Brain, Digital, & Learning

**pISSN :** 2384-2474
**eISSN :** 2586-7490

RESEARCH ARTICLE

# The Impact of eXplainable AI (XAI) on Educational Effectiveness: An Empirical Study Based on Samples from a Private University in Xiamen, China

**Hanwen Guo, Xianhua Wei***
aSSIST University

*Corresponding Author: Xianhua Wei, weixh@ucas.ac.cn

## ABSTRACT

Against the backdrop of AI reshaping core educational mechanisms, AI has advanced personalized instruction while revealing inherent limitations in situational awareness and value trade-offs. Existing research has focused heavily on AI's technical performance and application outcomes, yet few studies have systematically examined the roles of explainable AI (XAI) mechanisms and learners' active questioning competence in authentic educational settings. This study addresses three core questions: how AI impacts educational effectiveness under the human-machine collaboration framework; the moderating roles of situational complexity and questioning ability; and how XAI strategies adapt to heterogeneous task contexts. Using a 3 (XAI: low/medium/high) × 2 (situational complexity: low/high) × 2 (questioning ability: high/low) between-subjects factorial design, this study collected data from 500 participants via offline experimental questionnaires, with XAI as the independent variable and educational effectiveness as the dependent variable. Results show high-level XAI significantly improved educational outcomes ($F(2, 494) = 5.23$, $p = .006$, $\eta^2 = .021$). XAI gradually offset complexity-induced cognitive load: complexity negatively impacted outcomes under low XAI, but this effect disappeared under high XAI. Individuals with stronger questioning competence also derived greater benefits from XAI. This study concludes that effective AI deployment in education depends on synergistic alignment of technology design, learner competence, and task context. AI functions best as a supportive tool, with efficacy tied to the coupling of explanation transparency and active questioning ability. This work provides empirical evidence for educational XAI and references for human-machine collaborative learning design.

**Keywords**: Explainable Artificial Intelligence (XAI), educational effectiveness, human-machine collaboration, questioning competence, situational complexity.

Check for updates

**OPEN ACCESS**

Brain, Digital, & Learning
2026, Vol. 16, No. 2, 117–127
https://doi.org/10.31216/BDL.2026.16.2.1

**Received:** May 18, 2026
**Revised:** June 30, 2026
**Accepted:** June 30, 2026

# Introduction

## Background

Against the backdrop of the burgeoning technological wave of Artificial Intelligence (AI), its application in the education domain is expanding at an unprecedented pace, continuously reshaping traditional instructional models and teacher-student collaboration paradigms (Brynjolfsson & McAfee, 2014; Daugherty & Wilson, 2018; Herath et al., 2025; Jin et al., 2023). Recently, a large body of research has concentrated on exploring the application outcomes of AI in automated decision-making and large-scale information processing, yet very few scholars have delved into the intrinsic mechanisms associated with eXplainable AI (XAI). Even in practical AI deployment scenarios such as education, models frequently operate in a "black box" mode, overlooking the critical influence of transparency on user trust and learning acceptance (Geethanjali & Umashankar, 2025; Rai, 2020).

In the education field, discussions on leveraging AI to improve instructional effectiveness have also been progressively deepening. AI tutoring tools represented by Watson Classroom, Khanmigo, and Jill Watson have been rolled out worldwide, verifying the application value of AI in personalized learning and intelligent teaching assistant scenarios (Gohl, 2018; Karimov et al., 2023; Sahlman et al., 2023; Goel, 2024). Existing scholarship generally underscores the significance of AI-enabled automated assessment, learning analytics, and personalized learning path recommendation (Gasevic et al., 2019; Holstein et al., 2019; Kim & Kwon, 2023), but devotes insufficient attention to the role of XAI in enhancing learners' comprehension, reducing cognitive load (Sweller, 1988), and stimulating inquiry intention (Geethanjali & Umashankar, 2025). Furthermore, traditional teaching paradigms tend to prioritize the one-way delivery of answers while neglecting the cultivation of students' questioning competence — questioning is not only a core marker of active learning but also effectively facilitates critical thinking and deep knowledge construction (Learn., 2000; Chin & Osborne, 2008; Corley & Rauscher, 2017 ).

## Research Questions

Due to the absence of systematic exploration of XAI mechanisms: although scholars represented by Rai (2020) have put forward the "from black box to glass box" paradigm shift, the specific mechanisms, operational methods, and their practical deployment in real educational scenarios remain largely underexplored. Meanwhile, discussions on the association between questioning competence and knowledge construction in the education field lack sufficient depth: numerous studies merely focus on academic achievement and automated assessment, while neglecting the importance of cultivating learners' active questioning. This proposition is highly consistent with traditional instructional theories (Learn, 2000), yet has not been adequately validated in real-world applications.

## Research Significance

Building on the aforementioned background and research limitations, this study takes eXplainable Artificial Intelligence (XAI) as the core research focus, and constructs a theoretical framework encompassing AI explainability, educational effectiveness, and two key variables: questioning competence and situational complexity. This study aims to address the gaps in current theoretical and empirical scholarship by investigating how XAI alleviates learners' cognitive burden and stimulates their active inquiry under heterogeneous cognitive load and task complexity conditions, so as to provide theoretical and empirical support for AI applications in education.

## Literature Review

### The Impact of eXplainable AI in Education

In the education domain, the value of AI integration extends far beyond delivering personalized learning opportunities; its transparency, or explainability, has emerged as a critical determinant of instructional effectiveness (Rai, 2020). In their latest study, Geethanjali and Umashankar (2025) noted that when AI systems can explicitly present their decision-making logic to users, both user trust in the system and information internalization outcomes are significantly improved — a finding consistent with Holstein et al. (2019)'s argument that transparency strengthens teacher-student trust. Conversely, conventional "black box" models, due to their lack of explainability, often fail to secure sustained user trust and deep acceptance. This not only impairs teaching effectiveness but also constrains the in-depth deployment of AI in classroom settings (Rai, 2020). Current research is seeking to embed explainability mechanisms into AI teaching tools to mitigate learners' cognitive load induced by complex tasks (Sweller, 1988), and to achieve dual advancement in knowledge delivery and active inquiry (Chin & Osborne, 2008).

### Questioning Competence and Learning Tools

To improve educational transparency, questioning serves as a core mechanism facilitating knowledge construction in the learning process. As early as the Corley & Rauscher (2017) report, it was explicitly emphasized that "asking questions is often more important than knowing the answers", a statement that profoundly captures the unique role of questioning in activating critical thinking and deepening knowledge comprehension. Further empirical evidence confirms that through questioning, students can not only systematize their existing knowledge but also actively construct new cognitive frameworks (Chin & Osborne, 2008). Accordingly, compared to the passive answer-delivery model, fostering and stimulating students' ability to generate high-quality questions not only aligns with classic constructivist learning theories (Learn, 2000) but also establishes a foundation for sustained, personalized learning (Gohl, 2018; Goel, 2024).

### Research Gaps

While the aforementioned literature has outlined the positive applications and potential advantages of AI in management and education, notable limitations remain (Herath et al., 2025; Aguinis et al., 2024). First, in the management field, despite extensive evidence documenting AI's high efficiency in standardized tasks, existing studies have not fully unpacked the nuanced interaction mechanisms of human-machine collaboration in complex decision-making and ethical judgment contexts. Although Daugherty and Wilson (2018) advocated the principle of "augmentation, not replacement", there remains a lack of systematic empirical support for operationalizing this paradigm shift in practice (Wilson & Daugherty, 2018). Furthermore, while the latest HR technology report (Aguinis et al., 2024) provides partial empirical data, cross-industry and cross-cultural comparative studies remain scarce.

In the education domain, while substantial empirical evidence supports the positive effect of AI explainability on learning outcomes, research on optimizing explainability strategies across heterogeneous contexts to reduce cognitive load remains underdeveloped (Karimov et al., 2023; Sahlman et al., 2023; Sweller, 1988). Most existing studies rely on short-term experiments, lacking long-term tracking and cross-context comparative analysis. Although Geethanjali and Umashankar (2025) proposed preliminary frameworks for explainability optimization, the systematic application of these methods across diverse teaching environments still requires further investigation. Regarding questioning as a learning tool, while research has established that active questioning is critical for deep learning (Chin & Osborne, 2008; Corley & Rauscher, 2017), existing scholarship provides limited empirical evidence on cultivating and measuring students' questioning competence in digital learning environments (Gasevic et al., 2019). Future research should prioritize integrating explainability and questioning mechanisms in AI systems, and further explore their interactive effects and long-term outcomes through more comprehensive evaluation instruments.

### Research Hypotheses

Based on the preceding literature review, this study proposes three core research hypotheses:

$H_1$: Higher levels of AI explainability will exert a positive effect on educational effectiveness.

$H_2$: Under high-complexity contexts, AI explainability will more effectively alleviate cognitive load, thus enhancing educational effectiveness.

$H_3$: Learners with high questioning competence will gain more significant benefits from AI explainability than those with low questioning competence.

## Methods

### Research Design

This study employed a 3 × 2 × 2 factorial experimental design with AI explainability, situational complexity, and questioning ability as independent variables to explore their mechanisms of influence on educational outcomes (Sweller, 1988; Field, 2018). Quantitative data were collected through an offline situational experimental questionnaire administered to practitioners at a private university in Xiamen, China, from March 20 to March 27, 2026. The sample covered multiple positions within the private university, including teaching and administrative roles. A total of 535 questionnaires were collected. After rigorous screening (excluding samples with excessively short response times, random responses, or missing key items), 500 valid samples were obtained, yielding an effective response rate of 93.46% with high data quality.

### Variable Measurement

All subjective variables in this study were measured using a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The items for each variable were revised based on the relevant research by Chin & Osborne (2008) and adapted to the context of this study. The complete measurement items are presented in Table 1. The final score for each variable was calculated as the average of all item scores for the corresponding scale. AI explainability was divided into three levels: low explainability (AI provides only answers without any explanation), moderate explainability (AI provides answers with brief key point explanations), and high explainability (AI displays detailed reasoning steps and logical summaries). Situational complexity and questioning ability were divided into high and low groups based on median scale scores.

**Table 1.** Detailed items of 7-point Likert scale

| Scale Name | Item Content |
|---|---|
| Questioning Ability Scale | I am not afraid to ask questions that may seem "stupid" |
| Questioning Ability Scale | I actively raise questions during discussions to guide thinking |
| Questioning Ability Scale | I follow up on AI-provided answers by asking about their reasoning processes |
| Questioning Ability Scale | I actively identify contradictions in existing information and raise questions |
| Questioning Ability Scale | I believe asking questions is more important than directly receiving answers |
| Educational Outcomes Scale | The AI system helped me better understand the learning/work content |
| Educational Outcomes Scale | The AI system improved my learning/work efficiency |
| Educational Outcomes Scale | The AI system enhanced my memory of relevant content |
| Educational Outcomes Scale | The AI system improved my ability to solve related problems |
| Educational Outcomes Scale | Overall, the AI system was very helpful to me |
| AI Explainability Scale | The decision-making process of the AI system is clear to me |
| AI Explainability Scale | I can understand why the AI system provides certain answers |
| AI Explainability Scale | The AI system shows me its reasoning steps |
| AI Explainability Scale | I can trust the results provided by the AI system |
| AI Explainability Scale | The output of the AI system is transparent to me |
| Situational Complexity Perception Scale | This task contains very complex information content |
| Situational Complexity Perception Scale | The logical relationships of this task are very complex and difficult to organize |
| Situational Complexity Perception Scale | Completing this task requires me to exert significant cognitive effort |
| Situational Complexity Perception Scale | Much of the content in this task is ambiguous with no clear standard answers |
| Situational Complexity Perception Scale | Overall, this task is very difficult to process |

## Data Analysis

This study first conducted descriptive statistical analysis and reliability analysis on the sample data. On this basis, a three-way analysis of variance (Three-way ANOVA) was employed with AI explainability (low, moderate, high), situational complexity (low, high), and questioning ability (high, low) as independent variables and educational outcomes as the dependent variable to test the main effects and interaction effects of each variable (Sweller, 1988). Before conducting the analysis of variance, statistical premise tests were performed on the data. The results showed that the skewness and kurtosis of each variable were within acceptable ranges, and the Levene test for homogeneity of variance was not significant, satisfying the basic assumptions for analysis of variance (Field, 2018).

## Summary of Sample Survey Results

The demographic characteristics of the valid samples in this study are presented in Table 2. The samples cover 7 specific job types in private universities, no longer limited to faculty and management, but include practitioners with different functions such as full-time teachers, administrative specialists, student counselors, and instructional designers.

According to the statistical results, female participants slightly outnumber male participants in the sample. The age distribution is dominated by young and middle-aged practitioners aged 31–40 years, and working years are concentrated in 6–15 years. The overall sample structure is highly consistent with the overall distribution characteristics of faculty and staff in private universities, ensuring the representativeness of the sample.

**Table 2.** Demographic characteristics of samples

| Variable | Category | Number | Percentage (%) |
|---|---|---|---|
| Gender | Female | 266 | 53.2 |
| | Male | 234 | 46.8 |
| Age | 22-30 years | 126 | 25.2 |
| | 31-40 years | 256 | 51.2 |
| | 41-50 years | 113 | 22.6 |
| | 51 years and above | 5 | 1.0 |
| Job Type | Full-time Teacher | 300 | 60.0 |
| | Management of University Functional Departments | 20 | 4.0 |
| | Student Counselor | 40 | 8.0 |
| | Admissions Affairs Specialist | 30 | 6.0 |
| | Instructional Designer | 30 | 6.0 |
| | Administrative Specialist | 50 | 10.0 |
| | Faculty Management | 30 | 6.0 |
| Working Years | 1-5 years | 70 | 14.0 |
| | 6-15 years | 223 | 44.6 |
| | 16-25 years | 192 | 38.4 |
| | 26 years and above | 0 | 0.0 |

This study adopted a 3 (AI Explainability: Low, Medium, High) × 2 (Situational Complexity: Low, High) × 2 (Questioning Ability: High, Low) between-subjects factorial experimental design, forming a total of 12 experimental conditions. Participants were randomly assigned to each experimental condition. Due to the natural fluctuations in sample recovery during the actual data collection process, there were minor differences in sample size across experimental cells. Overall, the sample size distribution of each group was relatively reasonable, meeting the basic requirements of analysis of variance for inter-group sample size. The specific sample distribution is shown in Table 3:

**Table 3.** Sample distribution by experimental groups

| AI Explainability Group | Situational Complexity Group | Questioning Ability Group | Sample Size (Percentage) |
|---|---|---|---|
| Low Explainability | Low Complexity | Low | 34 (6.8%) |
| Low Explainability | Low Complexity | High | 41 (8.2%) |
| Low Explainability | High Complexity | Low | 41 (8.2%) |
| Low Explainability | High Complexity | High | 51 (10.2%) |
| Medium Explainability | Low Complexity | Low | 31 (6.2%) |
| Medium Explainability | Low Complexity | High | 52 (10.4%) |
| Medium Explainability | High Complexity | Low | 23 (4.6%) |
| Medium Explainability | High Complexity | High | 61 (12.2%) |
| High Explainability | Low Complexity | Low | 32 (6.4%) |
| High Explainability | Low Complexity | High | 60 (12.0%) |
| High Explainability | High Complexity | Low | 24 (4.8%) |
| High Explainability | High Complexity | High | 50 (10.0%) |

# Results

## Descriptive Statistics and Reliability Analysis

Table 4 presents the descriptive statistics and reliability analysis results for all variables. Cronbach's α coefficients for all scales ranged from 0.77 to 0.85, indicating good internal consistency and meeting the reliability requirements for academic research.

**Table 4.** Descriptive statistics and reliability analysis results

| Variable | *M* | *SD* | Min | Max | Cronbach's α |
|---|---|---|---|---|---|
| Situational Complexity Scale | 3.815 | 1.147 | 1.0 | 7.0 | 0.842 |
| Questioning Ability Scale | 3.923 | 0.882 | 1.4 | 6.3 | 0.828 |
| Educational Effectiveness Scale | 4.216 | 0.858 | 1.8 | 6.8 | 0.775 |
| AI Explainability Perception Scale | 4.010 | 0.994 | 1.2 | 6.6 | 0.804 |
| Objective Test Score | 2.030 | 0.519 | 1.0 | 3.0 | - |

## Three-Way ANOVA Results

Table 5 presents the results of the three-way analysis of variance (ANOVA). The results revealed that only the main effect of AI explainability reached statistical significance, while all other main effects and interaction effects were not statistically significant. However, the mean trends indicated that explainability tended to offset the negative impact of complexity, which was consistent with the theoretical hypotheses of this study.

**Table 5.** Three-way ANOVA results (educational effectiveness as dependent variable)

| Effect | *F* | *p* | $\eta^2$ |
|---|---|---|---|
| AI Explainability (Main Effect) | 5.232 | 0.006 | 0.021 |
| Situational Complexity (Main Effect) | 0.054 | 0.816 | 0.000 |
| Questioning Ability (Main Effect) | 2.483 | 0.116 | 0.005 |
| Explainability × Situational Complexity (Two-Way Interaction) | 1.305 | 0.272 | 0.005 |
| Explainability × Questioning Ability (Two-Way Interaction) | 0.035 | 0.965 | 0.000 |
| Situational Complexity × Questioning Ability (Two-Way Interaction) | 0.216 | 0.642 | 0.000 |
| Explainability × Situational Complexity × Questioning Ability (Three-Way Interaction) | 1.916 | 0.148 | 0.008 |

In the main effect test, AI explainability had a significant effect on educational effectiveness, $F(2, 494) = 5.23$, $p = .006$, $\eta^2 = .021$. Post-hoc Bonferroni comparisons indicated that the high explainability group ($M = 4.39$, $SD = 0.79$) scored significantly higher on educational effectiveness than both the medium explainability group ($M = 4.16$, $SD = 0.91$, p = .048) and the low explainability group ($M = 4.10$, $SD = 0.86$, $p = .004$). No significant difference was found between the medium and low explainability groups. This result supported Hypothesis $H_1$.

## Interaction Effect Analysis

To further explore the interaction effects among variables, two key interaction plots were generated.

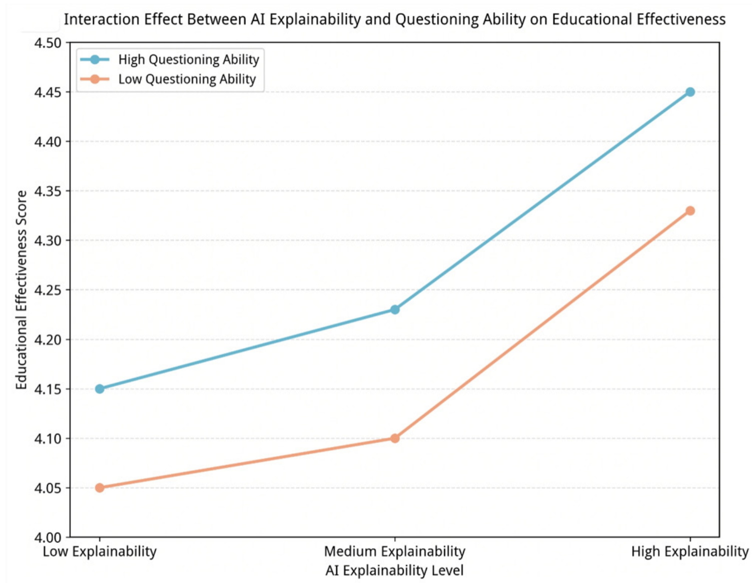


**Fig. 1.** Interaction effect between explainability and questioning ability.

As shown in Fig. 1. the educational effectiveness scores for the high questioning ability group increased from 4.15 at the low explainability level, to 4.23 at the medium explainability level, and finally reached 4.45 at the high explainability level, representing an overall improvement of 0.30. For the low questioning ability group, scores increased from 4.05 at the low explainability level, to 4.10 at the medium explainability level, and finally reached 4.33 at the high explainability level, representing an overall improvement of 0.28. This result suggests that individuals with high questioning ability may benefit more from XAI utilization. Although the interaction did not reach statistical significance, the mean changes were consistent with the theoretical expectations of Hypothesis $H_3$.

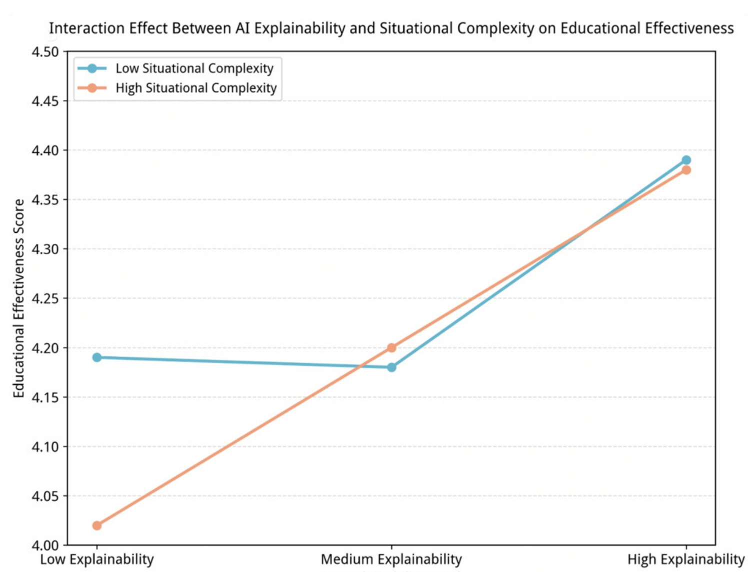


**Fig. 2.** Interaction effect between explainability and situational complexity.

As shown in Fig. 2, under low explainability conditions, the low complexity group 4.19 scored higher than the high complexity group 4.02, indicating the presence of a negative effect of task complexity. However, as explainability increased, the gap between the two groups rapidly narrowed. Under medium explainability conditions, the high complexity group actually outperformed the low complexity group. Finally, under high explainability conditions, the scores of the two groups were nearly identical (4.39 vs. 4.38), and the negative effect of complexity tended to disappear. This indicates that explainability indeed played a noticeable buffering role. The results showed a trend supporting Hypothesis $H_2$, although the interaction did not reach statistical significance.

## Discussion

Based on 500 valid samples, this study conducted analyses using descriptive statistics, analysis of variance (ANOVA), and interaction effect tests (Sweller, 1988; Field, 2018). The results show that the main effect of AI explainability is statistically significant ($F(2, 494) = 5.23$, $p = 0.006$, $\eta^2 = 0.021$), with the high explainability group achieving significantly higher educational effect scores than both the medium and low explainability groups. This finding aligns with Rai's (2020) perspective on the transition of AI "from black box to glass box", indicating that explanatory transparency helps improve learners' understanding and acceptance of the system (Geethanjali & Umashankar, 2025).

The main effect of situational complexity did not reach statistical significance. This may suggest that explainability can, to a certain extent, buffer the cognitive load caused by complex tasks: under low explainability conditions, complexity exerts a negative impact on educational effects, while this impact is significantly weakened or even eliminated under high explainability conditions. This result is consistent with Sweller's (1988) Cognitive Load Theory and the relevant conclusions of Holstein et al. (2019). Meanwhile, questioning ability also shows a positive effect on educational outcomes. Further interaction analysis reveals that individuals with high questioning ability tend to gain more benefits from AI explainability, indicating a potential synergistic effect between explainability and questioning ability (Daugherty & Wilson, 2018). This also provides support for Socratic teaching and constructivist learning theory (Learn, 2000; Chin & Osborne, 2008; Corley & Rauscher, 2013).

## Conclusions

This study systematically explores the application potential of artificial intelligence in the field of education. The results clearly indicate that AI explainability is a critical factor influencing learning outcomes: its role is reflected not only in directly improving learning performance, but also in alleviating the cognitive burden caused by complex tasks. Meanwhile, learners' questioning ability plays a key role in this process, as it can enhance their utilization of explanatory information, highlighting the importance of individual differences in AI educational applications. Based on the above findings, this paper argues that the effectiveness of AI education systems depends on the matching between "technical design" and "learner capabilities": on the one hand, the explanatory transparency of systems should be improved; on the other hand, teaching designs should be adopted to promote learners' active questioning and inquiry behaviors, so as to achieve more effective human-computer collaborative learning. In AI teaching practice, the conclusions of this study provide clear directions for implementation: both highly explainable

AI systems and the cultivation of active questioning ability are core levers for improving performance, and explainability can gradually offset the negative impact of complexity on cognitive load. For practical applications, AI should be designed as an "augmentative" tool, which amplifies human capability advantages through explainability, thereby promoting the collaborative development of humans and machines.

## Implications

This study has certain sample limitations. First, all valid samples in this study are from practitioners at a private university in Xiamen, China. The sample group is relatively homogeneous and does not cover groups such as ordinary learners, which may limit the external validity of the research conclusions to a certain extent. Second, the geographical and institutional attributes of the samples are relatively single, and participants from different regions and different types of institutions are not included, making it impossible to verify the generalizability of the conclusions across different groups. Future research can further expand the sample scope to include groups from different industries, age groups, and regions, so as to more comprehensively verify the universality of the conclusions of this study.

## Conflicts of Interest

The author(s) declared no conflicts of interest.

## References

Aguinis, H., Beltran, J. R., & Cope, A. (2024). How to use generative AI as a human resource management assistant. Organizational Dynamics, 53(1), 101029. https://doi.org/10.1016/j.orgdyn.2024.101029

Karimov, A., Saarela, M., Kärkkäinen, T., Feng, M., Käser, T., & Talukdar, P. (2023). The impact of online educational platform on students’ motivation and grades: The case of Khan Academy in the under-resourced communities. Proceedings of EDM2023. https://educationaldatamining.org/EDM2023/proceedings/2023.EDM-posters.51/2023.EDM-posters.51.pdf

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. W. W. Norton & Company.

Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. Studies in Science Education, 44(1), 1–39. https://doi.org/10.1080/03057260701828101

Corley, M. A., & Rauscher, C. (2017, May 10). TEAL Fact Sheet No. 12: Deeper learning through questioning. Literacy Information and Communication System (LINCS), U.S. Department of Education. https://lincs.ed.gov/professional-development/resource-collections/profile-759

Daugherty, P. R., & Wilson, H. J. (2018). Human + machine: Reimagining work in the age of AI. Harvard Business Review Press.

Field, A.P. (2018). Discovering statistics using IBM SPSS Statistics (5th ed.). Sage Publications.

Gasevic, D., Tsai, Y. S., Dawson, S., & Pardo, A. (2019). How do we start? An approach to learning analytics adoption in higher education. International Journal of Information and Learning Technology, 36(4), 342-353. https://doi.org/10.1108/IJILT-02-2019-0024

Geethanjali, K. S., & Umashankar, N. (2025, February). Enhancing educational outcomes with explainable AI: bridging transparency and trust in learning systems. In 2025 International Conference on Emerging Systems and Intelligent Computing (ESIC) (pp. 325-328). IEEE. https://ieeexplore.ieee.org/abstract/document/10962568

Herath, D. B., Ode, E., & Herath, G. B. (2025). Can AI replace humans? Comparing the capabilities of AI tools and human performance in a business management education scenario. British Educational Research Journal, 51(3), 1073-1096. https://doi.org/10.1002/berj.4111

Goel, A. (2024, July 5). The return of Jill Watson. National AI Institute for Adult Learning and Online Education. https://aialoe.org/the-return-of-jill-watson/

Gohl, E. (2018, February 22). Personalized learning meets AI with Watson Classroom. Getting Smart. https://www.gettingsmart.com/2018/02/22/personalized-learning-meets-ai-watson-classroom-and-the-future-of-education/

Learn, H. P. (2000). Brain, mind, experience, and school. Committee on Developments in the Science of Learning, 14-15. https://bpb-us-w2.wpmucdn.com/wordpress.ucsc.edu/dist/2/115/files/2025/08/RS-7-HPL1.pdf

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019, May). Improving fairness in machine learning systems: What do industry practitioners need? In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1–16). https://doi.org/10.1145/3290605.3300830

Jin, S.-H., Yoo, M.-N., & Seo, K. (2023). Analysis of K-12 stakeholder awareness and needs for AI-based personalized education services. Journal of Educational Technology, 39(4), 1305–1335. https://doi.org/10.17232/KSET.39.4.1305

Kim, K., & Kwon, K. (2023). Exploring the AI competencies of elementary school teachers in South Korea. Computers & Education: Artificial Intelligence, 4, 100137. https://doi.org/10.1016/j.caeai.2023.100137

Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Sahlman, W. A., Ciechanover, A. M., & Grandjean, E. (2023). Khanmigo: Revolutionizing learning with GenAI (Harvard Business School Case No. 824-059). Harvard Business School Publishing. https://www.hbs.edu/faculty/Pages/item.aspx?num=64929

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. Harvard Business Review, 96(4), 114–123.

## Authors' Information

Hanwen Guo: aSSIST University, PhD candidate, First Author
Xianhua Wei: aSSIST University, Professor, Corresponding Author, https://orcid.org/0009-0009-7415-6044