International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Development of Artificial Intelligence-Assisted Learning Satisfaction Scale: Validity and Reliability Study

Mithat Elcicek & Ata Pesen

To cite this article: Mithat Elcicek & Ata Pesen (26 Jan 2026): Development of Artificial Intelligence-Assisted Learning Satisfaction Scale: Validity and Reliability Study, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2026.2619619

To link to this article: https://doi.org/10.1080/10447318.2026.2619619

Published online: 26 Jan 2026.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2026.2619619

Development of Artificial Intelligence-Assisted Learning Satisfaction Scale: Validity and Reliability Study

Mithat Elcicek[a] and Ata Pesen[b]

[a]Computer Education and Instructional Technology, Siirt University, Siirt, Türkiye; [b]Curriculum and Instruction, Siirt University, Siirt, Türkiye

ABSTRACT
This research is targeted upon tracing the satisfaction levels of the students in universities with AI-assisted learning experiences besides to create a measurement tool, which is valid and reliable in this respect. The work on the Scale tool of the survey, designed to be developed in two stages, was executed in the academic year 2024–2025 and the number of the participants was 640 university students who said they had used different AI applications. After a thorough literature review, students' focus group interviews involving AI, and the discussions with experts in the academic field, the scale items were decided upon. The data were subjected to an Exploratory Factor Analysis (EFA) for the first sample group (n = 259), and it was revealed that the scale had one-factor nature. The results of EFA indicated that the 12-item draft scale covered 68.6% of the variance. A Confirmatory Factor Analysis (CFA) carried out on the data from the second sample group (n = 381) attests that the 12 items and the single-factor structure offered good fit indices. Later still, the second-level CFA aligned with this structure's validity. The Cronbach alpha internal consistency coefficient of the scale was reported as .96 and reliability was very high. The results demonstrate that the 12-item Learning Assistance with AI Satisfaction questionnaire is a valid and reliable measurement tool. The scale in question gives a strong foundation to researchers and practitioners to find out the satisfaction levels of students with the AI-assisted learning experiences and to make educational policies accordingly.

KEYWORDS
Artificial intelligence-assisted learning; university student; satisfaction; scale development

1. Introduction

Artificial Intelligence (AI) refers to the capacity of digital devices to perform complex cognitive functions by imitating the thinking, learning, and problem-solving processes of the human brain (Kaplan & Haenlein, 2019). Today, AI techniques such as computer vision, natural language processing, and machine learning have had a transformative impact not only in industrial fields but also in many academic disciplines, including education (Chiu, 2024; Crompton et al., 2024). The increasing use of AI in education offers multifaceted opportunities, which include individualized learning processes, enhanced teaching efficiency, and prediction of student success. Systematic reviews (Alfredo et al., 2024; Chiu et al., 2023), bibliometric analyses (Huang et al., 2023), scope scans (Joksimovic et al., 2023; Su & Yang, 2022) and meta-analysis studies (Hwang, 2022; Su et al., 2022) conducted in this direction reveal the transformative role of AI in education in a multifaceted way.

However, a major part of current studies focus on ways of applications of AI (Crompton & Burke, 2023; Huang et al., 2023; Oubibi et al., 2025), pedagogical foundations (Chai et al., 2021; McDonald et al., 2025; Moore et al., 2023; Yue et al., 2022) have future-oriented tendencies (Huang et al., 2023) while ignoring students' experiential and affective considerations regarding such technologies. However, it is of utmost importance how AI-based learning environments are designed and which pedagogical principles are taken as basis as well as how such environments are drawn upon and evaluated by the students (Luckin et al., 2022). In this context, student satisfaction is a key indicator in understanding

CONTACT Mithat Elcicek mithat_elcicek@siirt.edu.tr Computer Education and Instructional Technology, Siirt University, Siirt, Türkiye

the achievement of AI-assisted learning processes as well as their pedagogical compliance and sustainability.

The level of student satisfaction in terms of AI-assisted learning environments is associated with not only the functionality of technology but also with motivational tendencies of students toward learning processes. In this context, the Self-Determination Theory (SDT) offers a strong theoretical framework for the explanation of the psychological mechanisms underlying behind such satisfaction (Bureau et al., 2022; Howard et al., 2021; Ryan & Deci, 2020). According to the SDT, participation of individuals in an activity is internalized depending on their intrinsic and extrinsic motivation levels, which has a direct impact on the level of satisfaction experienced by individuals. Accordingly, the Human-Computer Interaction (HCI) theory helps explain students' satisfaction with and intention to use AI tools. Moreover, the Unified Theory of Acceptance and Use of Technology (UTAUT) model is a widely-used approach while seeking to understand individual attitudes and behaviors in using technology (Venkatesh et al., 2012). The UTAUT explains technology acceptance and intention to use through variables such as performance expectancy, effort expectancy, social influence, and facilitating conditions. These theoretical frameworks all together offer a comprehensive understanding about behavioral and experiential factors having an impact on AI-assisted learning in educational environments (Chatterjee & Bhattacharjee, 2020). In this direction, students' satisfaction with AI-assisted learning environments could be holistically assessed in terms of both motivational dynamics included in SDT and HCI and technological acceptance factors defined by the UTAUT.

Measurement of students' satisfaction with AI-assisted learning is critical in terms of effective and efficient integration of AI-assisted technologies into educational processes. Students' satisfaction-related data could help personalize learning experiences, enhance the interaction, and improve pedagogical designs (An et al., 2025; Kim et al., 2025; Yaseen et al., 2025). However, though current literature encompasses a myriad of AI-related scale development studies for measuring subscales including attitude (Grassini, 2023; Schepman & Rodway, 2023; Sindermann et al., 2021; Suh & Ahn, 2022), perception (Işık et al., 2024; Üzüm et al., 2025; Yang & Xu, 2025), literacy (Hornberger et al., 2023; Wang et al., 2023), technology acceptance (Arslankara & Usta, 2024; Wiss et al., 2025), threat (Kieslich et al., 2021), and competency (Wang et al., 2023; Wang & Chuang, 2024), it is observed that the literature lacks a valid and reliable instrument to directly measure AI-assisted learning satisfaction. This makes it hard to systematically investigate subjective evaluations of students' AI-assisted learning experiences and creates an empirical gap in the relevant field.

Therefore, there is a need for a novel instrument including both psychological and technological variables in an attempt to measure student satisfaction with AI-assisted learning. In light of the relevant requirement, the present study basically seeks to develop the "AI-assisted Learning Satisfaction Scale" along with its validity and reliability analyses. The relevant scale is expected to contribute to the holistic assessment of student satisfaction with AI-assisted educational practices as well as to empirically reveal pedagogical effects of educational technologies.

2. Literature review

2.1. AI in education

AI has turned into one of the most effective and transformative Technologies of the digital era with its capacity to imitate cognitive functions of human intelligence. Allowing the modeling of complex mental processes such as learning, reasoning, problem-solving, and decision-making, AI is not only a technological tool but also triggers social, cultural, and economic transformations (Abidoğlu, 2025; Joksimovic et al., 2023; Lv, 2023). Having a wide range of area of applications in several sectors including health, finance, transportation, and particularly education, AI stands out with its advanced cognitive capacities such as analysis of large-scale data sets, recognition of patterns, and inference from learning processes (Coppin, 2004; Peterson et al., 2021).

The transformative effect of AI is largely felt in many sectors, particularly in the field of education. AI-assisted systems offer adaptive learning ways suitable for students' readiness levels, learning paces, and interests by supporting personalized learning experiences (Alam & Mohanty, 2023;

Bearman & Ajjawi, 2023). Thus, students could have the opportunity to monitor learning behaviors in real time as well as personalize and organize learning strategies. This process mainly includes AI-based "learning analytics," "intelligent learning systems," "adaptive learning environments," and "recommendation systems" (Darwish, 2025; Guan et al., 2020).

AI-assisted systems create significant insights about the learning process by analyzing large data sets obtained from students' learning processes (Ahmad et al., 2024). These insights help restructure learning objectives, personalize learning content, and strengthen pedagogical interaction. Moreover, instant feedback mechanisms offered by AI allow students to constantly monitor learning processes, identify lacking aspects, and follow progress (Ruiz-Rojas et al., 2023). Accordingly, with the widespread adoption of e-learning in higher education, the need for automated assessment tools has increased, and studies aimed at improving the accuracy and consistency of automated essay scoring (AES) systems have gained importance. In this context, the study conducted by Beseiso et al. (2021) shows that AI-supported assessment processes can be reliably implemented in higher education.

AI applications offer crucial advantages particularly in teaching complex skills. Virtual simulators, digital instructors, augmented reality-assisted AI systems allow students to practise in real environments without taking high risks, resulting in reliable and effective learning opportunities in disciplines including health, engineering, and aviation (Somenko et al., 2023). In their study, Sami et al. (2025) reported that students in the field of medicine view artificial intelligence as effective and reliable learning agents and that AI helps them more efficiently learn complex medical skills than conventional methods. Such technologies back up the principles including active participation, experience-based learning, and social interaction emphasized by the constructive learning theory (Brooks & Brooks, 1999). Chan and Hu (2023) revealed that university students find AI as a valuable tool particularly for tasks such as brainstorming and analysis. In this context, AI is seen as a tool that accelerates the transition from conventional teaching to student-centred, interactive, and practice-based pedagogical models (Dixit et al., 2024).

A great number of countries systematically integrate AI-assisted learning technologies into educational processes. Applications including China-based SquirrelAI, the USA-based Watson, the UK-based Third Space Learning, and Sweden-based Sana Labs demonstrate how AI is effectively used in various contexts. In addition, platforms including ChatGPT, ALEKS, Duolingo, Coursera, Gemini, Calscraft, Utifen, and Assassin's Creed: Discovery Tour offer AI-assisted solutions in a wide range of applications such as language learning, problem-solving, personalized learning, and interactive learning (Strielkowski et al., 2025). Ellis and Slade (2023) emphasized that if used appropriately, ChatGPT could offer crucial advantages for both teachers and students. Accordingly, Vasconcelos and dos Santos (2023) reported that ChatGPT and Bing Chat improve students' critical and creative thinking skills and support problem-solving skills. This indicates that AI creates a holistic learning ecosystem that supports both pedagogical and technological innovations.

The effect of AI in education is directly associated with not only the development of technological tools but also the ways students interact with such systems and satisfaction with such experiences (Luckin et al., 2022; Zhang et al., 2025). It is observed that if students are satisfied at a high degree, they get involved in more interaction, voluntarily participate in learning processes, and obtain enhanced academic achievement rates (Capinding & Dumayas, 2024). Accordingly, correlating satisfaction scores with academic outcomes or motivational indicators will support broader construct validity of measurement tools, allowing for a more holistic assessment of the pedagogical effectiveness of AI-assisted learning experiences. Therefore, sustainable adoption of AI in education depends on integrating the technological infrastructure with pedagogical principles and taking into consideration affective dimensions of the learning experience.

As a result, AI in education does not only serve as a tool of innovation that transforms teaching and learning systems but also as a holistic system that personalizes learning experiences, guides pedagogical decisions based on data, and enhances learner participation. Thus, studies touching upon AI in education should be approached with a perspective that involves not only technological aspects but also student satisfaction.

2.2. Student satisfaction

Student satisfaction is a basic variable in the assessment of effectiveness of learning Technologies and the success of learning processes. Generally, satisfaction is associated with the level of meeting students' expectations from the learning experience, the significance of the learning process, and satisfaction with the outcomes (Taylor, 1996). The relevant concept is found to be an essential indicator of the quality of learning environments as it reflects both cognitive and affective reactions of the students (Knowles, 1970; Long, 1985).

Student satisfaction in AI-assisted learning environments is closely related not only to the functionality of technological innovations but also to how students integrate these technologies into their learning processes (Kim et al., 2025). Students' perceptions of the benefits obtained from AI-assisted systems, the ease of use, and pedagogical appropriateness of such systems are factors that directly affect their satisfaction levels. A study by Suchanek and Kralova (2025) revealed that students believe that AI improves the quality of education, enhancing their overall satisfaction levels. Akutay et al. (2024) reported that AI increases satisfaction by improving nursing students' case management performance, and the integration of AI into traditional nursing education curricula is recommended. In this context, satisfaction is considered a construct that reveals the extent to which students find AI-assisted applications meaningful, effective, engaging, and sustainable (An et al., 2025; Yaseen et al., 2025). At this point, it should also be considered that the concept of "satisfaction" in the context of Turkish higher education may carry some cultural and linguistic nuances. In Turkish, satisfaction often encompasses not only an emotional state of contentment but also a more holistic evaluation of how expectations have been met. Furthermore, cultural norms could have an impact on students' perceptions of authority figures and institutional structures, their ways of giving feedback, and their tendencies to Express.

Motivational processes play a significant role in the formation of satisfaction. According to SDT, the reasons why individuals participate in an activity vary depending on their levels of intrinsic and extrinsic motivation (Ryan & Deci, 2020). SDT emphasizes that meeting individual' needs for autonomy, competence, and relatedness increases satisfaction with the learning process (Bureau et al., 2022; Howard et al., 2021). In this context, the key factors that can increase satisfaction are that AI-assisted learning environments strengthen students' sense of autonomy by providing personalized learning experiences, support perceptions of competence through immediate feedback mechanisms, and foster social connectedness through collaborative AI tools. Similarly, HCI theory helps explain how students' satisfaction with AI tools and their intention to use them. Furthermore, the UTAUT model, developed in the context of technology acceptance and use, provides an important framework for understanding student satisfaction (Venkatesh et al., 2012). According to UTAUT, users' intention to adopt and use a technology is shaped by performance expectancy, effort expectancy, social influence, and facilitating conditions. Students' beliefs that AI will contribute to their learning (performance expectancy), their perceptions of the system's ease of use (effort expectancy), the influence of peers and faculty (social influence), and infrastructure support (facilitating conditions) are variables that directly influence satisfaction with AI-assisted learning (Oubibi et al., 2025; Su & Yang, 2022). Together, these theoretical frameworks provide a comprehensive understanding of the behavioral and experiential factors that influence AI-assisted learning in educational settings (Chatterjee & Bhattacharjee, 2020).

Consequently, student satisfaction in the context of AI-assisted learning can be considered a phenomenon at the intersection of both motivational (based on SDT and HCI) and technological acceptance (based on UTAUT) processes. Student satisfaction with AI-assisted learning experiences is a critical factor determining not only the quality of learning outcomes but also the sustainable use of these technologies in education. Therefore, developing tools that can validly and reliably measure student satisfaction with AI-assisted learning has become a crucial requirement for both theoretical clarity and applied research.

3. Methodology

3.1. Research design

The explanatory sequential design, one of the mixed research methods, was employed in the present study since the goal is to develop a scale aimed at measuring users' learning satisfaction with artificial

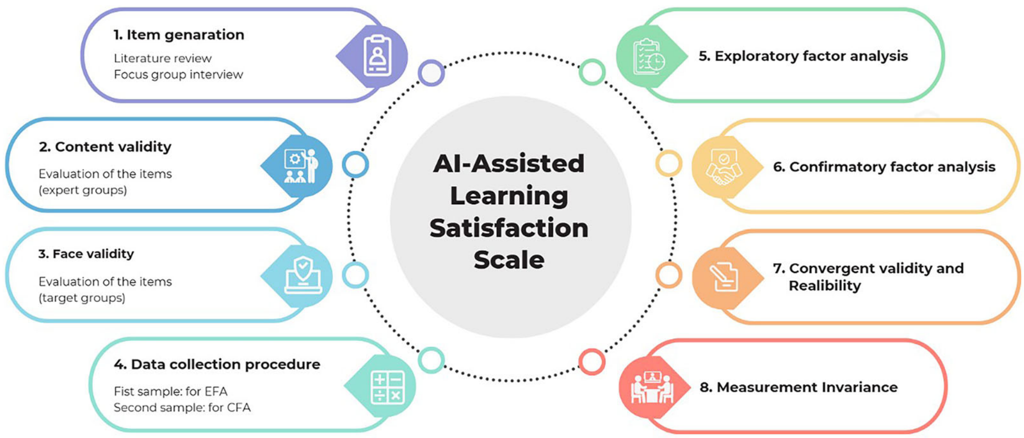

Figure 1. Scale development process.

intelligence. This method is generally advised for scale development studies in which researchers are supposed to combine qualitative and quantitative data with a deeper understanding (Creswell, 2021). The flow chart regarding the scale development process is given in Figure 1.

3.1.1. Item generation

Primarily the relevant literature was reviewed comprehensively while creating the item pool. In this context, SDT, UTAUT, and HCI theories were taken as basis to create the item pool. SDT emphasizes that meeting individual needs of autonomy, competence, and relatedness enhances satisfaction in the learning process (Ryan & Deci, 2020) while the UTAUT model suggests that performance and effort expectancy, social influence, and facilitating conditions are decisive in adoption of technological systems (Venkatesh et al., 2012). HCI theory explains the effect of learners" ways of interaction with technology, the applicability of the system, and user experience on perceived satisfaction (Chatterjee & Bhattacharjee, 2020). In light of these theoretical frameworks, the relevant item pool was created in a manner to include students' AI-assisted learning environments in reference to cognitive, affective, motivational, and technological perspectives. The item pool was created after similar scales in the literature were reviewed and recent studies related to AI's learning process integration and user satisfaction were taken into consideration (An et al., 2025; Oubibi et al., 2025; Yaseen et al., 2025). Besides, a focus group interview was conducted with 10 university students actively using AI applications in education. After the theoretical framework and qualitative data obtained from the participants were subjected to content analysis, a draft form of the item pool consisting of 29 items was created.

3.1.2. Content validity

For the content validity of the 29-item draft form, 10 PhD experts employed at different Turkish universities were consulted for their opinions along with an examination of the CVR (Content Validity Ratio) and CVI (Content Validity Index) values, which are important quantitative methods while evaluating the item pool with expert opinions. While experts use the CVR to show how "necessary" each item in the scale is, CVI is a more detailed content validity measure for experts to evaluate the suitability of an item for the structure it is intended to measure. First, experts were asked to rate the necessity of each scale item on a three-point scale as "3-necessary, 2-useful but not necessary, 1-unnecessary," and the CVR value was calculated for each item. The input from the experts was analyzed using the Lawshe (1975) technique. Lawshe (1975) suggested a cutoff rate of 0.62 for 10 expert opinions. Four items with CVR values below 0.62 were eliminated from the draft form, leaving a 25-item form. In the subsequent step, which involves determining the CVI, six different experts were approached to provide feedback on the suitability of each item. Five items with CVI scores less than 0.78 were deleted from the scale using Polit and Beck (2006) proposed cutoff value. Thus, a draft form consisting of 20 items was created after items deemed to have similar expressions in line with expert opinions, and items that were not directly related to learning satisfaction with artificial intelligence were eliminated.

3.1.3. Face validity

To evaluate whether the items of the scale are suitable in terms of language and expression, an academic in the field of the Turkish language was consulted for his opinion, and the necessary arrangements were made in line with suggestions taken from the academic. To test the comprehensibility of the items subject to the arrangement, the scale items were presented to 4 students (2 female, 2 male) who were not included in the study, and once feedback was provided for each item, obscure and ambiguous expressions were identified and the relevant items were rearranged. A five-point Likert-type rating scale was employed to determine the participants' opinions on the items in the scale. Accordingly, the responses to the items were graded as "Strongly Agree (5)," "Agree (4)," "Partially Agree (3)," "Disagree (2)" and "Strongly Disagree (1)." In line with this rating structure, the draft form of the scale was made ready for application.

3.1.4. Data collection procedure

This study followed all of the guidelines established in the Scientific Research and Publication Ethics Directive for Higher Education Institutions. During the investigation, no ethical violations, as described in the second section of the directive, titled "Actions Contrary to Scientific Research and Publication Ethics," occurred. The Siirt University Scientific Research and Publication Ethics Committee approved the research with ethical evaluation document number 4971/621 dated June 15, 2023. Data from the study were collected online via Google Form during the spring semester of the 2024–2025 academic year from a university located in the east of Turkey, on a voluntary basis and using the convenience sampling technique, one of the non-probability sampling methods. The participant group is made up of students who actively utilize AI-assisted programs (such as ChatGPT, Copilot, Gemini, GrammarlyGO, Duolingo Max, Khanmigo, Quizlet Q-Chat, Notion AI, and other education-oriented AI-assisted platforms). The data collection procedure was conducted twice. Exploratory Factor Analysis (EFA) was performed on a data set consisting of 259 (female = 175, male = 84) participants. At every run, the scale form was sent online to students enrolled at different faculties to achieve an appropriate sample size. Bentler and Chou (1987) and Tinsley and Kass (1979) suggested the measurement of the items five or tenfold, while Comrey and Lee (2013) and Tabachnick and Fidell (2013) reported that a good sample size should range from 300 to 500. Kline (2016) reported that acceptable sample size should be between 100 and 200. In this study, data collection was terminated once data saturation was considered. Following the EFA, the second data collection procedure was launched to reach a data set consisting of 381 (Female = 243, Male = 138) participants for the Confirmatory Factor Analysis (CFA). The demographic profile of the participants is presented in Table 1.

Table 1. Demographic profile of the participants.

		ECA Sampling		CFA Sampling	
Variable		n	%	n	%
Gender	Female	175	67.6	243	63.8
	Male	84	32.4	138	36.2
	Total	259	100	381	100
Age	18–21	177	68.4	275	72.2
	22–25	59	22.8	78	20.5
	26–29	18	6.9	16	4.2
	30+	5	1.9	12	3.1
	Total	259	100	381	100
Faculty	Education	35	13.5	60	15.7
	Arts and Science	33	12.7	45	11.8
	Fine Arts and Design	30	11.6	37	9.7
	Economics and Administrative Sciences	33	12.7	41	10.8
	Theology	32	12.4	52	13.6
	Engineering	31	12.0	40	10.5
	Health Sciences	33	12.7	55	14.4
	Veterinary	32	12.4	51	13.4
	Total	259	100	381	100

3.1.5. Exploratory factor analysis

Initially, an examination of the kurtosis and skewness coefficients of the data was performed to check the normality assumption, which is one of the conditions of EFA. As the kurtosis and skewness coefficients of the data collected from the first sample group ($N = 259$) were within the range of ±1, the data was deemed to meet the normality assumption (George & Mallery, 2010). To check whether the first sample size was sufficient for the EFA, the Kaiser–Meyer–Olkin (KMO) sample adequacy test was applied while the Bartlett Sphericity Test was employed to identify the significance of the correlation between the items (Kaiser & Rice, 1974). The analysis revealed that the KMO value was 0.966 and the results of the Bartlett test [χ^2=5163.866, $p < 0.01$] were found to be significant. Kaiser and Rice (1974) argue that a KMO value of 0.90 and above indicates a "perfect" level of adequacy. Therefore, it can be inferred that the sample size was extremely suitable for factor analysis along with a sufficient correlation between the items. Information on the findings is given in Table 2.

Cronbach's Alpha internal consistency coefficient and McDonald's Omega value were calculated to find out whether the 20-item scale was reliable or not, and the split-half test was performed to make a more conservative estimate of reliability (Guttman, 1945) and to cross-check the Cronbach's Alpha results (DeVellis, 2016). Information on the findings is given in Table 3.

Table 3 reveals that the internal consistency reliability of the 20-item scale was evaluated with Cronbach's Alpha and McDonald's Omega coefficients. Cronbach's Alpha value was calculated as 0.974 and McDonald's Omega value as 0.974, indicating an excellent internal consistency for the scale (Hair et al., 2019; Tavakol & Dennick, 2011). In addition, the split-half reliability analysis indicates that the Cronbach's Alpha values between the two halves of the scale were 0.958 and 0.953, respectively, and the correlation between the forms was 0.885. The Spearman-Brown coefficient (for equal and unequal lengths) was 0.939 and the Guttman Split-Half Coefficient was 0.939. These results reveal that both halves of the scale provide a high level of consistent measurement and that the overall structure of the scale is highly reliable (DeVellis, 2016; Guttman, 1945). Furthermore, the fact that the corrected item-total correlations of all items of the scale are $\geq$0.50 indicates that the scale consists of highly discriminatory and strong items (Nunnally & Bernstein, 1994). Hence, it is suggested that the item-total correlations of the 20-item scale are sufficient and the level of satisfaction with artificial intelligence can be measured reliably. As part of the EFA performed to identify the construct validity of the scale and the factors, the Maximum Likelihood method was used as the factor extraction method, and the rotation technique was not used because a single-factor structure was expected (Field, 2018; Osborne, 2014). The data had skewness and kurtosis values within ±1, hence the Maximum Likelihood approach was used to test factor loadings and generate fit indices (Fabrigar et al., 1999; Osborne, 2014). Principal Axis Factoring, a commonly recommended strategy for cases where normalcy is questionable, was not used in this study because it provided no further benefits based on the data structure. The first analysis revealed a scale with 20 items and a two-factor structure. However, the number of factors in the scale was identified through the Kaiser Criterion results and based on the Kaiser criterion, and the factor with an eigenvalue less than 1.0 (Extraction Sums of Squared Loadings = 0.852) was removed from the scale (Tabachnick & Fidell, 2013). Horn parallel analysis method (Horn, 1965), which provides more

Table 2. KMO and Bartlett's sphericity test results.

Kaiser–Meyer–Olkin measurement of sampling adequacy		0.966
Bartlett's Sphericity Test	Approximate Chi-square	5163,866
	Df	171
	Sig.	0.000

Table 3. EFA reliability results.

Cronbach's alpha			0.974
McDonald's Omega			0.974
Split-Half Test	Cronbach's Alpha	Part 1	0.958
		Part 2	0.953
Correlation between forms			0.885
Spearman–Brown Coefficient	Equal length		0.939
	Unequal length		0.939
Guttman Split-Half Coefficient			0.939

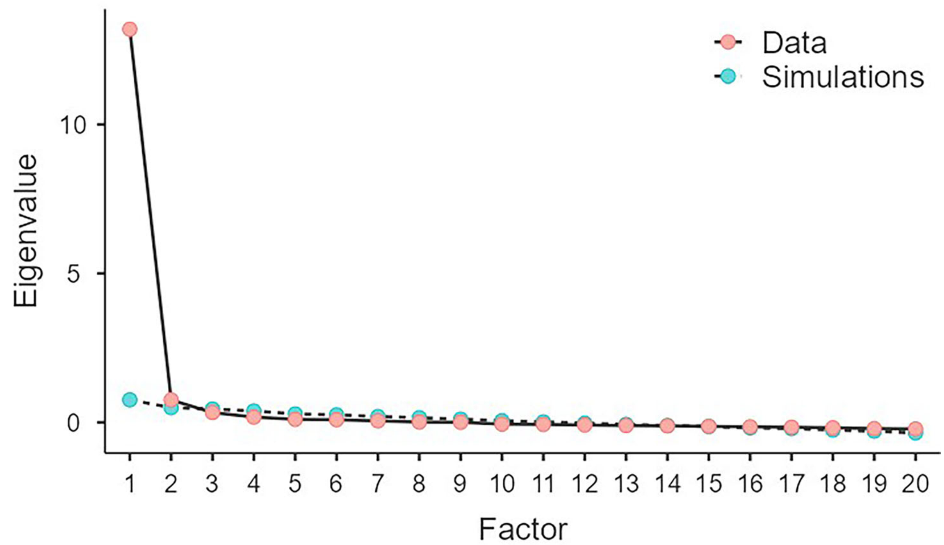

Figure 2. Horn parallel analysis graph of artificial ıntelligence-assisted learning satisfaction scale.

Table 4. EFA results of the artificial ıntelligence-assisted learning satisfaction scale.

Items	Mean	Std	Factor
Item 19	3.75	0.920	0.872
Item 7	3.78	0.915	0.870
Item 18	3.83	0.882	0.859
Item 3	3.64	0.905	0.856
Item 10	3.66	0.849	0.847
Item 9	3.66	0.872	0.834
Item 5	3.69	0.884	0.828
Item 6	3.48	0.921	0.826
Item 2	3.70	0.920	0.822
Item 4	3.69	0.954	0.815
Item 11	3.58	0.857	0.809
Item 15	3.56	0.866	0.806
Item 17	3.52	0.899	0.792
Item 16	3.52	0.925	0.792
Item 8	3.70	0.881	0.791
Eigenvalues		10,295	
Variance explained (%)		68,633	

reliable results than KMO and scree plot, was employed. Parallel analysis is based on the comparison of the magnitudes of the eigenvalues created in the data set (Data) with the eigenvalues obtained from a randomly generated data set (Simulations) that have the same magnitude (Fabrigar et al., 1999; Hayton et al., 2004). The parallel analysis results in Figure 2 highlight that the eigenvalue of only one factor was higher than the 95% confidence interval of the eigenvalues obtained from random data sets. This finding supports the single-factor structure of the scale (Hayton et al., 2004; Horn, 1965) (Figure 2).

As a result of EFA, five items (Item 1, Item 12, Item 13, Item 14, and Item 20) were removed from the scale to reach a 15-item single-factor structure. The 15-item and single-factor structure explained 68.633% of the variance. The averages of all items ranged from 3.48 to 4.56, with standard deviations of 0.84 to 0.95. In addition, factor loading values for each item in the scale were found to be above 0.700 (see Table 4). This rate exceeds the 60% threshold generally recommended for social science research (Hair et al., 2019; Tabachnick & Fidell, 2013). The high level of explained variance supports the power and unidimensionality of the scale.

3.1.6. Confirmatory factor analysis

The CFA was performed using the second data set of 381 people with the AMOS v24 program to examine the goodness of fit values. During confirmatory factor analysis (CFA), the model's fit indices were investigated. Modification indices revealed that several items had large correlation values between

Table 5. Fit index values of the scale at CFA.

Item fit indices	After modification	Acceptable fit	Good fit	Fit level
χ^2/df	2.637	≤ 5	≤ 3	Good fit
GFI	0.946	≥ 0.90	≥ 0.95	Acceptable fit
RMSEA	0.066	≤ 0.08	≤ 0.06	Acceptable fit
CFI	0.981	≥ 0.90	≥ 0.95	Good fit
IFI	0.981	≥ 0.90	≥ 0.95	Good fit
TLI	0.975	≥ 0.90	≥ 0.95	Good fit
NFI	0.970	≥ 0.90	≥ 0.95	Good fit
SRMR	0.024	≤ 0.08	≤ 0.05	Good fit

error terms. Specifically, three items (Items 2, 3, and 16) had a significant error covariance with numerous items, resulting in a local fit difficulty. This was considered item redundancy since the items in question used identical phrasing or had content overlap. To maintain the model's conceptual integrity and increase statistical fit, these three components were excluded from the analysis. The fit indices for the 12-item, single-factor validated model are shown in Table 5.

In this study, the normed chi-square index (χ^2/df) used to evaluate the model fit was 2.637, which is in the goodness-of-fit value range. The goodness of fit index (GFI) was 0.946, and the root mean square error of approximation (RMSEA) was 0.066, which is in the acceptable fit value range. The comparative fit index (CFI) was 0.981; the incremental fit index (IFI) was 0.981; the normed fit index (NFI) was 0.970; the Tucker–Lewis index (TLI) was 0.975; and the standardized root mean square measure (SRMR) was 0.024, all of which fall within the good fit value range (Hu & Bentler, 1999; Kline, 2016).

The CFA made it clear that the item factor loadings range between 0.77 and 0.85. CFA results: All 12 items in the scale supported the predicted single-factor structure of the AI-assisted learning satisfaction scale. Harman's single factor test was performed on all questions to see whether the explained variation was caused by Common Method Bias (Podsakoff et al., 2003). Unrotated factor analysis found that the first factor accounted for 67.3% of the total variation. Although the ratio exceeded the recommended 50% limit, confirmatory factor analysis showed acceptable fit values, indicating no significant measurement-related common method bias in the data. However, the variables may be conceptually related (Fuller et al., 2016; Podsakoff et al., 2003). As a result of CFA, to increase the model fit values of the AI-assisted learning satisfaction scale as well as to reach a more appropriate structure, covariances were allowed between the error terms of some items that were similar or consecutive in terms of item content based on the modification indices (Byrne, 2016; Hu & Bentler, 1999). To increase model fit indices in the CFA, several theoretically relevant error factors were linked. For example, because the items "Using AI was an easy and accessible experience for me" (e8) and "The instructions while working with AI were clear and understandable" (e9) have similar cognitive content, participants are likely to respond similarly to them. As a result, covariance was defined for the error terms of these items. Similarly, the items "Working with AI increased my curiosity about learning" (e15) and "AI made learning fun for me" (e17) have similar affective content; hence the error words are connected. The CFA results diagram for the AI-assisted learning satisfaction scale is shown in Figure 3.

3.1.7. Convergent validity ve reliability

Within the scope of the CFA, the Average Variance Explained (AVE) value (Hair et al., 2019) was calculated to determine the convergent validity of the scale, the Composite Reality (CR) value (Fornell & Larcker, 1981), Cronbach's Alpha (Tavakol & Dennick, 2011) and McDonald's Omega (McDonald, 1999) values were calculated to determine the internal consistency. The findings are given in Table 6.

The Average Variance Extracted (AVE) value, as a criterion for convergent validity, should be greater than 0.50 (Fornell & Larcker, 1981; Hair et al., 2019). Considering the item factor loadings obtained within the scope of CFA, the AVE value was calculated as 0.663. Since the resulting AVE value is above 0.50 (Hair et al., 2019), it can be argued that the AI-assisted learning satisfaction scale has convergent validity. Indeed, as highlighted by Hair et al. (2019), the fact that each item's factor loadings were greater than 0.50 and the average variance explained (AVE) value was found to be 0.663 indicates that the scale items sufficiently represent a shared construct. This implies that the construct "AI-assisted learning satisfaction" was conceptually quantified in a single dimension and consistently. The AVE is substantially above 0.50, indicating that the factor's explanatory power exceeds the

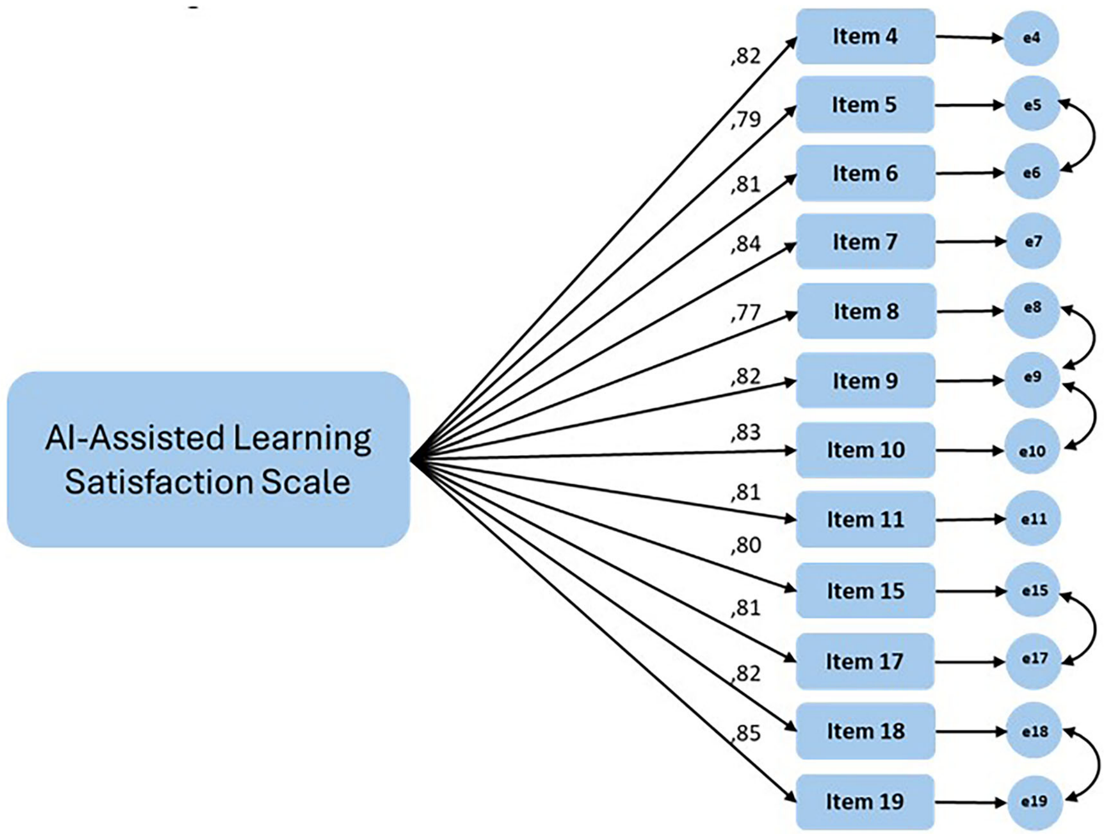

Figure 3. Confirmatory factor analysis of scale.

Table 6. CFA reliability results.

Convergent validity	Average variance extracted (AVE)			0.663
Reliability	Composite Reliability (CR)			0.959
	Cronbach's Alpha			0.961
	McDonald's Omega			0.961
	Split-Half Test	Cronbach's Alpha	Part 1 (8[a])	0.927
			Part 2 (7[b])	0.935
	Correlation Between Forms			0.882
	Spearman-Brown Coefficient	Equal Length		0.937
		Unequal Length		0.937
	Guttman Split-Half Coefficient			0.937

measurement margin of error and that the scale items have a high degree of shared variation (Fornell & Larcker, 1981; Hair et al., 2019). The internal consistency reliability of the scale was evaluated through Composite Reliability (CR), Cronbach's Alpha, and McDonald's Omega coefficients. The CR value was calculated as 0.966, the Cronbach's Alpha value as 0.961, and the McDonald's Omega value as 0.961, indicating that the scale has excellent internal consistency (Fornell & Larcker, 1981; Hair et al., 2019; Tavakol & Dennick, 2011). The literature (Clark & Watson, 1995; Streiner, 2003) suggests that item redundancy may occur when the α coefficient exceeds 0.95. Thus, inter-item correlations were investigated and are shown in Table 7. Correlation scores varied from 0.59 to 0.85. Correlations were mostly within the 0.30–0.80 range, with only two values above 0.80, indicating a conceptually uniform but not repetitive scale. As a result, the scale's high Cronbach's Alpha rating reflects its structural consistency rather than item redundancy. This finding backs up the scale's great internal consistency and comprehensive depiction of the construct it measures. Furthermore, the split-half reliability analysis revealed that Cronbach's Alpha values between the two halves of the scale were found to be 0.947 and 0.957, respectively, and the correlation between the forms was 0.882. The Spearman-Brown coefficient (for equal and unequal lengths) was 0.937 and 0.937, and the Guttman Split-Half Coefficient was 0.937.

Table 7. Inter-item correlation matrix for assessing ıtem homogeneity and redundancy.

		Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 15	Item 17	Item 18	Item 19
Item 4	*r*	1											
	p												
	n	381											
Item 5	*r*	0.746**	1										
	p	0.000											
	n	381	381										
Item 6	*r*	0.709**	0.768**	1									
	p	0.000	0.000										
	n	381	381	381									
Item 7	*r*	0.678**	0.682**	0.693**	1								
	p	0.000	0.000	0.000									
	n	381	381	381	381								
Item 8	*r*	0.612**	0.597**	0.613**	0.660**	1							
	p	0.000	0.000	0.000	0.000								
	N	381	381	381	381	381							
Item 9	*r*	0.669**	0.627**	0.673**	0.680**	0.776**	1						
	p	0.000	0.000	0.000	0.000	0.000							
	n	381	381	381	381	381	381						
Item 10	*r*	0.657**	0.642**	0.649**	0.716**	0.673**	0.806**	1					
	p	0.000	0.000	0.000	0.000	0.000	0.000						
	n	381	381	381	381	381	381	381					
Item 11	*r*	0.646**	0.599**	0.620**	0.662**	0.685**	0.693**	0.705**	1				
	p	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
	n	381	381	381	381	381	381	381	381				
Item 15	*r*	0.649**	0.633**	0.691**	0.635**	0.602**	0.632**	0.659**	0.659**	1			
	p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
	n	381	381	381	381	381	381	381	381	381			
Item 17	*r*	0.611**	0.602**	0.666**	0.675**	0.602**	0.646**	0.664**	0.654**	**0.815****	1		
	p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
	n	381	381	381	381	381	381	381	381	381	381		
Item 18	*r*	0.659**	0.642**	0.631**	0.690**	0.629**	0.678**	0.678**	0.673**	0.674**	0.707**	1	
	p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	n	381	381	381	381	381	381	381	381	381	381	381	
Item 19	*r*	0.684**	0.675**	0.662**	0.711**	0.627**	0.677**	0.697**	0.675**	0.728**	0.743**	**0.849****	1
	p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	n	381	381	381	381	381	381	381	381	381	381	381	381

**$p < 0.01$.

Table 8. Scale item analysis results.

Item no.	Corrected item-total correlation	Cronbach's alpha if item deleted	Lower Mean ± standard deviation	Upper Mean ± standard deviation	t
Item 4	0.793	0.958	2.61 ± 0.807	4.63 ± 0.610	20.253*
Item 5	0.783	0.958	2.74 ± 0.766	4.57 ± 0.587	19.284*
Item 6	0.800	0.958	2.48 ± 0.739	4.52 ± 0.624	21.500*
Item 7	0.812	0.957	2.83 ± 0.818	4.69 ± 0.505	19.578*
Item 8	0.763	0.959	2.87 ± 0.825	4.50 ± 0.655	15.626*
Item 9	0.819	0.957	2.72 ± 0.772	4.50 ± 0.624	18.257*
Item 10	0.818	0.957	2.78 ± 0.766	4.53 ± 0.623	18.058*
Item 11	0.786	0.958	2.77 ± 0.757	4.45 ± 0.668	16.893*
Item 15	0.801	0.957	2.61 ± 0.717	4.45 ± 0.653	19.204*
Item 17	0.801	0.958	2.57 ± 0.722	4.38 ± 0.756	17.534*
Item 18	0.815	0.957	2.68 ± 0.782	4.57 ± 0.604	19.441*
Item 19	0.842	0.956	2.67 ± 0.772	4.57 ± 0.587	19.909*

*$p < 0.01$.

These results reveal that both halves of the scale provide high-level consistent measurements and that the general structure of the scale is quite reliable (DeVellis, 2016; Guttman, 1945).

A value of ≥.50 for the corrected item-total correlations of all items of the scale indicates that the scale consists of highly distinctive and strong items (Nunnally & Bernstein, 1994). Hence, it is suggested that the item-total correlations of the 15-item scale are sufficient and that satisfaction with artificial intelligence can be measured reliably. In addition, item distinctiveness was assessed after the responses of the upper 27% and lower 27% of the participants were assessed based on the total scale scores (DeVellis, 2016; Kelley, 1939) along with independent sample t-tests performed for each item. The findings are given in Table 8.

The item-total correlations ranging between 0.800 and 0.837 indicate that all items in the scale have a high level of discriminatory power. Indeed, high item discrimination directly contributes to scale reliability because items that effectively discriminate between high and low scorers tend to correlate strongly with the overall scale score, thus increasing internal consistency (DeVellis, 2016; Tavakol & Dennick, 2011). The results of the item analysis including corrected item-total correlations and subjected to an examination in an attempt to evaluate how effectively the scale items can predict and discriminate the overall score, one may notice that significant t-values between 15.626 and 21.500) are observed for all items in the upper and lower 27% groups. Comparison of the upper and lower groups in terms of significant *t*-values shows the discriminatory power of the items.

3.1.8. Measurement ınvariance

Measurement invariance is an analysis that makes use of restrictions between groups to assess changes in the goodness of fit indices (Cheung & Rensvold, 2002). For measurement invariance, multi-group CFA, in which factor loadings, item constants, and error variances were freely estimated, was performed. Formal, metric, scalar, and error invariance of the AI-assisted learning satisfaction scale were checked based on gender. Chen (2007) reported that model fit is sufficient for groups if ΔCFI value is less than 0.01, ΔRMSEA value is less than 0.015 for metric and scalar invariance, and ΔSRMR value is less than 0.03 for metric invariance and 0.015 for scalar invariance. The findings are given in Table 9.

Findings from the configural invariance model show that the assessment tool's factor structure is consistent across genders. The model's fit indices (χ^2/df = 2.40, CFI = 0.969, TLI = 0.958, RMSEA = 0.061, SRMR = 0.031) are within the recommended threshold values (Hu & Bentler, 1999; Kline, 2016). The factor structure is also valid across both groups. As a result, the AI-assisted learning satisfaction scale fits the gender variable well and meets the configural invariance criterion. Metric invariance was tested using Model 1. A multi-group CFA was carried out by limiting the factor loadings of the scale items to be equal across gender. The AI-assisted learning satisfaction scale exhibits an adequate fit in terms of metric invariance, as evidenced by fit indices [$\chi^2(109) = 241.934$, RMSEA = 0.057, CFI = 0.970, SRMR = 0.033]. The metric invariance and formal invariance models did not show significant chi-square ($\Delta\chi^2(11) = 6.778$, $p = 0.817$, and ΔCFI = 0.001, ΔRMSEA = 0.004, and ΔSRMR = 0.001 values were within the recommended criterion values (Chen, 2007), and measurement invariance was achieved at the metric level between genders. After metric invariance was established, factor structures,

Table 9. Results regarding measurement invariance.

Gender												
Model	Invariance	χ^2	df	RMSEA	CFI	SRMR	$\Delta\chi^2$	Δdf	p	ΔCFI	ΔRMSEA	ΔSRMR
Unconstrained	Configural	235.156	98	0.061	0.969	0.031				–	–	–
Model 1	Metric	241.934	109	0.057	0.970	0.033	6.778	11	0.817	0.001	0.004	0.001
Model 2	Scalar	249.118	121	0.053	0.971	0.032	7.184	12	0.928	0.001	0.004	0.001
Model 3	Strict	287.465	138	0.053	0.966	0.036	38.347	17	0.012	0.003	0.000	0.004

factor loadings, and item constants were compared across groups, and scalar invariance was assessed. The fit indices for Model 2 [$\chi^2(121) = 249.118$, RMSEA = 0.053, CFI = 0.971, SRMR = 0.032] show that the AI-assisted learning satisfaction measure has sufficient scalar invariance across gender. The scalar invariance and metric invariance models showed no significant chi-square difference between groups ($\Delta\chi^2(12) = 7.184$, $p = 0.928$). The values of ΔCFI = 0.001, ΔRMSEA = 0.003, and ΔSRMR = 0.001 were within the recommended criterion values (Chen, 2007), indicating measurement invariance at the scalar level across gender. In the last stage, error invariance was assessed by equating the factor structures, factor loadings, item constants, and error variances across groups. The fit indices for error invariance [$\chi^2(138) = 287.465$, RMSEA = 0.053, CFI = 0.966, SRMR = 0.036] show that it has an adequate fit. When comparing error invariance with scalar invariance models, a significant $\Delta\chi^2$ was discovered [$\Delta\chi^2(17) = 38.347$, $p = 0.012$]. Although there was a substantial $\Delta\chi^2$ in the rigorous invariance model, where error variances were equal, the fact that ΔCFI (0.003) stayed below the recommended threshold value of 0.01 suggests that the error variance equality was at an acceptable level (Chen, 2007; Cheung & Rensvold, 2002). The model comparison results show that the AI-assisted learning satisfaction scale is measurement invariant across male and female participants at the formal, metric, scalar, and error levels, demonstrating a stringent measure of invariance. According to Byrne (2016), cross-construct group comparisons are valid when measurement invariance is accomplished at all levels. As a result, group means from this scale with male and female participants can be statistically and reliably compared.

4. Limitations and strengths

The most notable limitation of this study is that the sample was made up entirely of university students, with low demographic diversity. This may limit the findings' generalizability to student groups from other universities or cultural backgrounds. However, this constraint provides opportunity for future investigation. First, data collection from various higher education institutions, departments, and age groups would improve the scale's external validity. Second, relying on participant self-reporting may expose the data to measurement flaws such social desirability bias or response discrepancies. Third, the scale was tested in a specific cultural setting; therefore validation research in other cultures would improve the scale's generalizability and construct validity.

One of the study's main strengths is the complete literature review that was undertaken during the scale building process, as well as the establishment of content validity through expert judgments. Furthermore, the results of exploratory and confirmatory factor analyses corroborate the scale's construct validity. Cronbach's Alpha and McDonald's Omega values are high, indicating a solid scale structure. The AVE value (0.663) obtained for each factor exceeds the allowed threshold, indicating convergent validity. The CR score of 0.959 demonstrates adequate composite reliability. The scale's formal, metric, scalar, and error invariance across gender enables comparisons between groups. Finally, the scale's 12-item single-factor layout allows for easy administration and assessment.

5. Conclusion and recommendations

This study permitted the creation of a valid and reliable scale for assessing learner satisfaction with artificial intelligence. The item pool, which was created in accordance with current theoretical foundations, sought to assess satisfaction with AI-assisted learning as a multifaceted construct having cognitive, affective, motivational, and technological components. According to SDT, meeting individuals' needs for autonomy, competence, and relatedness increases satisfaction (Howard et al., 2021; Ryan & Deci, 2020), whereas the UTAUT Model reveals that users' satisfaction is shaped by performance and effort expectations, social influences, and facilitating conditions (Venkatesh et al., 2012). HCI theory highlights how technology experience affects user satisfaction in learning situations (Chatterjee & Bhattacharjee, 2020).

Originally constructed with a multidimensional structure in mind, the exploratory and confirmatory factor analyses for the scale found that the items had substantial loadings on one dominating factor. This shows that students rated autonomy, simplicity of use, interaction quality, and performance requirements in AI-assisted learning processes holistically, rather than independently, under a single "satisfaction" umbrella. Thus, the scale's single-factor structure is theoretically sound and statistically consistent.

Unlike previous scales in the literature that assess attitudes, acceptance, and competency toward artificial intelligence, this scale, which focuses on learning satisfaction, stands out for its ease of use and functionality. The results show that the scale has a good internal consistency, and its unidimensional structure is supported by both exploratory and confirmatory factor analyses. Furthermore, measurement invariance studies enable meaningful and consistent comparisons between male and female groups.

This scale offers researchers, AI developers, instructional designers, and educational policymakers a scientifically validated instrument for assessing and evaluating AI-assisted learning satisfaction. Educators can use the scale to examine how students interact with different AI-assisted learning applications and change content or interaction design based on their satisfaction levels. For example, instructional designers can utilize the scale data to evaluate which modules are the most effective and to optimize training materials. AI developers can use user feedback to improve the app experience and create new features. Education policymakers can use scale data to assess the efficacy of AI-assisted training programs and develop scientifically informed educational plans. In this context, the scale is an effective and instructive instrument for both academic research and application-based instructional design.

Future research should do correlation-based comparisons and study predicted correlations between the construct evaluated by the scale and existing AI attitude, acceptance, and competency scales. In addition, testing the measurement invariance between academic disciplines or levels of exposure to artificial intelligence in future studies will further strengthen the generalizability of the findings. Furthermore, comparison studies that allow for the examination of potential disparities in scale scores across characteristics such as gender, age, department, or degree of education can help to strengthen the scale's content validity. Furthermore, longitudinal studies can be carried out to assess the scale's stability over time, as well as validity analyses across cultures. Adaptation studies that incorporate validity and reliability analyses across cultures will improve the generalizability to multinational applications. While the scale is based on a sample of university students, applying it to students from various fields or user groups interacting with AI-based technologies provides an interdisciplinary perspective to the literature.

Finally, given the study's limited demographic diversity, conducting similar research with larger diversified groups might improve the findings' generalizability and external validity. These recommendations support the scale's use as a practical and guiding instrument in academic research and educational practice.

Ethical statement

This study was ethically reviewed by the "Siirt University Publication Ethics Committee" and was approved ethically with the; Date of Ethics Evaluation Document: 15.06.2023; Meeting No. 621; Issue Number of Ethics Evaluation Document: 4971/621.

AI assistance declaration

Artificial intelligence support was used during the translation and language editing stages of this study. Specifically, ChatGPT − 5 (developed by OpenAI) was employed to enhance linguistic accuracy and academic style. The AI tool was used solely for language refinement purposes and did not contribute to the generation or analysis of the research content. All text was subsequently reviewed and verified by the author(s), who take full responsibility for the final version of the manuscript.

Author contributions

CRediT: **Mithat Elcicek:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Visualization; **Ata Pesen**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors did not get any funds in any process of the present study.

ORCID

Mithat Elcicek http://orcid.org/0000-0003-1845-7271
Ata Pesen http://orcid.org/0000-0003-1676-7444

References

Abidoğlu, İ. (2025). Otonom zihinler, kutsal ruhlar: Yapay zekâ dünyasinda dinsel insanin anlam Arayışı. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, *27*(3), 1187–1204. https://doi.org/10.32709/akusosbil.1677070

Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Benhaddou, D., Ayyash, M., & Al-Fuqaha, A. (2024). Data-driven artificial intelligence in education: A comprehensive review. *IEEE Transactions on Learning Technologies*, *17*, 12–31. https://doi.org/10.1109/tlt.2023.3314610

Akutay, S., Kaçmaz, H. Y., & Kahraman, H. (2024). The effect of artificial intelligence supported case analysis on nursing students' case management performance and satisfaction: A randomized controlled trial. *Nurse Education in Practice*, *80*, 104142. https://doi.org/10.1016/j.nepr.2024.104142

Alam, A., & Mohanty, A. (2023). Educational technology: Exploring the convergence of technology and pedagogy through mobility, interactivity, AI, and learning tools. *Cogent Engineering*, *10*(2), 2283282. https://doi.org/10.1080/23311916.2023.2283282

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, *6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2025). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*, *38*(5–6), 987–1008. https://doi.org/10.1080/09588221.2023.2246519

Arslankara, V. B., & Usta, E. (2024). Generative artificial intelligence as a lifelong learning self-efficacy: Usage and competence scale. *Journal of Teacher Education and Lifelong Learning*, *6*(2), 288–302. https://doi.org/10.51535/tell.1489304

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1160–1173. https://doi.org/10.1111/bjet.13337

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Beseiso, M., Alzubi, O. A., & Rashaideh, H. (2021). A novel automated essay scoring approach for reliable higher educational assessments. *Journal of Computing in Higher Education*, *33*(3), 727–746. https://doi.org/10.1007/s12528-021-09283-1

Brooks, J. G., & Brooks, M. G. (1999). *In search of understanding: The case for constructivist classrooms*. ASCD.

Bureau, J. S., Howard, J. L., Chong, J. X., & Guay, F. (2022). Pathways to student motivation: A meta-analysis of antecedents of autonomous and controlled motivations. *Review of Educational Research*, *92*(1), 46–72. https://doi.org/10.3102/00346543211042426

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* (3rd ed.). Routledge.

Capinding, A. T., & Dumayas, F. T. (2024). Transformative pedagogy in the digital age: Unraveling the impact of artificial intelligence on higher education students. *Problems of Education in the 21st Century*, *82*(5), 630–657. https://doi.org/10.33225/pec/24.82.630
Chai, C. S., Lin, P. Y., Jong, M. S. Y., Dai, Y., Chiu, T. K., & Qin, J. (2021). Perceptions of and behavioral intentions towards learning artificial intelligence in primary school students. *Educational Technology & Society*, *24*(3), 89–101.
Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8
Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies*, *25*(5), 3443–3463. https://doi.org/10.1007/s10639-020-10159-7
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
Chiu, T. K. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence*, *6*, 100197. https://doi.org/10.1016/j.caeai.2023.100197
Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118
Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309
Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis* (2nd ed.). Psychology Press. https://doi.org/10.4324/9781315827506
Coppin, B. (2004). *Artificial intelligence illuminated*. Jones and Bartlett.
Creswell, J. W. (2021). *A concise introduction to mixed methods research.* (2nd ed.). SAGE Publications.
Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1), 1–22. https://doi.org/10.1186/s41239-023-00392-8
Crompton, H., Jones, M. V., & Burke, D. (2024). Affordances and challenges of artificial intelligence in K-12 education: A systematic review. *Journal of Research on Technology in Education*, *56*(3), 248–268. https://doi.org/10.1080/15391523.2022.2121344
Darwish, D. (2025). *Artificial intelligence implementation in education processes*. Deep Science Publishing.
DeVellis, R. F. (2016). *Scale development: Theory and applications.* (4th ed.). SAGE.
Dixit, A. C., Harshavardhan, B., Ashok, B. C., Sriraj, M. A., & Prakasha, K. N. (2024). Innovative pedagogical approaches for diverse learning styles and student-centric learning. *Journal of Engineering Education Transformations*, *37*(IS2), 178–188. https://doi.org/10.16920/jeet/2024/v37is2/24039
Ellis, A. R., & Slade, E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, *31*(2), 128–133. https://doi.org/10.1080/26939169.2023.2223609
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
Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.
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
Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. *Journal of Business Research*, *69*(8), 3192–3198. https://doi.org/10.1016/j.jbusres.2015.12.008
George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference.* (10th ed.). Pearson Allyn & Bacon.
Grassini, S. (2023). Development and validation of the AI Attitude Scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology*, *14*, 1191628. https://doi.org/10.3389/fpsyg.2023.1191628
Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, *4*(4), 134–147. https://doi.org/10.1016/j.ijis.2020.09.001
Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. https://doi.org/10.1007/BF02288892
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning, Harvard Unıversity Press.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, *5*, 100165. https://doi.org/10.1016/j.caeai.2023.100165

Howard, J. L., Bureau, J. S., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science: a Journal of the Association for Psychological Science*, *16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, *26*(1), 112–131. https://doi.org/10.30191/ETS.202301_26(1).0009

Hwang, S. (2022). Examining the effects of artificial intelligence on elementary students' mathematics achievement: A meta-analysis. *Sustainability*, *14*(20), 13185. https://doi.org/10.3390/su142013185

Işık, S., Çakır, R., & Korkmaz, Ö. (2024). Teachers' perception scale towards the use of artificial intelligence tools in education. *Participatory Educational Research*, *11*(H. Ferhan Odabaşı Gift Issue), 80–94. https://doi.org/10.17275/per.24.95.11.6

Joksimovic, S., Ifenthaler, D., Marrone, R., De Laat, M., & Siemens, G. (2023). Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence*, *4*, 100138. https://doi.org/10.1016/j.caeai.2023.100138

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, *30*(1), 17–24. https://doi.org/10.1037/h0057123

Kieslich, K., Lünich, M., & Marcinkowski, F. (2021). The threats of artificial ıntelligence scale (TAI) development, measurement, and test over three application domains. *International Journal of Social Robotics*, *13*(7), 1563–1577. https://doi.org/10.1007/s12369-020-00734-w

Kim, M., Kim, J., Knotts, T. L., & Albers, N. D. (2025). AI for academic success: Investigating the role of usability, enjoyment, and responsiveness in ChatGPT adoption. *Education and Information Technologies*, *30*(10), 14393–14414. https://doi.org/10.1007/s10639-025-13398-8

Kline, R. B. (2016). *Principles and practice of structural equation modeling.* (4th ed.). Guilford Press.

Knowles, M. S. (1970). *The modern practice of adult education: Andragogy versus pedagogy*. Association Press.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Long, H. B. (1985). Contradictory expectations? Achievement and satisfaction in adult learning. *The Journal of Continuing Higher Education*, *33*(3), 10–12. https://doi.org/10.1080/07377366.1985.10401035

Luckin, R., Cukurova, M., Kent, C., & Du Boulay, B. (2022). Empowering educators to be AI-ready. *Computers and Education: Artificial Intelligence*, *3*, 100076. https://doi.org/10.1016/j.caeai.2022.100076

Lv, Z. (2023). Generative artificial intelligence in the metaverse era. *Cognitive Robotics*, *3*, 208–217. https://doi.org/10.1016/j.cogr.2023.06.001

McDonald, N., Johri, A., Ali, A., & Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, *3*, 100121. https://doi.org/10.1016/j.chbah.2025.100121

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410601087

Moore, R. L., Jiang, S., & Abramowitz, B. (2023). What would the matrix do?: A systematic review of K-12 AI learning contexts and learner-interface interactions. *Journal of Research on Technology in Education*, *55*(1), 7–20. https://doi.org/10.1080/15391523.2022.2148785

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* (3rd ed.). McGraw-Hill.

Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. CreateSpace Independent Publishing Platform.

Oubibi, M., Hryshayeva, K., & Huang, R. (2025). Enhancing postgraduate digital academic writing proficiency: The interplay of artificial intelligence tools and ChatGPT. *Interactive Learning Environments*, *33*(6), 3940–3958. https://doi.org/10.1080/10494820.2025.2454445

Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D., & Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science (New York, N.Y.)*, *372*(6547), 1209–1214. https://doi.org/10.1126/science.abe2629
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. *The Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147
Ruiz-Rojas, L. I., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. *Sustainability*, *15*(15), 11524. https://doi.org/10.3390/su151511524
Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860
Sami, A., Tanveer, F., Sajwani, K., Kiran, N., Javed, M. A., Ozsahin, D. U., Muhammad, K., & Waheed, Y. (2025). Medical students' attitudes toward AI in education: Perception, effectiveness, and its credibility. *BMC Medical Education*, *25*(1), 82. https://doi.org/10.1186/s12909-025-06704-y
Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human–Computer Interaction*, *39*(13), 2724–2741. https://doi.org/10.1080/10447318.2022.2085400
Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI – Künstliche Intelligenz*, *35*(1), 109–118. https://doi.org/10.1007/s13218-020-00689-0
Somenko, D. V., Tryfonova, O. M., & Sadovyi, M. I. (2023). Artificial intelligence and neural networks in the educational process: Advantages and disadvantages. *Current Problems and Prospects of Technological and Professional Education*, 78–81.
Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2025). AI-driven adaptive learning for sustainable educational transformation. *Sustainable Development*, *33*(2), 1921–1947. https://doi.org/10.1002/sd.3221
Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence*, *3*, 100049. https://doi.org/10.1016/j.caeai.2022.100049
Su, J., Zhong, Y., & Ng, D. T. K. (2022). A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region. *Computers and Education: Artificial Intelligence*, *3*, 100065. https://doi.org/10.1016/j.caeai.2022.100065
Suchanek, P., & Kralova, M. (2025). Generative artificial intelligence expectations and experiences in management education: ChatGPT use and student satisfaction. *Journal of Innovation & Knowledge*, *10*(5), 100781. https://doi.org/10.1016/j.jik.2025.100781
Suh, W., & Ahn, S. (2022). Development and validation of a scale measuring student attitudes toward artificial intelligence. *Sage Open*, *12*(2), 1–12. https://doi.org/10.1177/21582440221100463
Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics.* (6th ed.). Pearson.
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd
Taylor, S. A. (1996). Consumer satisfaction with marketing education: Extending services theory to academic practice. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *9*, 207–220.
Tinsley, H. E., & Kass, R. A. (1979). The latent structure of the need satisfying properties of leisure activities. *Journal of Leisure Research*, *11*(4), 278–291. https://doi.org/10.1080/00222216.1979.11969406
Üzüm, B., Elçiçek, M., & Pesen, A. (2025). Development of teachers' perception scale regarding artificial intelligence use in education: Validity and reliability study. *International Journal of Human–Computer Interaction*, *41*(5), 1–12. https://doi.org/10.1080/10447318.2024.2385518
Vasconcelos, M. A. R., & dos Santos, R. P. (2023). Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2296. https://doi.org/10.29333/ejmste/13313
Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412
Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, *42*(9), 1324–1337. https://doi.org/10.1080/0144929X.2022.2072768

Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, *29*(4), 4785–4808. https://doi.org/10.1007/s10639-023-12015-w

Wiss, A., Joosten-Hagye, D., Pattershall-Geide, J., Showstark, M., Zschaebitz, E., Potter, K., Embry, E., Hageman, H., & Brooks, P. (2025). Development of the AI Acceptance Scale for Interprofessional Education (AAIPE) and collaborative practice settings. *Journal of Interprofessional Education & Practice*, *40*, 100752. https://doi.org/10.1016/j.xjep.2025.100752

Yang, Y., & Xu, H. (2025). Perception of AI creativity: Dimensional exploration and scale development. *The Journal of Creative Behavior*, *59*(2), 1–38. https://doi.org/10.1002/jocb.70028

Yaseen, H., Mohammad, A. S., Ashal, N., Abusaimeh, H., Ali, A., & Sharabati, A. A. A. (2025). The impact of adaptive learning technologies, personalized feedback, and interactive AI tools on student engagement: The moderating role of digital literacy. *Sustainability*, *17*(3), 1133. https://doi.org/10.3390/su17031133

Yue, M., Jong, M. S. Y., & Dai, Y. (2022). Pedagogical design of K-12 artificial intelligence education: A systematic review. *Sustainability*, *14*(23), 15620. https://doi.org/10.3390/su142315620

Zhang, W., Xiong, Y., Zhou, D., Liu, C., Gu, Y., & Yang, H. (2025). Balancing human and AI instruction: Insights from secondary student satisfaction with AI-assisted learning. *Interactive Learning Environments*, *33*(9), 5430–5445. https://doi.org/10.1080/10494820.2025.2482590

About the authors

Mithat Elcicek is an associate professor of Computer Education and Instructional Technology (CEIT) at Siirt University. He is interested in open and distance learning, educational technologies, information technology and education.

Ata Pesen is an associate professor of Curriculum and Instruction at Siirt University. He is an expert in assessment and evaluation, curriculum development, teacher education, flipped and blended learning environments.

Appendix A. Artificial ıntelligence-assisted learning satisfaction scale.

No.	Item (english)	Item (Turkish)	(5) Strongly agree	(4) Agree	(3) Partly agree	(2) Disagree	(1) Strongly disagree
1	AI helped me learn new information more easily	Yapay zekâ, yeni bilgileri daha kolay öğrenmemi sağladı.					
2	Content from AI was suitable for my learning needs	Yapay zekâ, öğrenme ihtiyacıma uygun içerikler sundu.					
3	The instructions were easy to follow and comprehend when studying with AI	Yapay zekâ ile çalışırken yönlendirmeler açık ve anlaşılırdı.					
4	AI allowed me to use my time more effectively while studying	Yapay zekâ, zamanı daha etkili kullanmamı sağladı.					
5	AI helped me enhance my academic performance	Yapay zekâ, akademik performansımın artmasına katkıda bulundu.					
6	AI facilitated the understanding of complex topics	Yapay zekâ, karmaşık konuların anlaşılmasını kolaylaştırdı.					
7	The feedback I received from AI was both fast and accurate.	Yapay zekâdan aldığım geri bildirimler hem hızlı hem de doğruydu.					
8	AI helped me have more curiosity for learning.	Yapay zekâ ile çalışmak, öğrenme merakımı artırdı.					
9	AI made learning fun for me	Yapay zekâ öğrenmeyi benim için eğlenceli hâle getirdi.					
10	Using AI was an easy and accessible experience for me.	Yapay zekâyı kullanmak benim için kolay ve erişilebilir bir deneyimdi.					
11	I would like to use AI in the future	Gelecekte de yapay zekayı kullanmak isterim.					
12	I would recommend others to use AI	Yapay zekâyı başkalarına da öneririm					