

Article

From Teacher to Algorithm: Teacher Endorsement and Student Acceptance of AI-Generated Content Within the Trust Transfer Theory Framework

Fawzia Omer Alubthane 

Department of Curriculum and Instruction, College of Education, Imam Mohammad Ibn Saud Islamic University, Riyadh 11432, Saudi Arabia; foalubthane@imamu.edu.sa

Abstract

The integration of AI-generated content into higher education has intensified interest in how students form and calibrate trust toward algorithmic outputs and whether pedagogical relationships can serve as conduits for that trust. Grounded in Trust Transfer Theory, this between-subjects randomized experimental study ($N = 320$) investigated whether teacher endorsement shapes students' perceptions of AI-generated educational content across four dimensions: Perceived AI Competence, Academic Integrity, Perceived Human-Mediated Reliability, and Behavioral Intention to adopt. Participants from Saudi Arabian universities were randomly assigned to an endorsed or non-endorsed vignette condition and responded to a validated 13-item Trust and Acceptance Scale. Independent-samples t -tests confirmed statistically significant differences across all four dimensions in favor of the endorsed condition, with effect sizes ranging from small to large, and findings remained robust after controlling for gender via ANCOVA. Within-condition regression analyses further established Perceived Human-Mediated Reliability as a structurally stable positive predictor of trust outcomes in both conditions, with predictive power consistently amplified under endorsement. Postgraduate students placed greater emphasis on human oversight, while no disciplinary differences emerged, confirming the cross-disciplinary universality of the trust transfer mechanism. These findings are consistent with positioning the teacher as a trust guarantor in AI-mediated learning environments and carry direct implications for pedagogical design and institutional AI governance.

Keywords: trust transfer; teacher endorsement; AI-generated content; higher education; human–AI interaction; generative AI; behavioral intention; academic integrity; randomized experiment



Academic Editors: Erin R. Pletcher and Travis R. Pollen

Received: 14 May 2026

Revised: 26 June 2026

Accepted: 9 July 2026

Published: 13 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The accelerating integration of generative artificial intelligence (AI) into higher education has fundamentally reshaped the epistemic landscape in which students encounter, evaluate, and engage with academic knowledge. As AI systems grow increasingly capable of producing coherent, contextually relevant educational content, a central pedagogical challenge has emerged: how do students determine whether algorithmically generated material is trustworthy, academically legitimate, and appropriate for integration into their scholarly practice? This question is not merely technical but deeply relational, implicating the social and institutional structures through which academic credibility is traditionally established and transmitted.

Trust in artificial intelligence differs fundamentally from trust in conventional technologies because intelligent systems are not passive tools but dynamic agents whose outputs resist straightforward verification by non-expert users (Siau & Wang, 2018). Student trust, in this context, refers to a willingness to rely on the guidance, information, and feedback provided by an AI system under conditions of uncertainty and potential vulnerability (Pitts & Motamedi, 2024). The multidimensional nature of this trust encompasses perceived capability, perceived benevolence, and perceived integrity, each representing a distinct evaluative orientation toward the AI as a knowledge source (Choung et al., 2023). Trust is, moreover, critically important for students' behavioral adoption of AI tools: without it, even technically superior systems fail to achieve uptake regardless of their objective quality (Nazaretsky et al., 2022a; Balaskas et al., 2025).

Despite the growing body of research on student attitudes toward generative AI, a significant gap persists in the literature regarding the mechanisms through which trust in AI-generated content is constructed, mediated, and transferred within pedagogical relationships. Students widely use or express willingness to use generative AI tools, yet they approach AI outputs with caution rather than unconditional trust, driven by concerns about accuracy, pedagogical relevance, and academic integrity (Johnston et al., 2024; Chan & Hu, 2023; Johri et al., 2024; Fošner, 2024; Black & Tomlinson, 2025). This persistent skepticism suggests that the adoption of AI-generated content is not simply a function of the content's technical quality but is mediated by relational and institutional factors that lie beyond the algorithm itself.

The present study addresses this gap by applying Trust Transfer Theory (Stewart, 2003) to examine how teacher endorsement functions as a trust transfer mechanism in students' evaluations of AI-generated educational content. Trust Transfer Theory posits that trust established in a primary, familiar source can be extended to a secondary, less familiar entity through perceived association or endorsement. Applied to the educational AI context, the theory predicts that instructor validation of AI-generated materials should activate and amplify student trust in those materials, thereby reducing the skepticism that otherwise impedes adoption. This theoretical proposition had not been systematically tested in controlled randomized research, leaving a significant empirical lacuna that the present study is designed to address.

Accordingly, this study pursues three research objectives: to determine whether teacher endorsement produces statistically significant and practically meaningful differences in students' trust perceptions across four dimensions, to examine the predictive power of Perceived Human-Mediated Reliability as a driver of remaining trust constructs within each experimental condition separately, and to explore whether academic level and major moderate the trust transfer process.

2. Theoretical Framework and Literature Review

2.1. Trust Transfer Theory: Origins, Scope, and Boundary Conditions

The present study adopts Trust Transfer Theory as its guiding conceptual framework. Originally formulated by Stewart (2003) in the context of World Wide Web commerce, the theory posits that trust is not formed in isolation but through relational chains: when a trust evaluator has established confidence in a primary entity, a degree of that confidence can be transferred to a secondary entity perceived as associated with or endorsed by the first. Critically, this transfer is partial rather than complete, as the secondary entity inherits some but not all of the credibility of the primary (Xu et al., 2023; Jøsang et al., 2006; Liu et al., 2021). The transfer mechanism has been operationalized across diverse domains, including e-commerce (Zhao et al., 2019; Xiao et al., 2019), peer-to-peer accommodation (Xu et al.,

2023), social commerce (Wiyata et al., 2022), social media influence (Wang & Chan-Olmsted, 2024), and human–robot interaction (Guo et al., 2023).

Before applying Trust Transfer Theory to the educational AI context, it is important to acknowledge its boundary conditions explicitly, as these define the interpretive scope of the present findings. First, Stewart’s (2003) original formulation was developed for iterative, choice-based online trust formation across multiple interactions, a process that differs in important respects from the one-time vignette exposure employed in the present study. Second, the theory presupposes a pre-existing positive trust relationship with the primary source; in the current design, participants’ baseline trust in the vignette instructor was not directly measured, meaning that the foundational condition of the theory was inferred from the experimental context rather than verified empirically. Third, the theory was not developed to account for situations where the primary source’s credibility is itself contested or low. Taken together, these boundary conditions mean that the present findings should be interpreted as tests of the trust transfer mechanism under conditions of assumed instructor credibility, and future research should incorporate instructor credibility as an explicit moderator variable. Because this second boundary condition was not empirically verified, the present results should be read as establishing the effect of explicit teacher endorsement on students’ trust-related perceptions, rather than as direct evidence that pre-existing trust in the instructor was transferred to the AI-generated content.

2.2. Trust Transfer in Educational AI Contexts

The application of Trust Transfer Theory to educational AI is a relatively recent development. Aly et al. (2025) provided direct empirical support for the trust transfer mechanism in an educational context, demonstrating that introducing an AI instructor through a human intermediary significantly increased student trust, perceived learning, and enjoyment relative to autonomous AI presentation. This borrowing credibility effect, whereby AI agents inherit the credibility of the human actors who introduce them, has been conceptually elaborated by Kizilcec (2023) and Omirali et al. (2025), who argue that the social legitimacy of AI in education is inherently co-constructed through the institutional and interpersonal relationships surrounding its deployment. A growing body of research identifies transparency in AI decision-making, prior experience with generative AI tools, psychological safety, and cultural and dispositional factors as key moderators of trust transfer magnitude (Nazaretsky et al., 2022a; Herdiani et al., 2024; Deric et al., 2025; Goldshtein et al., 2025; Song, 2024; Shahzad et al., 2024; Rana et al., 2024; Shen et al., 2025; Qin et al., 2020; Ayanwale et al., 2024; Viberg et al., 2024; Aladi, 2024), as summarized in Table 1.

Table 1. Sources and targets of trust shaping AI uptake in educational settings.

Trust Source	Trust Target	Educational Impact	Key Citations
Human instructor	AI instructor/chatbot	Increased trust, perceived learning, enjoyment	Aly et al. (2025); Pitts and Motamedi (2025); Goldshtein et al. (2025)
Institution/university	AI agents/MAS systems	Adoption readiness, moderated by ethics and privacy	Qin et al. (2020); Herdiani et al. (2024); Rana et al. (2024); Omirali et al. (2025)
Peers and colleagues	GenAI tools	Adoption by teachers and faculty	Nazaretsky et al. (2022a); Shata and Hartley (2025); Kizilcec (2023); Ayanwale et al. (2024); Aladi (2024)

2.3. The Teacher as a Trust Channel: Theoretical Operationalization

The present study positions the teacher not as a passive transmitter of knowledge but as an active trust channel within the Trust Transfer Theory framework, a conceptualization resting on the recognition that students enter pedagogical relationships with pre-existing trust in their instructors' expertise, academic integrity, and evaluative judgement. The teacher-as-trust-channel mechanism operates through three interrelated pathways that are theoretically derived from the framework and empirically operationalized in the study's vignette manipulation. The first is endorsement-based credentialing: when a teacher explicitly validates AI-generated content, the teacher's established attributes of competence and integrity are projected onto the content, reducing the student's cognitive and evaluative uncertainty. The second pathway involves perceived association: the mental link established between the teacher's judgement and the AI's output allows students to infer the quality of the latter from the reputation of the former, extending the teacher's credibility to the algorithm without requiring direct assessment of the algorithm's technical merits. The third pathway concerns the activation of human role awareness: explicit teacher endorsement heightens student awareness of the supervisory function that human oversight serves in quality assurance, transforming the AI output from a technically uncertain product into an academically certified one. These pathways are consistent with the broader literature on teacher support and technology adoption (Ouyang et al., 2025; Lin & Yu, 2023; Choi et al., 2022; Saleh et al., 2022).

2.4. Research Questions and Hypotheses

Based on the theoretical framework and literature review, the present study addresses the following research questions and corresponding hypotheses:

RQ1: Do teacher-endorsed and non-endorsed AI-generated content conditions produce statistically significant and practically meaningful differences in students' Perceived AI Competence (PAC), Academic Integrity perceptions (AI-scale), Perceived Human-Mediated Reliability (PHMR), and Behavioral Intention to adopt (BI)?

H1. *There are statistically significant differences ($p < 0.01$) with practically meaningful effect sizes (Cohen's $d \geq 0.30$) across all four trust dimensions in favor of the teacher-endorsed condition.*

RQ2: What is the predictive impact of Perceived Human-Mediated Reliability (PHMR) on Perceived AI Competence (PAC), Academic Integrity (AI-scale), and Behavioral Intention (BI) within each experimental condition separately?

H2. *Perceived Human-Mediated Reliability (PHMR) is a statistically significant positive predictor of PAC, AI-scale, and BI within both the endorsed and non-endorsed conditions independently.*

RQ3: Are there statistically significant differences in students' trust responses attributable to academic level (undergraduate vs. postgraduate) or academic major (administrative and economics, humanities, scientific)?

H3. *There are statistically significant differences in trust responses attributable to academic level and academic major.*

3. Materials and Methods

3.1. Research Design

The study employed a between-subjects randomized experimental design, a design widely used in social and educational science research to establish causal inferences through scenario-based random assignment (Mutz, 2011). Participants were randomly assigned

with equal probability (50:50) to one of two scenario-based vignette conditions: an endorsed condition in which AI-generated educational content was explicitly validated by a named instructor and a non-endorsed condition presenting identical content without any instructor validation. Randomization was implemented via an automated Google Apps Script procedure that directed participants to one of two equivalent Google Forms versions at study entry, with each allocation logged with a timestamp and group code in a secured spreadsheet to ensure transparency and auditability. Participants were unaware of the existence of the alternative condition, preserving the blind characteristic of the randomization. This design enables direct causal inference about the effect of teacher endorsement while holding all other content characteristics constant.

3.2. Participants and Sampling

The study population comprised university students from diverse disciplines and academic levels across Saudi Arabian universities. A convenience sample of 320 university students aged 18 years or older participated in the study, with 160 participants assigned to each experimental condition. The demographic characteristics of the sample are presented in Table 2. Significant group differences were observed for gender, with the endorsed group comprising a substantially higher proportion of female participants (81.3% vs. 60.0%). This imbalance is addressed via ANCOVA in Section 4.1 to ensure the robustness of the findings.

Table 2. Demographic distribution of participants by experimental condition.

Variable	Category	Endorsed <i>n</i> (%)	Non-Endorsed <i>n</i> (%)
Gender	Female	130 (81.3)	96 (60.0)
	Male	30 (18.7)	64 (40.0)
Academic Level	Undergraduate	120 (75.0)	118 (73.8)
	Postgraduate	40 (25.0)	42 (26.2)
Major	Admin & Economics	51 (31.9)	48 (30.0)
	Humanities	81 (50.6)	66 (41.3)
	Scientific	28 (17.5)	46 (28.7)
AI Experience	Intermediate	94 (58.8)	84 (52.5)
	Good	56 (35.0)	60 (37.5)
	Novice	10 (6.2)	16 (10.0)

3.3. Vignette Stimuli

The research instrument's first component was a scenario-based vignette presenting a realistic hypothetical learning situation. Both conditions received structurally identical vignettes in content type, subject matter, length, visual formatting, and platform presentation, with the sole manipulated element being the presence or absence of a teacher endorsement statement.

3.3.1. Endorsed Condition (Condition A)

Participants in the endorsed condition read the following scenario: *"Imagine that you are preparing for a final exam in a specialized course. Your course instructor uploaded a 'Comprehensive Summary' to the learning management system and sent the following message to students: 'I used artificial intelligence tools to generate this summary based on the lecture content. I have personally reviewed it word by word to verify the accuracy of the information, and I confirm that it is a reliable and comprehensive resource for your study preparation.'"*

3.3.2. Non-Endorsed Condition (Condition B)

Participants in the non-endorsed condition read: *"Imagine that you are preparing for a final exam in a difficult specialized course. Upon logging into the learning management system, you*

discovered a new feature called the ‘Smart Summary.’ When you clicked on it, the system automatically compiled the lecture content and instantly generated a summary for you. A system message appeared stating: ‘This summary was generated entirely by artificial intelligence based on the course content to ensure speed, without any human intervention or review by the course instructor.’

The structural design of the two conditions provides indirect evidence of successful manipulation: the vignettes were identical in all respects except for the presence or absence of the teacher endorsement statement, and the significant between-group differences observed across all four trust dimensions are consistent with successful perception of this manipulation. The absence of a formal manipulation check item is acknowledged as a limitation in Section 6.

3.4. The Trust and Acceptance Scale

The second instrument component was the Trust and Acceptance Scale comprising 13 Likert-type items (1 = strongly disagree to 5 = strongly agree) distributed across four theoretically derived subscales. The scale was developed based on Trust Transfer Theory and adapted from validated instruments in educational technology and AI trust research (Nazaretsky et al., 2022b; Choi et al., 2022; Pitts & Motamedi, 2024). All 13 items are reproduced in Table 3. Items were originally developed in Arabic and subjected to back-translation by an independent bilingual researcher to verify conceptual equivalence.

Table 3. Trust and acceptance scale—all 13 items by subscale.

#	Item Wording (English Translation)	Subscale
1	I believe this summary can effectively cover all concepts of the course.	PAC
2	I trust that the AI system can organize academic information in a logical manner.	PAC
3	The generated content appears to align with the standards required for success in the exam.	PAC
4	I trust that the information contained in this summary is accurate and scientifically sound.	AI-scale
5	I am not concerned about the presence of errors or “hallucinated” information in this content.	AI-scale
6	I believe this summary demonstrates academic honesty in conveying the lecture content.	AI-scale
7	I believe that the course instructor’s supervision is the primary guarantee of the accuracy of AI-generated information.	PHMR
8	I feel more comfortable using digital content when it is explicitly endorsed by a human expert.	PHMR
9	I believe that the teacher’s recommendation is what grants AI-generated content “legitimacy” for use in studying.	PHMR
10	The teacher’s educational filtering makes me trust the algorithm’s outputs more.	PHMR
11	I will rely primarily on this summary in my preparation for the final exam.	BI
12	I will share this summary with my peers as a reliable reference.	BI
13	I prefer to use this type of AI-generated summary in the future if it is under the teacher’s supervision.	BI

Note. PAC = Perceived AI Competence (Items 1–3); AI-scale = Academic Integrity subscale (Items 4–6); PHMR = Perceived Human-Mediated Reliability (Items 7–10); BI = Behavioral Intention (Items 11–13). Items originally developed in Arabic; English translation verified by back-translation.

3.5. Validity and Reliability

Content and face validity were established through expert review by a panel of specialists in educational technology and educational psychology, who evaluated item clarity, theoretical alignment, and vignette ecological realism. Internal consistency was assessed in a pilot study ($n = 40$, not included in the main sample). As shown in Table 4, all item–total correlations were statistically significant at $p < 0.01$, ranging from $r = 0.646$ to $r = 0.907$. Cronbach’s alpha coefficients, presented in Table 5, ranged from $\alpha = 0.771$ to $\alpha = 0.879$ at the subscale level, with overall scale reliability of $\alpha = 0.904$, meeting conventional thresholds for excellent internal consistency (Nunnally & Bernstein, 1994). It is noted that the PHMR

subscale comprises four items (Items 7–10), while the remaining subscales comprise three items each. It should be noted that the validity evidence reported here is restricted to expert-judged content and face validity, together with internal-consistency reliability; no confirmatory factor analysis or formal discriminant/convergent validity testing of the four-factor structure was conducted. The strong item–total correlations and subscale alphas support internal consistency but do not, on their own, establish the structural or discriminant validity of the scale, and this is noted as a limitation in Section 6.

Table 4. Pearson item–total correlation coefficients by subscale.

PAC	AI-Scale	PHMR	BI
Item 1: $r = 0.887^{**}$	Item 1: $r = 0.897^{**}$	Item 1: $r = 0.877^{**}$	Item 1: $r = 0.906^{**}$
Item 2: $r = 0.812^{**}$	Item 2: $r = 0.907^{**}$	Item 2: $r = 0.823^{**}$	Item 2: $r = 0.906^{**}$
Item 3: $r = 0.829^{**}$	Item 3: $r = 0.883^{**}$	Item 3: $r = 0.827^{**}$	Item 3: $r = 0.646^{**}$
	Item 4: $r = 0.907^{**}$	Item 4: $r = 0.907^{**}$	

Note. $^{**} p < 0.01$.

Table 5. Cronbach’s alpha reliability coefficients by subscale.

Subscale	No. of Items	α
Perceived AI Competence (PAC)	3	0.797
Academic Integrity (AI-scale)	3	0.871
Perceived Human-Mediated Reliability (PHMR)	4	0.879
Behavioral Intention (BI)	3	0.771
Total Scale	13	0.904

3.6. Data Analysis

All analyses were conducted in SPSS v.28. To address RQ1, independent-samples *t*-tests were conducted with Levene’s test for equality of variances; where Levene’s test indicated heteroscedasticity ($p < 0.05$), Welch-corrected *t*-values would have been reported. Cohen’s *d* was computed as the standardized mean difference, with $d \geq 0.20$, 0.50, and 0.80 representing small, medium, and large effects, respectively (Cohen, 1988). To address the observed gender imbalance, ANCOVAs were conducted with gender as the covariate; partial eta-squared (η^2p) values are reported for these models. To address RQ2, simple linear regression analyses were performed within each experimental condition separately, with PHMR as the predictor. To address RQ3, independent-samples *t*-tests examined differences by academic level, and one-way ANOVAs (with η^2p) examined differences by academic major. The overall significance threshold was $\alpha = 0.05$.

4. Results

4.1. Effect of Teacher Endorsement on Trust Perceptions (RQ1)

Table 6 presents the independent-samples *t*-test results and associated effect sizes. Statistically significant differences were found across all four dimensions in favor of the endorsed condition, confirming H1. Levene’s test indicated homogeneity of variances for all comparisons (all $p > 0.10$); standard *t*-values are therefore reported. Effect sizes ranged from small-to-medium ($d = 0.29$ for PHMR) to large ($d = 0.79$ for BI). ANCOVA analyses controlling for gender confirmed that endorsement effects remained statistically significant and largely unchanged across all dimensions, with gender not emerging as a significant predictor in any model (all $p > 0.12$).

Table 6. Independent-samples *t*-test results with effect sizes and ANCOVA partial η^2p (gender covariate).

Dimension	Group	N	M (SD)	t	p	d	η^2p (ANCOVA)
PAC	Endorsed	160	4.10 (0.73)	5.65	<0.001 **	0.63	0.084
	Non-endorsed	160	3.60 (0.85)				
AI-scale	Endorsed	160	3.75 (0.95)	4.94	<0.001 **	0.55	0.062
	Non-endorsed	160	3.20 (1.07)				
PHMR	Endorsed	160	4.53 (0.67)	2.46	0.014 *	0.29	0.018
	Non-endorsed	160	4.33 (0.74)				
BI	Endorsed	160	4.19 (0.84)	7.09	<0.001 **	0.79	0.124
	Non-endorsed	160	3.53 (0.84)				

Note: ** $p < 0.01$; * $p < 0.05$. d = Cohen's d. η^2p = partial eta-squared from ANCOVA with gender covariate. 95% CIs for d: PAC [0.40, 0.86]; AI-scale [0.33, 0.77]; PHMR [0.06, 0.52]; BI [0.56, 1.02]. PAC = Perceived AI Competence; AI-scale = Academic Integrity subscale; PHMR = Perceived Human-Mediated Reliability; BI = Behavioral Intention.

4.2. Predictive Role of PHMR Within Each Experimental Condition (RQ2)

To avoid conflating treatment-effect variance with the predictive relationship, simple linear regression analyses were conducted separately within each experimental condition. Results are presented in Tables 7–9. PHMR was a statistically significant positive predictor of all three dependent constructs in both conditions, confirming H2. Behavioral Intention was most strongly predicted across both conditions (endorsed: $R^2 = 0.356$, $\beta = 0.596$, $p < 0.001$; non-endorsed: $R^2 = 0.185$, $\beta = 0.430$, $p < 0.001$). Predictive power was consistently stronger in the endorsed condition across all three outcomes, suggesting that teacher endorsement not only elevates trust levels but also amplifies the coupling between perceived human oversight and trust outcomes.

Table 7. Within-group regression results—endorsed condition (n = 160).

Outcome/Predictor	B	SE	β	t	p
PAC ($R = 0.463$, $R^2 = 0.214$, $F(1, 158) = 43.086$, $p < 0.001$)					
Constant	1.839	0.349	—	5.269	<0.001 **
PHMR	0.501	0.076	0.463	6.564	<0.001 **
AI-scale ($R = 0.269$, $R^2 = 0.072$, $F(1, 158) = 12.314$, $p = 0.001$)					
Constant	2.047	0.492	—	4.161	<0.001 **
PHMR	0.377	0.108	0.269	3.509	0.001 **
BI ($R = 0.596$, $R^2 = 0.356$, $F(1, 158) = 87.255$, $p < 0.001$)					
Constant	0.823	0.365	—	2.257	0.025 *
PHMR	0.744	0.080	0.596	9.341	<0.001 **

Note: ** $p < 0.01$; * $p < 0.05$. B = unstandardized coefficient; SE = standard error; β = standardized coefficient.

Table 8. Within-group regression results—non-endorsed condition (n = 160).

Outcome/Predictor	B	SE	β	t	p
PAC ($R = 0.305$, $R^2 = 0.093$, $F(1, 158) = 16.240$, $p < 0.001$)					
Constant	2.077	0.384	—	5.401	<0.001 **
PHMR	0.353	0.088	0.305	4.030	<0.001 **
AI-scale ($R = 0.182$, $R^2 = 0.033$, $F(1, 158) = 5.395$, $p = 0.021$)					
Constant	2.047	0.501	—	4.083	<0.001 **
PHMR	0.265	0.114	0.182	2.323	0.021 *
BI ($R = 0.430$, $R^2 = 0.185$, $F(1, 158) = 35.854$, $p < 0.001$)					
Constant	1.400	0.360	—	3.887	<0.001 **
PHMR	0.491	0.082	0.430	5.988	<0.001 **

Note: ** $p < 0.01$; * $p < 0.05$. B = unstandardized coefficient; SE = standard error; β = standardized coefficient.

Table 9. Comparison of R^2 values across experimental conditions.

Outcome	Endorsed R^2	Non-Endorsed R^2
Behavioral Intention (BI)	0.356	0.185
Perceived AI Competence (PAC)	0.214	0.093
Academic Integrity (AI-scale)	0.072	0.033

Note: All regression models were statistically significant at $p < 0.05$ in both conditions.

4.3. Differences by Academic Level and Major (RQ3)

Table 10 presents the academic level comparison. Levene's test indicated homogeneity of variances across all four dimensions (all $p > 0.10$). Statistically significant differences were found for Academic Integrity ($t(318) = 3.17$, $p = 0.002$, $d = 0.41$), favoring undergraduate students, and for PHMR ($t(318) = 2.25$, $p = 0.025$, $d = 0.31$), favoring postgraduate students. No statistically significant differences were observed for PAC or BI, with trivial effect sizes ($d = 0.15$ and $d = 0.08$, respectively), providing partial support for H3.

Table 10. *t*-test results for differences in trust dimensions by academic level.

Dimension	Level	N	M	SD	t	p	d
PAC	Undergraduate	238	3.89	0.82	1.24	0.215	0.15
	Postgraduate	82	3.76	0.87			
AI-scale	Undergraduate	238	3.58	1.01	3.17	0.002 **	0.41
	Postgraduate	82	3.16	1.10			
PHMR	Undergraduate	238	4.38	0.74	2.25	0.025 *	0.31
	Postgraduate	82	4.58	0.60			
BI	Undergraduate	238	3.84	0.91	0.61	0.543	0.08
	Postgraduate	82	3.91	0.88			

Note: ** $p < 0.01$; * $p < 0.05$. d = Cohen's d .

One-way ANOVA revealed no statistically significant differences across any trust dimension as a function of academic major (all $p > 0.05$, all $\eta^2p < 0.02$), as detailed in Table 11, indicating negligible disciplinary effects.

Table 11. One-way ANOVA results for trust dimensions by academic major.

Dimension	Source	SS	df	MS	F	p	η^2p
PAC	Between	3.00	2	1.50	2.20	0.113	0.014
	Within	216.41	317	0.68			
AI-scale	Between	3.94	2	1.97	1.80	0.166	0.011
	Within	346.53	317	1.09			
PHMR	Between	1.29	2	0.65	1.28	0.279	0.008
	Within	160.06	317	0.50			
BI	Between	0.63	2	0.31	0.38	0.684	0.002
	Within	259.62	317	0.82			

Note: η^2p = partial eta-squared. All p -values > 0.05 (non-significant).

5. Discussion

The findings are broadly consistent with Trust Transfer Theory in educational AI contexts, suggesting that teacher endorsement may function as a meaningful mechanism through which student trust in AI-generated content is constructed and legitimized. The gradient of endorsement effects across the four dimensions is theoretically coherent rather than incidental: Behavioral Intention showed the largest effect while Academic Integrity showed the smallest, a pattern consistent with dual-process accounts of trust formation

(Chaiken & Trope, 1999; Kahneman, 2011) wherein adoption dispositions respond more readily to contextual legitimation signals than do deeply rooted moral–epistemic evaluations. This distinction carries a practical implication of direct relevance to educational policy: teacher endorsement is a powerful accelerant of AI adoption, but institutional mechanisms—transparent AI-use policies, provenance labeling, and academic integrity curricula—remain necessary to address the ethical concerns that interpersonal endorsement alone cannot resolve. It is also important to acknowledge that the observed effects are equally consistent with alternative, non-mutually exclusive mechanisms that this design cannot disentangle from trust transfer per se: an authority-compliance effect, whereby students defer to the instructor’s directive independent of any genuine shift in trust, and a social-cue or peripheral-route heuristic (Chaiken & Trope, 1999), whereby the mere presence of an authority signal is used as a low-effort proxy for content quality. Future research employing mediation designs that explicitly measure perceived authority and heuristic processing alongside trust constructs would help adjudicate among these accounts.

The high baseline score on Perceived Human-Mediated Reliability, even in the non-endorsed condition, reveals that students already hold a strong orientation toward human mediation as a prerequisite for trusting AI-generated content, independent of any explicit endorsement signal. Teacher endorsement does not create this orientation—it activates and amplifies one that students bring to the encounter. The within-condition regression analyses deepen this interpretation: Perceived Human-Mediated Reliability predicts Behavioral Intention with substantial explanatory power in both conditions independently, suggesting that the valuation of human oversight may be a structurally stable driver of AI adoption rather than merely an artifact of the experimental manipulation. The consistent amplification of predictive power in the endorsed condition further suggests that endorsement operates not merely by elevating trust levels but by strengthening the psychological coupling between perceived human oversight and trust outcomes—a pattern consistent with an extension of Trust Transfer Theory beyond its original commercial formulations (Stewart, 2003) rather than a direct confirmation of it. Because participants’ pre-existing trust in the vignette instructor was assumed rather than directly measured, these results are best interpreted as demonstrating the effect of teacher endorsement on students’ perceptions of AI-generated content, rather than as direct proof that trust was transferred from instructor to algorithm.

The moderating role of academic level reflects the epistemic maturation that advanced scholarly training produces: postgraduate students apply more stringent integrity standards born of heightened awareness of the specific risks that generative AI poses to academic knowledge production. The complete absence of disciplinary differences across all four dimensions is equally significant, confirming that the trust transfer mechanism operates with cross-disciplinary universality and directly substantiating the case for institution-wide rather than discipline-specific AI governance frameworks (Johnston et al., 2024). Taken together, the findings are consistent with positioning the teacher as a trust guarantor—understood here as a tentative, theory-derived interpretation rather than a directly demonstrated causal role—a sociotechnical intermediary whose credibility and explicit validation appear to serve as the primary legitimating mechanism through which algorithmically generated knowledge becomes academically usable.

6. Limitations

The scenario-based vignette methodology, while appropriate for experimental control, necessarily abstracts from the relational complexity of authentic classroom interaction. Real-world teacher endorsement accumulates across sustained pedagogical relationships—through vocal cues, iterative feedback, and the tacit authority built over an entire course—in

ways that a single written scenario cannot replicate. The observed effect sizes may therefore represent upper bounds that naturalistic settings would attenuate. Related to this, the study did not measure participants' baseline trust in the vignette instructor, leaving the foundational precondition of Trust Transfer Theory—a pre-existing positive trust relationship—inferred rather than verified. Future studies should manipulate instructor credibility as an independent variable to test whether the magnitude of trust transfer is contingent on prior relational trust.

The sample's restriction to Saudi Arabian universities, combined with the convenience sampling approach, limits cultural and population generalizability. The absence of a formal manipulation check, while methodologically defensible given the unambiguous structural difference between the two vignettes, means that individual variation in endorsement perception cannot be directly quantified. This constitutes a genuine threat to internal validity: although the between-group differences are consistent with the endorsement manipulation having been perceived as intended, it cannot be confirmed at the level of individual participants that the manipulation was interpreted as designed, nor can the magnitude of any unintended cross-condition variability in perceived endorsement be ruled out. Future replications of this design should therefore include an explicit manipulation-check item to verify, at the individual level, that participants perceived the endorsement condition as intended. Similarly, the validation evidence available for the Trust and Acceptance Scale was limited to expert-judged content and face validity together with internal-consistency reliability; in the absence of confirmatory factor analysis or discriminant/convergent validity testing, the structural validity of the four-factor model cannot be considered independently established. Finally, the cross-sectional design precludes inference about how trust in AI-generated content evolves across repeated exposure over a full academic semester—a trajectory that longitudinal research would be well-positioned to examine.

7. Conclusions

This study set out to determine whether teacher endorsement functions as a trust transfer mechanism in students' engagement with AI-generated educational content and to examine the conditions under which this mechanism operates most powerfully. Employing a between-subjects randomized experimental design with 320 university students, the findings answer this question affirmatively and with theoretical precision. Teacher endorsement produced statistically significant and practically meaningful differences across all four trust dimensions, with effects that were robust to gender covariate control and coherent with the theoretical architecture of Trust Transfer Theory. Behavioral Intention was most substantially influenced, Academic Integrity least so—a gradient that reflects the differential susceptibility of adoption orientations and moral-epistemic evaluations to interpersonal endorsement signals.

The regression analyses revealed that Perceived Human-Mediated Reliability is a structurally stable predictor of trust outcomes across both experimental conditions, with predictive power amplified in the endorsed condition—suggesting that endorsement strengthens rather than merely creates the coupling between perceived human oversight and trust. Academic level moderated the trust transfer process in theoretically interpretable ways, while the absence of disciplinary differences confirmed the cross-disciplinary universality of the mechanism and the consequent appropriateness of institution-wide rather than discipline-specific AI governance responses.

The central theoretical contribution of this study is an interpretive repositioning of the teacher as a trust guarantor within the sociotechnical ecology of educational AI—not a transitional figure whose role will diminish as AI improves, but a structurally necessary intermediary whose credibility and explicit validation appear to be essential conditions

for the legitimate integration of algorithmic knowledge production into the academic community, pending direct empirical verification of the assumed trust transfer pathway. Future scholarship should extend this inquiry through longitudinal, cross-cultural, and multi-method designs, with the aim of constructing a comprehensive and ecologically grounded account of how human–AI trust is formed, transferred, and sustained in the full diversity of global higher education contexts.

Funding: This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU), grant number IMSIU-DDRSP2604.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki. It was conducted using an anonymous electronic questionnaire involving no interventions and minimal risk, with no identifying data collected and participation was entirely voluntary. As such, it qualifies as exempt research. Informed consent was obtained from all participants prior to their participation.

Informed Consent Statement: Informed consent was embedded in the opening page of the survey instrument, prior to any study content or scale items. Participants were presented with a full disclosure of the study’s objectives, the voluntary nature of their participation, their unconditional right to withdraw at any time without consequence, and an explicit assurance that all responses would remain strictly anonymous and used solely for scientific research purposes. Proceeding to the survey by clicking the ‘Next’ button constituted an affirmative act of voluntary informed consent and confirmation that the participant was aged 18 or above.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to participant privacy considerations, as the survey responses—though anonymized—include demographic combinations that could potentially identify participants within small subgroups.

Acknowledgments: AI-assisted tools were used during the preparation of this manuscript solely for language editing, writing development, and literature search support. All AI-generated suggestions were critically reviewed and verified by the author, who bears full responsibility for the accuracy, integrity, and originality of the manuscript content. AI tools were not used in the generation of data, results, or scientific conclusions.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

PAC	Perceived AI Competence
PHMR	Perceived Human-Mediated Reliability
BI	Behavioral Intention
AI-scale	Academic Integrity subscale
TTT	Trust Transfer Theory
ANCOVA	Analysis of Covariance

References

Aladi, C. (2024). IT higher education teachers and trust in AI-enabled Ed-tech: Implications for adoption of AI in higher education. In *Proceedings of the 2024 computers and people research conference*. ACM. [CrossRef]

Aly, H., Volonte, M., Byrne, K. A., & Knijnenburg, B. P. (2025). Bridging the trust gap: Investigating the role of trust transfer in the adoption of AI instructors for digital privacy education. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1–22). ACM. [CrossRef]

- Ayanwale, M., Adelana, O., & Odufuwa, T. (2024). Exploring STEAM teachers' trust in AI-based educational technologies: A structural equation modelling approach. *Discover Education*, 3, 92. [\[CrossRef\]](#)
- Balaskas, S., Tsiantos, V., Chatzifotiou, S., & Rigou, M. (2025). Determinants of ChatGPT adoption intention in higher education: Expanding on TAM with the mediating roles of trust and risk. *Information*, 16, 82. [\[CrossRef\]](#)
- Black, R., & Tomlinson, B. (2025). University students describe how they adopt AI for writing and research in a general education course. *Scientific Reports*, 15, 8934. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. Guilford Press.
- Chan, C., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. [\[CrossRef\]](#)
- Choi, S., Jang, Y., & Kim, H. (2022). Influence of pedagogical beliefs and perceived trust on teachers' acceptance of educational artificial intelligence tools. *International Journal of Human-Computer Interaction*, 39, 910–922. [\[CrossRef\]](#)
- Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction*, 39, 1727–1739. [\[CrossRef\]](#)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Deric, E., Frank, D., & Milković, M. (2025). Trust in generative AI tools: A comparative study of higher education students, teachers, and researchers. *Information*, 16, 622. [\[CrossRef\]](#)
- Fošner, A. (2024). University students' attitudes and perceptions towards AI tools: Implications for sustainable educational practices. *Sustainability*, 16, 8668. [\[CrossRef\]](#)
- Goldshtein, M., Schroeder, N., & Chiou, E. (2025). The role of learner trust in generative artificially intelligent learning environments. *Journal of Engineering Education*, 114, e70000. [\[CrossRef\]](#)
- Guo, Y., Yang, J., & Shi, C. (2023). TIP: A trust inference and propagation model in multi-human multi-robot teams. In *Companion of the 2023 ACM/IEEE international conference on human-robot interaction*. ACM. [\[CrossRef\]](#)
- Herdiani, A., Mahayana, D., & Rosmansyah, Y. (2024). Building trust in an artificial intelligence-based educational support system: A narrative review. *Jurnal Sositologi*, 23, 101–119. [\[CrossRef\]](#)
- Johnston, H., Wells, R., Shanks, E., Boey, T., & Parsons, B. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity*, 20, 14. [\[CrossRef\]](#)
- Johri, A., Hingle, A., & Schleiss, J. (2024). Misconceptions, pragmatism, and value tensions: Evaluating students' understanding and perception of generative AI for education. In *2024 IEEE frontiers in education conference (FIE)* (pp. 1–9). IEEE. [\[CrossRef\]](#)
- Jøsang, A., Marsh, S., & Pope, S. (2006). Exploring different types of trust propagation. In K. Stølen, W. H. Winsborough, F. Martinelli, & F. Massacci (Eds.), *Trust management. iTrust 2006. Lecture notes in computer science* (Vol. 3986, pp. 179–192). Springer. [\[CrossRef\]](#)
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kizilcec, R. (2023). To advance AI use in education, focus on understanding educators. *International Journal of Artificial Intelligence in Education*, 34, 12–19. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lin, Y., & Yu, Z. (2023). Extending technology acceptance model to higher-education students' use of digital academic reading tools on computers. *International Journal of Educational Technology in Higher Education*, 20, 34. [\[CrossRef\]](#)
- Liu, Y., Liang, C., Chiclana, F., & Wu, J. (2021). A knowledge coverage-based trust propagation for recommendation mechanism in social network group decision making. *Applied Soft Computing*, 101, 107005. [\[CrossRef\]](#)
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University Press.
- Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022a). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, 53, 914–931. [\[CrossRef\]](#)
- Nazaretsky, T., Cukurova, M., & Alexandron, G. (2022b). An instrument for measuring teachers' trust in AI-based educational technology. In *Proceedings of the LAK22: 12th international learning analytics and knowledge conference* (pp. 56–66). ACM. [\[CrossRef\]](#)
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Omrali, A., Kozhakhmet, K., & Zhumaliyeva, R. (2025). Digital trust in transition: Student perceptions of AI-enhanced learning for sustainable educational futures. *Sustainability*, 17, 7567. [\[CrossRef\]](#)
- Ouyang, Z., Bai, X., Huang, L., Li, Q., & Li, Y. (2025). Self-motivated effects of teachers' supportive behaviors on students' intentions of online continuous learning. *PLoS ONE*, 20, e0324731. [\[CrossRef\]](#) [\[PubMed\]](#)
- Pitts, G., & Motamedi, S. (2024). Understanding human-AI trust in education. *arXiv*. [\[CrossRef\]](#)
- Pitts, G., & Motamedi, S. (2025). Understanding human-AI trust in education. *Telematics and Informatics Reports*, 20, 100270. [\[CrossRef\]](#)
- Qin, F., Li, K., & Yan, J. (2020). Understanding user trust in artificial intelligence-based educational systems: Evidence from China. *British Journal of Educational Technology*, 51, 1693–1710. [\[CrossRef\]](#)
- Rana, M., Siddiquee, M., Sakib, M., & Ahamed, M. (2024). Assessing AI adoption in developing country academia: A trust and privacy-augmented UTAUT framework. *Heliyon*, 10, e37569. [\[CrossRef\]](#) [\[PubMed\]](#)
- Saleh, S., Nat, M., & Aqel, M. (2022). Sustainable adoption of e-learning from the TAM perspective. *Sustainability*, 14, 3690. [\[CrossRef\]](#)

- Shahzad, M., Xu, S., & Zahid, H. (2024). Exploring the impact of generative AI-based technologies on learning performance through self-efficacy, fairness, ethics, creativity, and trust. *Education and Information Technologies*, 30, 3691–3716. [[CrossRef](#)]
- Shata, A., & Hartley, K. (2025). Artificial intelligence and communication technologies in academia: Faculty perceptions and the adoption of generative AI. *International Journal of Educational Technology in Higher Education*, 22, 12. [[CrossRef](#)]
- Shen, L., Qiu, N., & Wang, Z. (2025). Psychological safety and trust as drivers of teachers' continued use of AI tools in classrooms. *Scientific Reports*, 15, 9023. [[CrossRef](#)] [[PubMed](#)]
- Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31, 47–53.
- Song, D. (2024). How learners' trust changes in generative AI over a semester of undergraduate courses. *International Journal of Artificial Intelligence in Education*, 35, 1445–1464. [[CrossRef](#)]
- Stewart, K. J. (2003). Trust transfer on the World Wide Web. *Organization Science*, 14, 5–17. [[CrossRef](#)]
- Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., Wasson, B., Tømte, C., Spikol, D., Milrad, M., Coelho, R., & Kizilcec, R. F. (2024). What explains teachers' trust in AI in education across six countries? *International Journal of Artificial Intelligence in Education*, 35, 1288–1316. [[CrossRef](#)]
- Wang, R., & Chan-Olmsted, S. (2024). Brand communication through social media influencers: Trust building and trust transfer mechanisms. *International Journal of Business Communication*, 63, 761–782. [[CrossRef](#)]
- Wiyata, H., Hung, Y., & Chiu, C. (2022). Cross-border social commerce: From a trust transfer perspective. *Journal of Electronic Commerce Research*, 23, 185–202.
- Xiao, L., Zhang, Y., & Fu, B. (2019). Exploring the moderators and causal process of trust transfer in online-to-offline commerce. *Journal of Business Research*, 98, 214–226. [[CrossRef](#)]
- Xu, Y., Chen, X., Nicolau, J., & Luo, P. (2023). Trust transfer in peer-to-peer accommodation. *Annals of Tourism Research*, 100, 103603. [[CrossRef](#)]
- Zhao, J., Huang, J., & Su, S. (2019). The effects of trust on consumers' continuous purchase intentions in C2C social commerce: A trust transfer perspective. *Journal of Retailing and Consumer Services*, 53, 101975. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.