Xie et al. *Research and Practice in Technology Enhanced Learning* (2027) 22:19
https://doi.org/10.58459/rptel.2027.22019

Research and Practice in Technology Enhanced Learning

**RESEARCH** **Free and Open Access**

# Lightening the load: Effects of AI agent generated humor and question-asking type on cognitive load and learning in online instruction

Xinjie Xie [1]*, Chuwen Liu [2] and Xiangen Hu [3]

*Correspondence:
xinjiexie@foxmail.com
School of Psychology, Central China Normal University, No. 152, Luoyu Road, Wuhan, 430079, China
Full list of author information is available at the end of the article

**Abstract**

This study investigates the effects of humor and question-asking strategies used by generative AI agent on teaching in online learning environments. A 2 (Humor: humor vs. no-humor) × 2 (Question-asking Type: open-ended vs. close-ended) between-subjects factorial design was employed with 116 Chinese university students. Participants interacted with AI agent that delivered instructional content and question-asking consistent with the experimental conditions. Measured outcomes included perceived humor, perceived agent value, intrinsic motivation, positive emotion, cognitive load, and knowledge retention. Results indicated that humor significantly enhanced learners' intrinsic motivation, positive emotions, perceived agent value, and knowledge retention while also reducing intrinsic cognitive load (ICL). The findings underscore the effectiveness of incorporating humor in AI agent to positively influence both affective and cognitive dimensions of learning in online contexts.

**Keywords:** AI agent, humor, question-asking, cognitive load, knowledge retention, online learning

## Background

The integration of AI into education has enabled pedagogical agent–based instruction to provide scalable and adaptive learning experiences enriched with social and emotional cues, such as expressive feedback and conversational interaction (Li et al., 2019). In this study, the term AI agent refers specifically to a pedagogical agent powered by AI. Despite these advances, online learning environments continue to face challenges related to low engagement, reduced motivation, and negative emotional experiences, largely due to limited human interaction (Erdoğdu & Çakıroğlu, 2021). Improving online learning therefore requires instructional designs that simultaneously support learners' emotional engagement and manage their cognitive demands, both of which are critical for sustained motivation and effective learning (Frisby et al., 2024; Peng, 2025).

Instructional humor represents a promising yet underexplored design strategy in this context. Prior studies have shown that humor can enhance engagement and reduce anxiety, but this evidence is largely drawn from settings involving human instructors or scripted pedagogical agent–based systems (Banas et al., 2011; Buttussi & Chittaro, 2020). As a result, little is known about how humor functions in an AI agent that dynamically generates explanations, feedback, and questions in real time. At the same time, question-asking strategies—particularly open-ended questions—are widely regarded as beneficial for deeper learning, although they can also impose substantial cognitive load (Craig et al., 2012). It remains unclear whether humor can function as a pedagogical mechanism that offsets the cognitive demands induced by such question-asking strategies in AI agent–mediated instruction.

This study addresses these gaps by systematically examining how humor generated by an AI agent and question-asking type jointly influence learners' emotional, cognitive, and learning outcomes. Specifically, we investigate perceived humor, perceived agent value, intrinsic motivation, positive emotion, cognitive load, and knowledge retention. By integrating these variables within a single experimental framework, we aim to clarify not only whether humor enhances learning, but also how it operates as a pedagogical mechanism in AI agent–based instruction—an issue that has received limited empirical attention in prior research. In doing so, this study extends existing work on instructional humor by demonstrating its functional role in dynamic, AI agent–driven learning interactions rather than in static or scripted instructional contexts.

## Literature review

This section reviews to sort out the research on humor in education research, humor in AI research, question-asking in education research and related theories.

### Humor in education research

Humor has long been recognized as an effective pedagogical strategy for enhancing classroom climate, learner attention, motivation, and emotional engagement (Banas et al., 2011; Savage et al., 2017). In instructional contexts, humor is not inherently beneficial; its effects depend on how learners process and interpret humorous cues during learning.

Instructional Humor Processing Theory (IHPT) provides a framework for understanding these effects. According to IHPT, learners first detect humorous incongruity, then evaluate whether the humor is appropriate, and finally assess whether it is relevant to the instructional content (Wanzer et al., 2010). Importantly, IHPT distinguishes between content-related humor, which directly supports learning by reinforcing instructional material, and content-unrelated humor, which may distract learners and increase extraneous cognitive load. This distinction is central to instructional design, as only humor that is perceived as instructionally relevant is expected to facilitate learning.

Building on this framework, educational research has further categorized humor by direction, including affiliative, self-enhancing, self-defeating, and aggressive humor styles. Affiliative and content-related humor are generally considered most appropriate for educational settings because they promote social connection while supporting learning objectives (Robinson et al., 2024; Mendiburo-Seguel et al., 2015; Erdoğdu & Çakıroğlu, 2021). In contrast, aggressive humor is typically avoided, whereas self-defeating humor has shown limited but context-dependent benefits in pedagogical agent–based environments, such as increasing learner effort (Ceha et al., 2021).

Empirical studies increasingly support the positive role of humor in online learning. Humor embedded in instructional materials has been shown to increase learner interest and conceptual understanding (Miller et al., 2016), enhance behavioral, cognitive, and emotional engagement (Erdoğdu & Çakıroğlu, 2021), and help learners regulate stress and resilience in digital environments (Frisby et al., 2024). More recent work suggests that humor can also reduce cognitive load and affective filtering, thereby facilitating learning under cognitively demanding conditions (Peng, 2025).

Despite these findings, most prior research has examined humor delivered by human instructors or pedagogical agent–based systems, which typically rely on pre-scripted dialogue and limited interactional flexibility. These systems differ fundamentally from an AI agent, which can sustain open-ended instructional dialogue, dynamically respond to learner input, and actively guide learning through question-asking. Although studies on pedagogical agent–based systems have shown that humorous dialogue can enhance motivation, emotional connection, and perceived agent value (Buttussi & Chittaro, 2020; Ceha et al., 2021), they do not address how humor functions when it is generated and deployed dynamically within ongoing instructional interaction.

Accordingly, it remains unclear how instructional humor operates in an AI agent that not only explains content but also actively poses questions and steers the learning process. The present study addresses this gap by examining humor that is explicitly designed to be content-related and pedagogically embedded, consistent with IHPT, within AI agent–mediated instructional dialogue.

## Humor in AI research

Early research in artificial intelligence viewed humor as both a technical challenge and a window into human cognition (Ritchie, 2009). Initial systems such as JAPE demonstrated the feasibility of computational humor generation but were not designed for educational use or empirical evaluation (Binsted, 1996). Similarly, early theoretical discussions proposed potential benefits of humorous pedagogical agent–based instruction for learning, yet lacked supporting data (McKay, 2002). As a result, empirical research on AI-generated humor remained limited for many years.

The development of large language models (LLMs) has substantially changed this landscape. Recent studies show that LLMs can generate humor that is difficult for users to distinguish from human-created humor and, in some cases, comparable in perceived quality (Gorenz & Schwarz, 2024; Joshi, 2025). Although human humor is still often preferred for its contextual precision (Zhang et al., 2024a), these findings indicate that generative AI has reached a functional threshold at which humor can be reliably produced and recognized.

Outside educational contexts, AI-generated humor has been shown to influence user perceptions and behaviors. In customer service settings, humorous responses increase perceived warmth, competence, and tolerance for errors (Xu & Liu, 2022), while different humor styles affect continued use across hedonic and utilitarian contexts (Liu & Xu, 2023). Humor also enhances perceived social presence and entertainment value in chatbot interactions (Xie et al., 2024). These mechanisms—engagement, reduced anxiety, and social presence—are directly relevant to learning, particularly in online environments with limited human supervision.

From a Social Agency Theory (SAT) perspective, humor can function as a salient social cue that makes an AI agent appear more human-like and socially responsive (Martha & Santoso, 2019). When humor is generated dynamically and integrated into instructional dialogue, it becomes part of an AI agent's pedagogical persona rather than an isolated affective feature. However, empirical research has yet to examine how such generative humor operates within instructional interactions or how it interacts with other pedagogical strategies, such as question-asking. The present study addresses this gap by investigating humor as an active instructional mechanism in generative AI-based learning environments.

## Question-asking in education research

Question-asking is a core instructional strategy that supports comprehension, reasoning, and knowledge construction (Kuang et al., 2024). Educational research commonly distinguishes between closed-ended questions, which focus on specific answers and rapid knowledge checks, and open-ended questions, which encourage reflection and deeper cognitive processing (Pate, 2012). Both question-asking types serve pedagogical functions, although they differ in cognitive demands.

In practice, open-ended questions are used less frequently than their pedagogical potential would suggest. Classroom studies show that teachers rely predominantly on closed-ended questions, while students rarely initiate or sustain extended inquiry (Thompson et al., 2025). Moreover, the distinction between open-ended and closed-ended questions is often blurred in implementation, leading to inconsistent instructional effects

(Worley, 2015). These inconsistencies complicate efforts to isolate the instructional effects of question-asking format alone.

An AI agent provides an opportunity to standardize question-asking strategies across learners by delivering questions in a consistent and controlled manner. Prior studies using video-based pedagogical agent–based systems or intelligent tutoring systems have shown that pedagogical agent question-asking can support learning, particularly for learners with lower prior knowledge (Craig et al., 2012; Kuang et al., 2024). However, most of this evidence comes from systems with limited interactivity or pre-scripted dialogue. It therefore remains unclear how question-asking functions in an AI agent that engages learners in open-ended dialogue, dynamically responds to learner input, and actively guides the learning process through the AI agent's questioning.

Question-asking has direct implications for learners' cognitive load, particularly when question-asking requires learners to process complex information and generate responses. Open-ended questions may increase intrinsic cognitive load (ICL) due to the inherent complexity of the task, and may also increase extraneous cognitive load (ECL) when instructional guidance is insufficient or task demands are unclear (Craig et al., 2012). These effects are not uniform and can vary depending on task complexity, learner characteristics, and available instructional support.

Cognitive Load Theory (CLT) provides a framework for interpreting these effects by distinguishing among intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL) (Sweller et al., 2019). Within this framework, ICL is determined by the inherent complexity of the learning material, ECL arises from suboptimal instructional design, and GCL reflects the cognitive effort invested in schema construction. Effective instruction therefore aims to manage ICL, minimize unnecessary ECL, and support GCL—an objective that is particularly relevant when question-asking strategies impose substantial cognitive demands.

Taken together, prior research suggests that question-asking strategies intended to promote deeper processing may also increase learners' cognitive demands, with effects on intrinsic and extraneous cognitive load varying by task and learner characteristics. From a Cognitive Load Theory perspective, the role of humor in such contexts should therefore be understood as a potential means of regulating cognitive demand rather than merely enhancing affect. In the present study, instructional humor is deliberately designed to be content-related and tightly integrated into AI-generated explanations, feedback, and question-asking. Under this design, humor is expected to help learners interpret and organize complex information more efficiently, thereby reducing intrinsic cognitive load without introducing additional extraneous load. By examining humor and question-asking within a unified CLT-based framework, this study clarifies how an affective design element can function as part of the instructional process itself in AI-mediated learning.

## Research questions and hypotheses

Building on prior theoretical frameworks and empirical studies reviewed above, this study examines whether humor and question-asking strategies used by an AI agent can influence learners' affective, cognitive, and performance-related outcomes. We propose the following hypotheses.

Prior studies have demonstrated that large language models such as ChatGPT can generate humor comparable to that created by humans, and that users are able to recognize such humor as intended (Gorenz & Schwarz, 2024; Joshi, 2025).

H1: Humor generated automatically by an AI agent will be perceived as humorous by learners.

According to the IHPT, humor that is consistent with learning content directly promotes learning outcomes (Wanzer et al., 2010). Meanwhile instructional humor has been shown to increase attention and retention by lowering stress and enhancing cognitive elaboration (Banas et al., 2011; Wegener, 2022).

H2: A humorous AI agent will improve learners' knowledge retention.

Humor promotes emotional engagement and enjoyment in learning environments, contributing to higher intrinsic motivation and positive affect (Isen & Reeve, 2005; Buttussi & Chittaro, 2020).

H3: A humorous AI agent will enhance learners' intrinsic motivation and positive emotions.

According to SAT, expressive AI agent that exhibit humor are perceived as more engaging, credible, and human-like (Martha & Santoso, 2019; Xie et al., 2024).

H4: A humorous AI agent will increase learners' perceived agent value.

Empirical evidence indicates that humor can reduce intrinsic cognitive load by making complex learning content more approachable (Peng, 2025).

H5: A humorous AI agent is expected to reduce learners' intrinsic cognitive load.

Open-ended questions have been found to promote reflection and deep processing, which may enhance long-term knowledge retention (Pate, 2012; Kuang et al., 2024).

H6: An AI agent that poses open-ended questions will improve learners' knowledge retention.

Close-ended questions are more structured and may help reduce extraneous cognitive load by minimizing ambiguity and task complexity (Craig et al., 2012).

H7: An AI agent that poses close-ended questions will result in lower extraneous cognitive load.

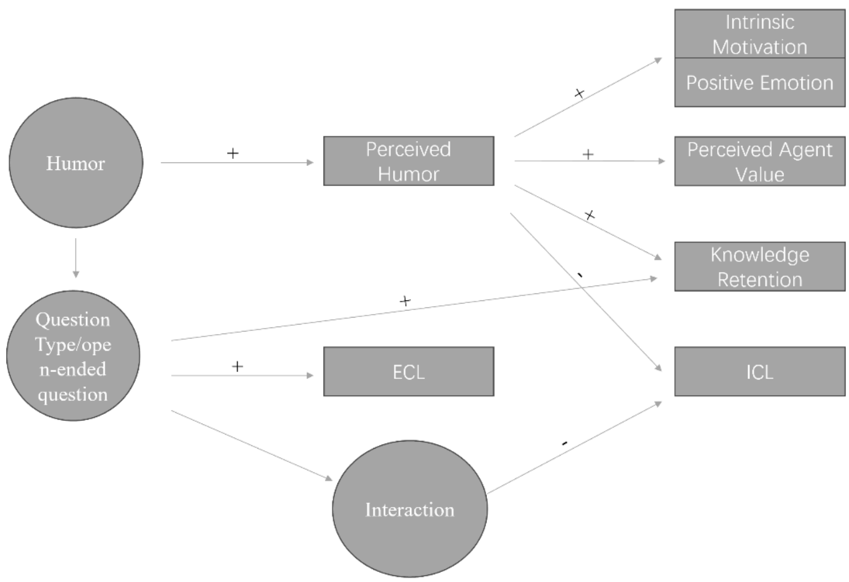
Given that humor may reduce cognitive load and open-ended questions may increase it, we propose the following hypothesis:

H8: Humor will moderate the relationship between open-ended questions and intrinsic cognitive load, such that the cognitive demands associated with open-ended questions may be attenuated when humor is present.

The causal relationship diagram is shown in Figure 1.

**Figure 1**

Causal relationship diagram




## Research method

We conducted a 2 (humor: no-humor vs. humor) × 2 (question-asking: close-ended vs. open-ended) between-subjects factorial experiment, resulting in four experimental conditions: (A) No humor × Close-ended questions; (B) Humor × Close-ended questions; (C) No humor × Open-ended questions; (D) Humor × Open-ended questions. The dependent variables included: perceived humor, perceived agent value, intrinsic motivation, positive emotions, cognitive load, and knowledge retention performance.

### Participant recruitment

We used G*Power to conduct an a priori power analysis. Based on general psychological research conventions, we estimated a medium-to-large effect size ($\eta^2$ = 0.25—0.40), resulting in a required sample size of 52 to 128 participants to ensure sufficient statistical power (0.80). This implies that each group should include 13 to 32 participants.

We recruited participants online, yielding a final sample of 116 Chinese higher education students (72 undergraduates and 44 postgraduates; 52.6% male; *M* = 21.59, *SD* = 2.28). They were enrolled in various non-psychology majors (e.g., accounting, mathematics, computer science) across 39 universities in China, meeting the key inclusion criterion of limited prior exposure to the experimental content (i.e., terminology and concepts related to experimental design flaws). This approach aligned with Kuang et al. (2024), who adopted the same learning materials and recruitment standards. Participants were randomly assigned to four experimental groups: A (n = 29), B (n = 28), C (n = 29), and D (n = 30).

The sample's disciplinary and institutional diversity reflects the broader population of Chinese students in higher education, aligning with the study's focus on online learning within this context.

## Experimental design

We developed an AI agent using the "Virtual Character Generation" module in Doubao, powered by Doubao 1.5-Pro—ByteDance's flagship Chinese-language model. This LLM is widely applied in Chinese educational platforms (e.g., Doubao AIXue) for its strong linguistic reasoning and culturally appropriate dialogue generation, making it suitable for delivering either humorous or neutral instructional content.

Instructional material was adapted from Kuang et al. (2024), which explains 12 common experimental design flaws. The full text was embedded as a system prompt; The AI agent generated all content—including explanations and questions—solely based on internal model capabilities, without access to external databases or retrieval tools, ensuring content consistency.

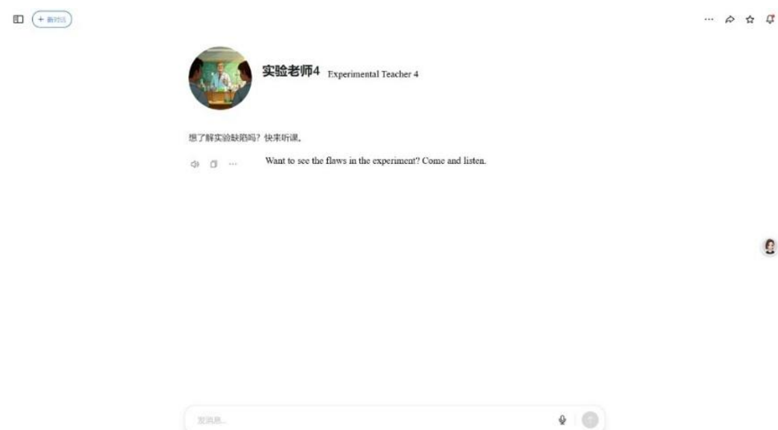
Humor style and question-asking type were controlled through uniform system prompts. Humor was manipulated at the level of instructional style rather than specific pre-authored content: The AI agent was instructed to adopt either a neutral or humorous teaching style, making its instructional discourse livelier and more engaging. Humor was implemented as a personality trait of the AI agent, and only content-relevant humor was used to minimize potential distraction, in line with IHPT. This approach preserves the generative nature of AI-based instruction and avoids reverting to pre-scripted humor commonly used in prior research.

Question-asking was fully agent-controlled across all conditions. After each concept explanation, AI agent posed two questions of the assigned type (close-ended or open-ended) as specified in the system prompt. The number and types of questions were fixed in advance and were not altered based on learners' responses, ensuring consistent manipulation regardless of participant behavior.

To verify instructional fidelity, two psychology graduate students and one educational technology professor pre-evaluated the AI agent on two dimensions: (1) consistency of humor style and question-asking format across groups, and (2) cognitive appropriateness of content for learners without psychology backgrounds. The evaluators reached a consensus through discussion, and no major inconsistencies were reported.

**Figure 2**

Learning interface 1




The following system prompts were used for the AI agent in each group:

Group A:

You are an excellent experimental psychology teacher. Please clearly explain each design flaw and its bracketed explanation, and then pose questions. Maintain a neutral and serious teaching style, avoiding any emotional expression. For each concept, ask two yes/no questions. After receiving answers, provide detailed feedback and neutral encouragement. Teach the 12 flaws in sequence, one at a time.

Group B:

Same as Group A, but with a humorous and entertaining teaching style, humor is closely related to the instructional content, including relevant internet slang and humorous feedback.

Group C:

Same as Group A, but ask two open-ended questions.

Group D:

Combines Group B's humorous style with Group C's open-ended questions and detailed feedback.

The study material text is omitted, which includes full term explanations and definitions.

AI agent was labeled as "Experimental Teacher 1–4," and all interactions took place in text format, as shown in

Figure **2**. See Appendix 1 for each group of AI explanation cases and feedback cases.

The experiment was conducted online via Tencent Meeting to align with the online learning context. Participants were informed of all experimental requirements during recruitment and were instructed to participate in a quiet, private environment. Cameras were required to remain on throughout the session, and the experimenter monitored compliance in real time. Before the experiment, participants received standardized instructions indicating that they should respond to the AI agent's questions following each concept explanation. If they preferred not to answer or wished to proceed, they could indicate "next concept" to continue. Session duration was limited, and the experimenter provided real-time reminders if learning time approached the predefined limit. Participants were instructed to report any interruptions or technical issues immediately and were informed of their right to withdraw at any point; data from withdrawn sessions were deleted and excluded from analysis.

Several design constraints inherent in AI-mediated instructional interaction should be noted. Although question-asking format was predefined at the AI agent level through system prompts, learner participation in answering questions and initiating additional queries remained voluntary. While participants were instructed to respond to questions posed by the AI agent, progression through the instructional material did not depend on response accuracy, and learners could proceed by indicating "next concept." Consequently, some learners actively initiated follow-up inquiries regardless of the assigned question-asking type, whereas others provided minimal or superficial responses. These interaction patterns may have attenuated experiential differences between open- and closed-ended question conditions.

To preserve ecological validity, learners were allowed to freely request clarification or proceed after each question, reflecting naturalistic AI agent-based learning behavior. Skipped questions, brief responses, and additional learner-initiated queries were therefore not restricted. While total learning duration was fixed across conditions, cumulative textual interaction could vary depending on individual engagement.

With respect to humor, although instructional humor was constrained to be content-related through system prompts, variation in humor realization was inherent to generative AI output. The AI agent adhered to platform-level safety policies that prevented the generation of aggressive or inappropriate humor. Response length followed the system's default generation range and was therefore consistent at the response level, although total exposure length could increase when learners initiated additional interaction.

## Experimental procedure

The total duration of the experiment ranged from approximately 30 to 40 minutes. The procedure consisted of three phases:

### *Pre-test phase*

Participants accessed a Wenjuanxing link to complete a demographic questionnaire. This phase lasted approximately 5 minutes.

### *Learning phase*

Based on participant ID, the experimenter provided each participant with a personalized Doubao link. Participants accessed a virtual AI teacher interface (see Figure 3) and engaged in typed conversations with the AI agent through a text-based chat interface within the Doubao app.

The instructional interaction followed a fixed, agent-controlled sequence across all conditions:

a. The AI agent explained a target concept (i.e., one experimental design flaw). Immediately after each explanation, the AI agent posed two questions.

b. Participants were required to respond to the questions via text input. If a response was correct, the AI agent proceeded directly to the explanation of the next concept.

c. If a response was incorrect, the AI agent provided additional explanation and then posed two new questions of the same assigned type before moving on.

All instructional content and interactions were completed within 15 to 25 minutes. Representative examples of the interactions are provided in Appendix 1.

### *Post-test phase*

Upon completion of the lesson, participants exited the Doubao app and accessed the post-test questionnaire via another Wenjuanxing link. This questionnaire measured all dependent variables, which will be detailed in the next section. This phase lasted approximately 10 minutes.

**Figure 3**

Learning interface 2


## Measurement instruments

The pre-test included only a demographic questionnaire, which asked participants about their ID number, group number, age, gender, university, educational level, and major.

All subjective scales are provided in the Appendix.

Perceived Humor was measured using the humor subscale from the Aroused Fear and Humor Questionnaire, adapted by Buttussi and Chittaro (2020) for studies on humor in pedagogical agent–based instruction, originally developed by Lee and Ferguson (2002). Reverse-worded items were excluded to reduce potential response bias among learners. The final scale consisted of five items rated on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree), replacing the original 10-point scale to improve response consistency. A sample item is “I found myself laughing during my studies.” The scale demonstrated good internal consistency (Cronbach’s $\alpha = 0.88$).

Agent Persona Instrument Revised (API-R), a revision of Ryu and Baylor (2005) instrument by Schroeder et al. (2018), was used to evaluate the pedagogical agent’s persona. The instrument included 20 items rated on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree) across four dimensions: Facilitating Learning (10 items, Cronbach’s $\alpha = 0.89$), Credible (5 items, Cronbach’ s $\alpha = 0.80$), Human-like (5 items, Cronbach’s $\alpha = 0.82$), and Engaging (5 items, Cronbach’s $\alpha = 0.83$). A sample item is “AI teachers are smart.” The overall scale showed high reliability (Cronbach’s $\alpha = 0.94$).

The Intrinsic Motivation Questionnaire, developed by Isen and Reeve (2005), has 8 items in total. The scale consisted of eight items rated on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree), consistent with the response format used for other subjective measures. A sample item is “This study aroused my curiosity.” The scale demonstrated excellent internal consistency (Cronbach’s $\alpha = 0.93$).

The Positive and Negative Affect Schedule, developed by Watson et al. (1988), includes 10 items for positive emotions (Cronbach’s $\alpha = 0.92$) and 10 items for negative emotions (Cronbach’s $\alpha = 0.91$). Items were rated on a 7-point scale (1 = extremely slight, 7 = extremely strong) to allow for neutral or absent emotional experiences. Example items include: “Irritable” and “Active.”

Cognitive load was measured using the scale adopted by Kuang et al. (2024), adapted from Leppink et al. (2013), comprising three dimensions: ICL (4 items, Cronbach’s $\alpha = 0.83$), ECL (4 items, Cronbach’s $\alpha = 0.84$), and GCL. To maintain conceptual coherence, one GCL item focusing on general mental effort investment was removed, while four items emphasizing knowledge understanding were retained (Cronbach’s $\alpha = 0.81$). A 10-point Likert scale (1 = not at all true, 10 = completely true) was used, consistent with the original instrument, to preserve sensitivity for cognitive load assessment. A sample item is “The content of learning materials is complex.”

Knowledge Retention Test was adapted from Kuang et al. (2024). It consisted of a single short-answer question worth 36 points: “Based on what you’ve just learned, please list and explain the 12 types of errors in experimental design, and describe how each can be modified or corrected in practice.” Responses were independently scored by two trained psychology postgraduate students using a standardized scoring rubric that specified reference answers and point allocations for each design flaw; the rubric was directly adopted from Kuang et al. (2024). The inter-rater correlation was 0.95, and the weighted kappa coefficient was 0.79.

## Results

The mean and standard deviation for each variable across the four groups are presented in Table 1. The Pearson correlation among the dependent variables is shown in

Table **2**. Visualizations of key variables are shown in Figure 4,

**Figure 5** and Figure 6.

**Table 1**

Mean and standard deviation of each variable

| | Close-ended | | | | Open-ended | | | |
|---|---|---|---|---|---|---|---|---|
| | Not humor (n=29) | | Humor (n=28) | | Not humor (n=29) | | Humor (n=30) | |
| | M | SD | M | SD | M | SD | M | SD |
| Perceived Humor | 4.26 | 0.84 | 5.00 | 0.80 | 4.06 | 1.07 | 4.83 | 0.85 |
| Perceived Agent Value | 4.50 | 0.57 | 4.93 | 0.73 | 4.46 | 0.69 | 4.84 | 0.68 |
| Facilitation learning | 4.82 | 0.52 | 5.10 | 0.72 | 4.87 | 0.71 | 5.02 | 0.68 |
| Credible | 4.88 | 0.66 | 5.21 | 0.74 | 5.00 | 0.64 | 4.99 | 0.69 |
| Human-like | 4.02 | 0.72 | 4.43 | 1.02 | 3.74 | 1.03 | 4.32 | 1.06 |
| Engaging | 4.28 | 0.83 | 5.01 | 0.81 | 4.21 | 1.01 | 5.03 | 0.68 |
| Intrinsic Motivation | 2.98 | 0.45 | 3.38 | 0.51 | 2.83 | 0.56 | 3.21 | 0.47 |
| Positive Emotion | 4.35 | 1.12 | 4.79 | 0.99 | 4.35 | 1.24 | 4.85 | 1.06 |
| Negative Emotion | 1.75 | 0.85 | 1.83 | 1.04 | 1.74 | 0.88 | 1.45 | 0.73 |
| ICL | 4.91 | 2.05 | 4.32 | 1.65 | 5.02 | 1.92 | 4.16 | 2.08 |
| ECL | 3.29 | 2.04 | 2.54 | 1.60 | 2.76 | 1.33 | 2.90 | 1.74 |
| GCL | 8.08 | 1.10 | 8.29 | 1.49 | 8.42 | 1.07 | 8.01 | 1.10 |
| Knowledge Retention | 10.89 | 7.24 | 13.04 | 8.12 | 11.77 | 6.87 | 16.63 | 9.92 |

**Table 2**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Perceived Humor | 1 | .758** | .614** | .535** | .670** | .729** | .965** | .449** | -.125 | -.127 | -.167 | .297** | .010 |
| Perceived Agent Value | .758** | 1 | .805** | .836** | .880** | .870** | .742** | .589** | -.221* | -.037 | -.304** | .406** | .041 |
| Facilitation learning | .614** | .805** | 1 | .774** | .558** | .534** | .612** | .441** | -.270** | .008 | -.306** | .519** | .043 |
| Credible | .535** | .836** | .774** | 1 | .581** | .597** | .527** | .431** | -.258** | -.019 | -.352** | .487** | -.004 |
| Human-like | .670** | .880** | .558** | .581** | 1 | .744** | .673** | .548** | -.068 | -.017 | -.142 | .166 | -.021 |
| Engaging | .729** | .870** | .534** | .597** | .744** | 1 | .683** | .552** | -.211* | -.084 | -.284** | .311** | .119 |
| Intrinsic Motivation | .965** | .742** | .612** | .527** | .673** | .683** | 1 | .430** | -.102 | -.140 | -.175 | .303** | -.034 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Positive Emotion | .449** | .589** | .441** | .431** | .548** | .552** | .430** | 1 | .038 | .097 | .025 | .304** | -.024 |
| Negative Emotion | -.125 | -.221* | -.270** | -.258** | -.068 | -.211* | -.102 | .038 | 1 | .220* | .341** | -.204* | -.177 |
| ICL | -.127 | -.037 | .008 | -.019 | -.017 | -.084 | -.140 | .097 | .220* | 1 | .522** | -.030 | -.177 |
| ECL | -.167 | -.304** | -.306** | -.352** | -.142 | -.284** | -.175 | .025 | .341** | .522** | 1 | -.374** | -.072 |
| GCL | .297** | .406** | .519** | .487** | .166 | .311** | .303** | .304** | -.204* | -.030 | -.374** | 1 | .032 |
| Knowledge Retention | .010 | .041 | .043 | -.004 | -.021 | .119 | -.034 | -.024 | -.177 | -.117 | -.072 | .032 | 1 |

PS: **. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Pearson correlation between variables

Statistical analyses were conducted using SPSS to explore the effects of humor and question-asking on various learning process, and results is shown in

**Table 3.**

**Table 3**

| Independent variable | dependent variable | Type III Sum of Squares | df | MS | F | p | η² |
|---|---|---|---|---|---|---|---|
| Humor | Perceived Humor | 16.654 | 1.000 | 16.654 | 20.669 | 0.000 | 0.156 |
| | Perceived Agent Value | 4.817 | 1.000 | 4.817 | 10.713 | 0.001 | 0.087 |
| | Facilitate learning | 1.298 | 1.000 | 1.298 | 2.975 | 0.087 | 0.026 |
| | Credible | 0.701 | 1.000 | 0.701 | 1.494 | 0.224 | 0.013 |
| | Human-like | 7.101 | 1.000 | 7.101 | 7.561 | 0.007 | 0.063 |
| | Engaging | 17.123 | 1.000 | 17.123 | 24.444 | 0.000 | 0.179 |
| | Intrinsic Motivation | 4.378 | 1.000 | 4.378 | 17.597 | 0.000 | 0.136 |
| | Positive Emotion | 6.391 | 1.000 | 6.391 | 5.207 | 0.024 | 0.044 |
| | Negative Emotion | 0.282 | 1.000 | 0.282 | 0.364 | 0.548 | 0.003 |
| | ICL | 15.080 | 1.000 | 15.080 | 4.018 | 0.047 | 0.035 |
| | ECL | 2.750 | 1.000 | 2.750 | 0.953 | 0.331 | 0.008 |
| | GCL | 0.307 | 1.000 | 0.307 | 0.213 | 0.645 | 0.002 |
| | Knowledge retention | 355.601 | 1.000 | 355.601 | 5.367 | 0.022 | 0.046 |
| Question-asking | Perceived Humor | 0.938 | 1.000 | 0.938 | 1.164 | 0.283 | 0.010 |
| | Perceived Agent Value | 0.147 | 1.000 | 0.147 | 0.326 | 0.569 | 0.003 |
| | Facilitate learning | 0.004 | 1.000 | 0.004 | 0.010 | 0.920 | 0.000 |
| | Credible | 0.077 | 1.000 | 0.077 | 0.165 | 0.686 | 0.001 |
| | Human-like | 1.110 | 1.000 | 1.110 | 1.182 | 0.279 | 0.010 |
| | Engaging | 0.018 | 1.000 | 0.018 | 0.025 | 0.874 | 0.000 |
| | Intrinsic Motivation | 0.732 | 1.000 | 0.732 | 2.942 | 0.089 | 0.026 |
| | Positive Emotion | 0.030 | 1.000 | 0.030 | 0.024 | 0.877 | 0.000 |
| | Negative Emotion | 1.078 | 1.000 | 1.078 | 1.390 | 0.241 | 0.012 |
| | ICL | 0.019 | 1.000 | 0.019 | 0.005 | 0.944 | 0.000 |
| | ECL | 0.210 | 1.000 | 0.210 | 0.073 | 0.788 | 0.001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | GCL | 0.033 | 1.000 | 0.033 | 0.023 | 0.880 | 0.000 |
| | Knowledge retention | 144.684 | 1.000 | 144.684 | 2.184 | 0.142 | 0.019 |
| Humor * Question-asking | Perceived Humor | 0.005 | 1.000 | 0.005 | 0.006 | 0.937 | 0.000 |
| | Perceived Agent Value | 0.017 | 1.000 | 0.017 | 0.039 | 0.844 | 0.000 |
| | Facilitate learning | 0.107 | 1.000 | 0.107 | 0.245 | 0.622 | 0.002 |
| | Credible | 0.826 | 1.000 | 0.826 | 1.761 | 0.187 | 0.015 |
| | Human-like | 0.220 | 1.000 | 0.220 | 0.234 | 0.629 | 0.002 |
| | Engaging | 0.057 | 1.000 | 0.057 | 0.081 | 0.776 | 0.001 |
| | Intrinsic Motivation | 0.002 | 1.000 | 0.002 | 0.008 | 0.931 | 0.000 |
| | Positive Emotion | 0.030 | 1.000 | 0.030 | 0.024 | 0.877 | 0.000 |
| | Negative Emotion | 1.002 | 1.000 | 1.002 | 1.293 | 0.258 | 0.011 |
| | ICL | 0.549 | 1.000 | 0.549 | 0.146 | 0.703 | 0.001 |
| | ECL | 5.853 | 1.000 | 5.853 | 2.030 | 0.157 | 0.018 |
| | GCL | 2.805 | 1.000 | 2.805 | 1.948 | 0.166 | 0.017 |
| | Knowledge retention | 53.212 | 1.000 | 53.212 | 0.803 | 0.372 | 0.007 |

Results of two-way ANOVA

Overall, the results indicate that humor exerted a consistent and meaningful influence across multiple dimensions of learning, whereas question-asking type did not yield distinguishable effects under the present conditions. The manipulation check confirmed that learners reliably perceived the humorous intervention, with a substantial effect size, indicating that humor was salient and recognizable within the AI-mediated instructional interaction.

Beyond perceptual recognition, humor produced moderate to large effects on learners' social and affective experiences. Notably, humor substantially enhanced perceived agent value, particularly in terms of engagement and human-likeness, suggesting that humor functioned as a strong social cue that shaped how learners evaluated the AI agent. Humor also showed a moderate effect on intrinsic motivation and a smaller but reliable effect on positive emotion, indicating that its primary affective contribution lay in sustaining learners' motivational involvement rather than merely inducing momentary enjoyment.

With respect to cognitive processing, humor demonstrated a small but meaningful reduction in intrinsic cognitive load. Although the effect size was modest, this finding suggests that humor can improve cognitive efficiency without introducing additional cognitive burden. Importantly, humor was also associated with a small-to-moderate improvement in knowledge retention, indicating that its benefits extended beyond subjective experience to measurable learning outcomes.

In contrast, neither question-asking type nor its interaction with humor produced significant effects, and the corresponding effect sizes were negligible. Taken together, these results suggest that, in AI agent-mediated instructional contexts, humor represents a robust and impactful design element.

**Figure 4**

Mean perceived humor scores across condition


**Figure 5**

Mean cognitive load across conditions

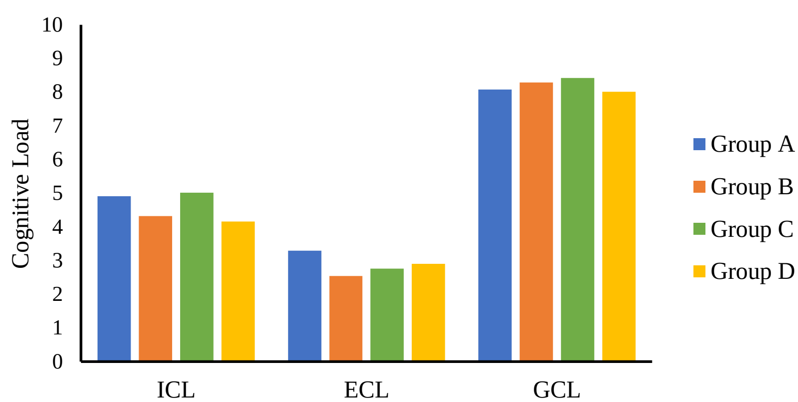
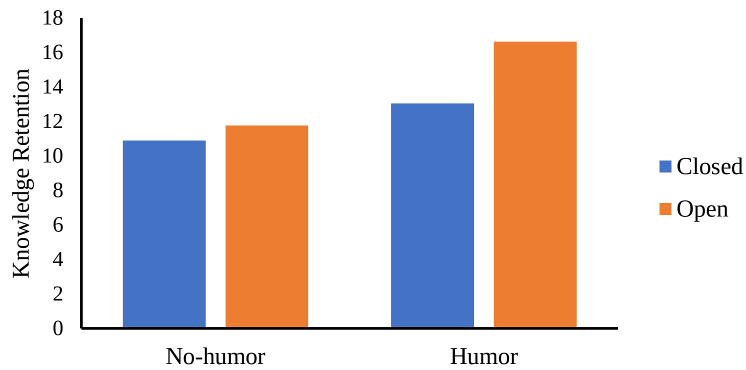



**Figure 6**

Mean knowledge retention scores across conditions




## Discussion

### Humorous AI agent and learning

The results supported H1, indicating that learners were able to reliably perceive humor generated by an AI agent. This finding extends prior work on humor recognition in AI agent (Gorenz & Schwarz, 2024; Joshi, 2025) by demonstrating that pedagogically embedded, content-related humor is also detectable in instructional contexts. This perceptual recognition provides a necessary foundation for examining humor's functional effects on learning.

Support for H2 showed that learners interacting with a humorous AI agent achieved higher knowledge retention. While previous studies on pedagogical agent–based systems have primarily reported affective or cognitive benefits—such as improved mood, engagement, or reduced mental effort (Buttussi & Chittaro, 2020; Ceha et al., 2021) — direct evidence for learning gains has been limited. The present findings address this gap by demonstrating that well-integrated humor can translate into measurable learning outcomes. Consistent with IHPT, humor that is closely aligned with instructional content appears to facilitate more efficient processing and retention of information.

H3 was supported, with a humorous AI agent eliciting higher intrinsic motivation and positive emotion. These findings align with prior research showing that instructional humor enhances learners' emotional engagement in digital learning contexts (Buttussi & Chittaro, 2020; Erdoğdu & Çakıroğlu, 2021; Frisby et al., 2024). Importantly, these affective gains provide a plausible account for why humor was associated with improved retention and reduced ICL: learners who feel more interested and positively engaged are more likely to sustain attention, invest effort, and persist through cognitively demanding material.

As an explicitly speculative account, prior work suggests that motivational states may temporarily support working-memory functioning and cognitive resources during learning (Schnotz et al., 2009; Mutlu-Bayraktar et al., 2019). Neurocognitive evidence further indicates that humor appreciation can engage reward- and affect-related regions implicated in motivation and emotional regulation (Chan et al., 2018). Although we did not directly measure these mechanisms, these converging accounts provide a theoretically coherent explanation for how humor-induced positive affect may translate into more efficient processing and stronger retention in AI agent-mediated instruction.

H4 further indicated that a humorous AI agent was perceived as more engaging, credible, and human-like. In line with SAT, humor appears to function as a salient social cue that enhances learners' perception of the AI agent as a socially meaningful instructional partner (Martha & Santoso, 2019). By using a multidimensional pedagogical agent value measure (Schroeder et al., 2018), this study extends prior work by showing that humor strengthens not only general impressions of warmth or likability, but also perceptions related to learning facilitation and credibility—factors that may indirectly support learning outcomes.

Finally, support for H5 demonstrated that a humorous AI agent reduced learners' ICL without increasing extraneous load or diminishing germane load. Unlike studies that treat cognitive load as a global construct (Peng, 2025), the present findings clarify that instructional humor selectively influences intrinsic load, consistent with CLT. This result addresses concerns that humor might increase cognitive demands when poorly integrated (Hu et al., 2017), and suggests that content-related humor, when pedagogically embedded, can enhance cognitive efficiency.

Taken together, the primary contribution of this study lies in demonstrating how instructional humor functions as an integrated pedagogical mechanism in AI agent-mediated instruction. By combining IHPT, SAT, and CLT within a single empirical framework, this study shows that content-related humor generated by an AI agent simultaneously supports affective engagement (motivation and positive emotion), enhances social perception of the AI agent (perceived agent value), and improves cognitive efficiency by reducing intrinsic cognitive load. Unlike prior research that has typically examined these mechanisms in isolation or within teacher-led or pedagogical agent–based contexts, the present findings provide converging evidence that humor in generative AI agent operates across affective, social, and cognitive dimensions to support learning.

## AI agent's question-asking and learning

H6 and H7 were not supported, as question-asking type did not produce significant effects on learning outcomes or cognitive load. Given the small effect sizes, these findings suggest that, under the present experimental conditions, question-asking format alone was insufficient to meaningfully differentiate learners' cognitive or performance outcomes.

This pattern is consistent with prior research reporting mixed and context-dependent effects of question-asking strategies (Pate, 2012; Sukor & Abdullah, 2022). Although open-ended questions are often associated with deeper processing, they may also increase task demands without reliably improving outcomes, particularly when adaptive scaffolding is limited (Craig et al., 2012; Kuang et al., 2024). Meta-analytic evidence further indicates that higher-level question-asking tends to yield small or inconsistent effects on achievement, suggesting that the null findings observed here are not anomalous (Samson et al., 1987).

Several design-related factors may have constrained the effectiveness of question-asking in this study. First, learner prior knowledge was not modeled as a moderating variable, despite evidence that effective question-asking depends on alignment with learners' cognitive states (Zhao et al., 2023). Second, to maintain experimental control, the AI agent posed questions in a fixed sequence following each explanation, limiting adaptiveness. Third, the AI agent was designed to be functionally consistent rather than persona-differentiated, which may have further reduced their capacity to support cognitively demanding question-asking strategies (Schroeder & Adesope, 2014).

More fundamentally, the role of question-asking may differ in AI agent-mediated instruction compared to traditional settings. Interaction with AI agent inherently involves an ongoing process of inquiry and response, in which learners actively seek explanations, clarification, and feedback. Within such an interactive dialogue, simple distinctions between open-ended and closed-ended questions may be less salient, as learners are already engaged in continuous cognitive elaboration through the interaction itself. This may attenuate the instructional impact of basic question-asking format manipulations, particularly when those manipulations are relatively coarse.

Additional research suggests that question-asking complexity poses an inherent tension for novice learners: overly complex questions may discourage participation, whereas overly simple questions may fail to stimulate elaborative processing (Tofade et al., 2013). In AI agent-mediated environments, where the AI agent often provides immediate feedback or correct answers regardless of response accuracy, this tension may be amplified. Although immediacy supports clarity, it may also reduce learners' incentive to engage deeply with subsequent questions, especially following incorrect responses (Degen, 2025; Tofade et al., 2013).

Taken together, the present findings suggest that question-asking format, when implemented in a generic and non-adaptive manner, may exert limited influence on learning outcomes in AI agent-mediated instruction. Importantly, this conclusion should not be interpreted as evidence against the pedagogical value of question-asking more broadly. Rather, it highlights the need to consider learner characteristics, adaptive scaffolding, and the inherently interactive nature of AI agent dialogue when designing question-asking strategies for online learning environments.

### Humorous AI agent and question-asking

H8 was also not supported. No significant interaction effects were found between the two independent variables. Given the non-significant interaction and small effect sizes, these results should be interpreted cautiously rather than as evidence that the two strategies cannot jointly influence learning.

From a theoretical perspective, humor and question-asking have been proposed to operate through partially distinct instructional mechanisms. Instructional humor is often discussed in relation to affective–motivational processes, such as attention, enjoyment, and emotional regulation (Wanzer et al., 2010; Craig et al., 2012), whereas question-asking is typically linked to cognitive processing and knowledge construction, which may be explained within the framework of CLT. However, the current data do not provide empirical support for an interaction between these mechanisms. Accordingly, this functional distinction should be understood as a theoretical framing rather than a direct explanation of the present findings.

Prior work has suggested that more complex or open-ended questions may increase intrinsic cognitive load, particularly for novice learners (Kuang et al., 2024). By contrast, the present study found that humor generated by an AI agent significantly reduced learners' ICL. Taken together at a conceptual level, these patterns imply that humor and question-asking may address different demands within the learning process, with humor primarily associated with cognitive regulation and motivational support, and question-asking more closely linked to cognitive elaboration demands. Although the current study did not demonstrate a statistical interaction, this conceptual alignment between theoretical accounts and the observed main effects helps clarify the conditions under which these strategies may be theoretically complementary, while emphasizing that such complementarities were not empirically verified in the present study.

## Limitations and future directions

First, although the system prompt ensured that the AI agent consistently adhered to the assigned humor style and question-asking format, it was not possible to completely isolate the instructional manipulation from additional learner-driven interactions. When learners posed follow-up questions or initiated additional dialogue beyond the designed instructional flow, these learner-initiated interactions effectively introduced an additional instructional variable that was not directly manipulated in this study. Rather than reflecting a failure of control by the AI agent, this phenomenon highlights an inherent feature of AI agent-mediated learning: the system affords opportunities for spontaneous dialogue that may meaningfully shape cognitive and affective processes. At present, it remains unclear whether such spontaneous extensions of interaction enhance or dilute the pedagogical impact of question-asking strategies. Future research could therefore systematically compare different AI agent–based question-asking paradigms, particularly Socratic question-asking (Zhang et al., 2024a) and student question-asking (Aflalo, 2018), to examine how these forms of dialogic engagement influence cognitive load, motivation, and learning outcomes.

Second, although the study focused primarily on main effects, prior research suggests that the impact of both humor and question-asking may be moderated by learner characteristics such as prior knowledge, baseline motivation, cognitive preferences, and learning strategies. The present design did not explicitly model these moderating factors, which may partly explain the variability and limited effect sizes observed for question-asking strategies. Future studies should incorporate measures of prior knowledge and learner profiles, and potentially adopt adaptive AI instructional mechanisms, such as Dialogue-based Intelligent Tutoring Systems powered by an AI model (Zhang et al., 2024b), to better determine for whom and under what conditions humor and question-asking strategies are most pedagogically effective.

Thirdly, this study focused on a Chinese higher education population, which means the findings were shaped by a cultural context that may view humor in education differently from Western settings. Prior research has shown that Eastern and Western cultures hold distinct attitudes toward humor, with some evidence suggesting that Eastern learners may perceive humor as less appropriate in formal learning environments (Yue et al., 2016). Such cultural norms may influence how students respond to a humorous AI agent and how willing they are to engage in interactive question-asking. Therefore, the generalizability of the present results should be treated with caution. Future research could conduct cross-cultural comparisons to examine whether humor generated by an AI agent functions similarly across cultural contexts, and could also extend to younger learners such as K–12 students, who may be less constrained by traditional cultural expectations. It is also important to recognize that cultural attitudes toward humor continue to evolve, meaning that the educational impact of humor generated by an AI agent may change over time.

In addition to the primary limitations discussed above, several methodological and analytic considerations warrant brief mention. First, knowledge transfer and delayed retention were not assessed due to practical constraints related to post-test length, which may have limited insight into longer-term learning effects. Second, although instructional fidelity was evaluated by domain experts through consensus discussion, independent inter-rater reliability indices (e.g., Cohen’s kappa) were not calculated. Third, the study relied on traditional experimental designs and group-level statistical analyses, which were appropriate for examining primary effects with a moderate sample size but did not capture more complex mediating or moderating relationships among psychological variables, nor underlying biological mechanisms. These considerations suggest promising directions for future research but do not undermine the validity of the present findings regarding primary effects.

## Conclusion

This study examined how humor and question-asking strategies employed by a generative AI agent influence learning outcomes. The findings demonstrate that learners reliably recognized content-related humor generated by an AI agent, and that such humor produced consistent benefits across affective, social, and cognitive dimensions. Specifically, a humorous AI agent enhanced intrinsic motivation, positive emotion, and perceived agent value, while simultaneously improving knowledge retention and reducing ICL. Together, these results show that content-related humor in AI agent-mediated instruction supports affective engagement, social perception, and cognitive efficiency, thereby extending prior instructional humor and social agency accounts to a generative AI agent.

Importantly, these findings carry clear implications for the design of AI agent–based instructional systems. Rather than treating humor as a peripheral or decorative feature, designers of the AI agent should consider humor as an integral component of instructional interaction—one that can simultaneously support motivation, social presence, and cognitive processing when it is content-related and pedagogically embedded. Incorporating well-aligned instructional humor offers a scalable and low-cost design strategy for improving learning efficiency in online environments, particularly in contexts where sustained engagement and cognitive load management are critical.

By contrast, the absence of significant effects for question-asking type and its interaction with humor highlights boundary conditions for question-asking–based instructional design in AI agent-mediated learning. In dialogic AI agent–mediated environments, where learners already engage in ongoing inquiry and feedback exchange, simple distinctions between open-ended and closed-ended questions may be insufficient to shape learning outcomes. Effective question-asking in an AI agent may therefore require more adaptive designs, such as sensitivity to learners' prior knowledge, dynamically adjusted question-asking strategies, or culturally responsive interaction patterns.

Overall, this study provides actionable guidance for the design of AI agents by identifying humor as a robust and empirically supported design lever for enhancing learning, while cautioning against overly simplistic implementations of question-asking strategies. Future research should extend this work by exploring adaptive humor and question-asking mechanisms across diverse learner populations and instructional contexts, further refining theory-driven and evidence-based instructional design for AI agents.

For AI agent design, the findings suggest that: (a) instructional humor should be content-aligned and embedded within explanations and feedback; (b) humor can be leveraged to improve learning efficiency without increasing cognitive load; and (c) question-asking strategies should emphasize adaptivity to learners over simple format distinctions.

## Appendix 1: AI agent generated snippets in teaching

Group A:

In the process of sample selection, the sample should be able to represent the population being tested. The selection of the sample should match the characteristics of the population. "Inappropriate Sample Selection" means that the sample selected in the research is far from the characteristics of the population, resulting in a deviation in the experimental results. To address this issue, relevant literature should be reviewed, the research question should be re - examined, the appropriate sample objects should be determined, and the sample data should be collected again.

Should the sample match the characteristics of the population?

If the sample is far from the characteristics of the population, will it cause a deviation in the experimental results?

Group B:

Alright, we're now onto “No Random Assignment”. Random assignment means that each subject has an equal chance or probability of being assigned to any group. They could end up in the experimental group or the control group with the same likelihood. This helps to cancel out the individual differences among subjects. For example, if you're doing an experiment on the effect of a new drug, you don't want all the healthy people in one group and all the sicker people in another by chance. If there's no random assignment, those individual differences might mess up your results. To fix this, you need to re - assign the subjects randomly, making sure everyone has an equal shot at being in each group.

First, if you're dividing people into two groups for an exercise - effectiveness study and you put all the young people in one group and all the old people in the other, is this a case of no random assignment? Second, if you randomly pick names out of a hat to assign people to groups, but you do it in a way that some people have a higher chance of getting picked, is this proper random assignment?

Group C:

Now let's move on to "Small Sample Size". The number of research subjects is called the sample size. The appropriate size of the sample is determined based on the size of the population. If one tries to generalize the situation of a large - scale population based on just a few samples, it is a case of small sample size error. To address this, more samples should be collected to ensure an adequate number of samples. Generally, in each experimental condition, the number of subjects in each group should be no less than 30.

Can you give an example of a field where small sample size might lead to inaccurate conclusions?

Why is the number 30 often considering a minimum in each experimental condition?

Group D:

Alright, now we're onto "Experimenter Bias". This is when the experimenter, that's the person running the experiment, might accidentally or on - purpose influence the subjects in some way. It could be through their body language, facial expressions, or the way they talk. They might make the subjects respond in a way that fits the experimenter's expectations. It's like the experimenter is secretly pulling the strings! For example, if an experimenter really wants a certain result in a taste - testing experiment, they might unconsciously give hints to the subjects.

To fix this, we can either pick someone outside the research team to be the experimenter. That way, we can be surer that any changes in the participants' behavior are because of the actual experiment, not the experimenter's influence. Or, the experimenter can be super careful to control their expressions and actions and treat all subjects the same.

First, in an experiment where you're testing if a new type of music makes people more creative, how could an experimenter accidentally show experimenter bias? Second, if you're the experimenter in this situation, what are some specific things you can do to control your expressions and actions?

## Appendix 2: The humor arousal dimension of the aroused fear and humor questionnaire

The following is your evaluation of the learning process just now. Please choose the option that best suits you according to your actual situation.

[1] I think the learning process I just had was very interesting.

[2] I found myself laughing during my studies.

[3] One of the reasons why I like this study is that it is very interesting.

[4] I really enjoyed the humor during this study.

[5] During this study, I felt very good.

## Appendix 3: The AI teacher persona instrument revised

The following is an evaluation of the credibility of AI teacher. Please choose the option that best suits you according to your actual situation (0 meaning not at all the case and 10 meaning completely the case). (Facilitate learning)

[1] The AI teacher made me think more deeply about the learning content.

[2] AI teachers make teaching fun.

[3] The AI teacher encouraged me to reflect on what I had learned.

[4] The AI teacher caught my attention.

[5] The AI teacher explained the learning content clearly.

[6] The AI teacher helped me focus on the learning content.

[7] The AI teacher made me focus on the information in the voice that was relevant to the learning content.

[8] The AI teacher helped me learn knowledge.

[9] AI teacher is very skilled at teaching.

[10] Learning from an AI teacher is easy.

The following is an evaluation of the credibility of AI teacher. Please choose the option that best suits you based on your actual situation. (Credible)

[1] AI teacher is knowledgeable.

[2] AI teacher is smart.

[3] AI teacher is necessary.

[4] The AI teacher is a teacher who can assist me.

[5] AI teacher are excellent teachers.

The following is an evaluation of AI teacher class humanity. Please choose the option that best suits you according to your actual situation. (Human-like)

[1] AI teacher is very personalized.

[2] The AI teacher's emotions are natural.

[3] AI teacher is like real people.

[4] The AI teacher expressed emotion.

[5] The AI teacher's tone is very natural.

The following is an evaluation of the attractiveness of AI teacher. Please choose the option that best suits you based on your actual situation. (Engaging)

[1] AI teacher is very attractive.

[2] The AI teacher is very enthusiastic.

[3] AI teacher is very interesting.

[4] AI teacher motivates me in my studies.

[5] AI teacher is easy to communicate with.

## Appendix 4: The internal motivation questionnaire

Here are some comments on this learning experience. Please choose the option that best suits you according to your actual situation.

[1] This study aroused my curiosity.

[2] This study was fascinating.

[3] This study was a lot of fun.

[4] I want to continue studying it.

[5] This study made me curious about the relevant content.

[6] This study was enjoyable.

[7] This study made me want to explore the relevant content further.

[8] I would also like to participate in future research that involves similar learning activities.

## Appendix 5: The cognitive load questionnaire

All the questions below refer to the learning activities you have just completed. Please select the option that matches you. Where 1 represents not at all true and 10 represents completely true.

[1] This study really increased my understanding of the content contained in the study materials. (GCL)

[2] The content of learning materials is complex. (ICL)

[3] The explanations and instructions in the study materials are very unclear. (ECL)

[4] This study really increased my understanding of the issues involved in the study materials. (GCL)

[5] The issues involved in the learning materials are very complex. (ICL)

[6] The explanations and instructions in the study materials are full of unclear language. (ECL)

[7] This study has increased my understanding of the concepts mentioned in the study materials. (GCL)

[8] The concepts mentioned in the study materials are very complex. (ICL)

[9] From a learning perspective, the interpretation and explanation of learning materials are highly ineffective.

[10] This study really enhanced my knowledge and understanding of the relevant content. (GCL)

[11] I put a lot of mental effort into learning the complex content of the material. (ICL)

[12] I put a lot of mental effort into the ineffective interpretation and explanation of learning materials. (ECL)

**Abbreviations**

AI: Artificial intelligence; LLMs: Large language models; JAPE: Joke Analysis and Production Engine; IHPT: Instructional Humor Processing Theory; SAT: Social Agency Theory; CLT: Cognitive Load Theory; ICL: Intrinsic cognitive load; ECL: Extraneous cognitive load; GCL: Germane cognitive load

**Acknowledgements**

We would like to thank Dr. ziyi kuang for providing learning materials and knowledge quizzes.

**Author's contributions**

Xinjie Xie: Conceptualization, Data curation, Software, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review and editing.

Chuwen Liu: Investigation, Resources.

Xiangen Hu: Funding acquisition, Project administration, Supervision, Writing – review and editing.

**Author's information**

Xinjie Xie: Graduated from the School of Psychology of Central China Normal University and currently holds a master's degree. His main research areas include Instructional humor, research on the application and cognition of generative AI in higher education, etc.

Chuwen Liu: Graduated from UCL and currently holds a master's degree.

Xiangen Hu: Professor Hu Xiang'en is a Chair Professor of Learning Sciences and Technologies at the Hong Kong Polytechnic University and the Dean of the Institute for Higher Education Research and Development. His main research areas include artificial intelligence, cognitive psychology, and computerized tutoring systems.

**Funding**

This research received no external funding.

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

## Declarations

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

[1] School of Psychology, Central China Normal University, Wuhan, China, xinjiexie@foxmail.com

[2] Childhood Nutrition Research Group, Population, Policy & Practice Department, UCL Great Ormond Street Institute of Child Health, London, UK, chuwen.liu.23@ucl.ac.uk

[3] Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, PR China, xiangen.hu@polyu.edu.hk

Received: 1 May 2025 Accepted: 29 January 2026

Published online: 1 January 2027 (Online First: 15 June 2026)

**References**

Aflalo, E. (2018). Students generating questions as a way of learning. *Active Learning in Higher Education*, 22(1), 63–75. https://doi.org/10.1177/1469787418769120

Banas, J. A., Dunbar, N., Rodriguez, D., & Liu, S. J. (2011). A review of humor in educational settings: Four decades of research. *Communication Education*, 60(1), 115–144. https://doi.org/10.1080/03634523.2010.496867

Binsted, K. (1996). *Machine humour: An implemented model of puns* (Doctoral dissertation, The University of Edinburgh). Edinburgh Research Archive. http://hdl.handle.net/1842/586

Buttussi, F., & Chittaro, L. (2020). Humor and fear appeals in animated pedagogical agents: An evaluation in aviation safety education. *IEEE Transactions on Learning Technologies*, 13(1), 63–76. https://doi.org/10.1109/tlt.2019.2902401

Ceha, J., Lee, K. J., Nilsen, E., Goh, J., & Law, E. (2021, May). Can a humorous conversational agent enhance learning experience and outcomes? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Article 685, pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445068

Chan, Y. C., Hsu, W. C., Liao, Y. J., Chen, H. C., Tu, C. H., & Wu, C. L. (2018). Appreciation of different styles of humor: An fMRI study. *Sci Rep*, 8(1), 15649. https://doi.org/10.1038/s41598-018-33715-1

Craig, S. D., Gholson, B., Brittingham, J. K., Williams, J. L., & Shubeck, K. T. (2012). Promoting vicarious learning of physics using deep questions with explanations. *Computers & Education*, 58(4), 1042–1048. https://doi.org/10.1016/j.compedu.2011.11.018

Degen, B. (2025). Resurrecting Socrates in the age of AI: A study protocol for evaluating a Socratic tutor to support research question development in higher education (Preprint). *arXiv*. https://arxiv.org/abs/2504.06294

Erdoğdu, F., & Çakıroğlu, Ü. (2021). The educational power of humor on student engagement in online learning environments. *Research and Practice in Technology Enhanced Learning*, 16(1). https://doi.org/10.1186/s41039-021-00158-8

Frisby, B. N., Sidelinger, R. J., Tatum, N. T., & McMullen, A. L. (2024). When synchronous online learning goes wrong: Using humor to cope and to build resilience. *Communication Research Reports*, 41(2), 94–104. https://doi.org/10.1080/08824096.2024.2337378

Gorenz, D., & Schwarz, N. (2024). How funny is ChatGPT? A comparison of human- and A.I.-produced jokes. *PLoS ONE*, 19(7), e0305364. https://doi.org/10.1371/journal.pone.0305364

Hu, D. L., Lefton, L., & Ludovice, P. J. (2017). Humour applied to STEM education. *Systems Research and Behavioral Science*, 34(3), 216–226. https://doi.org/10.1002/sres.2406

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion*, 29, 295–323. https://doi.org/10.1007/s11031-006-9019-8

Joshi, N. N. (2025, January). Evaluating human perception and bias in AI-generated humor. In *Proceedings of the 1st Workshop on Computational Humor (CHum)* (pp. 79–87). Association for Computational Linguistics. https://aclanthology.org/2025.chum-1.9.pdf

Kuang, Z., Jiang, X., & Zhang, Y. (2024). The effect of different question types on vicarious learning. *Educational Psychology*, 44(2), 123–140. https://doi.org/10.1080/01443410.2024.2325589

Lee, M. J., & Ferguson, M. A. (2002). Effects of anti-tobacco advertisements based on risk-taking tendencies: Realistic fear vs. vulgar humor. *Journalism & Mass Communication Quarterly*, 79(4), 945–963. https://doi.org/10.1177/107769900207900411

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Li, W., Wang, F., Mayer, R. E., & Liu, H. (2019). Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology*, 111(8), 1382–1395. https://doi.org/10.1037/edu0000352

Liu, J., & Xu, X. (2023). Humor type and service context shape AI service recovery. *Annals of Tourism Research*, 103, 103668. https://doi.org/10.1016/j.annals.2023.103668

Martha, A. S. D., & Santoso, H. B. (2019). The design and impact of the pedagogical agent: A systematic literature review. *Journal of Educators Online*, 16(1), Article 8. https://doi.org/10.9743/jeo.2019.16.1.8

McKay, J. (2002). Generation of idiom-based witticisms to aid second language learning. In *The April Fools' Day Workshop on Computational Humor* (pp. 77–87). University of Twente.

Mendiburo-Seguel, A., Paez, D., & Martinez-Sanchez, F. (2015). Humor styles and personality: A meta-analysis of the relation between humor styles and the big five personality traits. *Scandinavian Journal of Psychology*, 56(3), 335–340. https://doi.org/10.1111/sjop.12209

Miller, J. L., Wilson, K., Miller, J., & Enomoto, K. (2016). Humorous materials to enhance active learning. *Higher Education Research & Development*, 36(4), 791–806. https://doi.org/10.1080/07294360.2016.1238883

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, 141. https://doi.org/10.1016/j.compedu.2019.103618

Pate, R. S. (2012). Open versus closed questions: What constitutes a good question? In Educational research and innovations (pp. 29–40). Consortium for Educational Development, Evaluation and Research.

Peng, C. (2025). The power of humor: Its impact on cognitive load and affective filtering in EFL learning. *International Educational Research*, 8(2), 15. https://doi.org/10.30560/ier.v8n2p15

Ritchie, G. (2009). Can computers create humor? *AI Magazine*, 30(3), 71. https://doi.org/10.1609/aimag.v30i3.2251

Ryu, J. E. E. H. E. O. N., & Baylor, A. L. (2005). The psychometric structure of pedagogical agent persona. *Technology Instruction Cognition and Learning*, 2(4), 291

Robinson, H. L., Rose, S. E., Elliott, J. M., & Vivaldi, R. A. (2024). Teachers' humour use in the classroom: A scoping review. *Educational Psychology Review*, 36(73). https://doi.org/10.1007/s10648-024-09913-x

Samson, G. K., Strykowski, B., Weinstein, T., & Walberg, H. J. (1987). The effects of teacher questioning levels on student achievement: A quantitative synthesis. *The Journal of Educational Research*, 80(5), 290–295. https://doi.org/10.1080/00220671.1987.10885769

Savage, B. M., Lujan, H. L., Thipparthi, R. R., & DiCarlo, S. E. (2017). Humor, laughter, learning, and health! A brief review. *Advances in Physiology Education*, 41(3), 341–347. https://doi.org/10.1152/advan.00030.2017

Schnotz, W., Fries, S., & Horz, H. (2009). Motivational aspects of cognitive load theory. In M. Wosnitza, S. A. Karabenick, A. Efklides, & P. Nenniger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 69–96). Hogrefe & Huber Publishers.

Schroeder, N. L., & Adesope, O. O. (2014). A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education*, 46(3), 229–251. https://doi.org/10.1080/15391523.2014.888265

Schroeder, N. L., Yang, F., Banerjee, T., Romine, W. L., & Craig, S. D. (2018). The influence of learners' perceptions of virtual humans on learning transfer. *Computers & Education*, 126, 170–182. https://doi.org/10.1016/j.compedu.2018.07.005

Sukor, N. A. M., & Abdullah, N. A. (2022). Developing students' mathematical critical thinking skills using open-ended questions and activities based on student learning preferences. *Education Research International*, 2022, Article 3300363. https://doi.org/10.1155/2022/3300363

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Thompson, P., Paatsch, L., & Nolan, A. (2025). Understanding the talk patterns in secondary school classrooms. *Language and Education*, 39(2), 489–511. https://doi.org/10.1080/09500782.2024.2343292

Tofade, T., Elsner, J., & Haines, S. T. (2013). Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education*, 77(7), 155. https://doi.org/10.5688/ajpe777155

Wanzer, M. B., Frymier, A. B., & Irwin, J. (2010). An explanation of the relationship between instructor humor and student learning: Instructional humor processing theory. *Communication Education*, 59(1), 1–18. https://doi.org/10.1080/03634520903367238

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063. https://doi.org/10.1037/0022-3514.54.6.1063

Wegener, D. R. (2022). Information literacy: Making asynchronous learning more effective with best practices that include humor. *The Journal of Academic Librarianship*, 48, Article 102482. https://doi.org/10.1016/j.acalib.2021.102482

Worley, P. (2015). Open thinking, closed questioning: Two kinds of open and closed questions. *Journal of Philosophy in Schools*, 2(2), 17–29. https://doi.org/10.21913/jps.v2i2.1269

Xie, Y., Liang, C., Zhou, P., & Jiang, L. (2024). Exploring the influence mechanism of chatbot-expressed humor on service satisfaction in online customer service. *Journal of Retailing and Consumer Services*, 76, 103599. https://doi.org/10.1016/j.jretconser.2023.103599

Xu, X., & Liu, J. (2022). Artificial intelligence humor in service recovery. *Annals of Tourism Research*, 95, 103439. https://doi.org/10.1016/j.annals.2022.103439

Yue, X., Jiang, F., Lu, S., & Hiranandani, N. (2016). To be or not to be humorous? Cross-cultural perspectives on humor. *Frontiers in Psychology*, 7, 1495. https://doi.org/10.3389/fpsyg.2016.01495

Zhang, J., Jain, L., Guo, Y., Chen, J., Zhou, K. L., Suresh, S., Wagenmaker, A., Sievert, S., Rogers, T., Jamieson, K., Mankoff, R., & Nowak, R. (2024a). Humor in AI: Massive scale crowd-sourced preferences and benchmarks for cartoon captioning. *Advances in Neural Information Processing Systems*, 37, 125264–125286.

Zhang, L., Lin, J., Kuang, Z., Xu, S., & Hu, X. (2024b). SPL: A Socratic playground for learning powered by large language model (Preprint). *arXiv*. https://doi.org/10.48550/arXiv.2406.13919

Zhao, Y., Huang, T., Wang, H., & Geng, J. (2023). Personalized teaching questioning strategies study based on learners' cognitive structure diagnosis. *Behavioral Sciences*, 13(8), 660. https://doi.org/10.3390/bs13080660

## Publisher's Note