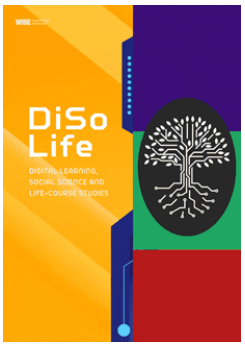
WISE Pendidikan Indonesia



DIGITAL LEARNING, SOCIAL SCIENCE, AND LIFE-COURSE STUDIES
VOL. 2 NO. 1 (2026)

ISSN: 3109-8096

Explainable AI, Academic Confidence, and Decision-Making Quality in Physics Learning

Nur Hidayah* and **Sodikin**

To cite this article. N. Hidayah and Sodikin, "Explainable AI, Academic Confidence, and Decision-Making Quality in Physics Learning," *Dig. Soc. Life.*, vol. 2, no. 1, pp. 28–45, 2026.
DOI: https://doi.org/10.70211/disolife.v2i1.777

To link to this article:

Published online: 30 June 2026

Submit an article to this journal

View crossmark data

OPEN ACCESS

https://doi.org/

Digital Learning, Social Science, and Life-course Studies Vol 2 No 1 (2026)

Check for updates

Research Article

Explainable AI, Academic Confidence, and Decision-Making Quality in Physics Learning

Nur Hidayah* and Sodikin

Received: 7 March 2026	Revised: 6 May 2026	Accepted: 26 May 2026	Online: 30 June 2026

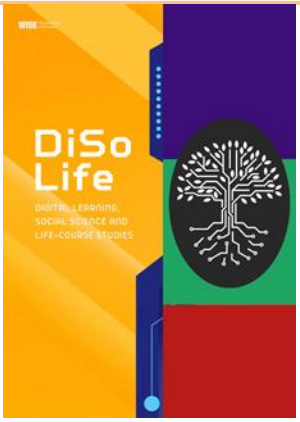
Abstract

Opaque artificial intelligence (AI) recommendations may undermine learners' ability to evaluate feedback and make informed decisions. This study examined whether an explainable artificial intelligence (XAI) learning system was associated with higher academic confidence and decision-making quality than a comparable AI system without explanatory output during undergraduate physics learning. A quantitative quasi-experimental pretest-posttest control-group design involved 120 students (mean age = 20.3 years, SD = 1.4; 54.2% female), with 60 students analyzed in each condition across eight weeks and treatment implemented through three intact courses. Academic confidence was measured using an 18-item study-adapted Academic Confidence Scale-Short Form (alpha = .87), whereas decision-making quality was assessed using a 16-item Learning Task Decision-Making Inventory (alpha = .83). ANCOVA controlling for corresponding pretest scores showed higher posttest academic confidence in the XAI condition, $F(1, 117) = 52.34$, $p < .001$, partial eta squared = .309, and higher decision-making quality, $F(1, 117) = 38.71$, $p < .001$, partial eta squared = .249. The results support an association between the explanatory-interface condition and both learner outcomes, but they do not establish confidence calibration, deliberative processing, or the proposed transparency mechanism. Interpretation is constrained by course-level allocation across only three clusters, unmodeled course dependence, a single-institution sample, study-adapted questionnaire outcomes, and the absence of a manipulation check or explanation-fidelity evidence.

Keywords: algorithmic transparency; academic confidence; AI literacy; higher education; learner agency

Publisher's Note:

WISE Pendidikan Indonesia stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



INTRODUCTION

Artificial intelligence (AI) has become a prominent component of contemporary higher education, spanning personalized support, automated feedback, assessment, conversational assistance, learning analytics, and institutional decision processes. Recent syntheses document a rapid expansion of AI research in universities while repeatedly identifying uneven methodological quality, limited longitudinal evidence, and a need to connect technical innovation with educational theory and learner outcomes [1], [2], [3]. Institutional responses have also moved beyond tool adoption toward questions of policy, governance, assessment, and responsible use, reflecting the speed with which generative AI has entered teaching and learning practices [4], [5]. This expansion makes it increasingly important to distinguish the mere availability of AI from pedagogically designed AI interactions that preserve students' capacity to understand, evaluate, and act on algorithmic support.

Student-facing research reinforces that distinction. University students often report substantial perceived benefits from generative AI, including rapid access to explanations, personalized assistance, idea generation, and efficiency, but they also identify concerns about accuracy, dependency, academic integrity, privacy, and loss of independent thinking [6], [7]. Empirical adoption studies further show that perceived benefit, trust, novelty, technology readiness, and contextual factors shape whether students engage with AI and how they use it [8], [9], [10]. Importantly, high use is not necessarily educationally beneficial: intensive or poorly regulated generative-AI use has been associated with procrastination and weaker academic patterns in some samples, while trust can produce both productive reliance and problematic over-reliance [11], [12]. These findings shift the pedagogical problem from whether students will use AI to whether AI interfaces help them make well-calibrated judgments while they use it.

AI literacy and competence are central to this problem because learners need enough conceptual and procedural understanding to interrogate AI outputs rather than treat them as authoritative answers. Recent higher-education research emphasizes AI literacy, critical thinking, and prompt competence as prerequisites for meaningful participation in AI-mediated learning [13], [14]. First-year students' AI competence and perceived benefits predict intended and actual learning-oriented AI use, indicating that learner characteristics materially shape engagement [15]. Related work on faculty AI self-efficacy and continuous use similarly demonstrates that capability beliefs, support, and perceived usefulness influence sustained adoption [16], [17]. Ethical competence is equally important: students and educators identify fairness, accountability, privacy, and trust as design requirements rather than peripheral concerns [18], [19], [20]. Consequently, an educational AI system should not only generate useful content; it should make its support sufficiently interpretable for users to evaluate it responsibly.

Recent research on AI-supported feedback and self-regulated learning offers a concrete pedagogical pathway for such interpretability. Students describe AI applications as potentially useful for cognitive, metacognitive, and behavioral regulation when those applications preserve learner activeness and positioning rather than replacing regulation [21]. In writing, AI-generated feedback can be clear and detailed, although controlled comparisons show that it does not automatically outperform human feedback, underscoring the importance of feedback design rather than source alone [22]. AI-enhanced peer assessment research likewise emphasizes feedback literacy, evaluative judgment, and active uptake as mechanisms through which automated support can become educationally productive [23]. A recent multisite experiment in university science courses provides especially relevant evidence: reflective and hybrid GenAI feedback designs

strengthened feedback uptake, self-regulated learning, conceptual learning, and delayed AI-free transfer more consistently than direct AI critique alone [24]. These findings suggest that explanatory support may matter when it prompts students to inspect reasons and retain ownership of decisions.

Human-AI collaboration studies also show that outcomes depend on how students interact with AI rather than on access alone. Students' experiences of collaborative content creation reveal both opportunities for ideation and concerns about authorship, control, and the distribution of cognitive work [25]. Dialogic engagement with generative chatbots can support creative problem solving and practical competence when students use the system iteratively rather than as a one-shot answer generator [26]. Recent feedback research also shows that students' perceptions of whether feedback is AI- or tutor-generated can influence their evaluations, while GenAI feedback may support selected self-regulated learning processes when embedded in instructional support [27]. Broader reviews of educational chatbots likewise conclude that immediate assistance and personalized learning are promising but that reliability, accuracy, and ethical risks remain persistent constraints [28]. Performance studies further show that generative AI can produce strong outputs on some assessments while remaining variable across task types and levels, reinforcing the need for learner verification rather than passive acceptance [29], [30].

Explainable artificial intelligence (XAI) provides a principled response to this verification problem by attempting to make the basis of an AI recommendation, classification, or feedback output intelligible to its user. In education, the XAIED framework argues that explanations should be evaluated according to pedagogical goals and stakeholder needs rather than technical transparency alone [31]. Recent human-AI research outside education further shows that explanations can affect decision accuracy, trust, and reliance, but their value depends on explanation quality, user capability, and whether the explanation enables meaningful verification rather than simply increasing confidence [38], [39], [40]. The present study therefore uses three complementary theoretical lenses with distinct roles. XAIED frames explanation as a learner-centered design feature [31]; social-cognitive theory motivates the academic-confidence outcome because interpreted performance information can shape capability judgments [34]; and dual-process theory motivates decision-making quality because visible reasons may create opportunities for more deliberate evaluation of AI advice [35]. These are interpretive mechanisms only: confidence calibration, deliberative processing, and cognitive transparency were not measured directly.

Despite the fast growth of AI research in higher education, controlled evidence linking explanatory AI features to learner-centered outcomes in discipline-based learning remains limited. Much recent work has examined perceptions, adoption, literacy, ethics, output quality, or general chatbot use [6], [13], [36], while fewer studies isolate explanatory versus non-explanatory AI as the instructional contrast and examine academic confidence together with decision-making quality. Physics provides a relevant context because students must coordinate conceptual, representational, and quantitative reasoning while deciding whether an external recommendation is consistent with disciplinary principles. The present study therefore examines whether an XAI-enhanced learning condition is associated with higher posttest academic confidence (H1) and decision-making quality (H2) than a conventional AI condition after adjustment for the corresponding pretest score. AI literacy is retained only as an ancillary exploratory variable and is not treated as a confirmatory moderator. Figure 1 summarizes only the relationships tested confirmatorily and separates them from the theoretical lenses used for interpretation.

Recent research further indicates that responsible educational AI requires alignment among system design, institutional policy, learner capability, and pedagogical purpose. Studies of AI ethics and assessment emphasize transparent responsibility and meaningful human oversight [18], [19], [20], while work on adoption and teaching practice shows that sustained AI use depends on perceived usefulness, competence, and contextual support [17], [37]. These converging findings reinforce the need to examine explainability as an instructional affordance rather than as a purely technical property. Accordingly, the present study contributes a learner-centered comparison of explanatory and non-explanatory AI conditions during structured physics learning; it does not claim to validate a particular XAI algorithm or to establish an unmeasured transparency mechanism.

Figure 1 presents the revised conceptual model used to organize the study. The model shows the experimental contrast an AI interface with explanatory output versus a comparable interface without explanatory output and the two confirmatory outcomes: academic confidence and decision-making quality. Social cognitive theory, dual-process theory, and the XAIED framework are shown as interpretive foundations rather than tested mediators. AI literacy is identified separately as an ancillary baseline variable because no moderation analysis was conducted.

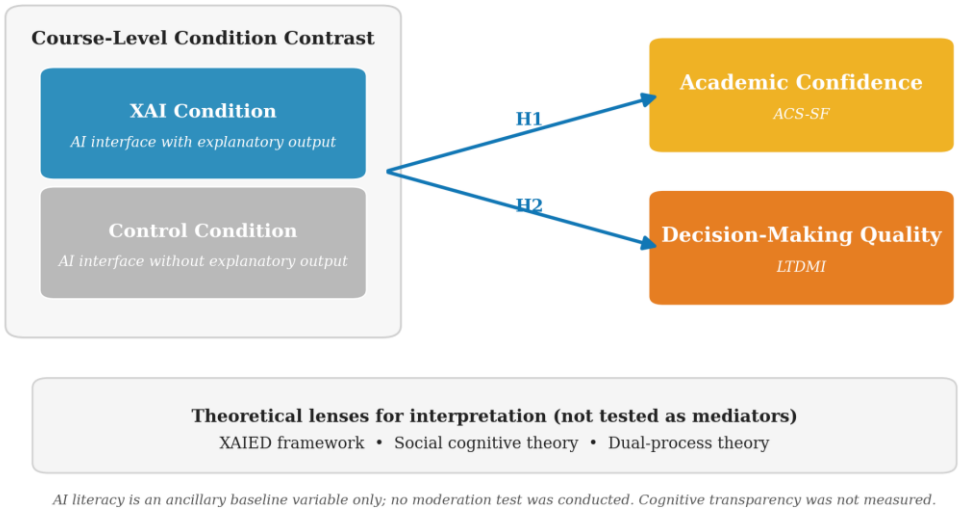


Figure 1. Conceptual model distinguishing the confirmatory condition-to-outcome pathways from theoretical interpretive lenses and the ancillary AI-literacy variable.

As depicted in Figure 1, H1 and H2 are the only confirmatory pathways evaluated in the reported analyses. The figure deliberately avoids mediation and moderation arrows because cognitive transparency was not measured and AI literacy was not tested in a condition-by-literacy interaction. This distinction keeps the conceptual rationale aligned with the empirical model and prevents untested mechanisms from being presented as demonstrated relationships.

METHODS

Research Design

The study employed a quantitative quasi-experimental pretest-posttest control-group design. Participants were recruited nonprobabilistically from three intact introductory physics courses, and condition allocation was implemented at the course level through random drawing. Because treatment was delivered through intact courses rather than through individual randomization, the

design was classified as quasi-experimental. The available study record does not preserve course-specific sample sizes or sufficient allocation detail to determine whether instructor or course effects were fully separated from treatment condition; these uncertainties are therefore treated as design limitations. The primary analyses compared posttest outcomes between the XAI and conventional-AI conditions while adjusting for the corresponding pretest score. This approach provides covariate-adjusted student-level comparisons while preserving the natural course-based instructional setting, but it does not remove the clustering limitation created by course-level delivery.

Population and Sample

Participants were undergraduate students taking physics education courses at a public Islamic university in Lampung Province, Indonesia. Eligibility required enrollment in a science-related course during the study period, regular access to university computing facilities, and no prior formal exposure to XAI tools or interfaces. The protocol excluded students who required modified assessment administration because the study measures had not been standardized for individualized administration. This exclusion was measurement-driven rather than a claim of educational ineligibility, and it may limit generalizability to students requiring assessment accommodations; the available record does not specify the exact modifications that would have been required. Of 129 eligible/enrolled students approached, 120 participated (93.0%): six withdrew from the course and three declined consent. The final sample comprised 65 women (54.2%) and 55 men (45.8%), with a mean age of 20.3 years (SD = 1.4; range = 18-24). Sixty students were analyzed in each condition.

Table 1 summarizes demographic and baseline characteristics. The table is included to establish the composition of the analytic sample and to document that the two conditions were closely balanced before the intervention on age, study year, gender distribution, and baseline AI literacy.

Table 1. Demographic and baseline characteristics of participants by condition

Variable	XAI (n = 60)	Control (n = 60)	Total (N = 120)
Gender			
Female	33 (55.0%)	32 (53.3%)	65 (54.2%)
Male	27 (45.0%)	28 (46.7%)	55 (45.8%)
Age, M (SD)	20.2 (1.3)	20.4 (1.5)	20.3 (1.4)
Study year			
2nd year	28 (46.7%)	29 (48.3%)	57 (47.5%)
3rd year	22 (36.7%)	21 (35.0%)	43 (35.8%)
4th year	10 (16.7%)	10 (16.7%)	20 (16.7%)
AI literacy, M (SD)	23.6 (4.8)	23.2 (4.6)	23.4 (4.7)

Note. AI literacy scores ranged from 10 to 50. Descriptive values were similar across conditions; exploratory baseline tests were nonsignificant (all p values > .50). These tests are not interpreted as evidence of statistical equivalence.

The descriptive balance shown in Table 1 indicates that the analyzed conditions were similar on the reported baseline characteristics, but these descriptive patterns and nonsignificant baseline tests should not be interpreted as proof of equivalence. They also do not address possible course-level dependence or unmeasured differences among the three intact courses. The study's single-institution convenience recruitment further means that treatment allocation should not be interpreted as random sampling from the wider population of Indonesian undergraduates.

An a priori G*Power 3.1 calculation was conducted for ANCOVA with one covariate, assuming a medium effect (f = 0.25), alpha = .05, and power = .80, and indicated a minimum of 52 participants per condition (N = 104). This calculation was performed at the individual-participant

level and did not incorporate a cluster design effect; it should therefore be interpreted as the original individual-level planning target rather than evidence of adequate power for course-level assignment. The achieved sample of N = 120 exceeded that target. All 120 analyzed participants completed both assessment occasions, so no missing-data imputation was reported.

Intervention and Control Conditions

The intervention lasted eight weeks. In Week 1, students completed baseline assessments. During the subsequent six instructional weeks, students completed four structured physics learning tasks covering kinematics, dynamics, waves, and electromagnetism; the task sequence and instructional period were the same across conditions according to the study protocol. The XAI condition used an AI-based learning system that displayed real-time natural-language explanations for its recommendations through an interface described in the protocol as LIME-inspired (Local Interpretable Model-agnostic Explanations). The available documentation does not permit reconstruction of the underlying model architecture, feature-weighting procedure, mapping from model output to natural-language rationale, explanation fidelity, or uncertainty calibration. The conventional-AI condition was described as comparable except for explanatory output, but the retained records are insufficient to verify equivalence in recommendation quality, response latency, or exposure time. Week 8 was devoted to posttest assessment under conditions matching the pretest administration. Accordingly, the intervention is interpreted operationally as an explanatory-interface versus non-explanatory-interface contrast, not as validation of LIME or any specific XAI algorithm.

Research Instruments

Academic confidence was assessed with an 18-item study-adapted Academic Confidence Scale-Short Form (ACS-SF). Items used a five-point Likert-type response format from 1 (strongly disagree) to 5 (strongly agree), producing composite scores from 18 to 90, with higher scores indicating greater academic confidence. Internal consistency in the study sample was alpha = .87. The available study record does not preserve a complete source and adaptation log for the exact short form administered; the measure is therefore treated as study-adapted rather than as independently cross-culturally validated. Decision-making quality was assessed using the 16-item Learning Task Decision-Making Inventory (LTDMI), which operationalized accuracy, analytical flexibility, and deliberateness in problem-solving decisions. Composite scores ranged from 16 to 80, with higher scores indicating better decision-making quality; internal consistency was alpha = .83. The retained protocol likewise does not establish the LTDMI as an independently validated external instrument. Both measures were pilot-tested with 20 nonparticipant students for cultural and linguistic appropriateness. The study record reports this pilot but does not retain an item-level revision log or a documented translation/back-translation procedure. A two-week stability pilot yielded r = .81 for academic confidence and r = .78 for decision-making quality. These indices support internal consistency and short-term stability only; the small pilot does not establish factor structure, cross-cultural measurement equivalence, or measurement invariance.

AI literacy was recorded as an ancillary baseline composite with a reported possible range of 10 to 50. The study protocol did not document item-level content or a separate validation source for this measure. For that reason, AI literacy is retained only for a hypothesis-generating ancillary analysis and is excluded from confirmatory claims about moderation, mechanism, or primary outcomes.

Table 2 consolidates the psychometric and descriptive information for the primary outcome measures. Presenting the measures together clarifies their scoring ranges and prevents reliability claims from being separated from the scales to which they apply.

Table 2. Psychometric and descriptive properties of the primary outcome measures

Scale	Items	Pretest M	SD	Observed range	α	Evidence in study
Academic Confidence Scale-Short Form (ACS-SF)	18	62.8	8.1	25-85	.87	Study-adapted; α = .87; pilot n = 20; factor validity not established
Learning Task Decision-Making Inventory (LTDMI)	16	58.9	8.9	20-76	.83	Study-adapted; α = .83; pilot n = 20; factor validity not established

Note. M and SD are aggregated across conditions at pretest. α = Cronbach's alpha coefficient. Pilot-test n = 20 nonparticipants.

As Table 2 shows, both primary scales demonstrated acceptable internal consistency in the analytic sample and encouraging short-term stability in the small pilot. These results are reliability evidence, not a substitute for construct validation. Claims about factor structure, cross-cultural validity, or measurement invariance therefore remain outside the evidence available in this study.

Data Collection Procedure and Ethics

Assessments were administered online during scheduled class sessions. Research assistants received standardized administration training in two sessions and were reported to remain unaware of participant condition during assessment administration. This blinding statement applies to assessment administration; the available records do not establish blinding during intervention delivery, when the explanatory interface itself would necessarily differ by condition. The study received approval from the Institutional Review Board of UIN Raden Intan Lampung (Protocol No. IRB-UINRIL-2024-087). All participants provided written informed consent, participation was voluntary, no course credit or tangible incentive depended on participation, and withdrawal was permitted without penalty. Data were anonymized with numeric codes; identifying information was stored separately in a password-protected repository accessible to the principal investigator.

Data Analysis

Primary hypotheses were evaluated with separate ANCOVA models for academic confidence and decision-making quality. In each model, posttest score was the dependent variable, condition (XAI vs. conventional AI) was the between-participant factor, and the corresponding pretest score was the covariate. The reported homogeneity-of-regression-slopes tests were nonsignificant for academic confidence ($p = .43$) and decision-making quality ($p = .61$). The archived analysis record also reported Shapiro-Wilk tests on posttest scores and Levene tests for homogeneity of variance; residual-based diagnostics, linearity checks, and influence diagnostics were not retained, so these tests provide only partial evidence regarding model assumptions. Effect magnitude is reported using partial eta squared from the ANCOVA models. Previously reported adjusted Cohen's d values were removed because the computational procedure could not be independently reconstructed from the retained output. Analyses used alpha = .05. Because both primary condition tests were $p < .001$,

their statistical significance would also remain under a simple two-outcome Bonferroni sensitivity threshold of .025; this is a sensitivity observation rather than a preregistered correction. A supplementary AI-literacy regression is retained only as a hypothesis-generating ancillary analysis and is not used to support the confirmatory conclusions. Analyses were reported as conducted in IBM SPSS Statistics 29.0. Participants were nested within three intact courses; because the student-level ANCOVA treated observations as independent and only three course clusters were available, course-level dependence could not be estimated reliably. The resulting standard errors and p values should therefore be interpreted as conditional on the independence assumption.

RESULTS AND DISCUSSION

Participant Flow and Preliminary Analysis

All 120 consenting participants contributed complete pretest and posttest data, yielding 60 observations per condition and no missing values. The archived preliminary analysis reported no significant deviation from normality for posttest academic confidence ($W = .982$, $p = .112$) or posttest decision-making quality ($W = .976$, $p = .073$), and Levene's tests were nonsignificant for both posttest outcomes ($p = .29$ and $p = .41$, respectively). These checks indicate no obvious gross distributional or variance problem in the reported scores but do not substitute for full residual diagnostics. Baseline comparisons were also nonsignificant for academic confidence, $t(118) = 0.49$, $p = .628$, $d = 0.09$, and decision-making quality, $t(118) = 0.32$, $p = .749$, $d = 0.06$. These baseline tests are presented descriptively and are not interpreted as proof of group equivalence.

Table 3 presents the unadjusted pretest and posttest distributions by condition. These descriptive values show the direction and magnitude of change before inferential adjustment and provide the empirical basis for the pre-post visualization in Figure 2.

Table 3. Descriptive statistics for primary outcomes by condition and time point

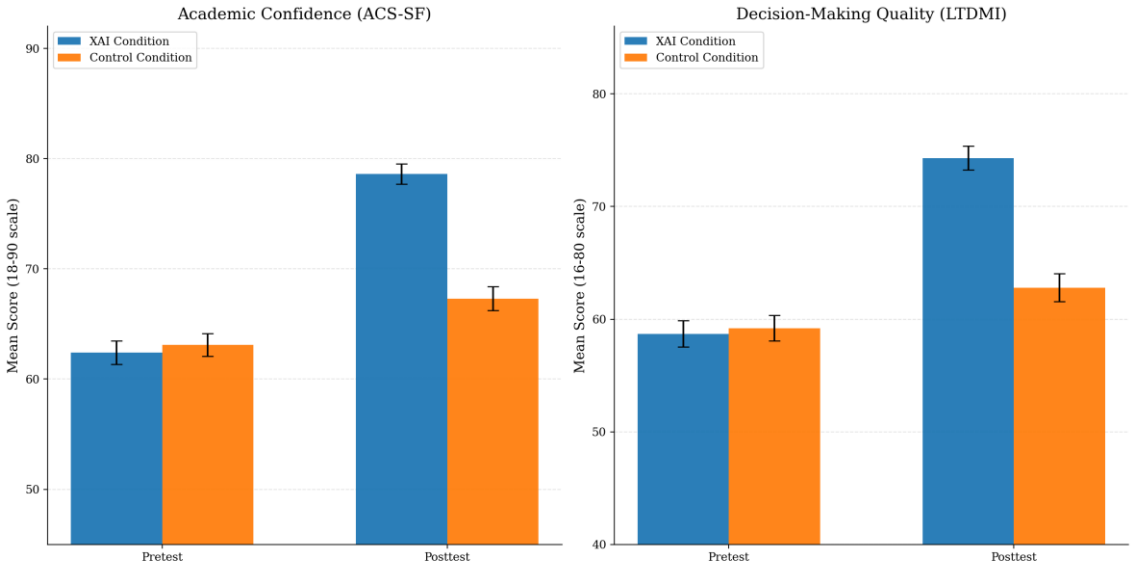
Measure / Time point	XAI M (SD)	XAI 95% CI	Control M (SD)	Control 95% CI
Academic confidence				
Pretest	62.4 (8.3)	[60.3, 64.5]	63.1 (7.9)	[61.1, 65.1]
Posttest	78.6 (7.1)	[76.8, 80.4]	67.3 (8.4)	[65.2, 69.4]
Decision-making quality				
Pretest	58.7 (9.1)	[56.4, 61.0]	59.2 (8.8)	[57.0, 61.4]
Posttest	74.3 (8.2)	[72.3, 76.3]	62.8 (9.6)	[60.4, 65.2]

Note. n = 60 per condition. M = mean; SD = standard deviation; CI = confidence interval.

The descriptive pattern in Table 3 is consistent across both outcomes. Academic confidence increased by 16.2 raw-score points in the XAI condition compared with 4.2 points in the control condition, while decision-making quality increased by 15.6 points in the XAI condition compared with 3.6 points in the control condition. The unadjusted difference in mean change happens to equal 12.0 raw-score points for both measures, but the instruments have different score ranges and represent different constructs; this numerical equality is therefore not interpreted as evidence of equivalent standardized effects. These change summaries are descriptive, and the primary inferential conclusions rely on ANCOVA rather than change-score testing.

Figure 2 visualizes the unadjusted pretest and posttest means with standard-error bars. It is included to show the descriptive trajectory of each condition across time. The figure is not an

inferential test and should be read alongside the covariate-adjusted ANCOVA results reported below.



Note. Error bars represent ±1 standard error of the mean. Values are unadjusted descriptives; inferential tests use ANCOVA controlling for pretest scores.

Figure 2. Unadjusted pre- and posttest means for academic confidence and decision-making quality by condition, with standard-error bars; inferential conclusions are based on ANCOVA.

Figure 2 shows minimal descriptive baseline separation and a larger unadjusted posttest difference favoring the XAI condition on both outcomes. Because the plotted means are unadjusted, the figure is used only to display the observed score pattern; statistical inference rests on the ANCOVA models, and the figure does not resolve the study's allocation or clustering limitations.

Academic Confidence and Decision-Making Quality

After adjustment for pretest academic confidence, the ANCOVA showed a statistically significant condition difference, $F(1, 117) = 52.34$, $p < .001$, partial eta squared = .309. The reported estimated marginal means were 78.4 (SE = 0.97) for the XAI condition and 67.5 (SE = 0.97) for the conventional-AI condition. H1 was supported at the level of the reported student-level ANCOVA comparison. Because course-level dependence was not modeled, the magnitude estimate is informative for the analyzed sample while the standard error and significance test remain conditional on the independence assumption.

The ANCOVA for decision-making quality likewise indicated a statistically significant condition difference after adjustment for pretest scores, $F(1, 117) = 38.71$, $p < .001$, partial eta squared = .249. Reported estimated marginal means were 74.1 (SE = 1.11) in the XAI condition and 63.0 (SE = 1.11) in the conventional-AI condition. H2 was supported at the level of the reported student-level ANCOVA comparison, subject to the same course-clustering limitation.

Table 4 brings the two ANCOVA models together so that the contribution of the baseline covariate and the condition term can be compared directly. Partial eta squared is retained as the reproducible model-based effect-size index available from the archived output; unreconstructable adjusted Cohen's d estimates have been removed.

Table 4. ANCOVA results for academic confidence and decision-making quality by condition

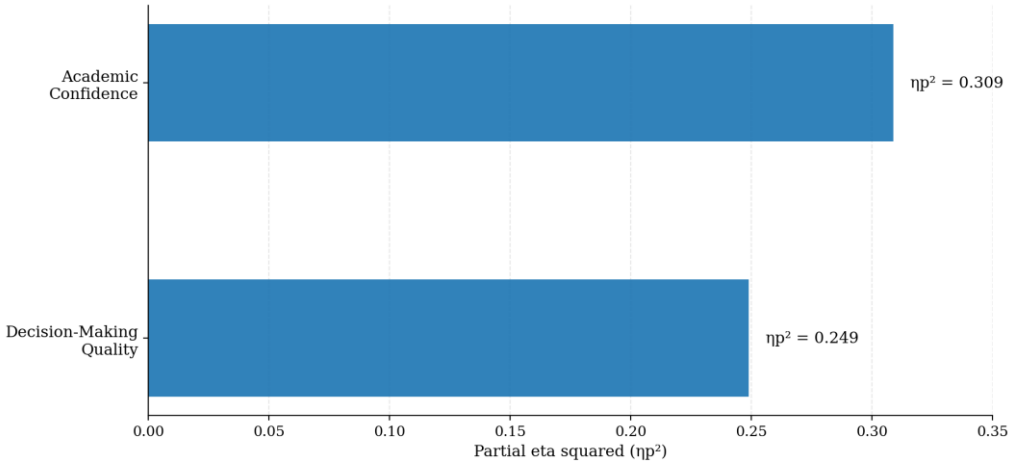
Outcome / Source	SS	df	MS	F	p	ηp²
Academic confidence						
Covariate (pretest)	2,847.3	1	2,847.3	51.26	< .001	.305

Outcome / Source	SS	df	MS	F	p	ηp²
Group (XAI vs. control)	2,908.1	1	2,908.1	52.34	< .001	.309
Error	6,500.4	117	55.6			
Decision-making quality						
Covariate (pretest)	2,103.7	1	2,103.7	34.52	< .001	.228
Group (XAI vs. control)	2,358.2	1	2,358.2	38.71	< .001	.249
Error	7,127.9	117	60.9			

Note. SS = sum of squares; MS = mean square; ηp² = partial eta squared. Each model adjusts for the corresponding pretest score. Partial eta squared is retained as the reproducible effect-size index available from the archived ANCOVA output.

As Table 4 indicates, the condition term accounted for substantial adjusted variance in both outcomes within the analyzed sample, with partial eta squared = .309 for academic confidence and .249 for decision-making quality. These values summarize model-based effect magnitude and should be interpreted alongside the course-level allocation and unmodeled dependence; the numerical difference between the two effect sizes is descriptive and was not formally tested.

Figure 3 compares the reported partial eta-squared values for the condition term across the two ANCOVA models. Using the same reproducible effect-size metric for both outcomes allows a direct descriptive comparison without relying on the previously unreconstructable standardized mean-difference calculations.



Reported ANCOVA condition-effect magnitudes. No confidence intervals or formal between-effect comparison were available.

Figure 3. Reported ANCOVA partial eta-squared values for the condition term across the two primary outcomes.

Figure 3 shows a larger reported partial eta-squared value for academic confidence than for decision-making quality. This is a descriptive comparison only: no confidence intervals for partial eta squared or formal between-effect test were available, and the figure does not correct for unmodeled course-level dependence.

Ancillary, Hypothesis-Generating Analysis of AI Literacy

For completeness, the originally reported supplementary regression examined posttest academic confidence using condition and baseline AI literacy as predictors, $F(2, 117) = 36.91$, $p < .001$, R squared = .387. This result is retained only as hypothesis-generating evidence because the AI-literacy scale is insufficiently documented and the model did not include pretest academic confidence, the key covariate used in the primary analysis. In addition, the archived coefficient labels do not permit unambiguous reconstruction of standardized versus unstandardized scaling, so

individual coefficient magnitudes are not compared here. The analysis does not constitute a moderation test and is not used to support H1, H2, or a claim that AI literacy changes the effect of explanatory output.

Interpretation of the Findings

The two primary ANCOVA comparisons indicate that students in the XAI condition had higher adjusted posttest academic confidence and decision-making quality than students using the non-explanatory AI condition. This pattern is educationally relevant because the study contrasts explanatory with non-explanatory AI rather than AI with no AI: the central question is how AI communicates support, not merely whether AI is present. Recent higher-education research similarly emphasizes learner agency, self-regulation, trust, and higher-order participation when evaluating AI-supported learning [2], [3]. Human-AI research also cautions that explanations can improve performance in some contexts while failing in others when they do not support verification or calibrated reliance [38], [39], [40]. Because participants were recruited from three intact courses and treatment was implemented at course level, the present findings are interpreted as controlled quasi-experimental evidence of an association between condition and outcomes rather than definitive individual-level randomized causal evidence.

The academic-confidence result is theoretically compatible with social-cognitive accounts in which interpreted information about performance and task capability can shape efficacy-related judgments [34]. Contemporary AI studies make this pathway plausible without demonstrating it in the present dataset. Faculty research links AI self-efficacy to technology-use patterns and professional-development needs [16], while first-year student research shows that AI competence and perceived benefits predict learning-oriented AI use [15]. Student acceptance studies also identify perceived usefulness and trust as important determinants of engagement [8], [9], [10]. An interpretable recommendation may therefore provide learners with a more usable basis for judging their understanding and next action. However, the outcome measured here is academic confidence, not a direct calibration index, and no comparison between subjective confidence and objective accuracy was collected. The result should therefore not be described as evidence that XAI improved confidence calibration.

The confidence finding also aligns with adjacent work on feedback and self-regulated learning. Students report that AI applications can support metacognitive, cognitive, and behavioral regulation when learners remain active in interpreting the support [21]. Controlled writing research shows that AI feedback can be valued for clarity and specificity yet does not automatically outperform human feedback [22], and AI-enhanced peer-assessment research foregrounds feedback literacy and evaluative judgment [23]. A recent multisite experiment likewise found that reflective and hybrid GenAI feedback arrangements promoted stronger uptake, self-regulated learning, and delayed transfer than direct GenAI critique alone [24]. These studies provide a pedagogical analogy rather than direct evidence for an XAI mechanism; they suggest why explanation may be useful when it promotes active evaluation, but the present study did not measure feedback uptake or reflective processing.

The decision-making result can be interpreted through a deliberative-processing lens. Dual-process theory distinguishes rapid intuitive responses from slower analytical evaluation [35], and explanatory AI can make reasons available for comparison with the learner's own disciplinary knowledge. Human-AI evidence is mixed in ways that are directly relevant here. Explanations improved decision accuracy in one quasi-experimental decision task, with effects depending on

users' task-related capacity [38], whereas a recent synthesis argues that explanations often fail to improve complementary performance when they do not enable users to verify the AI recommendation [39]. Natural-language explanations can also increase trust while leaving users unable to determine whether the AI is directing them toward the correct answer [40]. Thus, explanations may support better decisions when they stimulate verifiable scrutiny, but their mere presence cannot guarantee analytical engagement. The current study measured decision-making quality by questionnaire and did not observe the reasoning process directly.

Explainability, Trust Calibration, and Learner Agency

This qualification is important because recent evidence consistently warns against equating trust, acceptance, or fluent output with educational reliability. Students can develop both productive reliance and problematic over-reliance on generative AI [12], and intensive use can co-occur with procrastination or weaker academic patterns in some contexts [11]. Generative systems also vary in performance across assessment types and disciplinary tasks [29], [30]. Ethical studies consequently emphasize accountability, fairness, privacy, and transparent responsibility for AI-supported decisions [18], [19], [20]. The educational function of XAI should therefore be to support calibrated reliance: students should understand enough of the rationale to evaluate the output and remain responsible for the final decision.

The XAI-specific literature supports a distinction between explainability and persuasion. The XAIED framework conceptualizes educational explanation as a stakeholder- and goal-dependent design problem [31], while prescriptive learning-analytics research extends explainability from prediction toward actionable support [32] and domain-specific explanations have been shown to influence educators' trust and acceptance of AI-EdTech recommendations [33]. Human-centered work on calibrated XAI likewise emphasizes designing explanations to reduce over-trust and under-trust rather than simply maximizing acceptance [41], and recent metacognitive research argues that optimal human-AI decision making depends on sensitivity to when advice should be trusted [42]. The present findings complement this literature by focusing on academic confidence and decision-making quality. However, cognitive transparency was not measured, no manipulation check established that the XAI condition was perceived as more understandable, and explanation fidelity was not evaluated. The proposed mechanism therefore remains theoretically plausible but empirically untested in this dataset.

AI literacy remains a plausible boundary condition, but the present data cannot establish that role. Higher-education scholarship identifies AI competence, prompt literacy, and critical thinking as important for effective use [13], [14], and students differ in perceived benefits and patterns of engagement [6], [7]. The ancillary regression suggested an association between baseline AI literacy and posttest confidence, but the measure is insufficiently documented, pretest confidence was omitted from that model, and no condition-by-literacy interaction was tested. The result is therefore treated only as hypothesis-generating. Future work should use a validated AI-literacy instrument, include the corresponding baseline outcome, and prespecify interaction models before making moderation claims.

Practical Implications

The present study directly supports one practical inference: in this sample, the explanatory-interface condition was associated with higher adjusted academic-confidence and decision-making-quality scores than the non-explanatory interface. Broader implementation recommendations require support from the wider literature. Institutional policy frameworks are evolving as universities

respond to generative AI in teaching, assessment, research, and administration [4], [5], and research on educational chatbots and AI adoption continues to identify reliability, ethics, capability beliefs, and contextual support as important constraints [17], [28], [37]. On that broader basis, responsible XAI implementation should combine explanatory interfaces with explicit expectations for verification, opportunities to challenge AI recommendations, AI-literacy development, and assessment designs that preserve independent reasoning. Explanation should be treated as part of instructional design, not as a cosmetic transparency layer. In physics specifically, this means asking students to compare an AI rationale with governing principles, representations, assumptions, and calculation steps before accepting a recommendation.

Recent experimental evidence also supports the importance of feedback source and learner perception: GenAI feedback can support selected self-regulated learning processes, but its impact depends on how students perceive and use the feedback rather than on automation alone [27]. This reinforces the present study's implication that explanatory output should invite inspection, verification, and learner ownership instead of functioning as an authority cue.

Methodological Boundaries and Future Research

Future research should directly address the methodological weaknesses identified in the present study. Multi-institution studies should document the unit of allocation unambiguously, report course-specific sample sizes and instructor arrangements, use enough clusters when assignment occurs at class level, and analyze hierarchical data with models that account for course dependence. Outcome measurement should combine validated self-report scales with behavioral indicators such as explanation inspection, error detection, decision revision, justification quality, and AI-free transfer. Experimental manipulations should vary explanation modality, detail, uncertainty disclosure, and opportunities for learner contestation, while process measures test whether transparency, feedback uptake, metacognitive sensitivity, or self-regulation mediates outcomes [39], [41], [42]. Recent assessment and AI-literacy reviews likewise underscore the need for stronger evidence connecting design features to learning rather than relying on attitudes alone [36], [1]. The present study should therefore be interpreted as a controlled quasi-experimental comparison that motivates more rigorous XAIED experimentation rather than as a definitive test of a causal mechanism.

Design and sampling limitations constrain inference. Treatment was implemented across only three intact courses, course-specific allocation details and possible instructor confounding could not be fully reconstructed, and the student-level ANCOVA did not model course dependence. With only three clusters, a reliable multilevel correction could not be estimated from the retained analysis, so any non-negligible intraclass correlation could affect standard-error precision. Participants were recruited by convenience from one institution, limiting population generalizability, and the individual-level power calculation did not incorporate a cluster design effect. The eight-week duration also does not establish persistence of the observed differences.

Measurement, intervention, and analysis limitations are also important. Both primary outcomes relied on study-adapted questionnaires; internal consistency and short-term stability were encouraging, but the small pilot and incomplete adaptation records do not establish factor structure, cross-cultural validity, or measurement invariance, and decision-making quality was not measured behaviorally under genuine uncertainty. The LIME-inspired explanation interface could not be reconstructed at the level of model architecture, feature weighting, rationale generation, explanation fidelity, or uncertainty calibration, and no manipulation check verified that students perceived the

condition as more understandable or transparent. AI literacy was insufficiently documented and its ancillary regression omitted pretest confidence. Finally, the archived analysis did not retain full residual diagnostics. These limitations are why the manuscript reports partial eta squared from the ANCOVA output but removes the previously reported adjusted Cohen's d estimates whose computational procedure could not be reproduced.

CONCLUSION

This study provides controlled quasi-experimental evidence that, within the analyzed sample, an AI learning interface with explanatory output was associated with higher adjusted posttest academic confidence and decision-making quality than a comparable interface without explanatory output during undergraduate physics learning. The contribution is more specific than a general claim that AI is beneficial: it suggests that how AI communicates reasons may matter for learner-centered outcomes. The evidence does not, however, establish confidence calibration, deliberative processing, cognitive transparency, or a particular XAI algorithm as the mechanism. Course-level treatment across only three intact classes was not modeled hierarchically, the sample came from one institution, both outcomes were study-adapted questionnaires, intervention fidelity and perceived transparency were not directly verified, and the ancillary AI-literacy analysis was not sufficiently documented for confirmatory interpretation. Replication should therefore use clearly documented allocation, adequate numbers of clusters, multilevel analysis where appropriate, validated self-report and behavioral decision measures, direct transparency and verification measures, and AI-free transfer assessments. Under those conditions, XAI can be tested more directly as a pedagogical scaffold for calibrated reliance and deliberative judgment rather than assumed to be beneficial simply because explanations are displayed.

AUTHOR INFORMATION

Corresponding Author

Nur Hidayah – Faculty of Tarbiyah and Teacher Training,, Universitas Islam Negeri Raden Intan Lampung (Indonesia).
orcid.org/0009-0006-3221-6473
Email: nurhidayahpspf@gmail.com

Authors

Nur Hidayah – Faculty of Tarbiyah and Teacher Training,, Universitas Islam Negeri Raden Intan Lampung (Indonesia).
orcid.org/0009-0006-3221-6473
Email: nurhidayahpspf@gmail.com

Sodikin – Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Raden Intan Lampung, (Indonesia).
orcid.org/0009-0000-3994-3426
Email: sodikin@radenintan.ac.id

AUTHOR CONTRIBUTION

N.H. contributed to the conceptualization of the study, research design, data collection, data analysis, interpretation of findings, and preparation of the original manuscript draft. S. contributed to the development of the theoretical framework, validation of the research instrument, supervision of the research process, critical revision of the manuscript, and final approval of the version to be published. Both authors reviewed and approved the final manuscript and agreed to be accountable for all aspects of the work.

CONFLICT OF INTEREST

The authors declare no conflict of interest. This research received no external funding.

DECLARATION OF USE OF AI IN SCIENTIFIC WRITING

Generative AI (ChatGPT, OpenAI) was used as an editorial support tool to assist with language organization, academic wording, consistency checking, and manuscript formatting. The authors reviewed the revised scientific claims, retained statistical values, interpretations, and reference metadata and retain full responsibility for the final manuscript. Generative AI was not used to create, alter, or fabricate the research data reported in this study.

REFERENCES

[1] H. Crompton and D. Burke, “Artificial intelligence in higher education: The state of the field,” International Journal of Educational Technology in Higher Education, vol. 20, Art. no. 22, 2023. Available: https://doi.org/10.1186/s41239-023-00392-8

[2] M. Bond, H. Khosravi, M. de Laat, N. Bergdahl, V. Negrea, E. Oxley, et al., “A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour,” International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 4, 2024. Available: https://doi.org/10.1186/s41239-023-00436-z

[3] O. Zawacki-Richter, J. Y. H. Bai, K. Lee, P. J. Slagter van Tryon, and P. Prinsloo, “New advances in artificial intelligence applications in higher education?,” International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 32, 2024. Available: https://doi.org/10.1186/s41239-024-00464-3

[4] C. K. Y. Chan, “A comprehensive AI policy education framework for university teaching and learning,” International Journal of Educational Technology in Higher Education, vol. 20, Art. no. 38, 2023. Available: https://doi.org/10.1186/s41239-023-00408-3

[5] Y. An, J. H. Yu, and S. James, “Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration,” International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 10, 2025. Available: https://doi.org/10.1186/s41239-025-00507-3

[6] C. K. Y. Chan and W. Hu, “Students' voices on generative AI: Perceptions, benefits, and challenges in higher education,” International Journal of Educational Technology in Higher Education, vol. 20, Art. no. 43, 2023. Available: https://doi.org/10.1186/s41239-023-00411-8

[7] A. Yusuf, N. Pervin, and M. Román-González, “Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives,” International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 21, 2024. Available: https://doi.org/10.1186/s41239-024-00453-6

[8] H. Jo, “From concerns to benefits: A comprehensive study of ChatGPT usage in education,” International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 35, 2024. Available: https://doi.org/10.1186/s41239-024-00471-4

[9] M. F. Shahzad, S. Xu, and I. Javed, "ChatGPT awareness, acceptance, and adoption in higher education: The role of trust as a cornerstone," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 46, 2024. Available: https://doi.org/10.1186/s41239-024-00478-x
[10] H. Moradi, "Integrating AI in higher education: Factors influencing ChatGPT acceptance among Chinese university EFL students," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 30, 2025. Available: https://doi.org/10.1186/s41239-025-00530-4
[11] M. Abbas, F. A. Jam, and T. I. Khan, "Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 10, 2024. Available: https://doi.org/10.1186/s41239-024-00444-7
[12] F. Wang, N. Li, A. C. K. Cheung, and G. K. W. Wong, "In GenAI we trust: An investigation of university students' reliance on and resistance to generative AI in language learning," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 59, 2025. Available: https://doi.org/10.1186/s41239-025-00547-9
[13] Y. Walter, "Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 15, 2024. Available: https://doi.org/10.1186/s41239-024-00448-3
[14] D. Lee and E. Palmer, "Prompt engineering in higher education: A systematic review to help inform curricula," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 7, 2025. Available: https://doi.org/10.1186/s41239-025-00503-7
[15] J. Delcker, J. Heil, D. Ifenthaler, S. Seufert, and L. Spirgi, "First-year students AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 18, 2024. Available: https://doi.org/10.1186/s41239-024-00452-7
[16] D.-K. Mah and N. Groß, "Artificial intelligence in higher education: Exploring faculty use, self-efficacy, distinct profiles, and professional development needs," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 58, 2024. Available: https://doi.org/10.1186/s41239-024-00490-1
[17] M. I. Baig and E. Yadegaridehkordi, "Factors influencing academic staff satisfaction and continuous usage of generative artificial intelligence (GenAI) in higher education," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 5, 2025. Available: https://doi.org/10.1186/s41239-025-00506-4
[18] J. Kamali, M. F. Alpat, and A. Bozkurt, "AI ethics as a complex and multifaceted challenge: Decoding educators' AI ethics alignment through the lens of activity theory," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 62, 2024. Available: https://doi.org/10.1186/s41239-024-00496-9
[19] T. Lim, S. Gottipati, and M. Cheong, "What students really think: Unpacking AI ethics in educational assessments through a triadic framework," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 56, 2025. Available: https://doi.org/10.1186/s41239-025-00556-8
[20] T. Yang, J. Cheon, M.-H. Cho, M. Huang, and N. Cusson, "Undergraduate students' perspectives of generative AI ethics," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 35, 2025. Available: https://doi.org/10.1186/s41239-025-00533-1
[21] S.-H. Jin, K. Im, M. Yoo, I. Roll, and K. Seo, "Supporting students' self-regulated learning in online learning using artificial intelligence applications," International Journal of Educational Technology in Higher Education, vol. 20, Art. no. 37, 2023. Available: https://doi.org/10.1186/s41239-023-00406-5
[22] J. Escalante, A. Pack, and A. Barrett, "AI-generated feedback on writing: Insights into efficacy and ENL student preference," International Journal of Educational Technology in Higher Education, vol. 20, Art. no. 57, 2023. Available: https://doi.org/10.1186/s41239-023-00425-2
[23] K. J. Topping, E. Gehringer, H. Khosravi, S. Gudipati, K. Jadhav, and S. Susarla, "Enhancing peer assessment with artificial intelligence," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 3, 2025. Available: https://doi.org/10.1186/s41239-024-00501-1
[24] H. Ates, "Human-centered GenAI feedback design in higher education: A multisite experiment on direct, reflective, and hybrid approaches to scientific argumentation," International Journal of

Educational Technology in Higher Education, vol. 23, Art. no. 38, 2026. Available: https://doi.org/10.1186/s41239-026-00614-9
[25] Y. Hwang and J. H. Lee, "Exploring students' experiences and perceptions of human-AI collaboration in digital content making," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 44, 2025. Available: https://doi.org/10.1186/s41239-025-00542-0
[26] Y. Song, L. Huang, L. Zheng, M. Fan, and Z. Liu, "Interactions with generative AI chatbots: Unveiling dialogic dynamics, students' perceptions, and practical competencies in creative problem-solving," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 12, 2025. Available: https://doi.org/10.1186/s41239-025-00508-2
[27] M. Yilmaz, H. B. Temur, L. Emmungil, E. Çelik, A. Gauthier, and M. Cukurova, "Supporting self-regulated learning through generative AI feedback in online higher education: The importance of student perceptions of the source of feedback," International Journal of Educational Technology in Higher Education, vol. 23, Art. no. 16, 2026. Available: https://doi.org/10.1186/s41239-026-00592-y
[28] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI chatbots in education: Systematic literature review," International Journal of Educational Technology in Higher Education, vol. 20, Art. no. 56, 2023. Available: https://doi.org/10.1186/s41239-023-00426-1
[29] A. Williams, "Comparison of generative AI performance on undergraduate and postgraduate written assessments in the biomedical sciences," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 52, 2024. Available: https://doi.org/10.1186/s41239-024-00485-y
[30] K. Zdravkova and B. Ilijoski, "The impact of large language models on computer science student writing," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 32, 2025. Available: https://doi.org/10.1186/s41239-025-00525-1
[31] H. Khosravi, S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, et al., "Explainable artificial intelligence in education," Computers and Education: Artificial Intelligence, vol. 3, Art. no. 100074, 2022. Available: https://doi.org/10.1016/j.caeai.2022.100074
[32] T. Susnjak, "Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT," International Journal of Artificial Intelligence in Education, vol. 34, pp. 452-482, 2024. Available: https://doi.org/10.1007/s40593-023-00336-3
[33] Y. Feldman-Maggor, M. Cukurova, C. Kent, and G. Alexandron, "The impact of explainable AI on teachers' trust and acceptance of AI EdTech recommendations: The power of domain-specific explanations," International Journal of Artificial Intelligence in Education, vol. 35, pp. 2889-2922, 2025. Available: https://doi.org/10.1007/s40593-025-00486-6
[34] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," Psychological Review, 84(2), 191-215, 1977. Available: https://doi.org/10.1037/0033-295X.84.2.191
[35] D. Kahneman, Thinking, fast and slow. Farrar, Straus and Giroux, 2011.
[36] Q. Xia, X. Weng, F. Ouyang, T.-J. Lin, and T. K. F. Chiu, "A scoping review on how generative artificial intelligence transforms assessment in higher education," International Journal of Educational Technology in Higher Education, vol. 21, Art. no. 40, 2024. Available: https://doi.org/10.1186/s41239-024-00468-z
[37] R. Ellis, F. Han, and H. Cook, "Qualitatively different teacher experiences of teaching with generative artificial intelligence," International Journal of Educational Technology in Higher Education, vol. 22, Art. no. 33, 2025. Available: https://doi.org/10.1186/s41239-025-00532-2
[38] P. Wang and H. Ding, "The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-AI trust and decision performance," Information Processing & Management, vol. 61, no. 4, Art. no. 103732, 2024. Available: https://doi.org/10.1016/j.ipm.2024.103732
[39] R. Fok and D. S. Weld, "In search of verifiability: Explanations rarely enable complementary performance in AI-advised decision making," AI Magazine, vol. 45, no. 3, pp. 317-332, 2024. Available: https://doi.org/10.1002/aaai.12182
[40] S. Pareek, N. van Berkel, E. Velloso, and J. Goncalves, "Effect of Explanation Conceptualisations on Trust in AI-assisted Credibility Assessment," Proceedings of the ACM on Human-Computer Interaction, vol. 8, no. CSCW2, Art. no. 383, pp. 1-31, 2024. Available: https://doi.org/10.1145/3686922

[41] M. Naiseh, A. Simkute, B. Zieni, N. Jiang, and R. Ali, “C-XAI: A conceptual framework for designing XAI tools that support trust calibration,” Journal of Responsible Technology, vol. 17, Art. no. 100076, 2024. Available: https://doi.org/10.1016/j.jrt.2024.100076
[42] D. Lee, J. Pruitt, T. Zhou, J. Du, and B. Odegaard, “Metacognitive sensitivity: The key to calibrating trust and optimal decision making with AI,” PNAS Nexus, vol. 4, no. 5, Art. no. pgaf133, 2025. Available: https://doi.org/10.1093/pnasnexus/pgaf133