was evaluated using several statistical indicators. The Chi-square (Chisq) statistic was used to assess the difference between the observed covariance matrix and the covariance matrix estimated by the model, where a lower Chi-square value indicates a better fit. Degrees of freedom (df) were also considered, representing the difference between the number of independent elements in the observed covariance matrix and the number of parameters estimated in the model. The Standardized Root Mean Square Residual (SRMR) measured the average difference between the observed covariances and those predicted by the model, with values less than 0.08 generally indicating a good fit. The Root Mean Square Error of Approximation (RMSEA) penalized both lack of fit and model complexity, with values less than 0.05 indicating a good fit and values between 0.05 and 0.08 indicating a reasonable fit. The Tucker-Lewis Index (TLI) compared the fit of the specified model to a null model (with no structure), with values greater than 0.95 generally indicating a good fit. Similarly, the Comparative Fit Index (CFI) assessed the fit of the model relative to a null model, with values greater than 0.95 indicating a good fit. Together, these statistics helped determine how well the model fit the data, with lower values of RMSEA and SRMR and higher values of TLI and CFI indicating a better fit.

The analysis included two hypothetical models: one with four factors (Information Quality, System Quality, Chatbot Experience, Customer/User Satisfaction) and a second, derived from exploratory analysis, which condensed these into two factors (Chatbot Response Quality and Chatbot Experience and User Satisfaction). Each model was evaluated based on the percentage of variance explained and the aforementioned fit indices.

3. Results

In conducting the Exploratory Factor Analysis (EFA) to understand how variables group into factors, the correlation among variables and results from statistical tests are critical. The Kaiser-Meyer-Olkin (KMO) test yielded a high value of 0.98, and Bartlett's test of sphericity confirmed the variables' interrelatedness with a chi-square value of 27,743.78 and a p-value less than 0.001, indicating suitability for factor analysis. Given the non-orthogonality of the factors, a Promax rotation was utilized to better clarify the factor structure.

Analysis revealed that the four factors, based on the model proposed by Chung et al., accounted for 46.4% of the variance in the dataset. This model's factor associations are detailed in **Table 3**, highlighting the complex interactions and the degree to which these factors explain user experiences and satisfaction with educational chatbots. The specific contributions of each factor to the variance are further elaborated in **Table 4**.

Table 3. Correlation matrix between factors.

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	1.000	0.832	−0.397	0.858
Factor 2	0.832	1.000	−0.453	0.846
Factor 3	−0.397	−0.453	1.000	−0.447
Factor 4	0.858	0.846	−0.447	1.000

Table 4. Matrix of factorial loadings and explained variance.

	Factor 1	Factor 2	Factor 3	Factor 4
Charges	3.525	2.315	1.490	0.096
Proportion Var = 0.220		0.145	0.093	0.006
Cumulative Var = 0.220		0.365	0.458	0.464

Upon analyzing the factorial composition and the loadings of each variable, it is noted that the first factor includes variables SQ5, EWC1, EWC2, EWC3, CS1, CS2, CS3, and CS4, indicating a concentration of satisfaction and chatbot experience variables. The second factor consists of the variables IQ1, IQ2, IQ3, and IQ4, reflecting aspects related to information quality. The third factor groups system quality variables SQ1, SQ2, SQ3, and SQ4, suggesting a focus on the technical aspects of the chatbot system. The fourth factor does not associate with any variables due to significantly lower factor loadings compared to the other factors, indicating it does not significantly explain the variability in the data. These relationships and factor loadings are detailed in **Table 5**, providing a clear view of how each set of variables contributes to the different factors identified through the analysis.

The analysis provides a clear and structured insight into how the variables are grouped across the factors. The first three factors encompass a comprehensive set of variables that define distinct dimensions of chatbot interaction. It is observed that the loadings for CS2 and CS3 are comparatively lower than those in the other factors, indicating that these variables do not align well with the fourth factor. Consequently, CS2 and CS3 are not included in factor 4, highlighting their limited contribution to the variability explained by this factor. This delineation ensures a focused interpretation of the factors that significantly impact the user experience and satisfaction with the chatbot. **Table 6** presents the associations between the factors, illustrating that the two identified factors account for 55.2% of the data's variability. This result highlights an important aspect of factor analysis: a more streamlined model with fewer factors can often explain a significant portion of variability, indicating a robust representation of the underlying structure of the data. This finding suggests that the two-factor model provides a concise yet effective way to capture the essential dimensions of user interactions and satisfaction with chatbots.

Table 5. Factor measurements.

	Factor 1	Factor 2	Factor 3	Factor 4
IQ1	0.162	0.677	0.171	
IQ2	0.156	0.705	0.165	
IQ3	0.232	0.721		
IQ4	0.196	0.565	0.221	
SQ1	0.166	0.247	0.582	
SQ2	0.208	0.207	0.601	
SQ3	0.321	0.125	0.477	
SQ4	0.323	0.227	0.470	
SQ5	0.522	0.234	0.198	
EWC1	0.682	0.157	0.200	
EWC2	0.669	0.214	0.156	
EWC3	0.634	0.192	0.216	
CS1	0.655	0.215	0.153	
CS2	0.611	0.220	0.139	0.120
CS3	0.579	0.170	0.157	0.235
CS4	0.616	0.252	0.123	

Table 6. Correlation matrix between factors (Factor correlations).

	Factor 1	Factor 2
Factor1	1.000	
Factor2	−0.856	1.000

Table 7 reveals that Factor 1 explains 35% of the variance with a sum of squared loadings of 6.295, while Factor 2 explains an additional 20.3% of the variance, bringing the cumulative variance explained by both factors to 55.2%.

Table 7. Matrix of factorial loadings and explained variance.

	Factor 1	Factor 2
SS loadings	6.295	3.649
Proportion Var	0.350	0.203
Cumulative Var	0.350	0.552

In the factorial composition and variable loadings, it is observed that the first factor comprises variables IQ1, IQ2, IQ3, IQ4, SQ1, SQ2, SQ3, and SQ4, which together form the "Chatbot Response Quality" factor. This indicates a strong link between the quality of information provided and the system's functionality. The second factor includes variables SQ5, EWC1, EWC2, EWC3, CS1, CS2, CS3, and CS4, forming the "User Experience and Satisfaction with the Chatbot" factor, which reflects the overall satisfaction and emotional engagement of users with the chatbot. Both groupings are detailed in **Table 8**, highlighting the distinct dimensions they represent in evaluating chatbot effectiveness.

Table 8. Measurement of factors.

	Factor 1	Factor 2
IQ1	0.207	0.768
IQ2	0.201	0.801
IQ3	0.254	0.749
IQ4	0.314	0.686
SQ1	0.487	0.494
SQ2	0.537	0.448
SQ3	0.591	0.313
SQ4	0.578	0.404
SQ5	0.670	0.309
EWC1	0.818	0.181
EWC2	0.779	0.222
EWC3	0.780	0.231
CS1	0.779	0.235
CS2	0.740	0.269
CS3	0.726	0.276
CS4	0.727	0.282

The analysis provides a clear and structured view of how variables are organized into two critical factors: "Chatbot Response Quality" and "User Experience and Satisfaction with the Chatbot." These factors are crucial for assessing the impact of chatbot interactions on user satisfaction and the perceived quality of information and system functionality. The results of the study facilitate the exploration and implementation of new strategies that focus on user engagement, thereby enhancing the quality of service provided by the educational institution and improving its relationship with the users of the virtual assistant.

With the data obtained, confirmatory factor analysis (CFA) is applied to validate the appropriateness of the identified model. This step is crucial for ensuring that the model accurately reflects the intended constructs of chatbot experience, satisfaction, and response quality. CFA allows for rigorous testing of the model structure derived from exploratory analysis, enabling better decision-making regarding the implementation and improvement of chatbot functionalities in educational settings.

Based on the results in **Figure 1**, the four factors explain a high proportion of variance (greater than 90%) for each variable. This suggests a good model that effectively captures the relationships between the factors and the variables.

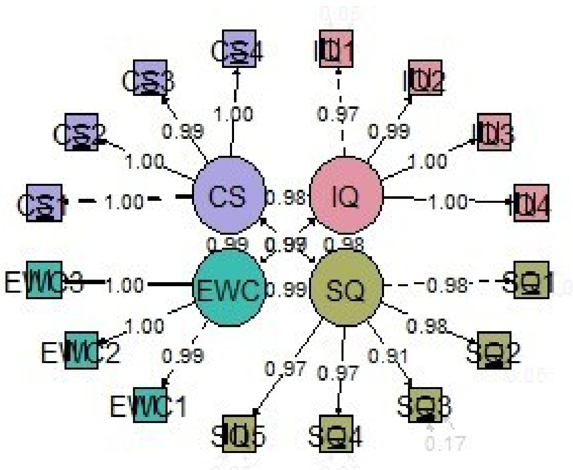

Figure 1. Confirmatory factor analysis model showing loadings and relationships between chatbot experience constructs (Four factors).

While the fit indices in **Table 9**, such as chi-square statistic and comparative fit index, suggest the Chung et al. model adequately explains user experience, satisfaction, and response quality for a chatbot system, it only captures 46.4% of the data's variability. This limited explanatory power is a significant weakness. Exploratory factor analysis of the 692 observations and 18 variables identified two key factors: CRC (factor one) and ESUC (factor two).

Table 9. Chi-square and indicators of evaluation of Model 1.

Chisq	Df	Srmr	rmsea	Tli	cfi
3.369	98.000	0.006	0.000	1.002	1.000

Analyzing **Figure 2** reveals a well-fitting model. The proportion of variance explained by the two factors for each variable is over 90%, indicating a strong relationship between the factors and the variables. Additionally, the figure depicts the degree of association between User Experience and Satisfaction with the Chatbot (ESUC) and Chatbot Response Quality (CRC).

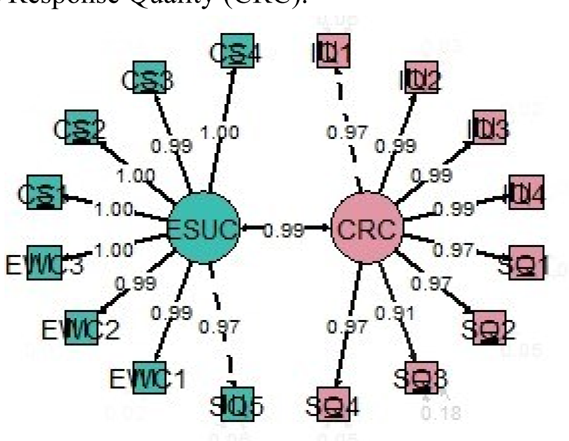

Figure 2. Confirmatory factor analysis model showing loadings and relationships between chatbot experience constructs (Two factors).

While the fit indices in **Table 10**, such as chi-square statistic and comparative fit index, suggest model 2 is acceptable for understanding user experience, satisfaction, and response quality of a chatbot system, it's important to note that only two factors explain 55.2% of the data's variance. This limited explanatory power, despite acceptable fit indices, suggests room for improvement in the model.

Table 10. Chi-square and model evaluation indicators 2.

Chisq	Df	Srmr	Rmsea	Tli	cfi
5.836	103.000	0.008	0.000	1.001	1.000

The results of the study offer a comprehensive understanding of the effectiveness of chatbots in educational environments by comparing the fit of two models. The four-factor model, adapted from Chung et al., included Information Quality, System Quality, Chatbot Experience, and Customer/User Satisfaction, but only explained 46.4% of the variance in the dataset. In contrast, the two-factor model, which emerged from exploratory factor analysis, focused on Chatbot Response Quality and User Experience and Satisfaction with the Chatbot, and explained 55.2% of the variance. This comparison underscores the importance of a more streamlined model that better captures the essential dimensions of user interactions and satisfaction. The high factor loadings associated with Chatbot Response Quality indicate that the quality of information provided and the system's functionality are crucial for user satisfaction. Similarly, the high loadings on User Experience and Satisfaction with the Chatbot highlight the significance of emotional engagement and overall satisfaction in influencing users' perceptions of chatbot effectiveness. These findings suggest that focusing on these two critical factors can significantly enhance the quality of service provided by educational institutions, thereby improving their relationship with users and optimizing the implementation of chatbot functionalities in educational settings.

4. Discussion

The growing trend of chatbots in higher education has spurred research into their effectiveness. This study investigated user experience and satisfaction with chatbots in this specific domain. Employing a user-centered approach, we distributed surveys to students and faculty at a university in Ecuador. The analysis revealed a two-factor model encompassing "Chatbot Response Quality" and "User Experience and Satisfaction with the Chatbot" as the most critical factors influencing user satisfaction. This streamlined model explained a significant portion of the data's variance (55.2%). Interestingly, a more complex, four-factor model based on research in the luxury retail sector (Chung et al., 2020) explained less variance (46.4%) in our educational context. This highlights the importance of tailoring evaluation models to specific domains.

Our findings resonate with the growing body of research on chatbots in education. Studies like Okonkwo and Ade-Ibijola (2021) emphasize the value of chatbots for delivering fast and personalized services, which aligns with our focus on user experience. However, our research takes a more user-centered approach compared to broader explorations of the research landscape.

Furthermore, studies on chatbots in other contexts provide valuable insights.

Chung et al. (2020) explored chatbots in luxury retail, demonstrating their potential across diverse environments. Similarly, Smutny and Schreiberova (2020) examined Facebook Messenger chatbots, reinforcing the notion that AI-powered chatbots are still evolving. This aligns with our observation of the need for further development in educational chatbots.

Building upon Chung et al.'s (2020) framework, our research emphasizes the unique needs of the educational sector. In contrast to broader models encompassing various industries, our findings suggest that user experience and the quality of chatbot responses are the most crucial factors for chatbot success in education. This focus on clarity and accuracy is paramount in educational settings where knowledge transmission and acquisition are central.

It is important to acknowledge a limitation of this study. Our data collection focused on a single university in Ecuador. Future research can explore these findings in a wider range of educational institutions and contexts. Additionally, delving deeper into user experience and response quality can help develop strategies for optimizing chatbot effectiveness in education. This could involve investigating specific aspects of user experience, such as ease of use, clarity of information, and efficiency of interaction with the chatbot. Similarly, exploring response quality could involve analyzing factors like accuracy, completeness, and the level of personalization offered by the chatbot.

This research underscores the importance of adapting existing chatbot evaluation models to specific contexts like education. Educational institutions implementing or improving chatbots should prioritize user experience and the quality of chatbot responses to maximize their effectiveness. Future research can build on these findings by exploring them in a wider range of settings and delving deeper into the core factors of user experience and response quality. Ultimately, by adapting evaluation models and prioritizing user experience and response quality, we can ensure that chatbots have a positive and transformative impact on educational settings.

5. Conclusion

This study investigated the factors influencing user experience and satisfaction with chatbots in higher education. Through a two-stage analysis, we identified "Chatbot Response Quality" and "User Experience and Satisfaction" as the most critical factors for chatbot success in this domain. This user-centered, two-factor model explained a significantly greater proportion of the data's variance compared to a more complex, four-factor model adapted from the business sector. These findings highlight the importance of tailoring evaluation models to specific contexts and prioritizing user experience and response quality when implementing or improving chatbots in educational settings.

Author contributions: Conceptualization, KPZ and JEGS; methodology, MFH; validation, KPZ, JEGS and DAPG; formal analysis, DAPG; writing—original draft preparation, PÁM and MFH; writing—review and editing, PÁM and DAPG; supervision, PÁM and MFH. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Ait Baha, T., El Hajji, M., Es-Saady, Y., et al. (2023). 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

Baskara, F. R. (2023). Chatbots and flipped learning: Enhancing student engagement and learning outcomes through personalised support and collaboration. International Journal of Recent Educational Research, 4(2), 223–238.

Cao, C. C., Ding, Z., Lin, J., & Hopfgartner, F. (2023). AI Chatbots as Multi-Role Pedagogical Agents: Transforming Engagement in CS Education. arXiv, arXiv:2308.03992.

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. Journal of Business Research, 117, 587–595.

Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. Sustainability, 15(4), 2940.

Fox, C. (2014). Chatbot takes on Turing. The Philosophers' Magazine, 66, 8–8.

Liu, L., Subbareddy, R., & Raghavendra, C. (2022). Ai intelligence chatbot to improve students learning in the higher education platform. Journal of Interconnection Networks, 22(Supp02), 2143032.

López-Aguado, M., & Gutiérrez-Provecho, L. (2019). How to perform and interpret an exploratory factor analysis using SPSS Statistics. Revista d'Innovació i Recerca En Educació, 12(2), 1.

Malhotra, N. K., Nunan, D., & Birks, D. F. (2020). Marketing research. Pearson UK.

Martínez, C. M., & Sepúlveda, M. A. R. (2012). Introduction to exploratory factor analysis (Spanish). Revista Colombiana de Psiquiatría, 41(1), 197–207.

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. Computers and Education: Artificial Intelligence, 2, 100033.

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.

Peyton, K., & Unnikrishnan, S. (2023). A comparison of chatbot platforms with the state-of-the-art sentence BERT for answering online student FAQs. Results in Engineering, 17, 100856.

Przegalinska, A., Ciechanowski, L., Stroz, A., et al. (2019). In bot we trust: A new methodology of chatbot performance measures. Business Horizons, 62(6), 785–797.

Sakulwichitsintu, S. (2023). ParichartBOT: a chatbot for automatic answering for postgraduate students of an open university. International Journal of Information Technology, 15(3), 1387–1397.

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. Computers & Education, 151, 103862.

Trivedi, J. (2019). Examining the customer experience of using banking chatbots and its impact on brand love: The moderating role of perceived risk. Journal of Internet Commerce, 18(1), 91–111.

Zarouali, B., Van den Broeck, E., Walrave, M., & Poels, K. (2018). Predicting consumer responses to a chatbot on Facebook. Cyberpsychology, Behavior, and Social Networking, 21(8), 491–497.

Zhang, G., & Preacher, K. J. (2015). Factor rotation and standard errors in exploratory factor analysis. Journal of Educational and Behavioral Statistics, 40(6), 579–603.