Investigating the Intrinsic Mechanisms by which AI-driven Educational Games Enhance Student Cognition

Zhuoran Zang*
Chinese Language and Culture College, Beijing Normal University, Beijing 100875, China
*Corresponding email: 5133248@qq.com

Abstract

Generative artificial intelligence is driving instructional games from static scripts toward intelligent forms that are generative, adaptive, and diagnostic. However, mechanistic evidence is still lacking on what kinds of AI-driven instructional games effectively promote learning and why. Grounded in a "design features - process mechanisms - learning outcomes" framework and targeting deep learning outcomes of understanding and transfer, this study collected 210 questionnaire responses from teachers and students who used AI instructional games during the past semester. Structural equation modeling (SEM) was employed to test pathways by which key design features - clarity of goals and rules, adaptive challenge, diagnostic feedback, and autonomy and control - affect learning outcomes via flow/immersion and cognitive engagement. Results indicate good scale reliability and validity and satisfactory model fit. Path analyses reveal a progressive, interlinked structure among AI instructional game design elements: clarity of goals and rules - diagnostic feedback - adaptive challenge - autonomy and control. These elements in turn enhance cognitive engagement by increasing flow/immersion, and ultimately significantly promote understanding and transfer. Based on these findings, the study offers classroom-oriented design implications: Prioritize transparent goals and rules, provide actionable diagnostic feedback, implement ability-matched adaptive support, and preserve learner autonomy and control to effectively convert immersive experiences into deep cognitive processing and transfer performance.

Keywords

AI educational games, Structural equation modeling, Student cognition, Intrinsic mechanisms

Introduction

Generative artificial intelligence is reshaping how educational games are developed and applied. Leveraging capabilities such as code generation, interface automation, and online deployment, instructors can design, iterate, and implement educational games in the classroom with lower barriers and shorter cycles. This drives a transition from "static rules and fixed scripts" to a generative, adaptive, and diagnostic intelligent paradigm [1]. Compared with traditional educational games, AI-driven educational games can more flexibly generate tasks and scenarios, dynamically adjust difficulty, and provide formative feedback, while also producing traceable learning process data. They are therefore widely regarded as having the potential to promote deep learning.

However, in terms of learning effectiveness, existing research has concentrated largely on how AI instructional games are generated and implemented, the construction of platforms and workflows, and prerequisite factors such as teacher AI literacy [2,3]. Few studies address a more central design question: what kinds of AI instructional games actually promote student learning. This gap leaves teachers without clear guidance when designing AI instructional games. AI instructional games are typically complex interventions composed of "intelligent functionality - gamified interaction - instructional sequencing" and their efficacy depends on how specific design features alter learners' key cognitive and behavioral processes. Therefore, it is necessary to anchor inquiry in learning outcomes and conduct mechanistic examinations of AI instructional games along the "design features - process mechanisms" pathway to produce reusable evidence and design principles. Building on this rationale, the present study focuses on deep learning outcomes such as understanding and transfer, systematically investigates the key design features of AI instructional games and their mechanisms of action. As a mainstream teaching model nowadays, the evaluation and improvement of deep learning capabilities in blended learning have become a research focus in the field of educational technology [4]. However, as an important interactive carrier of blended learning, AI-driven educational games still face an inherent mechanism issue that needs to be explored: How to promote the achievement of deep learning through scientific design. This paper poses the following research questions:

RQ_1: Which key design features of AI-enabled educational games can significantly enhance students' learning outcomes?

RQ_2: Through which learning process mechanisms do these design features exert their effects?

Literature review

Key design characteristics of AI-based educational games and their effects on learning outcomes

A substantial body of research indicates that carefully designed educational games can effectively enhance student learning outcomes. Educational games that integrate artificial intelligence embody a set of design features conducive to promoting deep learning [5].

Adaptive personalization: AI enables games to dynamically adjust task contexts and difficulty according to student performance, thereby providing appropriate challenges for learners at different levels. This individualized adaptation has been shown to enhance learning outcomes. Studies report that, compared with traditional instruction, such adaptive games significantly improve students' conceptual understanding and skill mastery and meaningfully enhance learning attitudes [6]. Notably, learners with low prior knowledge exhibit larger gains from adaptive games, indicating that personalized difficulty modulation is particularly beneficial for students with weaker foundations.

Intelligent feedback and scaffolding: AI-driven educational games can deliver timely, personalized feedback and guidance that help students correct errors and reflect on learning strategies during practice. Empirical work has found that cognitive feedback from virtual intelligent tutors can, in some cases, improve student performance more effectively than feedback from human instructors [7]. Meta-analytic evidence further indicates that embedding instructional scaffolds within games substantially improves learning outcomes and promotes transfer of acquired knowledge to real-world contexts. However, the design of feedback requires careful calibration: overly frequent feedback or feedback that makes excessive decisions on behalf of learners can be counterproductive. A higher-education experiment reported that adaptive feedback generated by large language models did not improve task performance and instead reduced students' interest in learning [8]. This effect was attributed to excessive feedback undermining learners' sense of autonomy and competence. Thus, while intelligent feedback should provide cognitive support, it must preserve student autonomy.

Gamified incentive mechanisms: AI-enhanced educational games commonly incorporate gamification elements to stimulate motivation and sustain engagement. Such incentive mechanisms positively influence students' affective and behavioral engagement, but their direct impact on cognitive achievement should be interpreted cautiously [9]. A study comparing badges and leaderboards in an online course found that neither element produced significant effects on quiz scores. However, most students reported favorable attitudes, perceiving these elements as motivating and wishing to retain them in the course. Consequently, gamified incentives primarily support learning indirectly by increasing enjoyment and willingness to engage. Their effects on learning outcomes often manifest only when integrated with other instructional design features.

These features constitute the intelligent capabilities and interaction design of AI-driven educational games. Overall, AI educational games that incorporate adaptive adjustment, timely feedback, and gamified incentives have been shown to better engage students and enhance learning performance. The combination of these key characteristics transforms instructional games from static scripts into intelligent interactive systems, thereby achieving significant improvements in student learning outcomes.

Mechanisms of design features: Intrinsic processes that promote deep learning

The ability of AI-based instructional games to foster deep learning outcomes such as understanding and transfer rests on how their design features positively influence students' cognitive and motivational processes. First, they promote immersion and sustained attention. Dynamic difficulty adjustment and clearly articulated, hierarchical objectives align game challenges with students' abilities, preventing boredom from tasks that are too easy and frustration from tasks that are too hard, thereby inducing a flow-like state of intense focus. In this state, students devote themselves fully to tasks, extending effective learning time and deepening knowledge processing and comprehension. Research shows that when instructional scaffolds are embedded within games, students may complete fewer levels on average but spend more time on each level and experience lower frustration. This pattern suggests that appropriate interventions can steer students toward deeper problem contemplation, even if immediate progress slows, such engagement ultimately supports improvements in knowledge transfer.

Second, AI instructional games enhance motivation and engagement. They evoke intrinsic motivation by satisfying students' basic psychological needs. Adaptive challenges and timely positive feedback sustain a sense of competence, while situational choices and open-ended exploration within the game strengthen autonomy. These factors align with self-determination theory and contribute to the emergence of students' intrinsic learning motivation and initiative [10]. Moreover, moderate competition and reward mechanisms typically elevate students' enthusiasm and persistence. Although extrinsic incentives do not necessarily directly raise academic performance, the enjoyment and sense of achievement they generate can alleviate the tedium of purely didactic tasks and thereby indirectly support cognitive engagement.

Third, they optimize cognitive processing and reflection. Intelligent feedback and instructional scaffolds act directly on students' cognitive processes. High-quality, timely feedback can correct misconceptions, provide cures for problem solving, and prompt students to reflect on and adjust their understanding. When games guide students to watch brief explanatory videos or summaries after actions, self-reflection is strengthened, which in turn fosters deeper comprehension and transfer [11]. Notably, AI agents embedded in learning games can detect and respond to students' emotional states, using encouraging dialogue or task adjustments to alleviate negative effects. Such emotional support helps students maintain a positive stance and learning resilience. Studies indicate that appropriate affective feedback within games can reduce off-task behavior and boredom, boost engagement. Through an extra layer of emotion-regulation mechanisms, it encourages students to experiment and tolerate errors in a safe, supportive atmosphere, viewing failure as chance for growth [12].

In summary, AI-driven educational games exert their instructional effects by enhancing engagement, eliciting intrinsic motivation, and optimizing cognitive and affective processes. Key design features operate synergistically. Adaptive challenge sustains continued investment, timely feedback and scaffolding deepen knowledge construction, and contextual incentives along with virtual agents safeguard motivation and emotional equilibrium. This multi-layered process optimization explains why AI educational games are more likely to foster deep understanding and transfer of knowledge. When well-designed intelligent functions are combined with gamification strategies, students "learn by playing", undergoing highly engaging, feedback-rich learning trajectories that yield outcomes superior to traditional methods. Attention must also be paid to balancing challenges and support, autonomy and guidance in design, to avoid undermining students' opportunities for autonomous reflection and interest through excessive intervention. Only design strategies grounded in mechanistic evidence can fully realize the potential of AI educational games to promote deep learning.

Research methods

Study design and technical approach

This study employed a questionnaire survey and validated the theoretical model using structural equation modeling (SEM). SEM was chosen because AI-driven instructional games constitute a composite intervention characterized by "intelligent functionality - gamified interaction - instructional design". Its effects often influence learning outcomes indirectly through learners' key process variables (e.g., flow/immersion, cognitive engagement). SEM enables simultaneous handling of multiple latent variables, measurement error, and mediation effects. Meta-analytic evidence from serious games research also indicates that the learning advantages of instructional games are often associated with changes in cognitive and motivational processes.

This supports an examination approach centered on process mechanisms.

Variable construction and operational definitions

To ensure that the design features are both theoretically grounded and pedagogically practicable, this study synthesizes and references three categories of frequently used frameworks to derive design dimensions:

Game attributes - learning classification evidence: The game-attributes taxonomy highlights that rules and goals, challenge, control, feedback and assessment, immersion, and social interaction are core attributes related to learning and can be adjusted through design.

Flow/experience-oriented models: Game Flow proposes eight elements including clear goals, challenges, control, feedback, immersion, and social interaction. EGame Flow further provides quantifiable dimensions applicable to educational games.

Mapping of instructional design to serious game mechanics: Related studies emphasize aligning learning activities, learning outcomes, and assessment/feedback mechanisms with game mechanics to support reusable instructional designs and classroom implementation.

Considering the salient features of AI-driven educational games and drawing on empirical research, intelligent agents' feedback and AI-generated adaptive feedback can produce a "double-edged effect" of learning gains or diminished interest. We therefore incorporate "adaptive challenge" and "diagnostic feedback/scaffolding" as key AI design dimensions.

Questionnaire design and model

This study employed a standardized questionnaire to measure the key design features of AI-based instructional games, the mediating learning-process mechanisms, and resulting learning outcomes. The questionnaire used "the most recent AI instructional game you encountered or used" as a fixed referent. Respondents were instructed to answer based on their actual experience with that specific game (for students) or their observations of typical student performance (for teachers), to reduce reference drift and recall bias. The instrument primarily used 5-point Likert items (1= strongly disagree, 5= strongly agree) and comprised seven latent variables measured by 21 items. Four design-feature dimensions include clarity of goals and rules (D1, 3 items), adaptive challenge (D2, 3 items), diagnostic feedback (D3, 3 items), and sense of autonomy and control (D4, 3 items). Two process-mechanism dimensions are flow/immersion experience (M1, 3 items) and cognitive engagement (M2, 3 items). Learning outcome is represented by deep learning effectiveness in understanding and transfer (Y, 3 items). Based on the theoretical pathway "design features - process mechanisms - learning outcomes" the structural model specified that D1-D4 each positively predict M1 and M2, M1 and M2 in turn positively predict Y, and direct effects from D1-D4 to Y were allowed to test for partial versus full mediation. Individual background variables were included as control variables to improve estimation robustness. After assessing scale reliability and validity via confirmatory factor analysis, the model can be tested using structural equation modeling to estimate path coefficients and mediation effects. This helps identify which design features of AI instructional games are effective and through which process mechanisms they operate.

Study population, sampling, and data collection

The study population comprised frontline teachers and students who had interacted with or used AI educational games within the most recent academic semester. Samples were collected using a combination of convenience sampling and snowball sampling, primarily through online questionnaires. At the beginning of the questionnaire respondents were instructed to use "the most recently encountered AI educational game" as a fixed reference point to mitigate recall bias. The survey was conducted anonymously and on the principles of informed consent and voluntary participation.

Data analysis method

Data analysis proceeded through data preprocessing, measurement model evaluation, and structural model assessment. First, missing values, outliers, and normality were examined; missing data were addressed using EM estimation or multiple imputation depending on the proportion and mechanism of missingness. Prior knowledge, gaming experience, frequency of AI use, and duration of use were included as control variables in the models. Second, confirmatory factor analysis (CFA) was conducted to assess the reliability and validity of the measurement model. Cronbach's α, composite reliability (CR), and average variance extracted (AVE) were reported to evaluate internal consistency and convergent validity, and discriminant validity was assessed using the Fornell-Larcker criterion. Conditional on acceptable measurement-model fit, structural model path coefficients were then estimated and overall model fit indices reported.

Data validation and analysis

A total of 210 valid questionnaires were collected, upon which data validation and analysis were conducted.

Data inspection

The Cronbach's alpha coefficient was 0.979, indicating excellent reliability of the questionnaire (Figure 1).

Table 1. Reliability test.

Cronbach's alpha	Standardized Cronbach's alpha	Number of items	Sample size (N)
0.979	0.980	21	210

The KMO value was 0.982, indicating a degree of suitability (Figure 2).

Table 2. Validity assessment.

Test	Statistic	Value
KMO measure of sampling adequacy	KMO	0.982
Bartlett's test of sphericity	Approx. Chi-Square	4623.338
	df	210
	Sig. (p-value)	0.000***

SEM structural equation analysis

This study used SEM to analyze validated data, verifying causal links and mediating effects to test the theoretical framework.

Table 3. Factor loading coefficients.

Factor	Item	Statement (English)	Unstandardized coefficient	Standardized loading	Z	Standard errors	p
Clarity of goals and rules	D1-1	I can clearly know the learning objectives to be achieved for each task/level.	1.000	0.794	/	/	/
	D1-2	The task rules and completion criteria are clear, and I do not need to repeatedly guess.	0.986	0.833	14.141	0.070	0.000***
	D1-3	I clearly understand how the system determines "correct/incorrect" or "completed/not completed".	0.975	0.811	13.627	0.072	0.000***
Adaptive challenge	D2-1	The system adjusts subsequent task difficulty or hint intensity based on performance.	1.000	0.805	/	/	/
	D2-2	Overall, the task difficulty matches my (or students') ability level.	0.884	0.768	12.888	0.069	0.000***
	D2-3	When I (or students) get stuck, the system provides a more appropriate next task or support.	0.930	0.784	13.264	0.070	0.000***
Diagnostic feedback	D3-1	I can receive timely feedback on whether my answer/action is correct.	1.000	0.889	/	/	/
	D3-2	The feedback explains the reasons for errors or key concepts, rather than only indicating right/wrong.	0.964	0.867	18.592	0.052	0.000***

Factor	Item	Statement (English)	Unstandardized coefficient	Standardized loading	Z	Standard errors	p
	D3-3	The feedback provides actionable suggestions for improvement (e.g., what to do next, how to adjust strategies).	0.774	0.876	19.005	0.041	0.000***
Autonomy and control	D4-1	I can choose different solution paths/strategies to complete the task.	1.000	0.869	/	/	/
	D4-2	The game allows me to explore and try at my own pace, rather than being forced to progress.	0.938	0.855	17.245	0.054	0.000***
	D4-3	I can review process information (records/hints/explanations) to adjust my learning strategies.	1.261	0.853	17.158	0.073	0.000***
Flow/Immersion	M1-1	During gameplay, I feel highly focused.	1.000	0.863	/	/	/
	M1-2	I often do not notice the passage of time.	0.982	0.857	17.100	0.057	0.000***
	M1-3	Even when I encounter difficulties, I am willing to keep trying and do not give up easily.	1.085	0.789	14.732	0.074	0.000***
Cognitive engagement	M2-1	I try to connect new information with prior knowledge to understand it.	1.000	0.879	/	/	/
	M2-2	I check whether I truly understand the underlying principles behind the task, rather than just completing it.	0.969	0.869	18.235	0.053	0.000***
	M2-3	After failures or errors, I reflect on the reasons and adjust my strategies.	1.052	0.864	18.041	0.058	0.000***
Understanding and transfer	Y-1	I can explain the key concepts/principles involved in the game in my own words.	1.000	0.801	/	/	/
	Y-2	I can summarize general methods or rules for solving this type of problem.	1.196	0.753	12.485	0.096	0.000***
	Y-3	When encountering new problems/situations, I can transfer what I learned in the game to solve them.	1.045	0.874	15.386	0.068	0.000***

The factor loading table indicates that all observed variables exhibit standardized loadings on their respective latent constructs above the recommended threshold of 0.700 (range 0.753-0.889), demonstrating satisfactory convergent validity for the scales. All unstandardized coefficients are significant ($p < 0.001$), with Z-values ranging from 12.485 to 19.005, indicating that each observed variable explains its latent construct at a statistically significant level. Notably, within the “diagnostic feedback” dimension, item D3-1 (feedback on whether my answer) has the highest standardized loading of 0.889, reflecting the central role of actionable guidance in the feedback mechanism. Conversely, in the “understanding and transfer” dimension, item Y-2 (summarizing general methods) has a relatively lower loading of 0.753. This may suggest that abstract generalization is more challenging than concrete knowledge transfer. Internal reliability of each dimension is supported by consistently high loadings across three indicators. For example, all three Standardized loading for the “cognitive engagement” dimension fall between 0.864 and 0.879. Standard errors are tightly controlled within a small range, attesting to precise parameter estimation. Model specification adequacy is reflected by: (1) Fixing the first indicator loading of each latent variable to 1 to establish the measurement scale. (2) All free-estimated parameters exhibiting standard errors below 0.1 and significance levels of $p<0.001$, indicating sufficient sample size and good model identification. The sole outlier is item D4-3 in the “autonomy and control” dimension, whose unstandardized coefficient reaches 1.261, possibly reflecting a particular contribution of retrospective functions to autonomy. However, its standardized loading (0.853) remains consistent within the dimension. Overall, the data supports that the measurement model possesses satisfactory reliability and validity metrics, with each latent construct being effectively measured by its observed indicators, thereby providing a sound basis for subsequent structural model analysis (Figure 3).

Table 4. Model regression coefficients.

Factor (latent variable) → Outcome (latent variable)	Unstandardized coefficient	Standardized coefficient	Standard errors	Z	p
Clarity of goals and rules → Diagnostic feedback	1.404	1	0.095	14.761	0.000***
Diagnostic feedback →Adaptive challenge	0.911	1	0.058	15.785	0.000***
Adaptive challenge → Autonomy and control	0.849	1	0.055	15.305	0.000***
Autonomy and control → Flow/Immersion	1.069	1	0.061	17.496	0.000***
Flow/Immersion → Cognitive engagement	0.892	1	0.050	17.955	0.000***
Cognitive engagement → Understanding and transfer	0.940	1	0.061	15.535	0.000***

The regression coefficients table for this structural equation model indicates that both unstandardized and standardized coefficients for all paths are highly statistically significant ($p = 0.000$). This demonstrates that the causal path relationships between latent and observed variables, as well as among latent variables, are strongly supported. The unstandardized coefficient show that a one-unit increase in the latent variable “clarity of goals and rules” corresponds to a 1.404-unit increase in “diagnostic feedback”. Subsequent path coefficients decrease sequentially but remain stable, with the effect of “flow/immersion” on “cognitive engagement” equal to 0.892 and the effect of “cognitive engagement” on “understanding and transfer” rising again to 0.940. The standardized coefficient is all 1, indicating that the model was standardized using a fixed-loading approach and that the paths exhibit identical fully standardized magnitudes. The Z-values for all paths greater than or equal to 14.761 (the critical value 2.580 for $p<0.001$), indicating highly precise parameter estimates, with standard errors controlled within a small range of 0.050-0.095, further validating the stability of the model estimates. Notably, the "flow/immersion → cognitive engagement" path attains the highest Z-value (17.955), representing the most significant relationship among all paths, whereas the initial path "clarity of goals and rules → diagnostic feedback" has the relatively lowest Z-value (14.761). The overall model exhibits a complete causal chain, with each stage from goal clarity to ultimate learning transfer highly significant. This confirms the effectiveness of the hypothesized multilevel transmission mechanism from environmental features to cognitive outcomes (Figure 4).

Table 5. Model fit indices.

Fit index	Recommended criterion	Result
χ^2 (Chi-square)	/	240.373
df	/	183.000
p	>0.05	0.003***
χ^2/df (Normed Chi-square)	<3.00	1.314
GFI	>0.90	0.950
RMSEA	<0.10	0.039
RMR	>0.05	0.271
CFI	>0.90	0.988
NFI	>0.90	0.950
NNFI (TLI)	>0.90	0.986

Model fit indices indicate a good fit between the data and the theoretical model. The chi-square test was significant ($\chi^2 = 240.373$, $df = 183.000$, $p = 0.003$***). However, because the chi-square statistics are sensitive to sample size, it should be interpreted in conjunction with other indices. The chi-square to degrees-of-freedom ratio was 1.314 (<3.00), indicating acceptable model parsimony. The Goodness of Fit Index (GFI) was 0.950 (>0.90), Root Mean Square Error of Approximation (RMSEA) was 0.039 (<0.10), and Root Mean Square Residual (RMR) was 0.271 (above the ideal threshold of 0.05), reflecting acceptable absolute fit. Comparative Fit Index (CFI) (0.988), Normed Fit Index (NFI) (0.950), and Non-Normed Fit Index (NNFI) (0.986) - all exceeded the 0.900 criterion, indicating excellent incremental fit. Although the RMR was somewhat elevated, the remaining indices met or exceeded recommended standards, overall suggesting a well-specified model with a high degree of correspondence between the data and the theoretical framework (Figure 5).

Discussion

This study, organized along the framework "design features - process mechanisms - learning outcomes", employed structural equation modeling on 210 valid samples to examine the internal mechanisms by which AI-driven instructional games promote deep learning (understanding and transfer). Overall, the measurement model demonstrated stable reliability and validity, and the structural model exhibited acceptable fit indices overall, indicating that the proposed mechanistic chain is supported empirically.

First, regarding "which key design features can significantly enhance learning outcomes" (RQ_1). The results indicate that the critical design elements of AI instructional games do not operate as independent, parallel factors. Instead, they tend to form a progressive, combinatorial relationship: clear goals and rules significantly predict diagnostic feedback; diagnostic feedback is further associated with adaptive challenge; adaptive challenge is linked to autonomy and control. Ultimately, through increased flow/immersion and cognitive engagement, this sequence facilitates comprehension and transfer. In other words, learners may experience these design elements on an experiential level as a coherent progression "from intelligible (goals and rules) to correctable (diagnostic feedback), then to matchable (adaptive challenge), and finally to controllable (autonomy and control)". Together, these elements constitute a learning context that sustains engagement and thereby supports deeper knowledge construction and transfer. This finding also suggests that in classroom implementations of AI instructional games, merely "stacking features" at isolated points may not yield equivalent benefits. What is more critical is organizing design elements according to the logic of the learning process, so that they form a continuous experiential chain perceptible to learners.

Second, regarding "through which process mechanisms effects are exerted" (RQ_2). This study validates the key transmission pathway of "flow/immersion - cognitive engagement", and further demonstrates that cognitive engagement is a more critical direct predictor of comprehension and transfer. In other words, flow/immersion functions more like a state-based precondition of attention and persistence that creates temporal and resource conditions for deep processing. The core link that actually drives improvements in comprehension and transfer is whether learners engage in active meaning construction and strategy adjustments, such as connecting prior knowledge, testing understanding, reflecting on errors and adjusting accordingly. Mechanistically: (1) Clear goals and rules reduce task uncertainty and trial-and-error costs, decreasing ineffective search and allowing cognitive resources to be allocated more to conceptual and principled processing. (2) Diagnostic feedback (especially actionable improvement suggestions) provides learners with causal cues about “why it was wrong and how to fix it”, prompting reflection and strategy updating. (3) Adaptive challenge keeps tasks within a “reachable yet demanding” range, suppressing boredom while alleviating frustration and thereby providing conditions for flow. (4) Autonomy and control enhance learners’ sense of mastery over pacing, pathways, and opportunities to revisit content, making them more likely to employ self-regulation strategies and thus convert immersive experience into effective cognitive engagement. Overall, AI functionalities do not directly “replace learning”, rather, they indirectly promote deep learning outcomes by optimizing learners’ attention, feedback processing, and opportunities for self-regulation.

Third, this study’s theoretical and empirical contributions are manifested in two main aspects. Initially, it incorporates design features of AI-based instructional games that are often “functionally described” into a mechanistic testing framework, thereby providing an explicable pathway from environmental features to deep learning outcomes. This helps translate “what AI can do” into “which design elements produce which learning changes”.

Next, the sequential chains revealed by the structural model offer a more parsimonious empirical account of the composite-intervention nature of instructional games: Design elements may exhibit temporal dependencies and mutual support relationships, while mechanistic variables (flow/immersion, cognitive engagement) serve as pivotal “transformative” mediators. These findings align with existing explanatory accounts emphasizing flow, self-determination, and scaffolding-reflection facilitation of transfer, and further reinforce through SEM evidence the design rationale that process variables function as critical levers.

Fourth, the practice-oriented implications can be summarized as “clarify first, diagnose second, adapt later, and preserve autonomy”. Specifically: (1) Prioritizing making goals, rules, and evaluation criteria transparent and visible to reduce students’ energy spent on “guessing how the system judges”. (2) Upgrading feedback from mere “right/wrong notification” to “explanation of reasons + next-step suggestions”, while controlling frequency and level of intervention to avoid undermining autonomy. (3) Adaptivity should not be limited to adjusting difficulty but should also include “appropriate supports and sequencing of subsequent tasks when a student is stuck”, in order to maintain challenge-ability alignment. (4) Providing students with options for pathways, pace control, and review/rewind tools to enable self-regulation. (5) At the classroom implementation level, align learning-process data (dwell time, error types, hint usage, review behaviors, etc.) with indicators of cognitive engagement to diagnose whether students are engaging in deep processing rather than merely “clearing the level”.

Fifth, the generalizability of the study’s conclusions should be treated with caution. This study has several limitations: (1) The sample was obtained through convenience and snowball sampling and relied primarily on self-report questionnaires, which may be subject to common method bias and social desirability effects. (2) Using the “most recently encountered AI educational game” as the reference point, while reducing reference drift, may still introduce recall bias and confounding due to differences in specific game types. (3) The study is cross-sectional, and although SEM path estimates support the theoretical directions, they cannot be directly equated with causal effects. (4) Very high Cronbach’s alpha coefficients indicate strong internal consistency but may also reflect some item redundancy; future work should refine items and test the robustness of discriminant validity. (5) Among the fit indices, an elevated RMR suggests that unexplained structure may remain at the residual level; future studies should additionally report SRMR, test alternative models, and diagnose local misfit.

Sixth, future research can proceed along three avenues: To begin with, conduct experimental or quasi-experimental studies that manipulate single design elements to test causal effects and compare heterogeneity across disciplines and task types. Subsequently, incorporate objective learning outcomes and transfer tests, process log data, and behavioral metrics to construct a multi-source evidence chain of “subjective experience - objective behavior - learning outcomes”. Ultimately, test key moderating variables and assess multi-group equivalence to clarify “for whom and under what conditions it is more effective”, thereby translating mechanistic conclusions into actionable guidelines for differentiated design.

Conclusion

This study investigates the mechanisms by which AI-based instructional games promote deep learning, and constructs and tests a structural model linking “design features - process mechanisms - learning outcomes”. The results indicate that critical design elements - clear goals and rules, diagnostic feedback, adaptive challenge, and the balance between autonomy and control - do not operate in isolation. Instead, they function through supportive and progressive interactions. These elements primarily enhance learners’ flow/immersion and cognitive engagement, which in turn facilitate improved comprehension and transfer. Overall, the study provides empirical support for shifting AI instructional game development from a “feature-stacking” approach to a “mechanism-oriented design” paradigm. It also offers actionable design and implementation recommendations for classroom practice.

Against the backdrop of ongoing digitalization and intelligent transformation in education, the value of AI-powered instructional games lies not only in enhancing the learning experience. They also facilitate learners’ higher-order cognitive processing and knowledge transfer in an interpretable and controllable manner. Future research must further deepen efforts in more rigorous causal inference, multi-source data triangulation, and validation across diverse contexts. This will help advance AI instructional games toward more robust and sustainable learning gains in authentic educational settings.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] French, F., Levi, D., Maczo, C., Simonaityte, A., Triantafyllidis, S., Varda, G. (2023) Creative use of OpenAI in education: case studies from game development. *Multimodal Technologies and Interaction*, 7(8), 81.

[2] Khine, M. S. (2024) AI in teaching and learning and intelligent tutoring systems. *Artificial Intelligence in Education: A Machine-Generated Literature Overview*, 467-570.

[3] McDonald, N., Johri, A., Ali, A., Collier, A. H. (2025) Generative artificial intelligence in higher education: evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3, 100121.

[4] Li, Y., Cheng, H., Qin, Q. (2025) Evaluations and improvement methods of deep learning ability in blended learning. *International Journal of e-Collaboration (IJeC)*, 21(1), 1-17.

[5] Rahimi, S., Shute, V. J., Fulwider, C., Bainbridge, K., Kuba, R., Yang, X., D’Mello, S. K. (2022) Timing of learning supports in educational games can impact students’ outcomes. *Computers & Education*, 190, 104600.

[6] Hooshyar, D., Malva, L., Yang, Y., Pedaste, M., Wang, M., Lim, H. (2021) An adaptive educational computer game: Effects on students’ knowledge and learning attitude in computational thinking. *Computers in Human Behavior*, 114, 106575.

[7] Arguedas, M., Daradoumis, T., Caballé, S. (2024) Measuring the effects of pedagogical agent cognitive and affective feedback on students’ academic performance. *Frontiers in Artificial Intelligence*, 7, 1495342.

[8] Bauer, E., Richters, C., Pickal, A. J., Klippert, M., Sailer, M., Stadler, M. (2025) Effects of AI-generated adaptive feedback on statistical skills and interest in statistics: a field experiment in higher education. *British Journal of Educational Technology*, 56(5), 1735-1757.

[9] Balci, S., Secaur, J. M., Morris, B. J. (2022) Comparing the effectiveness of badges and leaderboards on academic performance and motivation of students in fully versus partially gamified online physics classes. *Education and Information Technologies*, 27(6), 8669-8704.

[10] Laine, T. H., Lindberg, R. S. (2020) Designing engaging games for education: a systematic literature review on game motivators and design principles. *IEEE Transactions on Learning Technologies*, 13(4), 804-821.

[11] Yang, K. H., Chen, H. H. (2023) What increases learning retention: Employing the prediction-observation-explanation learning strategy in digital game-based learning. *Interactive Learning Environments*, 31(6), 3898-3913.

[12] Bernecker, K., Ninaus, M. (2021) No pain, no gain? Investigating motivational mechanisms of game elements in cognitive tasks. *Computers in Human Behavior*, 114, 106542.