Contemporary Educational Technology, 2026, **18**(1), ep627
ISSN: 1309-517X (Online)

OPEN ACCESS

Research Article

General chatbot acceptance, enjoyment, perceived risk, and value (G-CAVS): Scale development and validation

Seyat Polat [1*]
0000-0002-8570-2161

Günter Renner [1]
0009-0008-8290-9854

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Nürnberg, GERMANY
[*] Corresponding author: seyat.polat@fau.de

Citation: Polat, S., & Renner, G. (2026). General chatbot acceptance, enjoyment, perceived risk, and value (G-CAVS): Scale development and validation. *Contemporary Educational Technology, 18*(1), Article ep627. https://doi.org/10.30935/cedtech/17878

ARTICLE INFO

Received: 22 Sep 2025
Accepted: 16 Jan 2026

ABSTRACT

This study was conducted within the context of the KI meets vhb project funded by the Virtuelle Hochschule Bayern, which addresses the use of artificial intelligence applications in university-based teacher education. Despite the increasing use of chatbots in teacher education programs, there is a lack of comprehensive and psychometrically validated instruments to assess pre-service teachers' perceptions of different types of education chatbots. To address this gap, the present study reports the development and validation of a scale designed to measure pre-service teachers' perceptions of different types of chatbots used in educational contexts. The technology acceptance model (TAM, TAM 3) and the value-based adoption model (VAM) served as the theoretical foundation in the development of the scale items. Data were collected from 224 German pre-service teachers enrolled in university-based teacher education programs. Exploratory and confirmatory factor analyses supported a four-factor structure, with strong model fit indices. Criterion-related validity provided initial support for the scale, as significant associations with chatbot usage frequency were observed for all dimensions except perceived risk. The four-factor structure of the scale was further confirmed in an independent sample of 263 in-service teachers in Türkiye, demonstrating the robustness of the model across different teacher populations. Overall, the *G-CAVS* scale emerged as a valid and reliable instrument for assessing perceptions of chatbots in teacher education contexts, with implications for broader pre-service teachers' populations beyond the present samples.

Keywords: chatbot scale, scale development, technology acceptance, pre-service teachers, artificial intelligence in education

INTRODUCTION

Throughout history, humanity has been in constant development and has achieved significant progress in many fields. Especially in the last century, technological innovations have accelerated at a dizzying pace; one of the most distinctive features of human beings—intelligence—has begun to be simulated through machines. The process, shaped by the question "Can machines think?", laid the foundation for the birth of artificial intelligence (Akdogan, 2021; Özcan & Polat, 2023). The idea of the "Artificial Intelligence Summer Research Project," introduced at the Dartmouth Conference in 1956, has continued to grow exponentially to the present day (Mijwel, 2015; Simon, 1995).

Since November 2022, rapid developments in artificial intelligence have brought applications such as chatbots and custom artificial intelligence assistants (custombots) to the forefront (Dwivedi et al., 2023). These AI-based tools have received great interest in many domains, including education (Balcı, 2024; Şanlı, 2025). Accordingly, the use of AI assistants in teaching, the preparation of learning materials through such tools (e.g., platforms like Fobizz), and the discussions surrounding both the potential contributions and risks of chatbots in teacher education and broader higher education contexts have increasingly gained attention.

In this context, the potential benefits of chatbots in educational contexts, user acceptance, perceived value, and risk dimensions have begun to be addressed within the framework of theoretical models explaining technology adoption (Raiche et al., 2023; Sohn & Kwon, 2020). However, a valid and reliable instrument that comprehensively measures these intertwined dimensions—particularly for general chatbots[1] in educational contexts—remains underdeveloped. Within this scope, the aim of this study is to develop a comprehensive instrument for assessing chatbot acceptance in teacher education contexts by integrating enjoyment, perceived value, and perceived risk within a unified theoretical structure.

In contrast to existing scales that either focus on specific chatbot types or exhibit psychometric limitations, the *general chatbot acceptance, enjoyment, perceived risk, and value scale (G-CAVS)* offers a broad and theoretically grounded approach to measuring perceptions of chatbot use in teacher education contexts.

THEORETICAL FRAMEWORK

This section presents the theoretical framework of the study, drawing on the technology acceptance model (TAM and TAM 3) and the value-based adoption model (VAM).

Technology Acceptance Model in the Context of Chatbots

Computer-based information systems began to spread rapidly from the second half of the 20th century and soon entered educational settings (Valdez et al., 1999). In this process, Davis (1989) introduced his first model to explain technology acceptance in his 1985 doctoral dissertation and developed the TAM in 1989, which has become foundational in the literature. The model was designed with two main aims: first, to improve understanding of users' acceptance processes for new technologies and to provide a theoretical basis for the design and implementation of information systems; second, to offer system designers and practitioners a practical "acceptance test" methodology to evaluate user acceptance before implementation.

TAM is built upon three key components to explain user motivation: perceived usefulness, perceived ease of use, and attitude toward use (Davis, 1989). However, the rapid evolution of information technologies and the diversification of usage contexts have shown that TAM alone does not always provide sufficient explanatory power.

To address this limitation, Venkatesh and Davis (2000) developed TAM 2, incorporating social influence and cognitive processes. Later, Venkatesh and Bala (2008) further extended the model, proposing TAM 3. Based on extensive evidence from previous research, TAM 3 integrated new determinants influencing user acceptance—such as perceived enjoyment and computer anxiety—into the model. This framework enabled the study of factors influencing technology use at both the individual level and the pre-/post-implementation intervention level. In this respect, TAM 3 provides a strong theoretical foundation for understanding and enhancing the acceptance of chatbot technologies in educational contexts.

Value-Based Adoption Model in the Context of Chatbots

The rapid development of technology has not only led to the emergence of innovations but also enabled these innovations to reach broad audiences in a short period (Chambers, 2004; Gil-Garcia et al., 2014; Srinivasan, 2008). Especially with the spread of mobile communication technologies and smart devices, individuals' access to diverse technologies has become easier, accelerating their integration into daily life (Castells et al., 2004; Verhoef et al., 2017). Consequently, users have increasingly questioned not only the usability of technologies but also their perceived benefits and added value. This shift has heightened the need for new conceptual models to understand users' decision-making processes regarding technology. In this context, the VAM, developed by Kim et al. (2005), offers a robust theoretical framework that explains users' technology adoption decisions through the balance of perceived benefits and perceived sacrifices. The model

[1] In this study, the term "general chatbot" refers to general-purpose conversational AI systems (e.g., ChatGPT, DeepSeek, and Grok) that are not specifically designed for instructional purposes but are increasingly used by students and teachers within educational contexts.

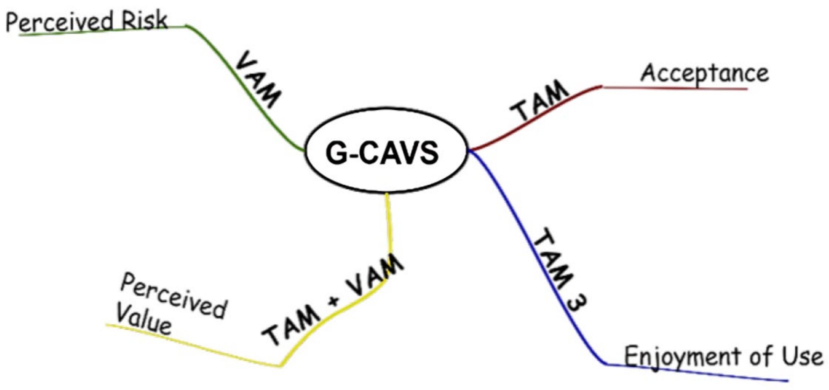

Figure 1. Conceptual framework of the G-CAVS (Authors' own illustration, created using Draw.io)

defines perceived value as the comparison between perceived benefits (e.g., time savings and increased efficiency) and perceived sacrifices (e.g., cost, risk, effort, and data security concerns) (Kim et al., 2007). Moreover, VAM places not only rational cost-benefit considerations but also emotional aspects such as enjoyment at the core of the adoption process (Liao et al., 2022).

A key implication of the model is that individuals are unlikely to invest time, effort, or money in a new technology without assurance of its benefits (Mathava, 2024). This approach provides a critical perspective for understanding users' perceptions of technology.

In the context of chatbots, users' acceptance processes are not limited to ease of use or functionality. Advantages such as learning efficiency (Chang et al., 2022), personalized support (Vashishth, 2024), and time savings (Labadze et al., 2023) are weighed against potential costs such as privacy concerns, misinformation, and technical failures (Gumusel et al., 2024; Gupta, 2022; Yang et al., 2023). Therefore, VAM provides a comprehensive theoretical foundation for understanding, predicting, and enhancing the value of chatbot technologies, particularly in educational settings.

G-CAVS Scale Dimension Selection

The G-CAVS scale was developed to capture the unique interaction dynamics of chatbot technologies among pre-service teachers by integrating the TAM (Davis, 1989), the extended TAM3 (Venkatesh & Bala, 2008), and the VAM (Kim et al., 2005). This theoretical integration allows a multidimensional understanding of user experience, as visualized in **Figure 1**. The four subscales—acceptance, enjoyment of use, perceived value, and perceived risk—were developed to provide a holistic assessment of users' perceptions of chatbots in educational contexts. Although the TAM/TAM 3 and the VAM include additional constructs such as perceived ease of use, attitude, anxiety, and perceived sacrifice, these dimensions were not included in the G-CAVS scale for conceptual and pragmatic reasons. The primary aim of the present study was to capture pre-service teachers' general perceptions of chatbots used in teacher education contexts, rather than to model detailed adoption intentions or emotional responses. Moreover, several of these constructs show substantial conceptual overlap with the selected dimensions (e.g., perceived ease of use with acceptance and perceived sacrifice with perceived value), and including all of them would have increased model complexity without necessarily enhancing interpretability. All relationships depicted in the framework are conceptual and non-causal. Explanations for each subscale are provided below.

As shown in **Figure 1**, the theoretical foundations and their associated scale dimensions are visualized. The arrows between constructs represent conceptual associations and do not imply causality.

The *acceptance* dimension reflects individuals' general perceptions regarding the inclusion of chatbot technologies in educational processes. According to TAM, individuals are more likely to adopt a technology if they perceive it as useful and easy to use. Güldal and Dinçer (2024) demonstrated that students found a rule-based educational chatbot useful and accessible, consequently exhibiting high acceptance levels despite technological limitations. This finding reveals that acceptance of chatbots in education is fueled not only by technical competence but also by perceived pedagogical value, integration into education, and ease of access.

Enjoyment of use refers to the individual's perception of the interaction process with a chatbot as enjoyable, motivating, and interesting. In the study by Borsci et al. (2022), this dimension was defined as the emotional component of user satisfaction and was emphasized to play a significant role in users developing positive perceptions towards the technology, especially in long-term and learning-focused interactions. Enjoyment of use encompasses not only the functional efficiency of the interaction but also motivation. This is particularly critical for chatbots, as dialogue-based interactions require emotional bonding, unlike traditional software. The pleasure, curiosity, and discovery of new learning that individuals derive from interacting with a chatbot reinforce their usage.

Perceived value, according to the VAM, posits that individuals make decisions to adopt a technology by balancing the benefits they obtain (usefulness & enjoyment) against the sacrifices or risks they undertake (perceived risk) (Al-Abdullatif, 2023). Accordingly, perceived value is the process of an individual's cognitive evaluation between the benefit derived from the technology and its potential risks. In other words, if an individual believes they gain personal benefits (time, feedback, ease of learning) from learning via a chatbot and that the risks (privacy, security, misinformation) are lower than these gains, they find the technology valuable.

The *perceived risk* dimension was considered a core dimension based on vulnerabilities specific to chatbots, particularly those related to data privacy and information security. These risks include reliability concerns regarding data privacy in chat interactions. Research shows that even if users are not explicitly aware of these risks, trust loss in cases of privacy breaches or misinformation dramatically affects chatbot acceptance (Chen et al., 2023). In this context, perceived risk refers to the evaluation of the student's perceptions regarding the negative outcomes they believe they might experience during interaction with the technology (Chen et al., 2023; Mokoena & Seeletse, 2025). All finalized items of the G-CAVS scale are presented.

Research Gap and Present Study

Several scale development studies on chatbots can be found in the literature (Borsci, 2022; Köhler & Hartig, 2024; Nemt-allah et al., 2024; Taktak & Bafrali, 2025). However, a systematic analysis reveals critical methodological limitations in existing instruments that constrain their applicability in educational contexts. Some of these studies reveal limited psychometric results (Al-Abdullatif, 2023; Borsci, 2022), while others focus exclusively on a single type of chatbot, such as ChatGPT-only approaches (Köhler & Hartig, 2024; Nemt-allah et al., 2024; Taktak & Bafrali, 2025). This narrow focus restricts their generalizability across the diverse range of chatbot technologies increasingly utilized in higher education, particularly within teacher education contexts involving pre-service and in-service teachers.

Furthermore, many instruments exhibit incomplete psychometric validation. For instance, Köhler and Hartig (2024) reported that while their knowledge scale demonstrated good model fit, the actual usage and use value scales showed only moderate fit values, underscoring the challenge of achieving comprehensive psychometric robustness across all dimensions. This pattern of inconsistent measurement quality is compounded by a general lack of evidence for discriminant validity and measurement invariance across different user groups in much of the existing literature.

Perhaps most critically, as Borsci (2022) explicitly acknowledged, existing scales "cannot be used as an off-the-shelf product for user research and usability tests," clearly indicating that current instruments are not sufficient for direct user research or usability testing in educational settings. These limitations highlight the necessity of developing a more holistic scale capable of evaluating different types of chatbots and applicable across diverse learning contexts.

In response to these limitations, the present study was conducted within the project [KI meets vhb], funded by the Virtuelle Hochschule Bayern, which focuses on artificial intelligence applications in teacher education, particularly among pre-service teachers. Based on this need, the G-CAVS was developed and validated. Its theoretical framework is built upon the TAM (Davis, 1989), the extended TAM (TAM3) (Venkatesh & Bala, 2008), and the VAM (Kim et al., 2005). An overview of the G-CAVS dimensions and their theoretical foundations is presented in **Table 1**.

Table 1. Dimensions of the G-CAVS and their theoretical foundations

Scale dimension	Originating model	Explanation
Acceptance	TAM (Davis, 1989)	Positive evaluation and endorsement of chatbots in educational contexts.
Enjoyment of use	TAM 3 (Venkatesh & Bala, 2008)	The extent to which interacting with chatbots is perceived as enjoyable, motivating, and curiosity-arousing.
Perceived value	TAM + VAM (Kim et al., 2005)	Combines TAM's perceived usefulness (e.g., efficiency, ease, learning support) with VAM's perceived value (added personal and educational benefit).
Perceived Risk	VAM (Kim et al., 2005)	Concerns about privacy, data security, and feelings of uncertainty when using chatbots.

METHOD

As this study is scale development research, it was conducted within the framework of the quantitative research paradigm. The theoretical framework suggested by DeVellis and Thorpe (2021), Gable and Wolf (2012), and Hinkin (1998) for the scale development process was taken as a basis. Accordingly, the main stages of item writing, expert evaluation, pilot testing, main application, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), and reliability analyses are detailed below.

Participants

Following the pilot study, a 19-item version of the G-CAVS was administered to a sample group of 224 teacher candidates. After data cleaning, 8 participants who completed the survey in less than 5 minutes were excluded to ensure data quality, resulting in a final main sample of N = 216. With 216 participants for 19 items, the sample size met established recommendations for EFA and CFA in scale development research (Hair et al., 2014). To ensure the independence of the exploratory and confirmatory stages, the main sample was randomly split into two non-overlapping subsamples. EFA was conducted on the first subsample, while CFA and criterion-related validity analyses were conducted on the second subsample (n = 184).

To further examine the robustness of the factor structure, an independent CFA was subsequently conducted using a Turkish teacher sample (N = 263), providing cross-sample validation. Participants' demographic characteristics are presented in **Table 2**. Participants' demographic characteristics for the Turkish sample are provided in **Appendix A**. To ensure a shared understanding of the chatbot concept, a brief definition and illustrative examples (e.g., ChatGPT, DeepSeek, and Grok) were provided to all participants at the beginning of the online survey. Throughout the adaptation process, strict adherence to ethical principles governing research involving human participants was maintained (APA, 2017). The scale items were designed to avoid causing psychological harm to participants, and during the online administration, measures were taken to ensure that participants did not feel pressured and could provide independent responses. The instructions of the online data collection tool and verbal explanations provided during the administration explicitly stated that participation was voluntary, responses would have no positive or negative consequences, personal information would be used solely for research purposes, and confidentiality would be ensured. No identifying information, such as names, was requested from participants.

Table 2. Demographic profile of German teacher candidates

Category	Subcategory	Frequency (n)	Percentage (%)
Educational degree	Bachelor	206	95.4
	Master	5	2.3
	Staatsexamen	5	2.3
Study semester	1-2	41	19.0
	3-4	95	44.0
	5-6	36	16.7
	7 or higher	44	20.4
Gender	Female	173	80.1
	Male	43	19.9

Scale Development

In the development process of the G-CAVS, studies were first reviewed in the Web of Science, Scopus, and ERİC databases using the keywords *"chatbot acceptance," "technology acceptance," "perceived usefulness," "perceived value,"* and *"risk perception."* Within this scope, sources such as Al-Abdullatif (2023), Borsci et al. (2022), Davis (1989), Ursavaş et al. (2014), Kim et al. (2005, 2017), Lemke et al. (2023), Özcan and Polat (2023), Raiche et al. (2023), Sohn and Kwon (2020), Venkatesh and Bala (2008), and Yurt (2025) were examined. Based on the literature, the TAM (Davis, 1989), the extended TAM 3 (Venkatesh & Bala, 2008), and the VAM (Kim et al., 2005) were determined as the theoretical framework for item pool creation. Considering the subdimensions of these models (*acceptance, enjoyment of use, perceived value,* and *perceived risk*), an initial item pool of 25 items was prepared (DeVellis & Thorpe, 2021). The expert panel consisted of eight academics, including four faculty members specializing in primary education and pedagogy, two faculty members from the field of educational measurement and evaluation, and two faculty members from German language and literature.

The experts evaluated the items in terms of meaning, structure, and language, and suggested removing two items (item 7 and item 17) that measured the same content with different wording. Based on these discussions and expert suggestions, these two items were excluded from the pilot testing and removed from the scale form. In cases where disagreements arose among experts—particularly regarding the use of synonymous expressions—the items were discussed jointly by the research team and a language specialist. Final decisions were reached through consensus, prioritizing conceptual clarity and linguistic appropriateness for the target population.

The pilot application of the G-CAVS scale, consisting of 23 items and a 5-point Likert-type response format (1 = strongly disagree, 5 = strongly agree), was carried out with a mixed group of 50 participants from different academic levels (bachelor's, master's, and doctoral). The survey also included an open-ended question inviting participants to evaluate the items in terms of meaning, structure, and language. Following the pilot study, both qualitative feedback and preliminary quantitative analyses (Cronbach's alpha and EFA) were evaluated. These analyses revealed that Items 1, 16, and 18 showed negative correlations, while Item 10 received negative feedback regarding clarity (e.g., "Item 10. New educational technologies–Unclear"). Despite recoding attempts, the correlations of these items remained unsatisfactory (DeVellis & Thorpe, 2021). Consequently, these four items were removed, resulting in a refined 19-item version of the G-CAVS.

The statistical procedures and criteria used for the EFA and CFA are described in detail in the Statistical Analyses section. Subsequently, the main study was administered to a new, larger sample of 216 participants. An EFA conducted on the first subsample confirmed a four-factor structure and led to the removal of five additional items, resulting in the final, parsimonious 14-item version of the G-CAVS.

In addition, for the purpose of cross-sample validation, the G-CAVS was translated into Turkish using a forward-backward translation procedure and reviewed by experts to ensure content and linguistic equivalence Şeker and Gençdoğan (2006). A small pilot test with 20 teachers confirmed the clarity and cultural relevance of the items prior to the main data collection.

As part of the finalization of the scale, item 12 and item 15 from the perceived risk subscale were reverse-coded due to their negative phrasing. This reverse-coding was applied because the items in the 'perceived risk' subscale (e.g., 'I am worried that my personal data might be shared …') are conceptually negative, measuring an undesirable outcome, whereas items in other subscales (e.g., acceptance and enjoyment) are positive and measure desirable outcomes. Aligning all items in the same conceptual direction is a standard procedure for calculating coherent total scores and reliability coefficients, as it prevents response bias and simplifies interpretation (DeVellis & Thorpe, 2021). Similarly, in the Turkish adaptation, item 5 and item 7, corresponding to the same subdimension, were also reverse coded to ensure all items were aligned in the same direction. This step was essential for accurate computation of total scores and reliability coefficients.

Statistical Analyses

The statistical analyses of the scale development process, regarding construct validity and reliability, were conducted step by step using the Jamovi program. First, in the original German version of the scale**,** in order to test the suitability of the dataset for factor analysis, the Kaiser-Meyer-Olkin (KMO) measure of sampling

Table 3. Factor loadings and uniqueness values of the G-CAVS items

Factor	Item	Factor Loading	Uniqueness
Acceptance	5. I believe it is appropriate that chatbots are used at universities.	0.814	0.23
	6. I think that the use of chatbots in schools, training, and education is meaningful.	0.809	0.30
	14. I believe that chatbots should be integrated into educational institutions.	0.831	0.24
	21. I think that chatbots are valuable for changes in the education sector.	0.697	0.32
Enjoyment of use	2. Interacting with chatbots is great fun for me.	0.666	0.45
	19. Using chatbots motivates me to discover more.	0.843	0.20
	20. When I interact with chatbots, my curiosity is sparked.	0.867	0.18
Perceived value	3. I intend to use chatbots for my future learning.	0.674	0.40
	4. Chatbots give me an advantage because they save me time and effort.	0.740	0.41
	22. Using chatbots makes my learning experience valuable.	0.617	0.44
	23. I think that chatbots can increase my personal success.	0.614	0.47
	24. Chatbots facilitate my learning processes.	0.742	0.35
Perceived risk	12. I am worried that my personal data might be shared when I use chatbots.	0.945	0.09
	15. I am concerned that my personal data may be misused when I use chatbots.	0.951	0.08

adequacy and Bartlett's test of sphericity were applied. The KMO value was found to be 0.837, which indicates a "good" level of adequacy for factor analysis. According to the classification proposed by Kaiser (1974) values of 0.90 and above are considered "excellent," 0.80-0.89 "good," 0.70-0.79 "medium," 0.60-0.69 "weak," and below 0.50 "unacceptable." In addition, Bartlett's test of sphericity was significant ($\chi^2 = 1439$, df = 91, $p < .001$), indicating a good level of correlation among the variables and that factor analysis could be conducted (Büyüköztürk, 2021; Tabachnick & Fidell, 2007).

With the sample of 216 participants, item analyses were first carried out, including Cronbach's alpha, McDonald's omega, and item-total correlations. In the literature, Cronbach's alpha values above .70 are considered acceptable, while values above .80 indicate strong internal consistency (DeVellis & Thorpe, 2021). The analyses showed that the scale items demonstrated strong internal consistency.

Subsequently, EFA was performed to test the construct validity of the scale. The extraction method was "minimum residual," and since the correlations between factors were low, the orthogonal Varimax rotation technique was chosen. During the factor analysis process, items with factor loadings below .30, items with high uniqueness values (> .60), and items that loaded above .30 on more than one factor with a difference smaller than .10 were removed from the scale (Büyüköztürk, 2021; Costello & Osborne, 2005).

The uniqueness value indicates the amount of variance in an item that is not explained by the factors. Ranging between 0 and 1, values above .60 are considered to indicate that an item contributes poorly to the model (Brown, 2015; Fabrigar et al., 1999). The items removed generally had uniqueness values above this threshold. The 14 remaining items in the scale had uniqueness values ranging between .19 and .49, which is considered an acceptable range (Brown, 2015; Fabrigar et al., 1999).

CFA was conducted using the maximum likelihood estimation method to test the fit of the four-factor, 14-item model identified through EFA. Model fit was evaluated using multiple goodness-of-fit indices, including the comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Following commonly accepted criteria, CFI and TLI values of .90 or above, RMSEA values below .08, and SRMR values below .08 were considered indicative of acceptable model fit (Brown, 2015; Hu & Bentler, 1999).

Criterion-related validity was examined by correlating the G-CAVS subscale scores with an external criterion reflecting chatbot usage frequency. Given the ordinal nature of the usage frequency variable, Spearman's rho correlation coefficients were computed to assess the associations between the scale dimensions and the external criterion.

Table 3 presents the factor loadings, uniqueness values, and factor assignments of the four-factor structure of the G-CAVS scale.

Table 4. Factor Eigenvalues and total variance values

Factor	Item	Factor loading	Eigenvalue	Variance (%)	Cumulative (%)
1	A5	0.814	2.83	20.2	20.2
	A6	0.809			
	A14	0.831			
	A21	0.697			
2	EU2	0.666	2.64	18.9	39.1
	EU19	0.843			
	EU20	0.867			
3	PV3	0.674	2.45	17.5	56.5
	PV4	0.740			
	PV22	0.617			
	PV23	0.614			
	PV24	0.742			
4	PR12	0.945	1.185	13.2	69.8
	PR15	0.951			

Note. Overall KMO value is 0.837

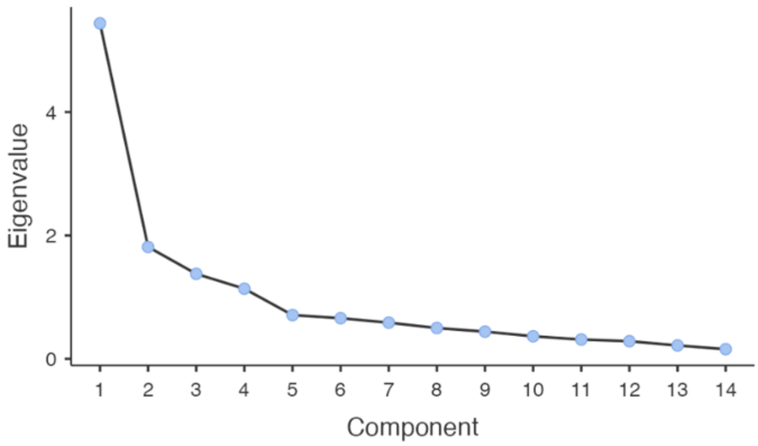

Figure 2. Scree plot of the four-factor structure (Authors' own analysis using jamovi)

RESULTS

Findings from Exploratory Factor Analysis

Before conducting the EFA for the G-CAVS, the suitability of the data for factor analysis was evaluated using the KMO measure of sampling adequacy. The calculations yielded a KMO value of 0.837, indicating that the dataset was at a good level of suitability for factor analysis. In addition, the result of Bartlett's test of sphericity, χ^2 (91) = 1439, $p < .001$, confirmed the appropriateness of the data for factor analysis (Büyüköztürk, 2021). These findings indicate that the necessary conditions for conducting factor analysis were met. **Table 4** presents the factor loadings, eigenvalues, and percentage of variance explained for the four-factor scale.

When **Table 4** is examined, it is observed that the EFA results revealed a four-factor structure, with each factor having an eigenvalue greater than 1. The total variance explained was found to be 69.8%, of which 20.2% was accounted for by the first factor, 18.9% by the second factor, 17.5% by the third factor, and 13.2% by the fourth factor. This result indicates that the first factor alone explains a large proportion of the variance. In addition, as shown in **Figure 2**, the Scree Plot revealed a clear inflection point after the third factor, which provides visual evidence supporting the four-factor structure of the scale.

As a result of the EFA, no items were found to show cross-loadings. Furthermore, the scree plot presented in **Figure 2** demonstrated that there were four factors with eigenvalues greater than 1.

Based on the findings of the EFA, the G-CAVS was confirmed to have a robust four-factor structure. These four factors meaningfully represent the dimensions of general acceptance, enjoyment of use, perceived risk, and perceived value of the scale. The factor analysis results demonstrate that the scale has a strong structure

Table 5. CFA results and reliability statistics (standardized factor loadings)

Factor	Item	FL	SE	CR	p	α	ω	CR	AVE	MSV
Acceptance	A5	0.822	0.052	12.83	***	0.852	0.857	0.857	0.601	0.29
	A6	0.707	0.069	10.40	***					
	A14	0.791	0.062	12.17	***					
	A21	0.775	0.056	11.77	***					
Enjoyment of use	EU2	0.622	0.064	8.76	***	0.816	0.826	0.829	0.622	0.27
	EU19	0.868	0.061	13.73	***					
	EU20	0.852	0.062	13.40	***					
Perceived value	PV3	0.707	0.060	10.21	***	0.804	0.808	0.808	0.461	0.29
	PV4	0.504	0.058	6.74	***					
	PV22	0.703	0.061	10.16	***					
	PV23	0.701	0.064	10.11	***					
	PV24	0.752	0.053	11.10	***					
Perceived risk	PR12	0.851	0.185	5.43	***	0.891	0.891	0.893	0.808	0.07
	PR15	0.944	0.196	5.54	***					

Note. ***p < .001; FL: Factor loading

supporting its validity, and that each factor shows high internal reliability. These findings indicate that the scale is a reliable and valid instrument for both scientific and practical use.

EFA revealed a four-factor structure consisting of 14 items (see **Appendix B** for the German versions). In determining the number of factors, the eigenvalue > 1 criterion and the scree plot were considered. The total variance explained by the four-factor solution was 69.8%, exceeding the recommended minimum range of 40-60% for social sciences (Hair, 2014). The variances explained by the individual factors were 20.02%, 18.9%, 17.52%, and 13.2%, respectively. All retained items showed factor loadings of .60 or higher, indicating strong contributions to the factor structure (Tabachnick & Fidell, 2007). Overall, the EFA results provide strong evidence for a stable four-factor structure of the G-CAVS scale.

Confirmatory Factor Analysis

CFA was conducted with data obtained from 184 participants to test the hypothesized four-factor structure of the G-CAVS. Model fit was evaluated according to widely accepted indices (Kline, 2023). The results indicated a strong model fit: $\chi^2/df = 1.58$, SRMR = .04, RMSEA = .06, CFI = .96, and TLI = .95. These fit indices meet the criteria recommended by Hu and Bentler (1999), supporting that the four-factor model provides a strong representation of the data structure. Accordingly, the findings suggest that the proposed four-factor structure of the G-CAVS demonstrates strong fit with the data. Factor loadings for each factor, as well as the results of discriminant and convergent validity analyses, are presented in **Table 5**.

The CFA findings demonstrate that the four-factor measurement model (acceptance, enjoyment of use, perceived value, and perceived risk) is statistically significant, reliable, and valid. All factor loadings ranged from .504 to .944, exceeding the recommended threshold of .50. Reliability coefficients (α, ω, and CR) were found between .80 and .89 across all factors, indicating strong internal consistency. In terms of convergent validity, three factors (acceptance, enjoyment of use, and perceived risk) were adequate with AVE values above .50, while the perceived value factor was marginal (.461). Nevertheless, the high composite reliability (CR = .808) and all factor loadings above .50 make this minor limitation tolerable. Discriminant validity was established according to the Fornell and Larcker (1981) criterion, as all MSV values were lower than the corresponding AVE values. In conclusion, the four-factor model presents strong evidence of validity and reliability, indicating that it is a suitable measurement instrument for future hypothesis testing.

Additionally, a separate CFA was conducted using the Turkish teacher sample (N = 263) to examine the stability of the four-factor structure across samples. The model showed an acceptable to good fit to the data ($\chi^2 = 173$, df = 71, $p < .001$, CFI = .95, TLI = .94, RMSEA = .07, 90% CI [.060-.088], SRMR = .03). Standardized factor loadings ranged from .679 to .915, all exceeding the recommended threshold of .50. Detailed item-level factor loadings and measurement statistics are presented in **Appendix C**.

Table 6. Means, standard deviations, and correlations among factors

Factors	Mean	Standard deviation	F1	F2	F3	F4
Acceptance	15. 04	3. 06	-			
Enjoyment of use	10. 00	2.46	0.468**	-		
Perceived value	19. 05	3. 17	0.555**	0.544**	-	
Perceived risk	5.24	2.24	0.155*	0.137*	0.081	-

Note. *p < .05; **p < .001

Table 7. Correlations and mean values

Dimension of the scale	Chatbot usage frequency	Standard deviation
Acceptance	.438**	3. 07
Enjoyment of use	.434**	2.46
Perceived value	.458**	3. 21
Perceived risk	.089(ns)	2.22

Note. Spearman's rho; N = 184; p < .01

Descriptive Statistics and Factor Correlations for G-CAVS

Table 6 presents the descriptive statistics and Pearson correlation coefficients among the four subdimensions of the G-CAVS. Descriptive statistics and Pearson correlation coefficients were calculated using the full development sample (N = 216) to ensure robust correlation estimates, while the CFA was performed on a subsample drawn from the same dataset (N = 184), as an additional validation step.

The correlation analysis results revealed that the positive perception dimensions—acceptance, enjoyment of use, and perceived value—were statistically significant, positive, and moderately strong (respectively, r = .555, p < .001; r = .544, p < .001; r = .468, p < .001). The relationships of the perceived risk dimension with the others were more complex. Perceived risk was found to have statistically significant but weak positive correlations with acceptance (r = .155, p = .022) and enjoyment of use (r = .137, p = .044). The most striking finding was the absence of a statistically significant relationship between perceived risk and perceived value (r = .081, p = .233). This result suggests that the value and benefits participants perceive from chatbots are not significantly associated with the risks they report.

In sum, while these findings confirm the internal consistency of the overall scale structure, they also provide important evidence that risk perception plays a distinct and independent role within the model. The perceived risk dimension clearly demonstrates discriminant validity within the scale.

Criterion-Related Validity Evidence

To evaluate the criterion-related validity of the G-CAVS, the scale was administered to a total of 184 pre-service teachers. Of the participants, 82% (n = 150) were female and 18% (n = 34) were male. In order to examine the relationship between the four subdimensions of the scale and chatbot usage frequency, the ordinal variable "chatbot usage frequency" (regularly = 1, occasionally = 2, never = 3) was used as an external criterion. Therefore, Spearman's rho coefficient was chosen for the correlation analysis.

The external criterion was considered as an indicator of criterion-related validity, and the findings are presented in **Table 7**.

Spearman's rho coefficients showed that the subdimensions of the G-CAVS scale (acceptance, enjoyment of use, perceived value) had moderate, positive, and statistically significant relationships with chatbot usage frequency (r = .434-.458, p < .001). The strongest relationship was observed for the *perceived value* subdimension (r = .458). In contrast, the relationship between the *perceived risk* subdimension and chatbot usage frequency was weak and statistically nonsignificant (r = .089, p > .05). These findings support the criterion-related validity of the scale, particularly in relation to the three dimensions of acceptance, enjoyment of use, and perceived value. In contrast, the perceived risk dimension showed no statistically significant association with usage frequency.

This result supports the validity of the scale in reflecting theoretically expected relationships. In other words, the scale accurately represents both the relationships that should exist and the ones that theoretically should not. When examining the correlations among the four subdimensions of the scale, moderate, positive, and statistically significant relationships were observed (r = .469-.549, p < .001). These findings indicate that, although the subdimensions belong to the same overarching construct, they measure theoretically distinct concepts. The fact that the correlations were not excessively high (> .85-.90) demonstrates that the subdimensions are distinguishable from one another, providing further evidence of discriminant validity (Fornell & Larcker, 1981).

DISCUSSION AND INTERPRETATION

In recent years, with the increasing popularity of artificial intelligence—particularly chatbots—there has been growing discussion about how and in what ways this technology should be used, especially in teacher education contexts. The main purpose of this study is to develop a new scale to measure perceptions of chatbot technologies in teacher education, particularly among pre-service teachers, focusing on the dimensions of acceptance, perceived value, perceived risk, and enjoyment of use.

The results obtained with the G-CAVS showed that the scale has a four-factor structure consisting of acceptance, enjoyment of use, perceived risk, and value. The data of the study were collected from different universities located in the Bavarian state of Germany. Before conducting EFA, the suitability of the data for factor analysis was tested, and the results showed that the data were at a good level of suitability for factor analysis.

Key findings obtained from the EFA revealed that the G-CAVS scale consists of four factors and that each factor has an eigenvalue above 1. The results of the EFA demonstrated that the scale has a strong structure supporting its validity and that each factor has high internal reliability. These results indicate that the scale is a reliable and valid tool for both scientific and practical use. Furthermore, these results obtained are consistent with the literature, especially in terms of the fact that the dimensions of acceptance, enjoyment, and perceived value show significant relationships with chatbot usage frequency (Davis, 1989; Kim et al., 2005). In contrast, the fact that the perceived risk dimension does not show a significant relationship with usage frequency is a remarkable finding. However, this situation is consistent with previous studies showing that risk perception often does not directly determine individuals' technology use behaviors (Lima et al., 2005; Venkatesh & Bala, 2008). Therefore, the four-factor structure presented by the G-CAVS scale confirms the theoretically expected relationships and supports nomological validity.

The CFA results showed that the four-factor model demonstrated acceptable to good fit (e.g., RMSEA = .06 and CFI = .96). All standardized factor loadings ranged between .50 and .94, indicating a significant and strong structure. These findings provide convincing evidence that the G-CAVS validly and reliably measures the structure proposed in theory.

When descriptive statistics and factor correlations for G-CAVS are examined, statistically significant, positive, and moderately strong relationships were found among the dimensions of acceptance, enjoyment of use, and perceived value. This consistent pattern of relationships indicates that these three dimensions are complementary structures in measuring the positive aspects of chatbot adoption. On the other hand, statistically significant but weak positive relationships were found between the risk dimension and acceptance and enjoyment of use. This unexpected finding suggests that participants associate risk perception related to chatbot use not with rejection of use, but with a certain level of adoption. This may point to a phenomenon that can be called the risk-adoption paradox (Pavlou, 2003). Indeed, in the studies of Lima et al. (2005) and Yang et al. (2023), it was reported that when users thought that the benefit of AI-based tools was very high, they could ignore their risk perceptions or consider these risks as a manageable cost. Moreover, no statistically significant relationship was found between perceived risk and perceived value. This shows that the benefits and value perceived by participants from chatbots and the risks they felt were independent of each other. This finding confirms the dynamics of the equation "perceived value = perceived benefit-perceived sacrifice," which forms the basis of the VAM (Kim et al., 2007). Perceived benefit and perceived risk (a type of sacrifice) can be considered as independent variables and interact to influence the final adoption decision. Labadze (2023), in his study conducted in the educational context, also found that benefits such as time saving and efficiency outweighed concerns about data privacy. However, alternative reasons should also be considered. First, users—particularly pre-service teachers—may not yet fully understand the potential risks associated with chatbot use, especially those related to accuracy, bias, or long-term dependency. Second, risk perceptions may increase after prolonged or more intensive use of chatbots, suggesting that risk awareness could be experience dependent. Finally, institutional contexts such as universities may create a sense of trust and perceived safety that temporarily masks risk concerns (Bulut, et al., 2025). In summary, the findings reveal that while the basic structure of the scale is consistent, risk perception plays a different and independent role from the other dimensions and has discriminant validity within the scale.

In order to evaluate the criterion-related validity of the G-CAVS scale, its relationship with chatbot usage frequency was examined using Spearman's rho coefficient. Analyses showed moderate, positive, and statistically significant relationships between the dimensions of acceptance, enjoyment of use, and perceived value and usage frequency, with the strongest relationship found in the perceived value dimension. In contrast, the relationship between the perceived risk dimension and usage frequency was weak and nonsignificant. These results support the criterion-related validity of the scale in terms of reflecting theoretically expected relationships. Particularly, the fact that the perceived value dimension showed the strongest relationship with usage frequency demonstrates that rational benefit-cost analyses of users play a central role in chatbot adoption and reinforces the explanatory power of the VAM model in this context (Sohn & Kwon, 2020). On the other hand, the nonsignificant relationship of perceived risk supports the view, also stated by Nemt-allah et al. (2024), that risk perception alone is not decisive for chatbot use and related behavioral outcomes, but should be evaluated together with other elements in the value equation.

The replication of the four-factor structure in the Turkish teacher sample further supports the robustness of the G-CAVS across educational and cultural contexts.

Study Limitations and Future Research Directions

Although this study provides robust evidence for the validity and reliability of the G-CAVS, certain limitations should be acknowledged in order to properly contextualize the findings and guide future research.

First, the primary data were collected through an online survey (Unipark) from university students in the Bavarian Region of Germany. This sample served as the main dataset for the scale development. In addition, a separate confirmatory analysis was conducted with a Turkish teacher sample to examine the stability of the factor structure. While these samples were adequate for initial validation, they still limit the generalizability of the findings to broader cultural and professional contexts. Considering that cultural factors may significantly influence TAMs (Schepers & Wetzels, 2007; Tarhini et al., 2017), future studies should examine the psychometric properties of G-CAVS in different countries and diverse learning environments (e.g., vocational training, corporate contexts, and different educational levels). Moreover, the applicability of the scale could be extended by testing it with in-service teachers, professionals from different sectors, or various age groups.

Second, a limitation concerns the perceived risk dimension. Initially, five items were developed for this subscale; however, during the scale refinement process, three items "Ich fühle mich bei der Nutzung von Chatbots nicht sicher" (*I do not feel safe when using chatbots*), "Ich befürchte, dass Chatbots meine kognitiven Fähigkeiten negativ beeinflussen" (*I fear that chatbots negatively affect my cognitive abilities*), "Ich denke, dass die Informationen von Chatbots unzuverlässig sind" (*I think that the information from chatbots is unreliable*) were removed due to low item-total correlations, leaving the subscale with only two items. Accordingly, in the present study, the perceived risk dimension mainly captures privacy-related concerns associated with chatbot use and should therefore be interpreted primarily within this scope. Although these two items demonstrated high reliability and factor loadings in the present sample, it is generally preferable to include a greater number of items to achieve a more robust and nuanced measurement of a construct (Borsci et al., 2022). Therefore, especially in cross-cultural adaptation studies, reintroducing these items or developing new ones addressing risks such as misinformation, academic integrity, or overreliance on chatbots, and testing them with larger samples, would strengthen this dimension.

Third, concerns content validity. Although the G-CAVS is grounded in established adoption models (TAM/TAM3 and VAM), certain constructs discussed in these frameworks—such as perceived ease of use, attitude, anxiety, and perceived sacrifice—were not included as separate dimensions. This decision was theoretically motivated to maintain a parsimonious and perception-focused scale structure aligned with the specific aims of the study. Nevertheless, the exclusion of these constructs may limit the coverage of all possible facets of chatbot adoption. Future research could address this limitation by integrating these constructs into extended versions of the scale or by examining their incremental contribution to chatbot acceptance.

Fourth, the sample size was sufficient for the exploratory (n = 216) and confirmatory (n = 184) factor analyses conducted with the German sample, but it remains relatively modest. Although an additional CFA was performed with a larger, independent Turkish teacher sample (n = 263), future studies should aim to validate the scale with larger and more diverse samples to enhance the stability and generalizability of the model (Kline, 2023).

Finally, in the present study, criterion-related validity was examined using self-reported chatbot usage frequency, which provides only an initial and indirect indicator of external validation. The study relied solely on self-report measures, which are subject to social desirability bias and common method variance (Podsakoff et al., 2003). A promising direction for future research would be to triangulate self-reports with behavioral data. For instance, G-CAVS scores could be correlated with actual usage metrics obtained from chatbot usage logs or with performance outcomes in learning tasks facilitated by chatbots. This approach would further strengthen the evidence for criterion-related validity. Future research could further extend this line of inquiry by examining criterion-related validity using a wider range of external variables and more in-depth research designs.

From a practical perspective, the G-CAVS scale offers several potential applications for institutions, policymakers, and practitioners in teacher education. Higher education institutions may use the scale to systematically assess pre-service teachers' perceptions of chatbot technologies prior to or following their integration into teaching and learning processes. Such assessments can inform institutional decisions regarding the adoption, regulation, or pedagogical framing of chatbot use in courses and teacher training programs.

For policymakers and educational administrators, aggregated G-CAVS results can provide empirical insights into perceived benefits and concerns related to chatbot use, supporting evidence-based policy development, data protection guidelines, and professional development initiatives. At the practitioner level, teacher educators may employ the scale as a diagnostic tool to identify areas where additional guidance, ethical discussion, or instructional support is needed, particularly with respect to perceived risks and responsible use of chatbots. Overall, the scale enables a structured and empirically grounded approach to evaluating chatbot acceptance in teacher education contexts. The scale is currently employed within the framework of the KI meets vhb project funded by the Virtuelle Hochschule Bayern (Virtual University of Bavaria) to evaluate teacher education courses, including Reformpädagogische und innovative Konzeptionen der Grundschule and Lehrgangsorientierte und lernwegsorientierte Konzeptionen des Schriftspracherwerbs.

In conclusion, despite these limitations, the G-CAVS emerges as a valid and reliable instrument for assessing perceptions of chatbot technologies in teacher education. Addressing these points in future research will further enhance the robustness of the scale and expand its applicability across diverse fields and cultural settings.

Author contributions: SP: conceptualization, writing – original draft, formal analysis; **GR:** funding acquisition, writing – review & editing. Both authors approved the final version of the article.

Funding: This project funded by the Virtuelle Hochschule Bayern and Department of Education, Institute for Primary Education Research, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany.

Acknowledgments: The authors would like to thank Prof. Bärbel Kopp, Dr. Simon Meyer, and all participants for their support in reviewing the scale items and for providing valuable feedback.

Ethics declaration: The study was conducted in accordance with ethical guidelines. Participation was voluntary, and informed consent was obtained.

AI statement: The authors used generative AI tools exclusively for language editing purposes. AI-based tools were not used for data analysis, interpretation, or the generation of scientific content.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Akdogan, E. (2021). Life and education built on trust. *Journal of Research in Social Sciences and Language, 1*(2), 128-137. https://doi.org/10.71514/jssal/2021.38

Al-Abdullatif, A. M. (2023). Modeling students' perceptions of chatbots in learning: Integrating technology acceptance with the value-based adoption model. *Education Sciences, 13*(11), Article 1151. https://doi.org/10.3390/educsci13111151

APA. (2017). Ethical principles of psychologist and code of conduct. *APA*. http://www.apa.org/ethics/code/index.aspx

Balcı, Ö. (2024). The role of ChatGPT in English as a foreign language (EFL) learning and teaching: A systematic review. *International Journal of Current Educational Studies, 3*(1), 66-82. https://doi.org/10.46328/ijces.107

Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The chatbot usability scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing, 26*(1), 95-119. https://doi.org/10.1007/s00779-021-01582-9

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Bulut, M. A., Adıgüzel, T., & Kaya, M. H. (2025). Exploring views and experiences of faculty members' participation in an asynchronous online program: Using a micro-learning format and CoP framework. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2025.2532056

Büyüköztürk, Ş. (2021). *Sosyal bilimler için veri analizi el kitabı* [A handbook of data analysis for the social sciences]. Pegem Publications.

Castells, M., Fernandez-Ardevol, M., Qiu, J. L., & Sey, A. (2009). *Mobile communication and society: A global perspective*. MIT Press. https://doi.org/10.1111/j.1944-8287.2008.tb00398.x

Chambers, C. (2004). Technological advancement, learning, and the adoption of new technology. *European Journal of Operational Research, 152*(1), 226-247. https://doi.org/10.1016/S0377-2217(02)00651-3

Chang, C. Y., Hwang, G. J., & Gau, M. L. (2022). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology, 53*(1), 171-188. https://doi.org/10.1111/bjet.13158

Chen, Q., Lu, Y., Gong, Y., & Xiong, J. (2023). Can AI chatbots help retain customers? Impact of AI service quality on customer loyalty. *Internet Research*, *33*(6), 2205–2243. https://doi.org/10.1108/INTR-09-2021-0686

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(1). https://doi.org/10.7275/jyj1-4868

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. SAGE.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Gable, R. K., & Wolf, M. B. (2012). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings*. Springer.

Gil-Garcia, J. R., Helbig, N., & Ojo, A. (2014). Being smart: Emerging technologies and innovation in the public sector. *Government Information Quarterly, 31*, I1-I8. https://doi.org/10.1016/j.giq.2014.09.001

Goli, M., Sahu, A. K., Bag, S., & Dhamija, P. (2023). Users' acceptance of artificial intelligence-based chatbots: An empirical study. *International Journal of Technology and Human Interaction, 19*(1), 1-18. https://doi.org/10.4018/IJTHI.318481

Güldal, H., & Dinçer, E. O. (2025). Can rule-based educational chatbots be an acceptable alternative for students in higher education? *Education and Information Technologies, 30*(3), 3979-4012. https://doi.org/10.1007/s10639-024-12977-5

Gumusel, E., Zhou, K. Z., & Sanfilippo, M. R. (2024). *User privacy harms and risks in conversational AI: A proposed framework*. arXiv. https://doi.org/10.48550/arXiv.2402.09716

Gupta, A. K. (2022). Ethical considerations in the deployment of AI chatbots: Lessons from ChatGPT. *Data Science Insights Magazine, 5*, 19-22.

Hair, J. F. (2014). Multivariate data analysis: An overview. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 904-907). Springer. https://doi.org/10.1007/978-3-642-04898-2_395

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104-121. https://doi.org/10.1177/109442819800100106

Hu, L. T., & 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

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36. https://doi.org/10.1007/BF02291575

Kim, H. W., Chan, H. C., & Gupta, S. (2007). Value-based adoption of mobile internet: An empirical investigation. *Decision Support Systems, 43*(1), 111-126. https://doi.org/10.1016/j.dss.2005.05.009

Kim, Y., Park, Y., & Choi, J. (2017). A study on the adoption of IoT smart home service: Using value-based adoption model. *Total Quality Management & Business Excellence, 28*(9-10), 1149-1165. https://doi.org/10.1080/14783363.2017.1310708

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.

Köhler, C., & Hartig, J. (2024). ChatGPT in higher education: Measurement instruments to assess student knowledge, usage, and attitude. *Contemporary Educational Technology, 16*(4), Article ep528. https://doi.org/10.30935/cedtech/15144

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), Article 56. https://doi.org/10.1186/s41239-023-00426-1

Lemke, C., Kirchner, K., Anandarajah, L., & Herfurth, F. N. (2023, October). Exploring the student perspective: Assessing technology readiness and acceptance for adopting large language models in higher education. In *Proceedings of the 22nd European Conference on e-Learning: ECEL 2023*. Academic Conferences and Publishing Limited. https://doi.org/10.34190/ecel.22.1.1828

Liao, Y. K., Wu, W. Y., Le, T. Q., & Phung, T. T. T. (2022). The integration of the technology acceptance model and value-based adoption model to study the adoption of e-learning: The moderating role of e-WOM. *Sustainability, 14*(2), Article 815. https://doi.org/10.3390/su14020815

Lima, M. L., Barnett, J., & Vala, J. (2005). Risk perception and technological development at a societal level. *Risk Analysis: An International Journal, 25*(5), 1229-1239. https://doi.org/10.1111/j.1539-6924.2005.00664.x

Mathavan, B., Vafaei-Zadeh, A., Hanifah, H., Ramayah, T., & Kurnia, S. (2024). Understanding the purchase intention of fitness wearables: Using value-based adoption model. *Asia-Pacific Journal of Business Administration, 16*(1), 101-126. https://doi.org/10.1108/APJBA-04-2022-0166

Mijwel, M. M. (2015). History of artificial intelligence. *Computer Science, 1*(1), 1-6. https://doi.org/10.13140/RG.2.2.16418.15046

Mokoena, O. P., & Seeletse, S. M. (2025). AI in rural classrooms: Challenges and perspectives from South African educators. *International Journal of Current Educational Studies, 4*(2), 30-52. https://doi.org/10.46328/ijces.199

Nemt-Allah, M., Khalifa, W., Badawy, M., Elbably, Y., & Ibrahim, A. (2024). Validating the ChatGPT usage scale: Psychometric properties and factor structures among postgraduate students. *BMC Psychology, 12*(1), Article 497. https://doi.org/10.1186/s40359-024-01983-4

Özcan, A., & Polat, S. (2023). Artificial intelligence and chat bots in academic research. *Journal of Research in Social Sciences and Language, 3*(2), 81-90. https://doi.org/10.71514/jssal/2023.111

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce—Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce, 7*, 69-103. https://doi.org/10.1080/10864415.2003.11044275

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), Article 879. https://doi.org/10.1037/0021-9010.88.5.879

Raiche, A. P., Dauphinais, L., Duval, M., De Luca, G., Rivest-Hénault, D., Vaughan, T., Proulx, C., & Guay, J. P. (2023). Factors influencing acceptance and trust of chatbots in juvenile offenders' risk assessment training. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1184016

Şanlı, C. (2025). Artificial intelligence in geography teaching: Potentialities, applications, and challenges. *International Journal of Current Educational Studies, 4*(1), 47-76. https://doi.org/10.46328/ijces.170

Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management, 44*(1), 90-103. https://doi.org/10.1016/j.im.2006.10.007

Şeker, H. ve Gençdoğan, B. (2006). *Psikolojide ve eğitimde ölçme aracı geliştirme* [Developing measurement tools in psychology and education]. Nobel Yayıncılık.

Simon, H. A. (1995). Artificial intelligence: An empirical science. *Artificial Intelligence, 77*(1), 95-127. https://doi.org/10.1016/0004-3702(95)00039-H

Sohn, K., & Kwon, O. (2020). Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. *Telematics and Informatics, 47*, Article 101324. https://doi.org/10.1016/j.tele.2019.101324

Srinivasan, R. (2008). Sources, characteristics and effects of emerging technologies: Research opportunities in innovation. *Industrial Marketing Management, 37*(6), 633-640. https://doi.org/10.1016/j.indmarman.2007.12.003

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Pearson.

Taktak, M. & Bafrali, G. (2025). ChatGPT usage scale in education: Validity and reliability study. *International Journal of Technology in Education, 8*(1), 193-207. https://doi.org/10.46328/ijte.1024

Tarhini, A., Masa'deh, R. E., Al-Busaidi, K. A., Mohammed, A. B., & Maqableh, M. (2017). Factors influencing students' adoption of e-learning: A structural equation modeling approach. *Journal of International Education in Business, 10*(2), 164-182. https://doi.org/10.1108/JIEB-09-2016-0032

Ursavaş, Ö., Şahin, S., & Mcılroy, D. (2014). Technology acceptance measure for teachers: T-TAM/Öğretmenler için teknoloji kabul ölçeği: Ö-TKÖ. *Eğitimde Kuram ve Uygulama, 10*(4), 885-917. https://dergipark.org.tr/tr/pub/eku/article/74152

Valdez, G., McNabb, M., Foertsch, M., Anderson, M., Hawkes, M., & Raack, L. (1999). *Computer-based technology and learning: Evolving uses and expectations*. NCREL.

Vashishth, T. K., Sharma, V., Sharma, K. K., Kumar, B., Kumar, A., & Panwar, R. (2024). Artificial intelligence (AI)-powered chatbots: Providing instant support and personalized recommendations to guests 24/7. In P. Thaichon, P. K. Dutta, P. R. Chelliah, & S. Gupta (Eds.), *Technology and luxury hospitality* (pp. 211-236). Routledge. https://doi.org/10.4324/9781003488248-15

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Verhoef, P. C., Stephen, A. T., Kannan, P. K., Luo, X., Abhishek, V., Andrews, M., Bart, Y., Datta, H., Fong, N., Hoffman, D. L., Hu, M. M., Novak, T., Rand, W., & Zhang, Y. (2017). Consumer connectivity in a complex, technology-enabled, and mobile-oriented world with smart products. *Journal of Interactive Marketing, 40*(1), 1-8. https://doi.org/10.1016/j.intmar.2017.06.001

Yang, J., Chen, Y. L., Por, L. Y., & Ku, C. S. (2023). A systematic literature review of information security in chatbots. *Applied Sciences, 13*(11), Article 6355. https://doi.org/10.3390/app13116355

Yurt, E. (2025). The self-regulation for AI-based learning scale: Psychometric properties and validation. *International Journal of Current Educational Studies, 4*(1), 95-118. https://doi.org/10.46328/ijces.176

APPENDIX A

Table A1. Demographic characteristics of Turkish teachers

		Frequency (n)	Percentage (%)
Field	Elementary school/preschool	48	18.3
	Verbal fields/humanities & social sciences	113	43.0
	Numerical fields/STEM fields	68	25.9
	Foreign language	34	12.9
Professional experience	15 years and above	119	45.2
	11-15 years	71	27.0
	6-10 years	40	15.2
	1-5 years	33	12.5
Age	41 years old and above	145	55.1
	31-40 years old	103	39.2
	26-30 years old	15	5.7
Gender	Female	150	57.03
	Male	113	42.97

APPENDIX B

Table B1. General chatbot acceptance, enjoyment, perceived risk, and value scale (G-CAVS) items

Factor	Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Acceptance (Akzeptanz)	5.I believe it is appropriate that chatbots are used at universities. (Ich finde es richtig, dass Chatbots an Universitäten eingesetzt werden.)					
	6. I think that the use of chatbots in schools, training, and education is meaningful. (Ich finde, dass der Einsatz von Chatbots in Schule, Ausbildung und Studium sinnvoll ist.)					
	14. I believe that chatbots should be integrated into educational institutions. (Ich finde, dass Chatbots in Bildungseinrichtungen integriert werden sollten.)					
	21. I think that chatbots are valuable for changes in the education sector. (Ich finde, dass Chatbots wertvoll für Veränderungen im Bildungsbereich sind.)					
Enjoyment of Use (Freude)	2. Interacting with chatbots is great fun for me. (Die Interaktion mit Chatbots macht mir großen Spaß.)					
	19. Using chatbots motivates me to discover more. (Die Nutzung von Chatbots motiviert mich, mehr zu entdecken.)					
	20. When I interact with chatbots, my curiosity is sparked. (Wenn ich mit Chatbots interagiere, wird meine Neugier geweckt.)					
Perceived Value (Wahrgenommener Wert)	3. I intend to use chatbots for my future learning. (Ich beabsichtige, Chatbots für mein zukünftiges Lernen zu verwenden.)					
	4. Chatbots give me an advantage because they save me time and effort. (Chatbots bringen mir einen Vorteil, da ich Zeit und Mühe sparen kann.)					
	22. Using chatbots makes my learning experience valuable. (Die Nutzung von Chatbots macht mein Lernen zu einem wertvollen Erlebnis.)					
	23. I think that chatbots can increase my personal success. (Ich denke, dass Chatbots meinen persönlichen Erfolg steigern können.)					
	24. Chatbots facilitate my learning processes. (Chatbots erleichtern meine Lernprozesse.)					
Perceived Risk (Wahrgenommenes Risiko)	12. I am worried that my personal data might be shared when I use chatbots. (Ich mache mir Sorgen, dass meine persönlichen Daten beim Nutzen von Chatbots weitergegeben werden.)					
	15. I am concerned that my personal data may be misused when I use chatbots. (Ich habe Bedenken, dass meine persönlichen Daten bei der Nutzung von Chatbots missbraucht werden.)					

Note: Stimme überhaupt nicht zu; Stimme eher nicht zu; Neutral; Stimme eher zu; Stimme voll und ganz zu

APPENDIX C

Table C1. CFA results and reliability statistics (Turkish version)

Factor	Item	FL	SE	CR	p	α	ω	CR	AVE	MSV
Acceptance	A3	0.783	0.052	14.97	***	0.881	0.881	0.919	0.739	0.71
	A4	0.843	0.055	15.22	***					
	A10	0.819	0.052	15.76	***					
	A14	0.790	0.049	15.82	***					
Enjoyment of use	EU1	0.821	0.057	13.12	***	0.816	0.822	0.848	0.651	0.65
	EU8	0.796	0.053	15.50	***					
	EU9	0.679	0.054	14.31	***					
Perceived value	PV2	0.679	0.051	13.27	***	0.874	0.875	0.909	0.627	0.71
	PV6	0.716	0.048	14.79	***					
	PV11	0.821	0.055	14.77	***					
	PV12	0.796	0.054	14.57	***					
	PV13	0.679	0.053	14.17	***					
Perceived risk	PR7	0.894	0.178	5.00	***	0.865	0.865	0.868	0.765	0.01
	PR5	0.915	0.183	4.98	***					

Note. FL: Factor loading