

AI-Infused Game Design Thinking: A Mixed-Method Study of Student Perceptions, Creativity Support and Verification Practices in Educational Game Development

SAGE Open
Vol. 16(3): 1–17
© The Author(s) 2026
DOI: 10.1177/21582440261486954
journals.sagepub.com/home/sgo
 Sage

Kee-Man Chuah¹ and Looi-Chin Ch'ng²

Abstract

This study investigates how undergraduate students perceive artificial intelligence (AI) integration during a structured Design Thinking cycle for educational game development, framing the AI-infused workflow as a blended learning environment. Eighty-two students in 14 teams completed a five-week project ending in a playable prototype and pitch. A convergent mixed-method design combined a 15-item perception scale, a three-item Design Thinking experience measure, team focus groups, reflective logs and artefact analysis. Exploratory factor analysis supported a four-factor structure: perceived usefulness, creativity support, trust and verification and ethical concerns. Results from multiple regression analyses showed that perceived usefulness and creativity support jointly accounted for 49.2% of variance in overall Design Thinking experience, while ethical concerns exerted a significant negative effect. AI use was significantly higher during prototyping and ideation than during empathise and define stages. Also, teams with higher AI reliance produced stronger prototypes but showed no advantage in reflective justification. These findings indicate that AI integration benefits are stage-dependent and that productivity gains do not automatically translate into deeper learning. The study contributes a conceptual model for AI-infused Design Thinking as a smart blended learning environment and empirically grounded design principles for responsible AI integration in project-based courses.

Plain Language Summary

This study explored how 82 university students used AI during a five-week Design Thinking project to create educational games. Students found AI most helpful for generating ideas and building prototypes, but less helpful for understanding users and defining problems. When students saw AI as useful and supportive of creativity, they reported a better learning experience. However, ethical concerns reduced their satisfaction. Teams that relied more on AI produced stronger prototypes, but they did not show better reflective justification for their design choices. The findings suggest AI can support game design learning, but it needs guidance, verification and stage-appropriate use.

Keywords

generative AI, design thinking, educational games, game development, blended learning

Received: 22 March 2026; Revised: 16 July 2026; Accepted: 26 August 2026

¹Faculty of Education, Language and Communication, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

²Academy of Language Studies, Universiti Teknologi Mara Sarawak Campus, Kota Samarahan, Malaysia

Corresponding author:

Kee-Man Chuah, Faculty of Education, Language and Communication, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Malaysia.

Email: kmchuah@unimas.my



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

1. Introduction

Educational game development is increasingly used in digital design and computing curricula because it combines authentic production with iterative feedback and collaboration (Kafai & Burke, 2016; Pacheco-Velazquez et al., 2024). Students must translate a real-world issue into a playable system that communicates or teaches, then justify design decisions in relation to users and evidence. Design Thinking is often used to structure this work through staged activities such as empathic inquiry, problem framing, ideation, prototyping and testing (Liedtka, 2018). Contemporary synthesis suggests Design Thinking interventions can yield positive learning effects across educational contexts, although outcomes depend on implementation quality and assessment alignment (Li & Zhan, 2022; Yu et al., 2024). Moreover, reviews such as by Lin et al. (2025), show that Design Thinking has become a sustained research and curriculum focus in schools and higher education, which increases the need to understand how emerging tools reshape its practice.

The diffusion of generative artificial intelligence (AI) tools in higher education has altered the material conditions of design work. Chatbots, code assistants and media generators can produce drafts of narratives, interface text, test plans, code snippets and design rationales within seconds. Institutions are now shifting from prohibition to managed integration, with guidance emphasising transparency, privacy, verification and development of AI literacies (Jin et al., 2025; UNESCO, 2023). Education research warns that large language models can support learning and feedback, but also introduce risks of misinformation, over-reliance and integrity ambiguity, which makes student verification practices and AI literacy critical learning outcomes (Giannakos et al., 2025; Kasneci et al., 2023; Long & Magerko, 2020).

In relation to the AI shift, previous studies indicate that student attitudes to AI are heterogeneous and shaped by perceived usefulness, perceived ease of use, trust and policy clarity. Acceptance and perception research reports both enthusiasm for productivity gains and concern about authorship, fairness and uncertainty about boundaries of acceptable assistance (Albayati, 2024; Ravšelj et al., 2025). Trust is particularly consequential because it influences when students treat AI outputs as suggestions that require checking versus as authoritative answers (Nazaretsky et al., 2025). These dynamics are intensified in design and development courses where outcomes include multimodal artefacts and teams are assessed on both product quality and design justification.

In educational game development, AI may function multiple roles. It can be used as an ideation partner, a prototyping accelerator and a technical scaffold (Ratican & Hutson, 2024; Yannakakis & Togelius, 2018). It may also compress the time learners spend on problem framing, push teams towards plausible but unverified user narratives and reduce collective diversity of solutions, even when individual outputs appear more novel. Evidence from creativity research suggests generative AI can enhance individual creativity while reducing diversity of novel content at the group level, which has implications for cohort-wide originality in project-based courses (Doshi & Hauser, 2024). Design research similarly frames AI as a co-creator and a design material that reshapes how options are generated and evaluated, sometimes encouraging dependency or superficial reasoning if judgement is outsourced (Ivcevic & Grandinetti, 2024; Yu, 2025). Within game contexts, emerging evidence highlights both benefits and risks, from faster iteration to concerns about originality and creative constraint (Alharthi, 2025), as well as the feasibility of embedding generative AI to provide scaffolding in educational role-play and scenario-based game designs (Chien et al., 2024; Xu et al., 2025).

Despite increasing use of generative AI in higher education, there is limited empirical understanding of how learners experience AI integration during structured, team-based Design Thinking processes that culminate in functional educational game prototypes. Much of the current research focuses on general chatbot use, single-session interventions (Xiaoyu et al., 2025; Zawacki-Richter et al., 2019), which overlooks the situated dynamics of collaborative design, iterative prototyping and the tensions between authenticity, creativity and responsible practice. Without such understanding, educators risk either under-utilising AI tools that could scaffold learning or allowing AI use to substitute for inquiry, judgement and reflection, which compromises learning and assessment validity.

Hence, this study aims to examine learners' perceptions of AI integration within a five-week Design Thinking cycle in a Digital Design and Development course. Specifically, it addresses the following questions:

RQ1: What relationship exists between students' perceptions of AI integration (perceived usefulness, creativity support, trust and verification, ethical concerns) and overall Design Thinking experience (learning value, engagement, satisfaction)?

RQ2: How does AI use vary across Design Thinking stages during educational game development?

RQ3: What differences emerge between teams with higher and lower AI reliance in prototype quality and reflective justification?

2. Conceptual and Empirical Background

2.1. Design Thinking as a Pedagogical Structure for Game Design Projects

Design Thinking is commonly operationalised in education via a five-stage cycle. While models differ, a shared emphasis remains on human-centred inquiry and iterative prototyping. In educational settings, Design Thinking is frequently paired with project-based learning and has been linked to development of competencies such as empathy, problem framing and creative problem-solving (Baltador et al., 2024; Cummings & Yur-Austin, 2022). A meta-analysis across 25 studies indicates that Design Thinking has a positive relationship with student learning, with moderators including stage model, team size and educational level (Yu et al., 2024). Research also suggests that online and collaborative implementations can build Design Thinking skills when supported by structured scenarios and artefact-based workflows (Sürmelioglu & Erdem, 2025).

For game design specifically, Design Thinking is attractive because it aligns with player-centred design and playtesting loops (Arnab et al., 2019; Chuah et al., 2021). It supports moving from an ill-structured real-world issue to a playable representation and iterating based on feedback. However, Design Thinking can be performed superficially, where empathy is reduced to assumptions, problem statements are generic and prototypes are treated as final products. These known pedagogical failure modes are important because AI tools, as highlighted by Saeidnia & Ausloos (2024), may either mitigate them through scaffolding or amplify them by providing plausible but unverified content.

2.2. AI-Infused Creative Work

A growing body of work frames AI not merely as automation but as a partner within creative cognition and production. In creativity research, AI is increasingly analysed as a creativity support tool that can augment ideation across levels of creativity and domains (Ivcevic & Grandinetti, 2024). In design research, AI can function as co-creator and as design material, reshaping what counts as an idea, how options are evaluated and how designers allocate attention (Sreenivasan & Suresh, 2024; Yu, 2025). These shifts matter educationally because students may not yet have stable professional judgement or disciplinary intuition to arbitrate AI suggestions.

Evidence from controlled studies also suggests that generative AI can increase perceived creativity and improve individual output quality, while also producing convergence in the space of ideas at the cohort level (Doshi & Hauser, 2024). In team-based classes, convergence can be pedagogically undesirable if it reduces diversity of solutions to real-world problems. It can also lead to prototype homogeneity, which can mask differences in learning outcomes and complicate assessment.

2.3. AI and Game Design

The integration of AI into game design and serious games is accelerating across professional practice and education. Designers report that generative AI can accelerate ideation and prototyping, while raising concerns about originality, authorship and creative constraint (Alharthi, 2025). Recent educational game work demonstrates that generative AI can support scaffolding and feedback interactions within role-playing and scenario-based formats, though design must attend to cognitive load, epistemic risk and learning alignment (Chien et al., 2024; Gruber & Faßbender, 2025; Xu et al., 2025).

As stipulated by Shen et al. (2025), game development education adds further complexity because students must coordinate multimodal production. Generative AI can draft dialogue, generate concept art and support code generation or debugging, but it can also introduce brittle code, inconsistent art direction and unverified claims in design rationales. Within learning game design settings, students' positive experiences with generative AI appear to depend on value orientations and prior attitudes. This suggests that AI infusion is not uniformly beneficial and may amplify differences in confidence and agency (Huang & Katz, 2025).

2.4. Working Model for AI-Infused Game Design Thinking

This study adopts a pragmatic working model that integrates technology acceptance, creativity support and trust-ethics perspectives, contextualised to Design Thinking stages. Technology acceptance constructs remain relevant because perceived usefulness and related expectancy constructs predict satisfaction and continued use across educational technologies (Davis, 1989; Venkatesh et al., 2003). Creativity support perspectives add constructs such as perceived creativity support and perceived autonomy, which are crucial in game design tasks that demand divergence, synthesis and aesthetic judgement (Ivcevic & Grandinetti, 2024). Trust and verification are positioned as core mediators rather than secondary factors, given the epistemic risks of confident but incorrect outputs (Kasneci et al., 2023; Nazaretsky et al., 2025). Ethical

concern is included because students often experience uncertainty about acceptable AI assistance and authorship boundaries in assessed, multimodal production (Albayati, 2024).

Figure 1 presents the conceptual model. The upper portion shows the Design Thinking stages as the pedagogical context within which AI integration occurs (RQ2). The middle portion shows the four perception constructs as predictors of overall Design Thinking experience (RQ1). The lower portion shows the team-level outcomes of prototype quality and reflective justification, compared across high and low AI reliance groups (RQ3). The model positions the AI-infused Design Thinking process as a smart learning environment (Hwang, 2014; Spector, 2014) in which generative AI tools, pedagogical structure and collaborative production interact to shape both individual perceptions and team-level outcomes.

3. Method

3.1. Research Design

A convergent mixed-method design was employed (Morse, 2016). Quantitative and qualitative data were collected during the same five-week period, analysed separately and then integrated through joint displays and narrative weaving. This approach is appropriate because the topic involves measurable perceptions and also complex, contextualised meaning-making.

3.2. Participants and Context

Participants were 82 undergraduate students enrolled in a Digital Design and Development course at a Malaysian university. The course was conducted in blended mode and included a project assessment requiring teams to create an educational digital game addressing a real-world problem. Students were divided into 14 groups (12 groups of six, two groups of five). The instructional design included a five-week Design Thinking process prior to the prototype pitch. Weekly activities aligned with the empathise, define, ideate, prototype and test stages.

AI tools (such as Claude AI, ChatGPT Codex, Google AI Studio) were made available and their use was permitted within a responsible-use policy. Students were instructed to document where AI was used, to verify outputs and to avoid presenting AI-generated user insights as real data. This policy aligns with common guidance emphasising transparency, verification and safeguarding of privacy (UNESCO, 2023) as well as the university guidelines on AI usage for assessments.

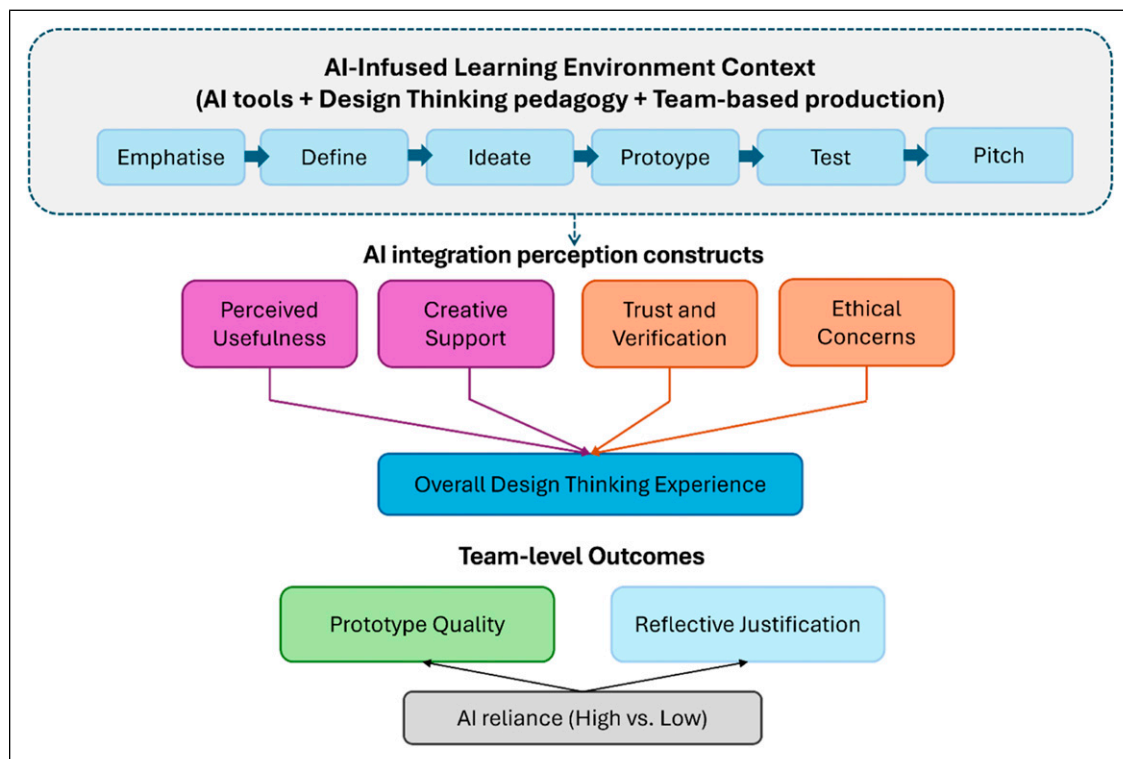


Figure 1. Conceptual model for AI-infused game design thinking

Part A of the survey profiled participants' background and AI use patterns. Prior game development experience was low ($M = 2.29$, $SD = 0.89$), whilst prior experience with generative AI tools was moderate ($M = 3.41$, $SD = 1.01$), indicating that most students had some familiarity with AI tools before the course but limited game development backgrounds. The most commonly used AI tool was ChatGPT (91.5% of participants), followed by image generators (45.1%), Google AI Studio (37.8%) and Claude AI (35.4%). Audio or video generators were used by 11.0% and other AI tools by 8.5%. On average, participants reported using about 3 tools ($M = 2.84$, $SD = 0.71$) across the project.

Furthermore, AI use frequency was moderate ($M = 3.20$, $SD = 0.67$ on a 1–5 scale), with the majority of participants (54.9%) reporting a frequency of 3 (sometimes), 30.5% reporting 4 (often) and 13.4% reporting 2 (rarely). No participant reported never using AI tools. When asked to estimate the proportion of their final submission that relied on AI-assisted work, participants reported the highest reliance for text ($M = 33.9\%$, $SD = 14.0\%$), followed by code ($M = 28.5\%$, $SD = 17.9\%$) and visual or audio assets ($M = 21.5\%$, $SD = 14.6\%$). These estimates suggest that AI was used as a supplementary resource rather than a primary production method, with text-based tasks receiving the most AI assistance.

3.3. Procedure

Over five weeks, teams progressed through the Design Thinking cycle aligned with empathise, define, ideate, prototype and test stages. In Week 1, they identified a real-world problem area, explored relevant stakeholders and gathered at least two forms of user-oriented evidence. AI was permitted for drafting interview prompts, summarising transcripts and generating alternative stakeholder perspectives, although students were required to retain raw notes and justify their synthesis. In Week 2, teams defined the problem by developing a problem statement, identifying a target audience and establishing success criteria. In doing so, they can consult AI tools to seek clarification about their chosen problem. In Week 3, they generated ideas for mechanics, narrative and learning design before selecting one concept for prototyping. AI could support brainstorming, story generation and microcopy drafting, but teams were expected to critique AI suggestions and document their reasons for selection. In Week 4, teams developed a playable prototype using the course toolchain, with AI permitted for coding support, debugging explanations and asset drafting. Students also tracked AI-assisted code and maintained a change log. In Week 5, teams conducted at least one playtest, revised the prototype and delivered a final pitch that included an evidence-informed justification, a demo and a statement of AI use.

3.4. Survey Instrument

The post-process survey comprised four parts administered immediately after the final pitch. Part A profiled AI use (tools, frequency, stage-specific and task-specific use, estimated proportion of AI-assisted work). Part B was a 15-item perception scale using a five-point Likert response format (1 = strongly disagree to 5 = strongly agree) covering four constructs: perceived usefulness (five items), perceived creativity support (four items, one reverse-coded), trust and verification (three items, one reverse-coded) and ethical concerns (three items). Part C included short open-ended prompts. Part D comprised three items measuring overall Design Thinking experience: learning value, engagement and satisfaction (five-point Likert scale).

3.5. Qualitative Data Sources

Qualitative data were gathered from three sources. First, one focus group was conducted per team after the final pitch. Each session lasted 20 to 30 minutes. All team members were invited and most contributed actively. Typically four to six members spoke in each session. Discussions covered AI use across stages, perceived benefits, recurring problems, ethical reasoning and stage dynamics. Sessions were audio-recorded with participant consent and transcribed verbatim for analysis. Second, students kept individual weekly reflective logs. The logs recorded design decisions, setbacks, AI use and verification practices. Third, team artefacts were collected and scored with a rubric. These artefacts included pitch decks, working prototypes, design logs and change logs.

3.6. Data Analysis

Quantitative analysis proceeded in four stages. First, item-level and subscale descriptive statistics were computed. Second, internal consistency was estimated using Cronbach's alpha and McDonald's omega. Third, exploratory factor analysis (EFA) was conducted on the 15 perception items using minimum residual extraction with oblimin rotation to assess dimensionality. Sampling adequacy was evaluated with Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) index. Factor retention was guided by parallel analysis at the 95th percentile criterion. Parallel analysis indicated a two-factor solution. We

extracted and inspected both the two-factor and the four-factor solutions. The two solutions were compared on interpretability and on alignment with the four theorised constructs. Section 4.3 reports this comparison in full. The four-factor solution was retained and used to derive composite subscales for subsequent analyses. Fourth, bivariate Pearson correlations were computed among the perception subscales and overall Design Thinking experience. For RQ1, standard multiple regression estimated the relationship between the four perception subscales and overall Design Thinking experience, with hierarchical regression used as a supplementary analysis to examine incremental variance explained by each subscale. For RQ2, a Friedman test compared AI use frequency across Design Thinking stages, followed by Bonferroni-corrected Wilcoxon signed-rank post-hoc tests with effect sizes reported as r . For RQ3, teams were classified into high and low AI reliance groups. A median split was applied to a composite reliance index. The index combined standardised AI use frequency with tool diversity. The median split reduces a continuous measure to two categories. It can also place teams with similar scores into different groups. The team-level comparison is therefore treated as exploratory throughout. Mann-Whitney U tests compared team-level prototype quality and reflective justification scores between groups, and additional Mann-Whitney U tests examined whether individual perception subscales differed by team AI reliance group, with effect sizes reported as Cliff's delta.

Qualitative analysis used reflexive thematic analysis following [Braun & Clarke \(2019\)](#). Two researchers coded the data independently. Coding began inductively and stayed close to participants' own words. A deductive lens was then applied. Codes were mapped to the Design Thinking stages and to the four perception constructs. This produced a hybrid inductive and deductive approach. Coding attended to contradictions and trade-offs. Stage-specific dynamics were also examined closely. The two coders met regularly to compare codes and resolve differences. Disagreements were discussed until consensus was reached. The codebook was refined accordingly. Reflective logs and artefact scores were analysed alongside the focus group data. Convergent and divergent patterns were recorded for each theme. Integration with the quantitative results used joint displays, as reported in Section 4.8. Each joint display aligns a quantitative pattern with the matching qualitative evidence. Trustworthiness was supported by this peer review of coding and by close grounding in participant accounts.

3.7. Ethics

Ethical procedures were implemented to safeguard participants' privacy and welfare. Ethics approval was obtained from the host institution before the study began. All procedures followed the relevant institutional guidelines. Participation was voluntary and was kept separate from course assessment. Students were told that taking part would not affect their grades. Refusal or later withdrawal carried no penalty. Written informed consent was obtained before any data were collected. Students received a plain-language information sheet describing the study purpose and the voluntary nature of participation, and they signed a consent form confirming that they understood their right to withdraw without penalty. Data collection took place after the project had been graded. Participation could not therefore influence assessment outcomes. The study posed minimal risk to participants. This risk was further reduced by de-identifying responses and restricting data access. Focus group recordings were stored securely and seen only by the research team. The potential benefits for AI-infused pedagogy were judged to outweigh these minimal risks. Students were also instructed not to enter identifiable personal or institutional information into third-party AI tools.

4. Results

4.1. Item-Level Descriptives

[Table 1](#) reports the item-level means and standard deviations for the 15-item perception scale and the three-item Design Thinking experience scale. Perceived usefulness items ranged from $M = 1.95$ ($SD = 0.53$) for AI supporting clearer problem statements (B5) to $M = 3.20$ ($SD = 0.63$) for AI helping generate ideas quickly (B1). Creativity support items ranged from $M = 2.57$ ($SD = 0.65$) for novel combinations (B8) to $M = 2.93$ ($SD = 0.58$) for exploring wider range (B6). The reverse-coded genericity risk item (B9) had a high raw mean ($M = 4.55$, $SD = 0.50$), indicating that students did perceive a tendency towards generic solutions. Trust and verification items showed moderate endorsement of checking behaviours (B10, $M = 2.87$, $SD = 0.50$), while the reverse-coded over-trust item (B12, $M = 4.73$, $SD = 0.45$) suggested students frequently encountered AI outputs that appeared correct when they were not. Ethical concerns items were moderate, ranging from $M = 2.08$ to $M = 2.47$.

Table 1. Descriptive Statistics for Survey Items (n = 82)

Item	Description	M	SD	Skew	Kurt
B1	AI helped generate ideas quickly	3.12	0.64	0.19	0.23
B2	AI improved prototype decisions	2.66	0.59	−0.46	0.19
B3	AI supported technical understanding	2.82	0.74	0.12	0.31
B4	AI supported faster prototyping	2.65	0.64	0.17	−0.36
B5	AI helped clearer problem statement	2.17	0.62	0.49	0.90
B6	AI helped explore wider range	2.83	0.68	−0.25	0.16
B7	AI increased confidence to propose ideas	2.77	0.65	0.00	−0.22
B8	AI helped generate novel combinations	2.38	0.70	0.23	−0.03
B9 ^R	AI pushed towards generic solutions	4.72	0.50	−1.57	1.63
B10	Trusted AI only after checking	2.78	0.65	−0.03	−0.15
B11	Could tell when AI was unreliable	2.35	0.60	−0.31	−0.64
B12 ^R	AI seemed correct when it was not	4.82	0.39	−1.67	0.81
B13	Uncertain about acceptable assistance	2.40	0.63	0.07	−0.18
B14	Concerns about originality/authorship	2.60	0.61	0.13	−0.39
B15	Concerned about privacy/consent	2.09	0.59	−0.02	−0.10
D1	Valuable learning experience	3.09	0.65		
D2	Felt engaged	2.99	0.68		
D3	Satisfied with process and prototype	2.94	0.69		

Note. ^R = reverse-coded item. Means reported in original direction. D1–D3 are overall Design Thinking experience items.

4.2. Scale Reliability and Subscale Descriptives

Table 2 reports internal consistency estimates and subscale descriptives. Perceived usefulness ($\alpha = .827$, $\omega = .827$) and DT experience ($\alpha = .820$, $\omega = .821$) showed good reliability. Ethical concerns showed adequate reliability ($\alpha = .690$, $\omega = .690$). Creativity support showed good reliability ($\alpha = .768$, $\omega = .759$), consistent with the inclusion of a reverse-coded item that may attenuate internal consistency. Trust and verification showed the lowest reliability ($\alpha = .646$, $\omega = .635$), which is noted as a limitation given the three-item scale and the heterogeneity of verification behaviours. All subscales were retained for analysis, with trust and verification results interpreted cautiously. Ethical concerns also showed only adequate reliability. Conclusions involving either subscale are therefore drawn with caution. A sensitivity analysis in Section 4.5 re-estimates the regression without the trust and verification subscale.

4.3. Exploratory Factor Analysis

Prior to EFA, sampling adequacy was assessed. Bartlett's test of sphericity was significant ($\chi^2 = 423.41$, $p < .001$). This confirmed sufficient inter-item correlations. The Kaiser-Meyer-Olkin measure was .734, which exceeds the commonly cited threshold of .70. The subject-to-item ratio was 5.47:1, based on 15 items and 82 respondents. This ratio sits below the often-recommended 10:1 standard. The EFA is therefore treated as exploratory rather than confirmatory. Parallel analysis at the 95th percentile criterion indicated retention of two factors. Both the two-factor and the four-factor solutions were extracted with minimum residuals and oblimin rotation. The two solutions were then compared on interpretability and theoretical fit.

Parallel analysis indicated two factors, so this solution was examined first. The two-factor result is reported in Table 3. Factor 1 absorbed all five usefulness items together with the creativity support items. Factor 2 combined the trust and

Table 2. Subscale Reliability and Descriptive Statistics

Scale	k	M	SD	α	ω
Perceived Usefulness	5	2.68	0.50	.827	.827
Creativity Support	4	2.31	0.49	.768	.759
Trust & Verification	3	2.11	0.43	.646	.635
Ethical Concerns	3	2.36	0.48	.690	.690
DT Experience	3	3.00	0.58	.820	.821

Note. k = number of items. M and SD are composite means. α = Cronbach's alpha. ω = McDonald's omega.

Table 3. Pattern Matrix From Exploratory Factor Analysis (Oblimin Rotation, Two Factors)

Item	F1 (benefit)	F2 (caution)
B1	.624	.061
B2	.561	.064
B3	.803	-.128
B4	.535	-.007
B5	.687	-.015
B6	.677	-.068
B7	.528	.054
B8	.577	.118
B9r	.382	-.008
B10	.409	.458
B11	.302	.596
B12r	-.043	.481
B13	-.120	.573
B14	-.124	.756
B15	-.101	.514

Note. Extraction: minimum residuals. Rotation: oblimin. Loadings $\geq .40$ indicate the dominant factor for each item. F1 is a general perceived-benefit factor combining usefulness and creativity items. F2 is a caution factor combining trust, verification and ethical-concern items. r = reverse-coded. The two-factor solution accounted for 37.5% of total variance.

verification items with the ethical concerns items. Item B10 cross-loaded on both factors (.41 and .46). This solution is parsimonious, yet it merges constructs that behave differently in later analyses. For example, usefulness and creativity support collapse into one general benefit factor. Creativity support, however, predicts the outcome independently of usefulness in the regression. Trust and verification also merge with ethical concerns under the two-factor solution. These two constructs relate to the outcome in opposite directions. One is positive and the other is negative, so merging them hides this contrast. The four-factor solution separates these constructs cleanly. It also accounts for more variance, rising from 37.5% to 47.5%. The four-factor structure was therefore retained on theoretical and interpretive grounds. This choice follows guidance that the clearest solution may exceed the number suggested by parallel analysis. The instrument is still treated as exploratory, and no strong psychometric claims are made.

Table 4 presents the pattern matrix. Factor 1 (perceived usefulness) showed loadings from B3 (.736), B1 (.703), B2 (.696), B5 (.651) and B4 (.623). Factor 2 (creativity support) loaded on B7 (.853), B6 (.714) and B8 (.683). Factor 3 (ethical concerns) loaded on B14 (.706), B13 (.657) and B15 (.555). Factor 4 (trust and verification) loaded on B11 (.915) and B10 (.525). Factor correlations were modest, with the strongest correlation between F1 and F3 ($r = .434$). The four-factor structure was broadly consistent with the a priori model, supporting the use of composite subscale scores in subsequent analyses.

4.4. Bivariate Correlations

Table 5 presents bivariate Pearson correlations among subscales and the outcome variable. Overall DT experience was significantly correlated with perceived usefulness ($r = .540, p < .001$), creativity support ($r = .646, p < .001$) and trust and verification ($r = .294, p = .007$). The correlation with ethical concerns was negative but non-significant ($r = -.155, p = .166$). Perceived usefulness and creativity support were moderately correlated ($r = .461, p < .001$), suggesting related but distinguishable constructs. Ethical concerns and trust and verification showed a positive correlation ($r = .410, p < .001$), indicating that students with stronger verification habits also reported more ethical uncertainty.

Table 4. Pattern Matrix From Exploratory Factor Analysis (Oblimin Rotation, Four Factors)

Item	F1 (PU)	F2 (PCS)	F3 (EC)	F4 (TV)
B1	.703	−.077	−.020	.143
B2	.696	−.036	.114	−.023
B3	.736	.157	−.098	−.020
B4	.623	−.029	.016	−.000
B5	.651	.095	−.028	.043
B6	.180	.714	−.067	−.040
B7	−.101	.853	−.015	.052
B8	.134	.683	.140	−.034
B9r	.020	.237	−.288	.349
B10	.132	.149	.084	.525
B11	.018	−.038	.033	.915
B12r	−.097	.072	.355	.184
B13	.069	−.048	.657	−.029
B14	−.122	.119	.706	.123
B15	.103	−.109	.555	.011

Note. Extraction: minimum residuals. Rotation: oblimin. Loadings $\geq .40$ are in bold in the original analysis. PU = perceived usefulness, EC = ethical concerns, PCS = perceived creativity support, TV = trust and verification. r = reverse-coded.

4.5. Students' Perceptions of AI Integration and Design Thinking Experience

To address RQ1, a multiple regression model predicted overall Design Thinking experience from the four perception subscales. Table 6 presents the results. The model was significant, $F(4, 77) = 20.86$, $p < .001$, $R^2 = .520$, adjusted $R^2 = .495$, representing a large effect (Cohen's $f^2 = 1.084$). Multicollinearity was not a concern, with all VIF values below 1.44. The f^2 value of 1.084 is very large and far exceeds the .35 benchmark for a large effect. Such a high value, with four predictors and 82 cases, warrants caution about overfitting. The adjusted R^2 of .495 shows only modest shrinkage from the R^2 of .520. The effect is best read as substantial rather than precisely estimated.

Perceived usefulness ($\beta = .273$, $t = 2.99$, $p = .004$) and perceived creativity support ($\beta = .466$, $t = 5.10$, $p < .001$) were both significant positive predictors. Ethical concerns was a significant negative predictor ($\beta = -.178$, $t = -2.00$, $p = .049$). Trust and verification ($\beta = .150$, $t = 1.58$, $p = .117$) did not reach significance. Hierarchical regression confirmed that perceived usefulness alone accounted for 29.2% of variance (Step 1, $p < .001$), with creativity support contributing an additional 20.0% (Step 2, $\Delta R^2 = .200$, $p < .001$). Trust and verification did not add significantly ($\Delta R^2 = .003$, $p = .465$), while ethical concerns made a marginally significant contribution ($\Delta R^2 = .025$, $p = .049$).

A sensitivity analysis addressed the low reliability of the trust and verification subscale. The regression was re-estimated without that subscale. Perceived usefulness and creativity support remained strong and significant predictors. The reduced model still explained almost half of the variance ($R^2 = .504$, adjusted $R^2 = .485$). The main conclusions were therefore unchanged. Ethical concerns became non-significant once trust and verification was removed ($\beta = -.113$, $p = .161$). This

Table 5. Bivariate Pearson Correlations Among Subscales ($n = 82$)

	1	2	3	4	5
1. PU	—				
2. PCS	.461**	—			
3. TV	.289**	.298**	—		
4. EC	−.052	−.052	.410**	—	
5. DT Exp	.540**	.646**	.294**	−.155	—

Note. * $p < .05$. ** $p < .01$. PU = perceived usefulness, PCS = perceived creativity support, TV = trust and verification, EC = ethical concerns, DT Exp = overall Design Thinking experience.

Table 6. Multiple Regression Predicting Overall Design Thinking Experience From Perception Subscales

Predictor	B	SE	β	t	p	95% CI
(Constant)	0.964	0.368	—	2.62	.011	[0.23, 1.70]
PU	0.316	0.106	.273	2.99	.004	[0.11, 0.53]
PCS	0.549	0.108	.466	5.10	<.001	[0.34, 0.76]
TV	0.204	0.129	.150	1.58	.117	[−0.05, 0.46]
EC	−0.215	0.108	−.178	−2.00	.049	[−0.43, −0.00]

Note. $R^2 = .520$, Adjusted $R^2 = .495$, $F(4, 77) = 20.86$, $p < .001$, Cohen's $f^2 = 1.084$.

shift reflects the positive correlation between the two constructs. Findings for ethical concerns are therefore interpreted with added caution.

4.6. AI Use Across Design Thinking Stages

To address RQ2, Table 7 reports AI use frequency across the six Design Thinking stages. A Friedman test indicated significant differences in AI use across stages, $\chi^2(5) = 63.76$, $p < .001$. Bonferroni-corrected Wilcoxon signed-rank tests showed that AI use during prototyping ($M = 3.83$, $SD = 0.84$) and ideation ($M = 3.83$, $SD = 0.72$) was significantly higher than during empathise ($M = 2.99$, $SD = 0.97$; $z = 4.97$, $p < .001$, $r = .549$ for empathise vs prototype; $z = 5.42$, $p < .001$, $r = .599$ for define vs ideate). The difference between ideation and prototyping was not significant ($z = 0.02$, $p = 1.000$). Early stages (empathise, define) showed the lowest AI use, with no significant difference between them ($z = 1.61$, $p = 1.000$).

4.7. Team Comparisons by AI Reliance

To address RQ3, teams were classified into high ($n = 7$ teams) and low ($n = 7$ teams) AI reliance groups using a median split on the composite reliance index. Table 8 presents team-level prototype quality and reflective justification scores.

High AI reliance teams achieved significantly higher prototype quality scores ($M = 16.3$, $SD = 1.3$) than low reliance teams ($M = 12.0$, $SD = 1.5$), $U = 48.0$, $p = .003$, Cliff's $\delta = 0.959$, representing a large effect. However, reflective justification scores did not differ significantly between groups (high $M = 7.3$, $SD = 1.6$; low $M = 8.3$, $SD = 1.5$; $U = 21.0$, $p = .688$, Cliff's $\delta = -0.143$). Individual-level perception subscales showed no significant differences between students in high versus low reliance teams for any subscale (all $p > .05$), suggesting that AI reliance was a team-level phenomenon rather than a reflection of individual attitude differences. These team results rest on only 14 teams, with seven in each group. A single unusual team could shift the outcome. The median split also reduces a continuous measure to two categories. The comparison is therefore tentative and needs replication.

4.8. Qualitative Findings

Reflexive thematic analysis of focus group transcripts, reflective logs and artefact coding produced four themes that both converge with and extend the quantitative patterns. A joint display in Table 9 maps each theme onto the matching quantitative result. The display records how the qualitative evidence confirms or complicates that result.

Table 7. AI Use Frequency by Design Thinking Stage

Stage	M	SD	Median
Empathise	2.99	0.97	3.0
Define	2.89	1.05	3.0
Ideate	3.83	0.72	4.0
Prototype	3.83	0.84	4.0
Test	3.33	1.04	3.0
Pitch	3.44	0.86	4.0

Note. Friedman $\chi^2(5) = 63.76$, $p < .001$. Scale: 1 = never, 5 = very often.

Table 8. Team-Level Prototype Quality and Reflective Justification by AI Reliance Group

Team	Group	Proto.	Justif.	FC	AL	LD	IT
1	High	15	9	3	3	3	3
2	High	17	4	3	3	4	3
3	High	14	6	2	3	3	3
4	Low	13	7	3	3	3	2
5	High	16	8	3	4	3	3
6	Low	14	11	3	2	3	3
7	High	17	8	4	3	3	4
8	Low	14	8	2	3	3	3
9	Low	10	7	2	2	2	2
10	Low	11	7	3	2	2	2
11	Low	11	10	2	2	2	2
12	High	17	8	4	3	3	4
13	High	18	8	4	4	4	3
14	Low	11	8	2	2	2	3

Note. Proto. = prototype quality total (max 20). Justif. = reflective justification total (max 12). FC = functional completeness, AL = alignment, LD = learning design, IT = iteration evidence. All criterion scores 0–4.

4.8.1. Theme 1: AI as Productivity Scaffold, Not Design Partner

Across all 14 teams, the most frequently cited benefit of AI was speed. Students described using AI to generate initial code, draft dialogue and troubleshoot errors. However, teams consistently distinguished between productivity uses (where AI saved time) and design uses (where AI shaped decisions). Most teams reported that AI was treated as a tool for execution rather than as a contributor to design reasoning. As one student noted in their reflective log, the AI gave them working code but did not help them understand why certain mechanics suited their learning objectives.

Two excerpts capture this distinction.

“The AI was great for getting unstuck. It gave us working code in minutes. But it never told us why a mechanic fitted our learning goal. That part was still on us” (Team 4, focus group).

“I used it like a faster search engine. It speeds up the building, not the thinking” (Participant 23, reflective log).

4.8.2. Theme 2: Stage-Dependent Integration Patterns

Consistent with the quantitative finding that AI use was lowest during empathise and define stages, focus groups revealed deliberate avoidance of AI for user research. Several teams expressed concern that AI-generated personas would lack the specificity needed for their chosen real-world problems. In contrast, AI use during ideation and prototyping was described as natural and efficient. Teams reported that AI was most valuable when the task was well-defined and the team already had criteria for evaluating outputs.

Table 9. Joint Display Integrating Quantitative and Qualitative Findings

Theme	Quantitative signal	Qualitative insight	Integrated inference
1. Productivity scaffold	AI use frequency highest in prototyping and ideation.	AI valued for speed and execution, not design reasoning.	Confirms and explains: AI serves production more than judgement.
2. Stage-dependent use	AI use lowest in the empathise and define stages.	Teams avoided AI personas as too generic for real users.	Expands: low early use was deliberate, not incidental.
3. Verification	Trust scores moderate; subscale reliability low.	Checking ranged from systematic to skipped near deadlines.	Complicates: verification is uneven and learned, not default.
4. Ethical awareness	Ethical concerns predict lower DT experience.	Students used AI yet felt unease over authorship and privacy.	Confirms: ethical awareness adds friction without stopping use.

Note. Each row aligns a qualitative theme with its corresponding quantitative result. The final column records the integrated inference from the convergent design.

Focus group accounts made this boundary explicit.

“We did not trust AI for the empathise part. The personas it gave us felt generic. Our users were specific, so we spoke to real people instead” (Team 9, focus group).

“Once we knew the problem, AI was perfect for prototyping. We could judge what was good because we had set our own criteria” (Team 2, focus group).

4.8.3. Theme 3: Verification as Learned Practice, Not Default Behaviour

Despite moderate quantitative scores on trust and verification items, qualitative data revealed that verification practices were unevenly developed. Some teams described systematic checking strategies, including running AI-generated code in isolation, cross-referencing AI design suggestions with playtest feedback and maintaining a change log of AI-assisted modifications. Other teams acknowledged that they often accepted AI outputs without checking, particularly under time pressure. This unevenness suggests that verification is a skill that develops through practice and explicit instruction rather than emerging spontaneously.

The unevenness was visible across teams.

“We ran every AI snippet in a separate file before adding it. If we could not explain it, we did not use it” (Team 6, reflective log).

“Honestly, near the deadline we just pasted what worked. There was no time to check everything” (Participant 51, focus group).

4.8.4. Theme 4: Ethical Awareness Without Resolution

Students expressed awareness of ethical tensions around AI use but often lacked frameworks for resolving them. Common concerns included uncertainty about how much AI assistance was acceptable, discomfort with submitting AI-generated assets as original work and worry about data privacy when entering user research information into AI tools. These concerns did not prevent AI use but created an undertone of ambivalence that shaped how students discussed their work during pitches.

Students voiced this ambivalence directly.

“I felt uneasy submitting AI art as ours. It looked polished, but it did not feel fully mine” (Team 11, focus group).

“Nobody told us where the line was. How much help is too much? We were guessing” (Participant 38, reflective log).

5. Discussion

5.1. AI Integration and Design Thinking Experience

The regression analysis showed that perceived usefulness and perceived creativity support were the two significant predictors of overall Design Thinking experience, jointly accounting for approximately 49% of the variance. This result is consistent with technology acceptance research, which has long identified perceived usefulness as a central determinant of positive user experience and continued engagement with digital tools (Albayati, 2024; Davis, 1989). It also fits emerging higher education research on generative AI, where students tend to respond positively when AI is experienced as practically helpful for brainstorming, drafting and academic support (Chan & Hu, 2023). In the present study, however, usefulness alone was not sufficient. Perceived creativity support made an additional independent contribution, suggesting that students valued AI not only as an efficiency tool but also as a partner in idea generation and concept development. This interpretation is in line with recent work showing that AI can strengthen creative thinking and support ideation, although its benefits depend on how thoughtfully it is integrated into learning tasks (Habib et al., 2024; Sreenivasan & Suresh, 2024).

In contrast, trust and verification did not make a significant independent contribution, despite the emphasis placed on verification as a key mechanism in responsible AI use (Nazaretsky et al., 2025). One plausible interpretation is that trust and verification function less as direct predictors of positive experience and more as regulatory conditions that shape the quality of AI-supported learning. In other words, they may not directly increase students' satisfaction with the Design Thinking process, but they may help prevent over-reliance, shallow acceptance of AI outputs or errors in judgement. This interpretation is consistent with qualitative studies showing that students often recognise the need to question AI-generated content, yet apply verification unevenly and in highly individual ways (Hadinejad et al., 2025). A second interpretation is developmental in which students may already regard checking, cross-referencing and validating AI outputs as necessary but effortful practices, and therefore not as features that enhance their immediate learning experience. This null result should be read with caution. The subscale's reliability was low ($\alpha = .646$). Measurement error can attenuate coefficients and mask a real effect. A

sensitivity analysis without this subscale left the main findings unchanged. The role of trust and verification therefore remains an open question.

The significant negative coefficient for ethical concerns ($\beta = -.178$, $p = .049$) further suggests that ethical awareness introduces a form of cognitive and affective friction that slightly depresses reported satisfaction. Qualitative findings indicated that ethical concern did not prevent AI use but instead created ambivalence that tempered otherwise positive evaluations. This pattern is consistent with recent studies showing that students often continue to use generative AI while simultaneously expressing concern about academic integrity, reliability, fairness and appropriate boundaries of use (Hadinejad et al., 2025; Johnston et al., 2024). Rather than indicating resistance to AI, the negative association may therefore reflect a more cautious and reflexive stance: students who are more alert to ethical issues may experience AI-supported Design Thinking as productive, but also as more complicated and morally demanding.

5.2. Stage-Dependent AI Use in Educational Game Development

The significant Friedman test and post-hoc comparisons showed that AI use was concentrated in the ideation and prototyping stages, with substantially lower use during empathise and define stages. This stage-dependent pattern is pedagogically meaningful rather than merely procedural. In design thinking, the early stages are intended to foreground users' lived needs, support careful interpretation and enable problem framing before teams move towards solution generation. These phases depend on observation, listening and reframing, whereas later stages such as ideation and prototyping are more closely aligned with the affordances of generative AI (Verganti et al., 2021). This pattern is similar to what was reported by Saeidnia & Ausloos (2024) where AI tools are more prevalent in the ideation and prototyping stages. In fact, user feedback may matter even more in game development. The games are made for real players, so authentic input is essential. This helps explain the lower use of AI in the early stages. It also echoes work on building empathy through serious game design (Techanamurthy et al., 2025).

From this perspective, the relatively low AI use in empathise and define stages may also reflect productive pedagogical restraint. The qualitative finding that teams deliberately avoided AI for user research suggests that many students recognised an important boundary condition of AI support in human-centred work. This is notable because design thinking has been criticised when it shifts too quickly into solutionism, reducing complex human situations into problems that seem immediately solvable (Richterich, 2024). Generative AI may intensify that risk because it produces fluent and plausible responses rapidly, which can encourage premature closure if students begin treating those outputs as substitutes for direct user understanding and interpretive inquiry.

The qualitative finding that teams deliberately avoided AI for user research suggests the former interpretation has some support. However, the modest mean for B5 (AI helped make a clearer problem statement, $M = 2.17$) indicates that even when students did use AI in early stages, they did not find it particularly helpful for the kind of human-centred framing that these stages require. On the whole, these findings suggest that AI may be most pedagogically valuable in educational game development when it supports creative expansion after learners have grounded their work in authentic user insights.

5.3. Prototype Quality Versus Reflective Justification

The RQ3 comparison revealed that high AI reliance teams produced significantly stronger prototypes (Cliff's $\delta = 0.959$) but showed no advantage in reflective justification. This pattern corresponds most closely to what the original conceptual framework described as a scenario where AI acts as production scaffolding while compressing learning about design rationale. High reliance teams may have used AI to produce more polished and functionally complete prototypes, but this did not translate into stronger evidence-based reasoning, AI attribution or critical judgement in their pitch narratives (Du et al., 2025).

This finding carries direct implications for assessment design. If prototype quality alone were the assessment criterion, AI reliance would appear beneficial. When reflective justification is included, the picture becomes more nuanced. The absence of a justification advantage for high reliance teams suggests that AI use did not inherently deepen metacognitive engagement (Tankelevitch et al., 2024). Thus, explicit pedagogical scaffolding for reflection and verification appears necessary to convert productivity gains into learning gains.

6. Implications and Limitations

Several design principles for courses that include AI-infused educational game development emerge from these findings. First, assessment should include marks for evidence trails documenting what was generated, what was checked and what was revised. The disparity between prototype quality and justification quality in high reliance teams suggests that verification and

attribution need to be rewarded, not merely permitted. Second, early Design Thinking stages require deliberate protection. Requiring at least one form of non-AI user evidence and penalising ungrounded synthetic empathy can preserve the epistemic integrity of the empathise and define stages. Third, prompt literacy should be taught as design literacy, focusing on generating counterexamples, surfacing assumptions and exploring constraints rather than merely obtaining outputs. Fourth, convergence mitigation strategies such as divergence documentation and originality criteria linked to user evidence can address the risk that AI nudges teams towards generic solutions.

The study contributes a context-specific model connecting technology acceptance and creativity support constructs within a stage-based Design Thinking process for educational game development. The four-factor structure supported by EFA provides empirical support for distinguishing perceived usefulness, creativity support, trust and verification, and ethical concerns as separable dimensions of student experience with AI. The finding that creativity support predicts DT experience independently of usefulness extends acceptance models by foregrounding a construct that is particularly relevant in creative production contexts.

However, this study acknowledges some limitations that constrain interpretation. First, the study was conducted in a single course at one Malaysian university with a modest sample size ($n = 82$), which limits generalisability and the stability of EFA results. The sample, while adequate for the analyses conducted ($KMO = .734$), remains modest for confirmatory purposes. With 15 items and 82 respondents, the subject-to-item ratio was 5.47:1. This falls below the often-recommended 10:1 ratio. The factor structure should be read as exploratory rather than confirmed. Factor loadings could differ in a new sample. Replication with larger and more diverse samples is needed. Second, self-report measures may be influenced by social desirability (Banihashem et al., 2024), especially regarding AI use under assessment conditions. Third, the trust and verification subscale showed relatively low internal consistency ($\alpha = .646$), suggesting that the three items may not fully capture the construct. Fourth, the team-level analysis rested on only 14 teams, with seven per group. The median split reduced a continuous reliance measure to two coarse categories. Both features make the team comparison fragile and sensitive to individual teams. These results are exploratory and require replication with more teams. Fifth, teams selected their own real-world problems, introducing variability in domain complexity and stakeholder access. Finally, AI tools evolve rapidly and differ in capability. This study used a specific set of tools. For example, Claude AI, ChatGPT Codex and Google AI Studio were the main tools available. The findings may not generalise to other AI tools or to later versions.

7. Conclusion

This study examined undergraduate students' perceptions of AI integration across a five-week Design Thinking cycle for educational game development, conceptualising the AI-infused process as a smart blended learning environment. The findings show that the value of AI was not uniform across the design process. Students viewed AI as most useful during ideation and prototyping, where it appeared to support creative expansion and production efficiency, but far less useful during empathise and define, where human-centred interpretation and problem framing remained central. Perceived usefulness and creativity support predicted a more positive Design Thinking experience, while ethical concerns introduced measurable friction. At the team level, higher AI reliance was associated with stronger prototypes but not with stronger reflective justification, suggesting that AI supports making more readily than it supports explaining, critiquing or justifying. All in all, the study argues that the educational value of AI in game development lies not in replacing core Design Thinking practices, but in being orchestrated as a stage-sensitive scaffold within smart learning environments that preserve empathy, strengthen verification and reward reflective judgement alongside product quality. Therefore, in educational game development, AI appears most valuable not when it accelerates everything, but when it is used deliberately at the right stage, for the right purpose, and under conditions that keep learning and human-centred design firmly in view.

ORCID iD

Kee-Man Chuah  <https://orcid.org/0000-0001-7499-5889>

Ethical Considerations

Ethics approval was obtained from the host institution before the study began. All procedures followed the relevant institutional guidelines and regulations. The study posed minimal risk to participants. This risk was further limited by de-identifying responses and restricting data access to the research team. The potential benefits of improving AI-infused pedagogy for students were judged to outweigh these minimal risks.

Consent to Participate

Written informed consent was obtained from all participants before data collection. Participation was voluntary and was kept separate from course assessment. Students were informed that participation would not affect their grades and that refusal or later withdrawal carried no penalty of any kind. Surveys and focus groups were administered after final grades had been released, which kept the data separate from assessment.

Consent for Publication

Not applicable. No identifiable individual data, images or personal details are included in this manuscript.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data that support the findings of this study are not publicly available due to privacy and ethical restrictions involving participant information. De-identified data may be made available from the corresponding author upon reasonable request, subject to institutional approval.

References

- Albayati, H. (2024). Investigating undergraduate students' perceptions and awareness of using ChatGPT as a regular assistance tool: A user acceptance perspective study. *Computers and Education: Artificial Intelligence*, 6, Article 100203. <https://doi.org/10.1016/j.caeai.2024.100203>
- Alharthi, S. A. (2025). Generative AI in game design: Enhancing creativity or constraining innovation? *Journal of Intelligence*, 13(6), Article 60. <https://doi.org/10.3390/jintelligence13060060>
- Arnab, S., Clarke, S., & Morini, L. (2019). Co-creativity through play and game design thinking. *Electronic Journal of E-Learning*, 17(3), 184–198. <https://doi.org/10.34190/jel.17.3.002>
- Baltador, L. A., Grecu, V., Panța, N. D., & Beju, A. M. (2024). Design thinking in education: Evaluating the impact on student entrepreneurship competencies. *Education Sciences*, 14(12), Article 1311. <https://doi.org/10.3390/educsci14121311>
- Banihashem, S. K., Dehghanzadeh, H., Clark, D., Noroozi, O., & Biemans, H. J. A. (2024). Learning analytics for online game-based learning: A systematic literature review. *Behaviour & Information Technology*, 43(12), 2689–2716. <https://doi.org/10.1080/0144929X.2023.2255301>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676x.2019.1628806>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chien, C. C., Chan, H. Y., & Hou, H. T. (2024). Learning by playing with generative AI: Design and evaluation of a role-playing educational game. *Journal of Research on Technology in Education*, 57(4), 1–20. <https://doi.org/10.1080/15391523.2024.2338085>
- Chuah, K. M., Mohamad, F. S., Minoi, J. L., & Kabilan, M. K. (2021, November). Of marbles and Minecraft: Designing STEM educational games for culturally diverse young learners in Malaysia. In *2021 World Engineering Education Forum/Global Engineering Deans Council (WEEF/GEDC)* (pp. 44–49). IEEE.
- Cummings, C., & Yur-Austin, J. (2022). Design thinking and community impact: A case study of project-based learning in an MBA capstone course. *Journal of Education for Business*, 97(2), 126–132. <https://doi.org/10.1080/08832323.2021.1887795>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28), Article eadn5290. <https://doi.org/10.1126/sciadv.adn5290>
- Du, X., Du, M., Zhou, Z., & Bai, Y. (2025). Facilitator or hindrance? The impact of AI on university students' higher-order thinking skills in complex problem solving. *International Journal of Educational Technology in Higher Education*, 22(1), Article 39. <https://doi.org/10.1186/s41239-025-00534-0>
- Giannakos, M. N., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2025). The promise and challenges of generative AI in education: A critical review. *Behaviour & Information Technology*, 44(11), 2518–2544. <https://doi.org/10.1080/0144929X.2024.2394886>

- Gruber, M., & Faßbender, M. (2025). Digital educational escape game design for STEM higher education. *Frontiers in Education*, 10, Article 1497291. <https://doi.org/10.3389/feduc.2025.1497291>
- Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, 34(1), Article 100072. <https://doi.org/10.1016/j.yjoc.2023.100072>
- Hadinejad, N., Sperling, K., & McGrath, C. (2025). Generative AI chatbots in higher education: Student experiences and perceived ethical challenges. *Computers and Education Open*, 9, Article 100311. <https://doi.org/10.1016/j.caeo.2025.100311>
- Huang, D., & Katz, J. E. (2025). GenAI learning for game design: Both prior self-transcendent pursuit and material desire contribute to a positive experience. *Big Data and Cognitive Computing*, 9(4), Article 78. <https://doi.org/10.3390/bdcc9040078>
- Hwang, G. J. (2014). Definition, framework and research issues of smart learning environments: A context-aware ubiquitous learning perspective. *Smart Learning Environments*, 1(1), Article 4. <https://doi.org/10.1186/s40561-014-0004-5>
- Ivcevic, Z., & Grandinetti, A. (2024). Artificial intelligence as a tool for creativity. *Journal of Creativity*, 34(2), Article 100079. <https://doi.org/10.1016/j.yjoc.2024.100079>
- Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies. *Computers and Education: Artificial Intelligence*, 8, Article 100348. <https://doi.org/10.1016/j.caeai.2024.100348>
- Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity*, 20(1), Article 2. <https://doi.org/10.1007/s40979-024-00149-4>
- Kafai, Y. B., & Burke, Q. (2016). *Connected gaming: What making video games can teach us about learning and literacy*. MIT Press.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Li, T., & Zhan, Z. (2022). A systematic review on design thinking integrated learning in K-12 education. *Applied Sciences*, 12(16), Article 8077. <https://doi.org/10.3390/app12168077>
- Liedtka, J. (2018). Why design thinking works. *Harvard Business Review*, 96(5), 72–79. <https://hbr.org/2018/09/why-design-thinking-works>
- Lin, L., Shen, S., Shadiev, R., & Yu, M. (2025). Review of research on design thinking in K-12 education. *Educational Research Review*, 47, Article 100682. <https://doi.org/10.1016/j.edurev.2025.100682>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Morse, J. M. (2016). *Mixed method design: Principles and procedures*. Routledge.
- Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2025). The critical role of trust in adopting AI-powered educational technology for learning. *Computers and Education: Artificial Intelligence*, 8, Article 100368. <https://doi.org/10.1016/j.caeai.2025.100368>
- Pacheco-Velazquez, E., Salinas-Navarro, D. E., & Mejía-Argueta, C. (2024). Developing learning skills through game-based learning in complex scenarios. *Journal of Technology and Science Education*, 14(2), Article 2219. <https://doi.org/10.3926/jotse.2219>
- Ratican, J., & Hutson, J. (2024). Video game development 3.0: AI-driven collaborative co-creation. *Metaverse*, 5(2), 1–18. <https://doi.org/10.54517/m2904>
- Ravšelj, D., Keržič, D., Tomažević, N., Umek, L., Brezovar, N., Iahad, N. A., Aristovnik, A., Akopyan, A., Aldana Segura, M. W., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydın, F., Badran, A., Balbontín-Alvarado, R., & Ben Saad, H. (2025). Higher education students' perceptions of ChatGPT: A global study. *PLOS ONE*, 20(2), Article e0315011. <https://doi.org/10.1371/journal.pone.0315011>
- Richterich, A. (2024). Can't fix this? Innovation, social change, and solutionism in design thinking. *Media and Communication*, 12, Article 7427. <https://doi.org/10.17645/mac.7427>
- Saeidnia, H. R., & Ausloos, M. (2024). Integrating artificial intelligence into design thinking: A comprehensive examination of the principles and potentialities of AI for design thinking framework. *InfoScience Trends*, 1(2), 1–9. <https://doi.org/10.61186/ist.202401.01.09>
- Shen, Y., Jalil, H. A., & Jamaluddin, R. (2025). Exploring adolescents' multimodal literacy as critical and creative prosumers in digital game-based multimodal composition. *Educational Technology & Society*, 28(1), 1–17. <https://www.jstor.org/stable/48810705>
- Spector, J. M. (2014). Conceptualizing the emerging field of smart learning environments. *Smart Learning Environments*, 1(1), Article 2. <https://doi.org/10.1186/s40561-014-0002-7>
- Sreenivasan, A., & Suresh, M. (2024). Design thinking and artificial intelligence: A systematic literature review exploring synergies. *International Journal of Innovation Studies*, 8(3), 297–312. <https://doi.org/10.1016/j.ijis.2024.05.001>

- Sürmelioglu, Y., & Erdem, M. (2025). Enhancing teachers' design thinking skills in the online instructional process. *Humanities and Social Sciences Communications*, 12, Article 5131. <https://doi.org/10.1057/s41599-025-05131-0>
- Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI' 24)* (pp. 1–24). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642902>
- Techanamurthy, U., Chuah, K. M., & Yee, I. V. (2025). Developing engineering students' empathy skills using serious games in project management: An exploratory study. *Australasian Journal of Engineering Education*, 30(2), 109–123. <https://doi.org/10.1080/22054952.2025.2538329>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Verganti, R., Dell'Era, C., & Swan, K. S. (2021). Design thinking: Critical analysis and future evolution. *Journal of Product Innovation Management*, 38(6), 603–622. <https://doi.org/10.1111/jpim.12610>
- Xiaoyu, W., Zainuddin, Z., & Hai Leng, C. (2025). Generative artificial intelligence in pedagogical practices: A systematic review of empirical studies (2022–2024). *Cogent Education*, 12(1), Article 2485499. <https://doi.org/10.1080/2331186X.2025.2485499>
- Xu, S., Lo, C. K., Ling, M. H., & Chen, G. (2025). Leveraging generative AI in a 3D scenario-based game for vocabulary acquisition. *Cogent Education*, 12(1), Article 2560059. <https://doi.org/10.1080/2331186X.2025.2560059>
- Yannakakis, G. N., & Togelius, J. (2018). *Artificial intelligence and games*. Springer.
- Yu, W. F. (2025). AI as a co-creator and a design material: Transforming the design process. *Design Studies*, 97, Article 101303. <https://doi.org/10.1016/j.destud.2025.101303>
- Yu, D., Yu, L., & Lin, X. (2024). A meta-analysis of the effects of design thinking on student learning. *Humanities and Social Sciences Communications*, 11(1), Article 742. <https://doi.org/10.1057/s41599-024-03237-5>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), Article 39. <https://doi.org/10.1186/s41239-019-0171-0>