BMC Psychology

https://doi.org/10.1186/s40359-026-04679-z

Article in Press

Trust in AI: scale development and validation for online distance learners

Received: 11 March 2026

Accepted: 29 April 2026

Published online: 07 May 2026

Cite this article as: Üstün A.G., Yavuz M., Kayalı B. *et al.* Trust in AI: scale development and validation for online distance learners. *BMC Psychol* (2026). https://doi.org/10.1186/s40359-026-04679-z

Ayşin Gaye Üstün, Mehmet Yavuz, Bünyami Kayalı, Hasan Uçar, Erdem Erdoğdu & Aras Bozkurt

Trust in AI: Scale Development and Validation for Online Distance Learners

Ayşin Gaye ÜSTÜN[1*], Mehmet YAVUZ[2†], Bünyami KAYALI[3†], Hasan UÇAR[4†], Erdem ERDOĞDU[4†], Aras BOZKURT[4†]

[1*]Department of Computer Technology, Sinop University, Street, Sinop, 57000, State, Türkiye.
[2]Management Information Systems, Bingöl University, Street, Bingöl, 12000, State, Türkiye.
[3]Department of Computer Technology, Bayburt University, Street, Bayburt, 69000, State, Türkiye.
[4]Open and Distance Learning Department, Anadolu University, Street, Eskişehir, 26470, State, Türkiye.

*Corresponding author(s). E-mail(s): aysingaye.ustun@gmail.com;
Contributing authors: myavuz@bingol.edu.tr;
bunyamikayali@bayburt.edu.tr; hasanucar@anadolu.edu.tr;
erdeme@anadolu.edu.tr; arasbozkurt@gmail.com;
†These authors contributed equally to this work.

Abstract

Trust in artificial intelligence (AI) has emerged as a critical factor influencing students' interactions with AI-supported learning environments. However, validated instruments for measuring student trust in AI within educational contexts remain limited. This study developed and validated a multidimensional scale to assess student trust in AI systems in online distance education. Scale development followed a sequential multi-stage design, including item generation based on literature review, expert evaluation for content validity, and psychometric validation through exploratory and confirmatory factor analyses.A total of 837 distance learning students participated in the study (EFA = 412; CFA = 307; validation sample = 118). Exploratory factor analysis supported a five-factor structure explaining 63.20% of the total variance. Confirmatory factor analysis with an independent sample indicated good model fit (χ^2/df = 2.17; CFI = .95; TLI = .94; RMSEA = .06). The scale demonstrated high internal consistency (α = .94; ω = .94). Multi-group confirmatory factor analysis supported metric invariance

across gender, indicating comparable measurement across groups. No significant gender differences were observed, and trust in AI was not significantly associated with grade point average or system engagement frequency. These results suggest that trust represents a distinct perception that may influence how students engage with AI-supported learning environments. The resulting 21-item scale provides a reliable instrument for investigating student trust in AI and offers practical implications for the design and implementation of AI-supported learning in higher education.

Keywords: artificial intelligence, trust, technology acceptance, online distance education, higher education, scale development

1 Introduction

Trust is a fundamental socio-psychological construct characterized by positive expectations toward an agent in conditions of uncertainty and risk [1]. Early research on technological trust originated in human-computer and human-automation interaction, where trust was viewed as a rational, performance-based evaluation of technical competence and predictability [2]. This reductionist perspective prioritized cognitive judgments of functional accuracy but proved insufficient for explaining the affective and relational dimensions of modern AI systems. As interaction paradigms shifted, scholars recognized that trust necessitates a broader conceptualization involving comprehensibility and personal commitment [3].

Contemporary frameworks now treat trust in AI as a multidimensional construct influenced by perceived competence, helpfulness, and risk [4–6]. Emerging psychometric evidence suggests that traditional automation-based scales may not remain stable when applied to autonomous agents. Specifically, treating trust and distrust as opposite poles of a single dimension can lead to significant measurement errors [7]. While AI-specific tools have begun to emerge [8, 9], many remain confined to specific domains, leaving a gap for highly generalizable instruments.

The modern consensus further integrates normative elements such as algorithmic fairness, transparency, and data privacy into the trust architecture [10, 11]. With the rise of generative and chat-based AI, human-AI interaction has acquired a human-like, relational character. Features such as personalized feedback and empathetic language enable users to form emotional bonds, suggesting that anthropomorphism is now a decisive factor in trust formation [4, 12].

This study addresses these theoretical gaps by developing a psychometrically robust tool that integrates functional, normative, and relational dimensions.

2 Related Literature

The evolution of trust measurement in technology has moved from a performance centric paradigm to a holistic socio technical perspective. Early models rooted in human computer and human automation interaction reduced trust to a cognitive evaluation of technical competence and predictability [2]. However, this reductionist view fails to

capture the affective and relational complexities inherent in modern AI interactions. Recognizing these limitations, [3] introduced conceptual expansion by incorporating comprehensibility and personal commitment as core components. Consequently, trust in AI is now recognized as a multidimensional construct shaped by perceived competence, benevolence, and risk [4–6].

Despite the proliferation of AI specific research, current measurement instruments face significant psychometric and theoretical challenges. First, emerging evidence suggests that the factor structures of traditional automation scales lack stability when applied to autonomous AI agents. Notably, treating trust and distrust as opposite poles of a single continuum can lead to measurement errors, necessitating their assessment as distinct constructs [7]. Second, while recent scale development efforts have emerged [8, 9], many remain domain specific or fail to integrate the normative dimensions of algorithmic fairness and data privacy [10, 11, 13].

Furthermore, the rise of generative AI has introduced a relational dimension previously absent in traditional systems. As chat-based systems employ human-like language and personalized feedback, users develop emotional bonds, making anthropomorphism a critical determinant of trust [4, 12]. Most existing scales fail to account for this human-like character as a primary factor in trust.

This study addresses these gaps by developing and validating a psychometrically robust instrument tailored to the educational context. By integrating fragmented dimensions, specifically perceived justice, ethics, and anthropomorphism, this research provides the necessary psychometric discrimination to accurately measure students' trust in AI.

3 Method

This study employed a systematic scale development design to construct and validate an instrument to measure user trust in AI. The development process adhered to the methodological framework established by [14], which encompasses defining the construct, item pooling, expert review, pilot testing, and rigorous factor analysis. This structured approach ensures both the theoretical alignment and the psychometric integrity of the final scale.

3.1 Participants

The study involved three distinct participant groups to support the various stages of validation and application. Data were collected from students enrolled at the Open Education Faculty of Anadolu University. Following the removal of invalid responses identified by control item checks, the final analysis included a total of 412 participants. The demographic characteristics, including gender distribution, age range, academic performance, and engagement levels with the learning management system, are detailed in Table 1.

Participant characteristics across the EFA, CFA, and application samples are summarized in Table 1. The samples were broadly comparable, with female participants forming the majority. Age composition differed slightly, with the EFA group skewing younger and the CFA/application groups including more middle-aged and older

Table 1: Demographic profile of the participants

		Group 1 (EFA)		Group 2 (CFA)		Group 3 (App.)	
Variable	Category	N	%	N	%	N	%
Gender	Female	254	61.65	159	51.79	72	61.02
	Male	158	38.35	148	48.21	46	38.98
Age	19–25	134	32.52	39	12.70	15	12.71
	25–31	75	18.20	33	10.75	6	5.08
	31–37	50	12.14	50	16.29	7	5.93
	37–43	35	8.50	37	12.05	20	16.95
	43–49	49	11.89	41	13.36	24	20.34
	49–55	32	7.77	44	14.33	23	19.49
	55–61	23	5.58	31	10.10	10	8.47
	61–67	9	2.18	23	7.49	10	8.47
	67–73	5	1.21	8	2.61	3	2.54
Grade Average	0–1	45	10.92	38	12.38	17	14.41
	1–2	121	29.37	77	25.08	33	27.97
	2–3	165	40.05	129	42.02	51	43.22
	3–4	81	19.66	62	20.20	17	14.41
LMS Login Count	0–30	196	47.57	144	46.91	60	50.85
	30–60	118	28.64	73	23.78	35	29.66
	60–90	46	11.17	35	11.40	12	10.17
	90–120	20	5.10	124	40.39	4	3.39
	>120	30	7.28	30	9.77	7	5.93
Number of LMS Material Usages	0–110	178	43.20	137	44.63	61	51.69
	110–220	121	29.37	81	26.38	24	20.34
	220–330	55	13.35	34	11.07	15	12.71
	330–440	25	6.07	20	6.51	9	7.63
	>440	33	8.01	34	11.07	9	7.63

learners. Group 3 profiles were similar across groups, and LMS engagement varied, supporting validation across diverse learner profiles in distance education.

3.1.1 Definition of the construct to be measured

The primary construct of this study is students' perceptions of trust in artificial intelligence. While traditional frameworks establish that technological trust dictates usage intentions and behaviours, the autonomous nature of AI necessitates an expansion beyond these classical models.

Synthesizing contemporary research [3, 7, 9, 15–20], this study treats trust as a complex multidimensional structure. A critical theoretical addition is the dimension of anthropomorphism, as evidence suggests that trust in AI is increasingly shaped by relational attributes such as empathy and human-like communication rather than mere functional accuracy. Consequently, an eight-dimensional framework was established to capture the breadth of the construct, including perceived fairness, comprehensibility, benefit, trust, risk, ethical and data privacy-based trust, anthropomorphism, and personal commitment.

3.1.2 Creation of the item pool

The second phase generated a comprehensive item pool to ensure balanced coverage of all theoretical dimensions. We conducted an extensive review of research on trust in AI, technology acceptance, data security, and ethical perceptions, and refined the items with input from experts in measurement, educational technology, psychology, and AI. To address a noted gap in human–AI interaction research, we added anthropomorphism as a distinct dimension. This process produced an initial pool of 47 items for content validation and psychometric testing.

3.1.3 Determination of the item format

Items were rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) to measure students' trust in AI. This format is widely used for assessing latent constructs such as attitudes and perceptions and supports comparability with prior trust research [21].

3.1.4 Examination of content validity via expert opinion

A multidisciplinary panel of 13 experts (measurement/evaluation, educational technology, psychology, and AI) reviewed the 47-item pool for relevance and clarity. Content validity was quantified using [22]'s CVR, with a critical value of .54 based on 13 experts. Items below the cutoff were removed, and wording was refined based on qualitative feedback. This process reduced the pool to 30 items with established content validity. The values obtained within this scope are presented in Table 2.

3.1.5 Implementation of the pilot study

The draft scale was piloted with 13 volunteer university students who were not included in the main study samples. Participants were asked to complete the scale and provide written feedback regarding item clarity, wording comprehensibility, and response format. Based on this feedback, minor revisions were made to the phrasing of several items to improve readability, and the response instructions were refined for greater clarity. No items were added or removed at this stage. These adjustments confirmed the instrument's readiness for full psychometric evaluation.

3.1.6 Statistical analysis of the items

An EFA was performed to determine the instrument's latent factor structure and internal reliability. The study sample included students from the Open Education Faculty of Anadolu University. The initial instrument consisted of 30 items plus a single control item to ensure data quality. Following the exclusion of participants who failed the control item check, the final analysis was conducted with 412 valid responses.

The suitability of the data for factor extraction was assessed using the Kaiser Meyer Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity [23]. The KMO value was calculated at .96, indicating excellent sampling adequacy. Bartlett's Test of Sphericity yielded significant results with $\chi^2(378) = 6376.37, p < .001$, confirming that the correlation matrix was appropriate for factoring.

Table 2: Item pool and CVR analysis

Items	Sources	N	S	SR	NS	CVR	Action	EFA	CFA
Perceived Justice									
AI provides responses without bias.	Developed by the authors	13	12	1	0	0.85	Retained	Retained	Retained
AI adopts an objective approach when generating responses.	Developed by the authors	13	10	3	0	0.54	Retained	Retained	Retained
AI provides responses based solely on data.	Developed by the authors	13	8	5	0	0.23	Removed	–	–
AI treats everyone equally when generating responses.	Developed by the authors	13	11	0	2	0.69	Retained	Retained	Retained
Perceived Understandability									
I am familiar with AI.	Jian et al. (2000); Perrig et al. (2023)	13	7	5	1	0.08	Removed	–	–
I know how AI works.	Madsen & Gregor (2000); Ribeiro et al. (2016)	13	5	7	1	-0.23	Removed	–	–
I understand the rationale behind the responses provided by AI.	Developed by the authors	13	6	4	3	-0.07	Removed	–	–
I know which data AI uses when generating responses.	Developed by the authors	13	10	3	0	0.54	Revised	Removed	–
Perceived Barriers									
The decision-making processes of AI are not transparent.	Nazaretsky et al. (2025)	13	8	5	0	0.23	Removed	–	–
It is unclear who is responsible for the decisions made by AI.	Nazaretsky et al. (2025)	13	9	4	0	0.38	Removed	–	–
The data sources used by AI are not clearly identified.	Nazaretsky et al. (2025)	13	8	4	1	0.23	Removed	–	–
The data collection methods employed by AI are not clearly specified.	Nazaretsky et al. (2025)	13	8	4	1	0.23	Removed	–	–
Perceived Usefulness									
AI provides recommendations that support my decision-making.	Madsen & Gregor (2000)	13	10	3	0	0.54	Retained	Retained	Retained
The services offered by AI have a low error rate.	Chi et al. (2021)	13	10	3	0	0.54	Retained	Retained	Retained
I believe that AI has the functions I need.	Wang et al. (2025)	13	10	3	0	0.54	Retained	Retained	Retained
The responses provided by AI meet my needs.	Dang & Li (2025)	13	10	3	0	0.54	Revised	Retained	Retained
AI provides personalized responses based on my level of learning.	Nazaretsky et al. (2025)	13	6	7	0	-0.08	Removed	–	–
By identifying my strengths and weaknesses, AI helps teachers assign tasks appropriate to me.	Nazaretsky et al. (2025)	13	6	7	0	-0.08	Removed	–	–
AI provides responses that meet my expectations.	Gulati et al. (2019); Wang et al. (2025)	13	10	3	0	0.54	Revised	Removed	–
Perceived Trust									
AI provides assistance when needed.	Gulati et al. (2019)	13	11	2	0	0.69	Retained	Retained	Retained
I believe that I can use AI effectively.	Chi et al. (2021); Madsen & Gregor (2000); Nazaretsky et al. (2025)	13	11	2	0	0.69	Retained	Retained	Retained
I consider AI's recommendations to be as valuable as those of a reliable person.	Nazaretsky et al. (2025)	13	10	2	1	0.54	Revised	Removed	–
As the amount of data used by AI increases, I trust its recommendations more.	Nazaretsky et al. (2025)	13	10	3	0	0.54	Retained	Removed	–

6

ARTICLE IN PRESS

Table 2: Item pool and CVR analysis (continued)

Items	Sources	N	S	SR	NS	CVR	Action	EFA	CFA
Perceived Trust									
The recommendations provided by AI constitute a reliable basis for my decision-making.	Nazaretsky et al. (2025)	13	12	1	0	0.85	Retained	Retained	Retained
I trust the responses generated by AI.	Wang et al. (2025); Gulati et al. (2019); Jian et al. (2000); Perrig et al. (2023)	13	11	2	0	0.69	Retained	Retained	Retained
AI can explain the reasons behind the results it produces.	Developed by the authors	13	9	2	2	0.38	Removed	–	–
AI operates in a reliable manner.	Madsen & Gregor (2000)	13	8	4	1	0.23	Removed	–	–
The responses provided by AI are consistent.	Jian et al. (2000); Perrig et al. (2023); Madsen & Gregor (2000)	13	10	3	0	0.54	Retained	Retained	Retained
Perceived Risk									
I am cautious about the potential risks of AI.	Jian et al. (2000); Perrig et al. (2023); Gulati et al. (2019)	13	12	1	0	0.85	Retained	Removed	–
AI can be misleading.	Jian et al. (2000); Perrig et al. (2023)	13	8	5	0	0.23	Removed	–	–
Interacting with AI is risky.	Gulati et al. (2019)	13	7	4	0	0.08	Removed	–	–
AI may lead to unexpected outcomes.	Developed by the authors	13	11	2	0	0.69	Retained	Retained	Removed
Ethics- and Data Privacy-Based Trust									
The responses generated by AI raise ethical concerns.	Nazaretsky et al. (2025); Gulati et al. (2019); Jian et al. (2000); Perrig et al. (2023)	13	10	3	0	0.54	Retained	Retained	Removed
The data used by AI create privacy concerns for me.	Nazaretsky et al. (2025)	13	12	1	0	0.85	Retained	Retained	Retained
AI acts in accordance with ethical principles.	Floridi & Cowls (2022); Nazaretsky et al. (2025)	13	10	2	1	0.54	Retained	Retained	Retained
The intentions and funding sources of AI providers are not transparent.	Rzepka et al. (2022)	13	7	5	1	0.08	Removed	–	–
AI avoids misusing user data.	Developed by the authors	13	10	3	0	0.54	Retained	Retained	Retained
Anthropomorphism									
Interacting with AI provides a human-like experience.	Developed by the authors	13	12	1	0	0.85	Retained	Retained	Retained
AI uses human-like humor in its responses.	Developed by the authors	13	11	2	0	0.69	Retained	Retained	Retained
AI employs an empathetic tone in its responses.	Developed by the authors	13	12	1	0	0.85	Retained	Retained	Retained
AI possesses human-like qualities such as politeness and understanding.	Developed by the authors	13	10	3	0	0.54	Retained	Retained	Retained
Personal Attachment									
I use AI tools to personalize my learning process.	Nazaretsky et al. (2025)	13	10	3	0	0.54	Retained	Retained	Retained
I enjoy using AI.	Madsen & Gregor (2000)	13	11	2	0	0.69	Retained	Retained	Retained
I consider the absence of AI to be a deficiency.	Madsen & Gregor (2000)	13	10	3	0	0.54	Revised	Removed	–
I feel a sense of attachment to AI.	Chi et al. (2021); Madsen & Gregor (2000)	13	8	4	1	0.23	Removed	–	–
I prefer AI when making decisions.	Madsen & Gregor (2000)	13	8	5	0	0.23	Removed	–	–
The responses provided by AI are compatible with my way of thinking.	Madsen & Gregor (2000)	13	10	2	1	0.54	Retained	Removed	–

Note. N = Number of experts; S = Essential; SR = Should be revised; NS = Not suitable; CVR = Content validity ratio; EFA = Exploratory factor analysis; CFA = Confirmatory factor analysis.

To further evaluate the psychometric quality of the items, item total correlations were computed, and internal consistency was assessed using Cronbach's alpha. These preliminary statistical indicators provided the necessary empirical justification to proceed with the identification of the underlying sub dimensions of the trust scale.

3.1.7 Testing construct validity and reliability

The five-factor structure identified through the exploratory phase was subsequently tested for structural stability using Confirmatory Factor Analysis, an approach increasingly adopted in educational research to validate complex latent constructs through rigorous SEM-based procedures [24, 25]. The instrument was administered to a new sample of 336 students from the Open Education Faculty of Anadolu University. Following the removal of participants who failed the attention check items, the final analysis proceeded with 307 valid responses.

The reliability of the scale demonstrated high internal consistency as evidenced by Cronbach's alpha ($\alpha = .94$) and McDonald's omega ($\omega = .94$) coefficients. To ensure the robustness of the latent constructs, validity analyses examined composite reliability and average variance extracted across the factors. Composite reliability values ranged from .78 to .91, while average variance extracted values ranged from .56 to .70, both exceeding the recommended thresholds for convergent validity. Furthermore, the data distribution was assessed for normality, with skewness values between -1.17 and .68 and kurtosis values between -1.00 and 1.62, indicating that the items were suitable for maximum likelihood estimation. These results collectively confirm that the five-factor structure of the scale is theoretically sound and psychometrically robust for assessing trust in artificial intelligence.

3.1.8 Development of the final scale

The confirmatory factor analysis supported the stability of the five-factor structure. During final refinement, items with weak loadings or limited contribution to their intended constructs were removed to strengthen discrimination and theoretical coherence.

To ensure transparency, key decision points across all stages of the item reduction process are summarized. The initial pool of 47 items was reduced to 30 following expert panel review, with items falling below the CVR threshold of .54 excluded. EFA subsequently eliminated 7 items (2 due to cross-loadings and 5 due to insufficient factor loadings or non-significant correlation matrix values), yielding 23 items. In the final CFA phase, 2 additional items were removed for failing to meet the .50 standardized loading threshold, resulting in the 21-item final scale.

The resulting AI Trust Scale comprises 21 items across five sub- dimensions (Appendix 1). The scale demonstrates acceptable evidence of validity and reliability for assessing students' trust in AI systems. Its concise format improves practical usability for research and educational settings while still capturing key aspects of the socio-technical relationship between learners and AI agents.

3.1.9 Implementation of the developed scale

The findings from the implementation phase indicate that, within the present sample, trust in AI was not significantly associated with GPA or LMS engagement frequency, suggesting that trust scores were not meaningfully influenced by these particular variables. An independent sample t-test conducted with 118 students revealed no statistically significant difference between the trust scores of male and female participants [$t(116) = -0.166$; $p = .868$], with a negligible effect size. This outcome is consistent with the previously established metric invariance, confirming that the scale provides equivalent measurement across genders and is free from gender-based bias.

Furthermore, Spearman rho correlation analysis indicated that trust in AI was not significantly associated with overall grade point average ($r = -.161; p = .082$) or the number of LMS logins ($r = -.055, p = .554$). These results suggest that trust is not merely a byproduct of academic success or routine exposure to the system. Instead, students' trust in AI constitutes a distinct psychological evaluation that remains consistent regardless of their academic standing or technical engagement habits. This independence further validates the scale as a specialized tool for measuring trust as a primary socio technical construct.

4 Results

4.1 EFA

The item reduction process was conducted to maintain a parsimonious set of indicators that represent the core construct validity of the scale while minimizing participant cognitive load [26]. Initial diagnostics confirmed the superior suitability of the data for factor analysis, with a KMO value of .96 and a significant Bartlett's Test of Sphericity, $\chi^2(378) = 6376.37, p < .001$. These metrics demonstrate that the inter-item correlations are sufficiently robust for factor extraction.To determine the optimal number of factors, Scree Plot analysis was corroborated by parallel analysis. The convergence of these methods, along with eigenvalues exceeding 1.00, statistically supported a five-factor structure. This multidimensional configuration ensures that the scale captures the complexity of trust in artificial intelligence without compromising efficiency. The analysis results are shown in Figure 1.

Analysis of the eigenvalues confirms a five-factor solution, with all retained factors exceeding the 1.00 threshold. The primary factor accounts for 41.96% of the total variance, followed by four additional factors contributing 6.34%, 5.64%, 5.18%, and 4.09% respectively. Collectively, this structure explains 63.20% of the total variance, meeting the established criteria for a robust measurement model in the social sciences. The scree plot, generated based on the 30 items entered into the analysis, provided visual confirmation of this structure, with a clear inflection point emerging after the fifth factor. Factor extraction was performed using the Generalized Least Squares method. To account for the theoretical interconnectedness of trust dimensions, the Direct Oblimin oblique rotation technique was applied [27]. Post rotation, the factors showed a balanced distribution of variance, with the primary factor explaining 40.50% and the remaining factors ranging from 2.64% to 4.95%. These analyses, conducted via SPSS

9

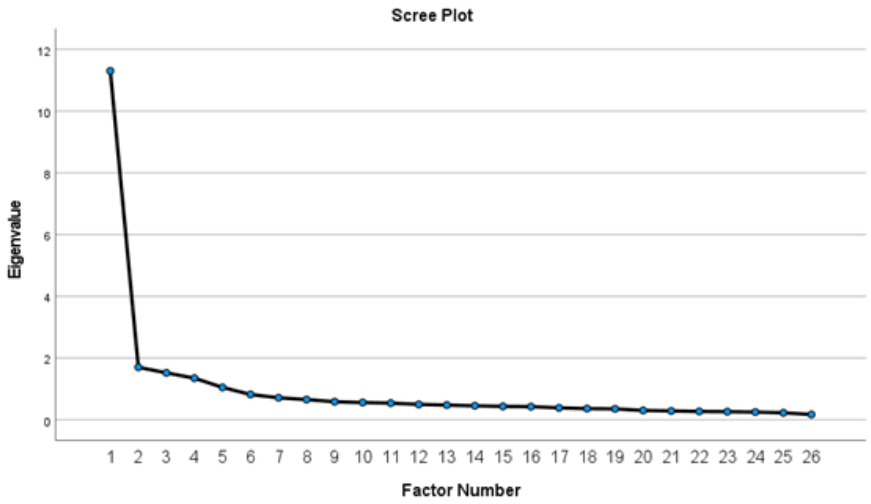

Fig. 1: Scree plots

26, verify that the five-factor configuration is the most parsimonious and theoretically sound representation of the data. Detailed factor loadings and item distributions are presented in Table 3.

The factor extraction and rotation procedures resulted in a stable five-factor solution that converged in 11 iterations. To ensure the reliability of the latent structure, items were subjected to rigorous purification based on established psychometric criteria. Specifically, items that exhibited cross loadings on multiple factors or displayed weak factor loadings below .40 were systematically eliminated. This threshold aligns with the recommendations of [23, 28], who suggest that loadings above .40 are ideal for ensuring that an item contributes meaningfully to its assigned factor. Several items within Factor 2 (Items 1, 2, and 3) and Factor 5 (Items 6, 14, 15, and 16) yielded negative factor loadings. These items do not contain reverse-worded statements; rather, the negative signs reflect the arbitrary direction of factor extraction in exploratory analysis. As the sign of a factor loading does not affect its magnitude or substantive interpretation, these items were retained based on the absolute value of their loadings, consistent with standard psychometric practice [29]. Based on these criteria, Items 9 and 12 were removed due to significant cross loadings that threatened the discriminant validity of the sub dimensions. Furthermore, Items 4, 13, 19, 29, and 30 were excluded because they failed to meet the .40 loading threshold. Item 17 was also removed because it did not achieve a statistically significant value within the correlation matrix. Following these refinements, the scale was reduced from 31 items including the control item to a final set of 23 items. This purification process enhances the instrument's structural integrity and ensures that each remaining item provides a distinct and robust measurement of the intended trust dimensions.

Table 3: Results of EFA

Factors	Items	N	X	SD	Skewness	Kurtosis	1	2	3	4	5
Perceived Justice	I1	412	3,10	1,21	-,15	-,99	-0,76				
	I2	412	3,17	1,14	-,27	-,87	-0,90				
	I3	412	3,28	1,16	-,31	-,91	-0,71				
Perceived Benefit and Usability	I5	412	3,64	1,010	-,97	-,69		0,52			
	I7	412	3,28	1,010	-.72	-,23		0,48			
	I8	412	3,40	,960	-,84	,34		0,55			
	I9	412	3,78	,935	-1,17	1,49		0,64			
	I11	412	3,74	,954	-,96	1,05		0,73			
	I27	412	3,71	1,024	-1,06	,89		0,69			
	I28	412	3,74	1,042	-,98	,74		0,61			
Perceived Trust	I6	412	2,63	1,063	,00	-,96			-0,47		
	I14	412	3,05	1,092	-,32	-,67			-0,41		
	I15	412	3,02	1,000	-,41	-,52			-0,53		
	I16	412	2,90	1,056	-,35	-,81			-0,48		
Ethics and Data Privacy	I18	412	2,10	1,002	,78	,06				0,58	
	I19	412	2,73	1,085	,05	-,69				0,66	
	I20	412	2,25	1,126	,66	-,36				0,66	
	I21	412	2,89	1,060	-,26	-,53				0,56	
	I22	412	2,68	1,013	-,08	-,36				0,54	
Anthropomorphism	I23	412	2,76	1,236	-,02	-1,25					0,63
	I24	412	3,08	1,164	-,37	-,94					0,83
	I25	412	3,34	1,136	-,61	-,51					0,67
	I26	412	3,19	1,225	-,46	-,85					0,61

4.2 CFA

The structural integrity of the factor model identified in the exploratory phase was tested using a Confirmatory Factor Analysis on a separate and independent sample. This methodological progression ensures that the latent structure is not sample specific but represents a stable and generalizable measurement framework [23, 24]. For the execution of these analyses, AMOS 22 and SmartPLS software were utilized to provide comprehensive cross validation of the results. The validation process involved a rigorous examination of model fit indices, standardized factor loadings, and overall construct validity. By evaluating the degree of correspondence between the hypothesized five factor structure and the observed empirical data, the analysis confirms the robustness of the scale. The specific goodness of fit indices employed, alongside their established theoretical reference ranges, are presented in Table 4.

According to the results of the CFA analysis presented in Table 4, the chi-square value was statistically significant ($\chi^2 = 382.923, df = 176, p < .001$). However, since the chi-square statistic is sensitive to sample size [24], other fit indices are taken into consideration when evaluating model fit. The χ^2/df ratio was calculated as 2.176, which falls within the range of $0 \leq \chi^2/df \leq 3$ suggested by [24], indicating a perfect fit. The values for CFI (.95), TLI (.94), IFI (.95), and NFI (.91) meet the $\geq .90$ criterion

11

Table 4: Model fit indices for CFA

	Fit Criteria[1]		Evaluation		
Indices	Value	Acceptable Fit	Perfect Fit	Reference	Decision
χ^2/df	2.17	$3 \leq \chi^2/df \leq 5$	$0 \leq \chi^2/df < 3$	Kline (2016)	Perfect fit
CFI	.95	$0.90 \leq CFI < 0.95$	$0.95 \leq CFI \leq 1$	Hu & Bentler (1999)	Perfect fit
TLI	.94	$0.90 \leq TLI < 0.95$	$0.95 \leq TLI \leq 1$	Brown (2015)	Acceptable fit
GFI	.89	$0.80 \leq GFI < 0.90$	$0.90 \leq GFI \leq 1$	Brown (2015)	Acceptable fit
AGFI	.86	$0.80 \leq AGFI < 0.90$	$0.90 \leq AGFI \leq 1$	Brown (2015)	Acceptable fit
IFI	.95	$0.90 \leq IFI < 0.95$	$0.95 \leq IFI \leq 1$	Kline (2016)	Perfect fit
NFI	.91	$0.90 \leq NFI < 0.95$	$0.95 \leq NFI \leq 1.00$	Bentler & Bonett (1980)	Acceptable fit
RMSEA	.06	$0.05 \leq RMSEA \leq 0.10$	$0.00 \leq RMSEA < 0.05$	Hu & Bentler (1999)	Acceptable fit

Fit criteria indicate commonly used thresholds for evaluating model fit in structural equation modeling.

[1]CFI: Comparative Fit Index; TLI: Tucker–Lewis Index; GFI: Goodness-of-Fit Index; AGFI: Adjusted Goodness-of-Fit Index; IFI: Incremental Fit Index; NFI: Normed Fit Index; RMSEA: Root Mean Square Error of Approximation.

for good model fit as recommended by [24, 30]. Finally, the RMSEA value was found to be .06, which falls within the range of $0.05 \leq \text{RMSEA} < 0.10$ proposed by [30], indicating an acceptable level of model fit.

The collective evidence from the confirmatory phase indicates that the measurement model demonstrates a superior fit, aligning consistently with the rigorous statistical thresholds established in the literature. These findings verify that the five-factor structure identified in the exploratory analysis provides an accurate and stable representation of the data, thereby reinforcing the instrument's construct validity.

The robust performance of the goodness of fit indices confirms that the latent dimensions effectively capture the complexities of trust in artificial intelligence. Consequently, the model provides a theoretically sound framework for examining the interconnectedness of these constructs. The standardized factor loadings and the specific inter factor correlation values, which further illustrate the internal consistency and discriminant validity of the five-dimensional structure, are presented in Figure 2.

The structural evaluation presented in Figure 2 reveals that the standardized factor loadings for all items range from .50 to .91. These robust values indicate that each item strongly represents its respective latent construct, exceeding the common psychometric requirement for meaningful factor contribution.

To further optimize the model fit, a limited number of error covariance modifications were implemented. These adjustments were justified by the need to account for shared variance among items with similar conceptual content within the same sub-dimension. Specifically, error covariances were permitted between Item 27 and Item 28, and between Item 7 and Item 10 within the Perceived Benefit and Usability factor, as these items share overlapping content related to functional utility and user experience. Additionally, an error covariance was introduced between Item 14 and Item 16 within the Perceived Trust factor, reflecting their conceptual proximity in assessing the reliability of AI responses. All modifications were confined to items within the same sub-dimension and were theoretically justified prior to implementation [31, 32] This refinement process enhanced the overall goodness of fit without compromising the theoretical integrity of the scale. Collectively, these results provide definitive evidence of strong construct validity and confirm that the five-dimensional structure is a psychometrically sound representation of trust in artificial intelligence. Detailed statistical outputs for the Confirmatory Factor Analysis are provided in Table 5.

The skewness and kurtosis values provided in Table 5 were analyzed to evaluate the distributional characteristics of the scale items. Skewness values ranging from -1.17 to .68 and kurtosis values between -1.00 and 1.62 fall within the acceptable parameters defined by [24, 33]. These results confirm that the dataset satisfies the assumption of univariate normality, justifying the use of maximum likelihood estimation in the confirmatory factor analysis. To assess the discriminatory power of the items, corrected item total correlation values were calculated. The resulting coefficients, ranging from .41 to .80, exceed the thresholds suggested by [14, 23]. These values indicate that each item contributes significantly to its respective factor and the overall scale score, demonstrating high conceptual consistency and internal validity. The standardized factor loadings from the confirmatory phase were evaluated against a minimum threshold of .50. Items I18 and I19 yielded loadings of .49 and .47 respectively and were

Table 5: Results of CFA

Factors	Items	N	X	SD	$\uparrow 27\% + SD$	$\downarrow 27\% + SD$	CITC	Sk	K	Loading
Perceived Justice	I1	307	3.21	1.13	$3.99 \pm .87$	2.35 ± 1.15	.57	-.37	-.75	.83
	I2	307	3.25	1.08	$4.05 \pm .70$	2.33 ± 1.10	.62	-.50	-.54	.89
	I3	307	3.38	1.19	$4.23 \pm .84$	2.38 ± 1.17	.62	-.46	-.79	.78
Perceived Benefit and Usability	I5	307	3.57	.99	$4.32 \pm .54$	2.54 ± 1.06	.74	-.98	.59	.79
	I7	307	3.32	.98	$4.15 \pm .47$	$2.33 \pm .95$	.74	-.73	-.07	.83
	I8	307	3.43	.96	$4.24 \pm .53$	2.45 ± 1.01	.79	-.80	.51	.86
	I10	307	3.73	.92	$4.41 \pm .52$	2.78 ± 1.04	.79	-1.17	1.62	.88
	I11	307	3.63	.96	$4.34 \pm .61$	2.93 ± 1.08	.63	-.78	.50	.69
	I27	307	3.63	.99	$4.27 \pm .61$	2.83 ± 1.18	.60	-.95	.70	.62
	I28	307	3.52	1.11	$4.43 \pm .68$	2.43 ± 1.08	.70	-.64	-.18	.75
Perceived Trust	I6	307	2.79	1.08	$3.46 \pm .88$	$1.98 \pm .99$	.50	-.07	-.80	.55
	I14	307	3.17	1.03	$4.06 \pm .61$	$2.10 \pm .91$	.76	-.47	-.30	.82
	I15	307	3.03	.99	$3.99 \pm .57$	$1.96 \pm .72$	.80	-.25	-.53	.91
	I16	307	3.08	1.04	$3.83 \pm .76$	$2.11 \pm .99$	.65	-.46	-.37	.74
Ethics and Data Privacy	I20	307	2.30	1.07	2.91 ± 1.14	$1.74 \pm .84$	.41	.68	-.14	.50
	I21	307	2.91	1.00	$3.68 \pm .71$	$2.02 \pm .95$	.68	-.31	-.33	.85
	I22	307	2.79	1.03	$3.70 \pm .76$	$1.91 \pm .85$	.67	-.06	-.42	.85
Anthropomorphism	I23	307	2.99	1.19	$3.94 \pm .79$	1.91 ± 1.01	.67	-.27	-1.00	.76
	I24	307	3.11	1.13	$3.96 \pm .76$	2.16 ± 1.04	.60	-.34	-.80	.76
	I25	307	3.39	1.02	$4.16 \pm .67$	2.67 ± 1.07	.58	-.51	-.34	.76
	I26	307	3.31	1.12	$4.07 \pm .73$	2.40 ± 1.09	.59	-.57	-.43	.73

CITC = corrected item-total correlation; Sk = skewness; K = kurtosis; Loading = standardized factor loading.

14

ARTICLE IN PRESS

consequently removed to ensure the structural integrity of the factors [23, 24]. Furthermore, an analysis of modification indices revealed shared error variance between specific items with overlapping conceptual content. Following the recommendations of [31, 32], error covariances were correlated between these theoretically aligned items. This strategic modification accounts for shared method variance rather than model error, thereby improving fit indices while preserving the construct's structural validity. These refinements confirm that the remaining items strongly represent their factors, resulting in a robust model fit as detailed in Table 6.

Table 6: Validity and reliability analysis results

Factors	Number of Items	Cronbach's α	McDonald's ω	CR	AVE
Perceived Justice	3	.87	.87	.87	.70
Perceived Benefit and Usability	7	.92	.92	.91	.61
Perceived Trust	4	.85	.85	.85	.60
Ethics and Data Privacy	3	.76	.78	.78	.56
Anthropomorphism	4	.84	.84	.84	.57
Total	21	.94	.94		

CR = Composite Reliability; AVE = Average Variance Extracted.

Following the confirmatory factor analysis, internal consistency coefficients were calculated for each factor. Current methodological recommendations suggest reporting both Cronbach's α and McDonald's ω and supporting reliability decisions primarily based on ω [34]. As shown in Table 6, according to Cronbach's α and McDonald's ω analyses, the following values were obtained: for Perceived Justice, $\alpha = .87$ and $\omega = .87$ for Perceived Justice; $\alpha = .92$ and $\omega = .92$ for Perceived Benefit and Usability; $\alpha = .85$ and $\omega = .85$ for Perceived Trust; $\alpha = .76$ and $\omega = .78$ for Ethics and Data Privacy; and $\alpha = .84$ and $\omega = .84$ for Anthropomorphism. For the overall scale, $\alpha = .94$ and $\omega = .94$ were calculated. These findings indicate that all factors, and the overall scale, demonstrate good to very good levels of internal consistency (mostly $\geq .85$), confirming that the scale is reliable [33]. Furthermore, the very close correspondence between Cronbach's α and McDonald's ω values further supports the consistency and stability of the measurement instrument.

The psychometric robustness of the scale was further evaluated through composite reliability and convergent validity indices. According to [23, 35], a Composite Reliability (CR) value of .70 or higher signifies adequate internal consistency. In this study, CR values ranged from .78 to .91, demonstrating satisfactory reliability across all latent constructs. Furthermore, the Average Variance Extracted (AVE) values exceeded the .50 threshold for all factors, ranging from .56 to .70. These results indicate that the constructs explain a significant proportion of the variance in their respective indicators, thereby establishing strong evidence for convergent validity [36]. To ensure that the measurement instrument effectively distinguishes between different constructs, a discriminant validity analysis was conducted using the Heterotrait Monotrait (HTMT)

ratio [37]. The HTMT ratio provides a more sensitive assessment of discriminant validity compared to traditional methods by comparing inter construct correlations to intra construct correlations. Following the criteria established by [37], all HTMT values remained below the .90 threshold for theoretically related constructs and below .85 for distinct dimensions. These findings confirm the discriminant validity of the five-factor structure, as detailed in Table 7.

Table 7: Discriminant validity matrix

	A	EDP	PBU	PJ	PT
Anthropomorphism (A)					
Ethics and Data Privacy (EDP)	0.66				
Perceived Benefit and Usability (PBU)	0.77	0.68			
Perceived Justice (PJ)	0.50	0.64	0.62		
Perceived Trust (PT)	0.66	0.79	0.84	0.67	

Note. A = Anthropomorphism; EDP = Ethics and Data Privacy; PBU = Perceived Benefit and Usability; PJ = Perceived Justice; PT = Perceived Trust.

The HTMT ratios presented in Table 7 confirm that the measurement instrument achieves a high level of differentiation among its latent constructs. All obtained values remain below the rigorous .90 threshold, indicating that each sub dimension captures a distinct facet of trust in artificial intelligence. The highest HTMT coefficient was observed between Perceived Benefit and Usability (PBU) and Perceived Trust (PT) at .84, suggesting a strong but distinct theoretical relationship. Conversely, the lowest value was identified between Anthropomorphism (A) and Perceived Justice (PJ) at .50, highlighting a clear divergence between relational perceptions and ethical evaluations. These results provide empirical evidence that the constructs are sufficiently independent and that the scale effectively distinguishes between the various psychological dimensions it is designed to measure. Consequently, the findings verify that each factor within the five-dimensional model is assessed with high reliability and validity. Although all HTMT values remained below the recommended thresholds, the relatively higher coefficient observed between Perceived Benefit and Usability and Perceived Trust (HTMT = .84) warrants careful interpretation. These two constructs share conceptual proximity in the trust literature, as perceived usability is frequently theorized as an antecedent of trust rather than a wholly independent dimension. Nevertheless, the empirical distinctiveness of these factors was supported by the HTMT criterion, and their separation is consistent with the theoretical framework guiding this study.

4.2.1 Measurement invariance via multi-group CFA across genders

To test whether the structural validity of the developed scale is invariant across gender groups (male and female), a Multiple-Group Confirmatory Factor Analysis (MG-CFA) was conducted. The analysis was carried out in four progressively constrained stages: configural invariance, metric invariance, structural covariance invariance, and strict invariance. The results obtained are presented in Table 8.

16

Table 8: Measurement invariance results across genders for the five-factor model

Model	χ^2 (CMIN)	*df*	$\Delta\chi^2$	Δdf	*p*	RMSEA	CFI	TLI	SRMR
Configural Invariance	1631.376	916	–	–	–	0.047	0.944	0.933	0.0635
Metric Invariance	1645.935	932	14.559	16	0.557	0.046	0.944	0.936	0.0632
Structural Covariances	1670.046	947	38.670	31	0.162	0.046	0.942	0.936	0.0757
Strict Invariance	1736.023	971	104.647	55	0.000	0.048	0.932	0.930	0.0749

χ^2 = chi-square statistic; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; SRMR = Standardized Root Mean Square Residual.

17

Upon examining Table 8, the configural invariance model—testing whether the basic factor structure is the same across both gender groups—was first evaluated. The results indicated that the model demonstrated good fit to the data ($\chi^2/df = 1.78, \text{RMSEA} = .047, \text{CFI} = .944, \text{SRMR} = .0635$). The metric invariance model, in which factor loadings were constrained to be equal across groups, was subsequently tested. The findings revealed no statistically significant difference between the configural and metric models ($\chi^2/df = 1.78$, RMSEA = .047, CFI = .944, SRMR = .0635). Furthermore, the ΔCFI value for the metric model was below the .01 criterion (.000), and the RMSEA value (.046) remained stable [38]. These results indicate that metric invariance across gender groups has been established.

At the structural covariance invariance stage, the model demonstrated acceptable fit ($\Delta\text{CFI} = .002$, $p = .162$). The strict invariance model, however, yielded a ΔCFI of .010 alongside a statistically significant chi-square difference (p ¡ .001), suggesting that while the value remains at the boundary of the recommended criterion, full strict invariance could not be unequivocally claimed. Given that metric invariance was firmly established ($\Delta\text{CFI} = .000$, $p = .557$), meaningful comparisons between gender groups remain statistically justified [39, 40].

4.2.2 Post-scale development implementation

Following the establishment of measurement invariance across gender groups, an independent sample t-test was conducted to evaluate whether student trust in artificial intelligence differs by gender. The results of this analysis are detailed in Table 9.

Table 9: Independent samples t-test results for Trust by gender

Variable	Group	*N*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Trust	Male	46	66.95	15.20	-0.166	116	.868	-0.031
	Female	72	67.38	12.78				

The independent samples t-test indicated that there was no statistically significant difference between male ($M = 66.95$, $SD = 15.20$) and female ($M = 67.38$, $SD = 12.78$) students' total trust scores [$t(116) = -0.166$, $p = .868$]. The calculated effect size ($d = -0.031$) suggests that any observed difference between the groups is negligible. This finding is fully consistent with the established metric invariance ($p = .557$) and confirms that the scale provides equivalent measurement across gender groups.

To examine the relationship between student trust in AI and academic performance variables, a Spearman rho correlation analysis was performed. Specifically, the analysis investigated correlations with overall grade point average and the frequency of LMS logins. The results are presented in Table 10.

The analysis revealed no statistically significant relationship between AI trust levels and grade point average (r = -0.161; p = .082) or the number of LMS logins (r = -0.055; p = .554). These findings indicate that, in the present sample, trust in AI was not

18

Table 10: Correlation analysis results among the variables

Variable	1	2	3
1. AI trust	1		
2. Grade point average	-0.161	1	
3. Number of LMS logins	-0.055	0.255**	1

**$p < .01$.

significantly associated with academic performance or system engagement frequency. While this pattern is consistent with the view that trust may function as a relatively independent construct, the absence of significant correlations alone does not constitute evidence of construct distinctiveness. Future research employing designs that explicitly examine relationships with theoretically proximal constructs would provide stronger support for this interpretation.

Interestingly, a low but statistically significant positive relationship was identified between grade point average and LMS login frequency (r = .255; p = .005). While this indicates that high achieving students interact more frequently with the LMS, it also demonstrates that such behavioural patterns do not directly influence or determine their trust perceptions regarding AI. This reinforces the value of the scale in capturing a unique psychological dimension that is not confounded by academic performance or system usage.

5 Discussion

The AI Trust Scale validated in this study provides a robust five factor structure that represents a significant departure from traditional unidimensional models. While [9] identified four factors in educational contexts and others like [15] or [41] proposed three-dimensional frameworks for robotics and healthcare, the present study reveals a more nuanced latent structure. By isolating perceived justice and anthropomorphism as distinct factors, this scale captures the unique socio technical nature of generative AI that previous automation-based instruments often overlook [42].

A critical finding of this research is the relative divergence between perceived benefit and ethical trust. Although participants recognized the high utility and functionality of AI, their scores regarding ethics and data privacy were notably more cautious. This discrepancy confirms that trust is not a linear outcome of system performance [1, 5]. In AI-based learning environments, students may perceive a tool as highly useful yet remain skeptical of its underlying algorithmic transparency [9]. This suggests that trust operates as a deep psychosocial evaluation rather than a simple reaction to instrumental effectiveness.

These results demand a critical reconsideration of established technology acceptance frameworks. Traditionally, the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) have prioritized performance expectancy and ease of use. However, the current study demonstrates that in

19

the context of autonomous systems, trust serves as a foundational evaluative mechanism that may precede or even govern perceived usefulness. By integrating dimensions such as data privacy and algorithmic reliability, this research provides empirical support for calls to evolve UTAUT to include trust as a central determinant in AI adoption.

The significance of this study lies in its provision of a sensitive analytical framework for educational practitioners. Unlike general scales developed for human robot interaction [15], this instrument is tailored to the opaque decision-making processes of online educational AI. It allows institutions to move beyond monitoring mere usage rates and instead assess the quality of the user relationship. Identifying specific trust vulnerabilities enables developers to strengthen explainability and accountability mechanisms, which are essential for transforming temporary usage into sustained and durable trust.

Despite these contributions, the single country context and student-based sample require caution regarding broad generalization. Future research should employ longitudinal designs to observe how trust evolves as students move from initial novelty to long term AI integration. Furthermore, as AI governance matures, future iterations of the scale should explicitly incorporate emerging constructs such as algorithmic explainability to maintain its predictive power in a rapidly shifting technological landscape.

6 Conclusion and Implications

This study establishes a psychometrically validated multidimensional framework for assessing students' trust in artificial intelligence. The five-factor structure, encompassing perceived justice, benefit and usability, perceived trust, ethics and data privacy, and anthropomorphism, offers a significant advancement over existing unidimensional or automation-based tools. By validating this scale through rigorous exploratory and confirmatory factor analyses, the research confirms that trust in autonomous educational systems is a complex construct where functional utility and normative concerns, such as algorithmic fairness, are deeply intertwined.

The practical implications of these findings for educational institutions are twofold. First, the scale serves as a diagnostic instrument for identifying specific trust deficits within digital learning environments. Administrators can utilize these metrics to move beyond surface level adoption rates and understand the underlying psychosocial barriers to AI integration. Second, the findings provide a roadmap for AI developers to prioritize transparency and ethical governance. The divergence between perceived benefit and ethical trust suggests that high performance alone cannot sustain user acceptance. Instead, technical competence must be reinforced by robust data protection and explainability mechanisms to foster a reliable user experience.

From a theoretical perspective, this research challenges traditional technology acceptance models by positioning trust as a multidimensional regulator of human AI interaction. By integrating anthropomorphism and perceived justice, the scale accounts for the increasingly relational nature of generative AI. This provides a foundation for future scholars to explore how these social dimensions influence long term academic engagement and decision-making processes.

20

All in all, as AI shifts from a passive tool to an agentic participant in the classroom, so our understanding of trust must evolve as well. This study underscores that trust is not a static state to be achieved, but a dynamic and fragile equilibrium between human vulnerability and machine autonomy. If educational institutions fail to systematically monitor and nurture this trust, they risk creating a landscape of digital alienation where the efficiency of AI comes at the cost of student agency and institutional integrity. The proliferation of intelligent systems should not be measured by the sophistication of the code, but by the degree to which users can safely and critically rely on the algorithms that increasingly shape their intellectual growth.

Ethical approval declarations (only required where applicable) Any article reporting experiment/s carried out on (i) live vertebrate (or higher invertebrates), (ii) humans or (iii) human samples must include an unambiguous statement within the methods section that meets the following requirements:

Acknowledgements. Not applicable.

Funding. The authors did not receive support from any organization for the submitted work.

Data availability. Data will be made available from the corresponding author on reasonable request.

7 Declarations

Ethics approval and consent to participate. This study was approved by the Anadolu University, Human Research Ethics Committee, Türkiye. All procedures followed established ethical standards, and informed verbal and written consent was obtained from participants in accordance with the Declaration of Helsinki (1964). Informed consent was obtained from all participants on a voluntary basis after clearly explaining the study's purpose, confidentiality measures, and anonymous nature. Participants faced no risks or compensation, and returning the completed questionnaire constituted informed consent.

Consent for publication. Not applicable.

Competing interests. The authors declare no competing interests.

Authors' Disclosures Language Translation and Localization. For the translation and localization of content, [DeepL and ChatGPT 5.2, March 2026] was employed. Human translators subsequently reviewed and adjusted the translations to ensure accuracy, cultural appropriateness, and contextual relevance.

Appendix A AI Trust Scale Items

This appendix presents the final items of the AI Trust Scale used in this study.

21

Factor	Item	Statement
Perceived Justice	1	AI provides responses without bias.
Perceived Justice	2	AI adopts an objective approach when generating responses.
Perceived Justice	3	AI treats everyone equally when generating responses.
Perceived Benefit and Usability	4	AI provides recommendations that support my decision-making.
Perceived Benefit and Usability	5	I believe that AI has the functions I need.
Perceived Benefit and Usability	6	The responses provided by AI meet my needs.
Perceived Benefit and Usability	7	AI assists me when I need it.
Perceived Benefit and Usability	8	I believe that I can use AI effectively.
Perceived Benefit and Usability	9	I use AI tools to personalize my learning process.
Perceived Benefit and Usability	10	I enjoy using AI.
Perceived Trust	11	The services offered by AI have a low error rate.
Perceived Trust	12	The recommendations provided by AI constitute a reliable basis for my decision-making.
Perceived Trust	13	I trust the responses generated by AI.
Perceived Trust	14	The responses provided by AI are consistent.
Ethics and Data Privacy	15	The data used by AI create privacy concerns for me.
Ethics and Data Privacy	16	AI acts in accordance with ethical principles.
Ethics and Data Privacy	17	AI avoids misusing user data.
Anthropomorphism	18	Interacting with AI provides a human-like experience.
Anthropomorphism	19	AI uses human-like humor in its responses.
Anthropomorphism	20	AI employs an empathetic tone in its responses.
Anthropomorphism	21	AI possesses human-like qualities such as politeness and understanding.

Note. Appendix 1 presents the final items of the AI Trust Scale used in this study.

References

[1] Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Acad. Manage. Rev. **20**(3), 709–734 (1995) https://doi.org/10.2307/258792

[2] McKnight, D.H., Carter, M., Thatcher, J.B., Clay, P.F.: Trust in a specific technology: An investigation of its components and measures. ACM Trans. Manage. Inf. Syst. **2**(2), 1–25 (2011) https://doi.org/10.1145/1985347.1985353

[3] Madsen, M., Gregor, S.: Measuring human-computer trust. In: Proceedings of the 11th Australasian Conference on Information Systems, vol. 53, pp. 6–8 (2000)

[4] Glikson, E., Woolley, A.W.: Human trust in artificial intelligence: Review of empirical research. Acad. Manage. Ann. **14**(2), 627–660 (2020) https://doi.org/10.5465/annals.2018.0057

[5] Hoff, K.A., Bashir, M.: Trust in automation: Integrating empirical evidence on factors that influence trust. Hum. Factors **57**(3), 407–434 (2015) https://doi.org/10.1177/0018720814547570

[6] Shin, D.: The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. Int. J. Hum.-Comput. Stud. **146**, 102551 (2021) https://doi.org/10.1016/j.ijhcs.2020.102551

[7] Perrig, S.A.C., Scharowski, N., Brühlmann, F.: Trust issues with trust scales: Examining the psychometric quality of trust measures in the context of ai. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing

22

Systems, pp. 1–7 (2023). https://doi.org/10.1145/3544549.3585808

[8] Larasati, R., De Liddo, A., Motta, E.: Human and ai trust: Trust attitude measurement instrument development. Human-Computer Interaction (2023) https://doi.org/10.48550/arXiv.2510.21535

[9] Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., Käser, T.: The critical role of trust in adopting ai-powered educational technology for learning: An instrument for measuring student perceptions. Comput. Educ. Artif. Intell. **8**, 100368 (2025) https://doi.org/10.1016/j.caeai.2024.100368

[10] Afroogh, S., Akbari, A., Malone, E., Kargar, M., Alambeigi, H.: Trust in ai: Progress, challenges, and future directions. Humanit. Soc. Sci. Commun. **11**(1), 1568 (2024) https://doi.org/10.1057/s41599-024-04044-8

[11] Bozkurt, A.: The three laws of artificial intelligence: Re-evaluating human-ai agency and interaction in a time of the generative and agentic ai renaissance. Open Praxis **17**(3), 421–428 (2025) https://doi.org/10.55982/openpraxis.17.3.794

[12] Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative agents: Interactive simulacra of human behavior. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1–22 (2023). https://doi.org/10.1145/3586183.3606763

[13] Bozkurt, A., Sharma, R.C.: Trust, credibility and transparency in human-ai interaction: Why we need explainable and trustworthy ai and why we need it now. Asian Journal of Distance Education **19**(2) (2024)

[14] DeVellis, R.F., Thorpe, C.T.: Scale Development: Theory and Applications. SAGE, Thousand Oaks, CA (2021)

[15] Chi, O.H., Jia, S., Li, Y., Gursoy. D.: Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (ai) social robots in service delivery. Comput. Hum. Behav. **118**, 106700 (2021) https://doi.org/10.1016/j.chb.2020.106700

[16] Dang, Q., Li, G.: Unveiling trust in ai: The interplay of antecedents, consequences, and cultural dynamics. AI & Society, 1–24 (2025) https://doi.org/10.1007/s00146-025-02477-6

[17] Gulati, S., Sousa, S., Lamas, D.: Design, development and evaluation of a human-computer trust scale. Behav. & Inf. Technol. **38**(10), 1004–1015 (2019) https://doi.org/10.1080/0144929X.2019.1656779

[18] Jian, J.-Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated systems. Int. J. Cogn. Ergon. **4**(1), 53–71 (2000) https://doi.org/10.1207/S15327566IJCE0401_04

23

[19] Lee, J.D., See, K.A.: Trust in automation: Designing for appropriate reliance. Hum. Factors **46**(1), 50–80 (2004) https://doi.org/10.1518/hfes.46.1.50.30392

[20] Ribeiro, M.T., Singh, S., Guestrin, C.: Why Should I Trust You? explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016). https://doi.org/10.1145/2939672.2939778

[21] Koo, M., Yang, S.-W.: Likert-type scale. Encyclopedia **5**(1), 18 (2025) https://doi.org/10.3390/encyclopedia5010018

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

[23] Hair, J.F.J., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate Data Analysis, 7th edn. Pearson, Upper Saddle River, NJ (2010)

[24] Kline, R.B.: Principles and Practice of Structural Equation Modeling. Guilford Publications, New York (2023)

[25] Rehman, N., Huang, X., Mahmood, A.: Enhancing mathematical problem-solving and 21st-century skills through pjbl: A structural equation modelling approach. Educational Studies, 1–26 (2025) https://doi.org/10.1080/03055698.2025.2514691

[26] Hinkin, T.R.: A brief tutorial on the development of measures for use in survey questionnaires. Organ. Res. Methods **1**(1), 104–121 (1998) https://doi.org/10.1177/109442819800100106

[27] Lewis-Beck, M.S., Bryman, A., Liao, T.F.: The Sage Encyclopedia of Social Science Research Methods. Sage Publications, London (2003). https://doi.org/10.4135/9781412950589

[28] Field, A.: Discovering Statistics Using IBM SPSS Statistics. Sage Publications Limited, London (2024)

[29] Tabachnick, B.G., Fidell, L.S., Ullman, J.B.: Using Multivariate Statistics, 5th edn. Pearson, Boston, MA (2007)

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

[31] Byrne, B.M.: Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming. Routledge, London (2013)

[32] Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method

24

biases in behavioral research: A critical review of the literature and recommended remedies. J. Appl. Psychol. **88**(5), 879–903 (2003) https://doi.org/10.1037/0021-9010.88.5.879

[33] George, D., Mallery, P.: IBM SPSS Statistics 29 Step by Step: A Simple Guide and Reference. Routledge, New York (2024)

[34] Hayes, A.F., Coutts, J.J.: Use omega rather than cronbach's alpha for estimating reliability. but... Commun. Methods Meas. **14**(1), 1–24 (2020) https://doi.org/10.1080/19312458.2020.1718629

[35] Nunnally, J.C., Bernstein, I.H.: Psychometric Theory, 3rd edn. McGraw-Hill, New York (1994)

[36] Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. **18**(1), 39–50 (1981) https://doi.org/10.1177/002224378101800104

[37] Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. J. Acad. Mark. Sci. **43**(1), 115–135 (2015) https://doi.org/10.1007/s11747-014-0403-8

[38] Cheung, G.W., Rensvold, R.B.: Evaluating goodness-of-fit indexes for testing measurement invariance. Struct. Equ. Model. **9**(2), 233–255 (2002) https://doi.org/10.1207/S15328007SEM0902_5

[39] Putnick, D.L., Bornstein, M.H.: Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental Review **41**, 71–90 (2016) https://doi.org/10.1016/j.dr.2016.06.004

[40] Vandenberg, R.J., Lance, C.E.: A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods **3**(1), 4–70 (2000) https://doi.org/10.1177/109442810031002

[41] Xie, L., Guo, T., Yang, Y., Yu, L., Chen, Y., Peng, X., Zheng, Z., Tu, H.: Trust in artificial intelligence-based follow-up in hospital information systems: development and validation of a new scale. BMC psychology **14**(1), 95 (2026) https://doi.org/10.1186/s40359-025-03855-x

[42] McGrath, M.J., Lack, O., Tisch, J., Duenser, A.: Measuring trust in artificial intelligence: Validation of an established scale and its short form. Front. Artif. Intell. **8**, 1582880 (2025) https://doi.org/10.3389/frai.2025.1582880

25

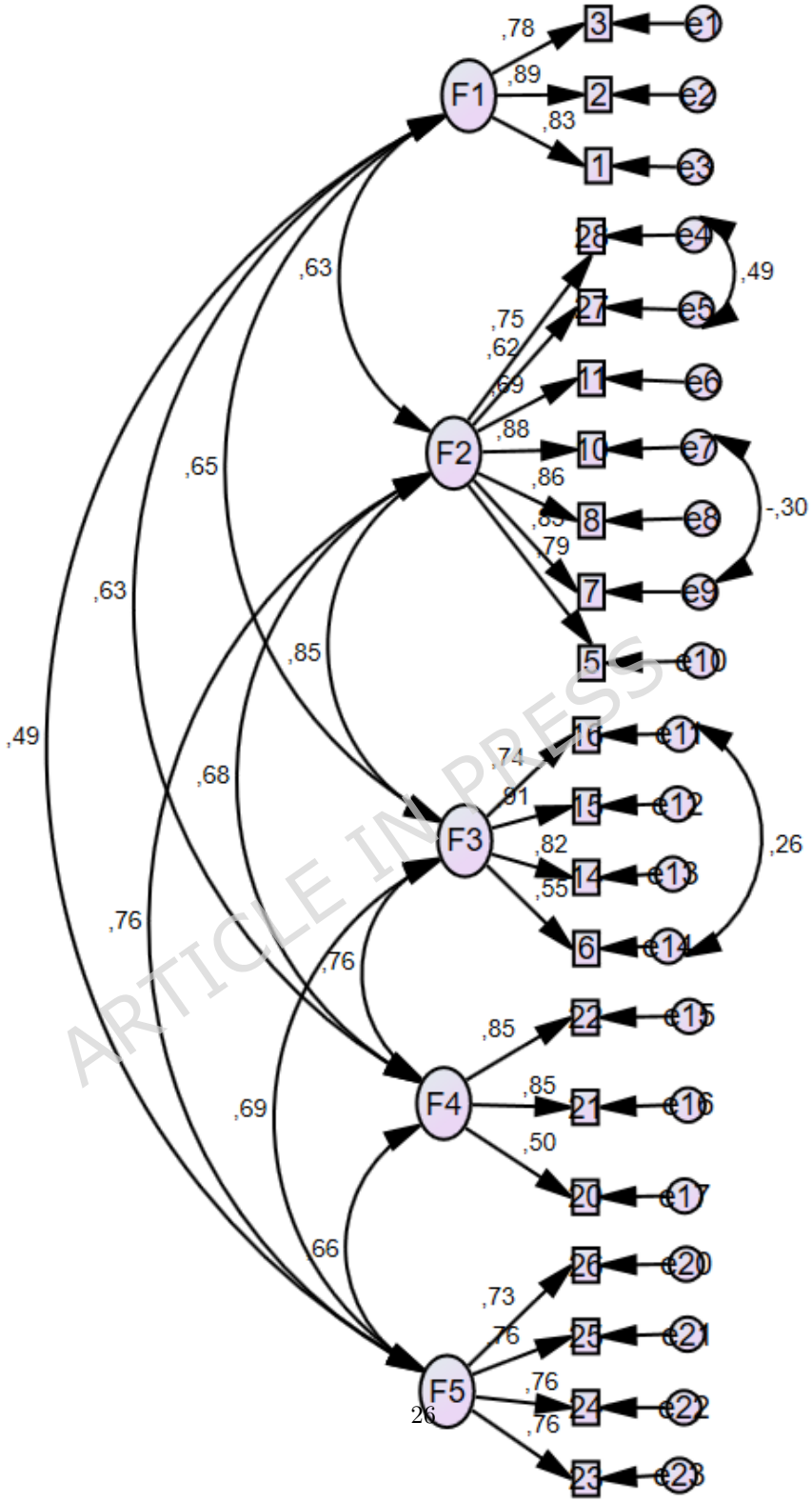

Fig. 2: A factor structure model of the scale