



Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu



Using ChatGPT to bring non-player characters to life: Effects on students' storyline-driven game-based writing learning

Ya Xiao ^{a,b}, Danling Li ^c, Kai Guo ^{d,*}

^a School of Education Science, Nanjing Normal University, Nanjing, China

^b Adolescent Education and Intelligence Support Lab of Nanjing Normal University, Laboratory of Philosophy and Social Sciences at Universities in Jiangsu Province, China

^c College of International Studies, Shenzhen University, Guangdong, China

^d Faculty of Education, The University of Hong Kong, Hong Kong, China



ARTICLE INFO

Keywords:

Storyline
ChatGPT
Non-player characters
Game-based learning
EFL writing

ABSTRACT

This study introduces a novel design approach to game-based learning by integrating ChatGPT into the development of non-player characters (NPCs) within storylines. This integration seeks to enhance interactivity with student players and foster an immersive, dynamic learning environment. We applied this approach in English as a Foreign Language (EFL) writing classrooms to examine its impact on both the affective and cognitive aspects of student writing learning, as well as their writing performance. Utilizing a quasi-experimental design, the study involved two classes of Chinese undergraduate students. One class ($n = 42$) engaged with ChatGPT-powered NPCs, while the other class ($n = 43$) interacted with conventional NPCs in a storyline-based game designed to enhance argumentative writing skills. Quantitative data were collected through questionnaires assessing students' intrinsic motivation, situational interest, cognitive load, and effort regulation. Additionally, student essays were evaluated to compare writing performance across the two conditions. Qualitative data, including students' chat histories with ChatGPT-powered NPCs, were gathered to investigate their interactions. The results indicated that engaging with ChatGPT-enhanced NPCs positively influenced students' intrinsic motivation and situational interest, leading them to invest greater effort in completing learning tasks. Notably, interactions with ChatGPT-enhanced NPCs did not increase students' cognitive load during the learning process. Moreover, students who interacted with ChatGPT-powered NPCs exhibited superior writing performance, producing essays that were significantly clearer, more elaborated, and more persuasive, while also more effectively addressing opposing viewpoints. Additionally, students employed various conversational strategies while interacting with these NPCs. This study not only advances our understanding of the role of generative artificial intelligence in educational contexts but also provides valuable insights for educators seeking to enhance student engagement and learning outcomes through innovative instructional strategies.

1. Introduction

Game-based learning has gained significant traction in educational contexts due to its numerous benefits, including increased

* Corresponding author. Room 901, T.T. Tsui Building, The University of Hong Kong, Pokfulam Road, Hong Kong, China.
E-mail addresses: yaxiaoivy999@gmail.com (Y. Xiao), danling0125@gmail.com (D. Li), kaiguo@connect.hku.hk (K. Guo).

engagement, enhanced motivation, and improved learning outcomes (Bado, 2022). Different from gamification, which adds game elements to existing learning activities, game-based learning uses games as the medium through which students acquire knowledge (Lester et al., 2023). The storyline, a prominent element for game design, plays a pivotal role in game-based learning by immersing students and fostering meaningful connections to the content (Malone, 1981; Malone & Lepper, 1987). The storyline-driven game-based learning approach has been employed to enhance student learning experiences (for a review, see Novak, 2015).

A key component of these storylines is non-player characters (NPCs), which refer to interactive entities within the game (Barz et al., 2024; Uludağlı & Oğuz, 2023). NPCs are vital for fostering engagement and facilitating narrative progression. However, in conventional game-based learning, NPCs are often simplistic, scripted entities that lack the intelligence and sophistication needed for meaningful communication with student players (Chen & Wu, 2021; Lawrence & Sherry, 2021; Liew et al., 2024; Lin et al., 2018; Song & Sparks, 2019a, 2019b; Yang et al., 2022). Their interactions are typically predefined and limited in scope, resulting in inadequate immersion and interactivity.

With the recent advancements in generative artificial intelligence (AI), particularly large language models (LLMs) like ChatGPT, game developers are increasingly exploring the integration of these technologies into NPC design. This allows NPCs to engage in more dynamic and free-flowing conversations with players (Anand, 2023; Chen & Hou, 2024; Gao et al., 2024). This innovation can offer valuable insights to address the limitations in student-NPC interactions in storyline-driven game-based learning. As ChatGPT has captured popular imagination, ChatGPT is used as a catchall phrase for LLMs in our discussion of pertinent literature in this article.

In the realm of language learning, while ChatGPT may not yet be the primary learning tool for most students, it serves a distinctive role in their learning strategies (Stojanov et al., 2024). Survey studies have indicated high levels of student acceptance of ChatGPT, driven by its perceived usefulness and ease of use (Pasupuleti & Thiyyagura, 2024; Tram et al., 2024). ChatGPT has shown promise in evaluating linguistic accuracy in second language writing, achieving results comparable to human raters (Pfau et al., 2023), and is particularly cost-effective when well-trained human raters are not available (Steiss et al., 2024). Experimental studies have further demonstrated ChatGPT's effectiveness in enhancing students' writing motivation and improving their writing skills (Zare et al., 2025). Despite these advancements, few studies have explored the integration of ChatGPT in NPC design for game-based language learning. Enhancing NPCs with ChatGPT could significantly enrich students' language learning experiences by enabling more dynamic and personalized interactions (Iarovoii et al., 2024).

Thus, in this study, we integrate ChatGPT into the design and development of NPCs within storyline-driven game-based learning, specifically in the context of English as a Foreign Language (EFL) writing. Writing can often be a challenging task for EFL learners due to various factors, including anxiety and lack of motivation (De Smedt et al., 2018; Gürvendir & Uzun, 2023). Previous studies have investigated the use of the game-based learning approach to enhance students' writing skills (see Guo, Zhong, Zainuddin, & Chu, 2023 for a review). Some of these studies have utilized storyline-driven game-based learning to boost students' motivation, engagement, and performance in writing (e.g., Song & Sparks, 2019a, 2019b). To further enhance this approach, this study incorporates ChatGPT to empower NPCs within storylines, fostering a more interactive and immersive learning environment for EFL learners in the writing classroom. Guided by the Cognitive-Affective Model of Immersive Learning (CAMIL; Makransky & Petersen, 2021), this study aims to examine the effects of this innovative instructional design on students' affective and cognitive learning aspects, as well as their learning performance. Additionally, the study explores students' interactions with ChatGPT-powered NPCs during their learning processes. This study contributes to the fields of educational technology and language education by exploring innovative and effective applications of generative AI in game-based learning.

2. Literature review

2.1. Storyline-driven game-based learning

The storyline—the sequence of events and actions that shape a narrative—has long been recognized as a prominent game design element that motivates players to continue engaging with the game (Malone, 1981; Malone & Lepper, 1987). A well-crafted storyline effectively connects players to the game world, its characters, and the challenges they encounter. This powerful tool immerses players in the gameplay, sustaining both their emotional and cognitive engagement throughout the experience (Hilderbrand et al., 2024). Since the early 1990s, researchers have acknowledged the potential of integrating storylines in game-based learning, a concept known as storyline-driven game-based learning (Iuppa & Borst, 2007). Much like how storylines draw players into gameplay, storyline-driven game-based learning aims to create an immersive educational environment that fully engages students in the learning process, both affectively and cognitively.

From an affective perspective, storylines provide a meaningful context, allowing students to pursue concrete goals, rather than abstract concepts (Kua et al., 2024). As students observe the direct consequences of their actions within the storyline, they develop a strong identification with the characters they portray and feel related to their characters' fates (Chan et al., 2023). This engagement can lead to increased enjoyment as students interact with unfolding plots (de Almeida & dos Santos Machado, 2021).

From a cognitive perspective, storylines can serve as memory enhancers. Human memory and experiences are organized as storylines (Bruner, 1991). Comprehending a storyline involves connecting new information with existing knowledge, which promotes deeper understanding and facilitates the application of newly acquired knowledge to real-world situations. Thus, using storylines to present instructional content aids in information recall, comprehension, and application (Zuo et al., 2023).

Despite these advantages, the use of storylines in educational contexts has also faced criticism, with some researchers dismissing this approach as merely 'sugar-coated instruction.' Critics argue that poorly designed storylines may distract learners and fail to produce meaningful learning outcomes. For example, incorporating irrelevant or poorly integrated stories into math word problems

demotivates students and imposes additional cognitive load as they struggle to translate the story into equations. In such cases, storylines become a counterproductive distraction (Parker & Lepper, 1992). A systematic review by Novak (2015) found that six out of seven experimental studies manipulating storyline game design elements reported either non-significant or lower learning gains associated with storyline-driven game-based instruction compared to non-storyline conditions. However, Novak argued that a well-crafted storyline, one that aligns naturally with learning objectives, can still have a powerful motivational effect and enhance learning. He emphasized that future research should improve the design of storylines to ensure they support, rather than hinder, educational outcomes.

Several attempts have experimentally examined the effectiveness of storyline-driven game-based learning in writing education, consistently reporting its positive influence. This approach has been shown to improve argumentation performance (Lam et al., 2018; Song & Sparks, 2019a), foster more on-topic online contributions in the discussion forum of a writing class (Lam et al., 2018), boost business writing performance (Lin et al., 2018), and enhance complexity and fluency of narrative writing (Sánchez & Sunesson, 2023). Additionally, in some single-group case studies, students have reported increased interest and attention (Song & Sparks, 2019b), enhanced creativity (Lee, 2019), and heightened curiosity and plausibility (Dickey, 2011) in storyline-driven game-based writing learning. They also noted that receiving in-story feedback was beneficial for mastering argumentation skills (Lawrence & Sherry, 2021).

These findings underscore the significant potential of storyline-driven game-based learning to enhance writing education. They also emphasize the need for further research into more carefully designed implementations to fully realize its educational benefits.

2.2. NPC design in game storylines

Games with embedded storylines feature various types of characters, with players typically controlling the main protagonists at one end of the spectrum, while on the other end, the computer controls other characters to populate the environment and provide ample fodder for the protagonists to develop their identities and achieve their goals (Daviault, 2012). The characters that are controlled by the game's algorithms, rather than players, are defined as NPCs, also known as digital agents (Barz et al., 2024; Uludaglı & Oğuz, 2023). NPCs are particularly indispensable in narrative-driven genres such as role-playing games, action-adventure games, and first-person shooter games (Adams, 2013). NPCs facilitate game progression often by acting in a manner that *pushes/guides* the player in a certain direction. They serve multiple key functions that not only advance storylines but also enhance players' experience in gameplay, two aspects that ultimately determine players' reception and enjoyment of a game (Daviault, 2012).

Flow refers to the “process of optimal experience” and describes a state of deep focus, enjoyment, and immersion when individuals are fully engaged in an activity (Csikszentmihalyi & LeFevre, 1989, p. 816). According to flow theory, activities that are most likely to induce the flow state share key components, including clear goals, a balance between challenge and skill, social and emotional connections, a sense of control, and immersion (Nakamura & Csikszentmihalyi, 2009). Consistent with the flow theory, NPCs in digital games can afford these components by assuming various roles to enhance gameplay (Liew et al., 2024). Specifically, NPCs can act as *quest givers*, offering players quests/missions through immersive cutscenes, providing clear objectives crucial for maintaining flow (Warpefert & Verhagen, 2015). In addition to guiding players, NPCs can take on the role of *enemies*, presenting combat challenges that test players' skills. Such scenarios balance challenge and mastery by requiring players to apply their abilities strategically. Success in these encounters often results in rewards – such as valuable loot or improved combat proficiency – that reinforce player progression and sustain the flow state (Hubble et al., 2021). NPCs also frequently serve as *allies/companions*, assisting by either autonomously fighting alongside players or engaging in cooperative gameplay, whose presence can add to flow through interactions and teamwork (Wehbe et al., 2017). In non-combat games, NPCs often serve as *information providers*, offering guidance, tutorials, hints, and resources to help players learn game mechanics or navigate challenges usually through conversations (Ahlquist & Novak, 2008). These interactions can prevent the confusion or sense of losing control that might disrupt flow. Lastly, in some games, NPCs exist solely to enhance immersion. For example, townsfolk in a virtual village may not directly interact with players but enrich the environment by creating a believable and authentic game world. This authenticity deepens immersion, a key element of flow (Warpefert & Verhagen, 2015).

2.3. NPC design in storyline-driven game-based learning

Game-based learning seeks to harness the entertainment value of games to achieve educational objectives. In this context, the roles assumed by NPCs during gameplay can be adapted to the learning process to enhance students' flow states (Liew et al., 2024). Prior studies have examined their impacts, demonstrating largely positive educational results linked to this enhancement. Specifically, the *quest givers* function of NPCs has been employed to assign tasks to students and evaluate their performance upon submission of answers. This implementation has been linked to enhanced skill development, such as improved argumentation skills (Song & Sparks, 2019a, 2019b) and critical thinking skills (Chen & Wu, 2021). Meanwhile, studies have highlighted students' positive attitudes toward learning activities supported by such NPCs (Chen & Wu, 2021; Lawrence & Sherry, 2021; Lin et al., 2018).

The role of NPCs as *enemies* has been employed to challenge students, encouraging them to enhance their speed and accuracy in problem-solving and completing learning tasks. This approach has been shown to significantly enhance learning performance when compared to non-game conditions (Jaguš et al., 2018).

As *allies/companions*, NPCs have been adopted to provide emotional support rather than direct problem-solving assistance. For example, they offer encouragement when students face failure (Yang et al., 2022) or simulate peer interactions in online classes (Yuan & Gao, 2024). These NPCs have been shown to significantly afford playfulness and flow, which, in turn, can boost learning interest and

continuous learning intention (Yang et al., 2022). Furthermore, they can significantly improve immersion and social presence compared to non-NPC conditions (Yuan & Gao, 2024).

The *information providers* function of NPCs has been used to support instruction by offering hints and guidance that help students navigate learning materials. Liew et al. (2024) revealed that NPCs in this role significantly afforded social presence, immersion, social support, and story progression in educational games.

Lastly, NPCs designed to populate the game world have been adapted as “idle sitting” classmates in virtual classrooms to increase the realism of the learning environment. However, this specific use has shown limited contributions to learning experiences and performance (Kasapakis et al., 2023).

Despite the largely promising effects of NPC design in storyline-driven game-based learning, the level of interactivity between NPCs and students in these instructional designs remains limited. Specifically, when engaging in ‘conversations’ with NPCs, students’ input was limited. They can only respond by ‘clicking’ (Lin et al., 2018; Song & Sparks, 2019a, 2019b) or ‘dragging’ (Lawrence & Sherry, 2021). At the same time, NPCs’ responses are typically pre-scripted, delivering uniform content to all students. For instance, NPCs may reward or deny points based on predefined correct or incorrect answers (Chen & Wu, 2021; Lawrence & Sherry, 2021; Lin et al., 2018; Song & Sparks, 2019a, 2019b), or provide identical encouragement messages (Liew et al., 2024; Yang et al., 2022). These scripted dialogues restrict personalization, improvisation and flexibility, straying from real-world interactions. Such limitations diminish NPCs’ ability to create a more immersive game world and evoke meaningful responses from students (Iarovoii et al., 2024). Therefore, enhancing the design of NPCs to enable more dynamic and personalized interactions is highly desirable. This improvement can significantly elevate the level of interactivity and immersion of the storyline-driven game-based learning, leading to a more engaging learning experience and improved learning outcomes.

2.4. A ChatGPT-powered approach to NPC design

ChatGPT, a state-of-the-art LLM, is revolutionizing the AI field with its human-like skills. ChatGPT’s essential advantage lies in its ability to communicate in natural language by understanding and answering questions (Patil & Gudivada, 2024). ChatGPT particularly excels in language processing and conversation by delivering relevant responses drawn from extensive training on diverse sources, including Wikipedia, books, articles, and the internet. By analyzing users’ inputs and offering feedback tailored to individual understandings of a topic, it effectively meets users’ needs. Users are quickly captivated, enjoying free input and dynamic interactions that feel more like conversations with a friend or colleague than with a machine. Additionally, its willingness to engage tirelessly in discussions on any topic and to listen attentively further enhances its appeal. These qualities have contributed to the phenomenal success of ChatGPT, generating significant interest and accelerating efforts to explore its applications.

In the gaming industry, ChatGPT holds great potential for creating unique NPCs by enhancing their conversational abilities. This advancement could lead to more personalized and dynamic interactions, resulting in richer storytelling and immersive experiences for players (Huang & Huang, 2023; Makridakis et al., 2023; Moon et al., 2024). However, efforts to integrate ChatGPT-powered NPCs into game design are still in the early stages. Some initiatives for entertainment have offered players a selection of traditional dialogue options (Huang, 2024; Iarovoii et al., 2024) or unstructured free audio/textual input (Nagarkar, 2024; Obedkov, 2023) when interacting with these NPCs. A limited number of attempts have proposed or implemented integrating ChatGPT-powered NPCs to non-entertainment games, such as anxiety disorder treatment (Gao et al., 2024), training in verbal de-escalation strategies – aimed at reducing tension and abusive behavior while promoting empathy, boundaries, and cooperation among police officers (Anand, 2023), and employee working ethics training (Chen & Hou, 2024). However, the results of these initiatives have been inconsistent. While some participants reported that this method was effective for training (Anand, 2023), others found no significant effects on their intrinsic motivation, or anxiety during the training (Chen & Hou, 2024).

Implementations of ChatGPT-powered NPCs in game-based learning is even more sparse and limited (Chen et al., 2024; Guo & Liu, 2024; Karaosmanoglu et al., 2024; Song et al., 2024; Wang et al., 2025). Most existing studies have primarily focused on presenting the design process (Guo & Liu, 2024; Song et al., 2024) or tested prototypes with a small sample size (Karaosmanoglu et al., 2024). For instance, Karaosmanoglu et al. (2024) piloted ChatGPT-powered NPCs to help learners practice French vocabulary and grammar, receiving positive feedback on their effectiveness. Few studies have experimentally examined the impact of ChatGPT-powered NPCs on educational outcomes. Chen et al. (2024) utilized ChatGPT-powered NPCs to practice oral English with students, offering scores and feedback on their English spoken sentences. A quasi-experiment revealed that this approach significantly improved students’ intrinsic motivation and self-efficacy in English speaking. However, it did not produce significant improvements in learning performance compared to the conventional approach using non-AI NPCs. Wang et al. (2025) adopted ChatGPT-powered NPCs to help students learn knowledge about plants, an essential aspect of science education. Their quasi-experiment demonstrated that this approach significantly enhanced both students’ flow experience and learning performance.

Inspired by the significant potential of ChatGPT-powered NPCs and the initial attempts described above, this study seeks to investigate their implementation in the context of EFL writing education. EFL writing is often perceived as a challenging and monotonous task by learners, which can result in decreased motivation and engagement (Solhi et al., 2024). Specifically, our research focuses on leveraging ChatGPT-powered NPCs to enhance EFL students’ learning of argumentative writing. In argumentative writing, to construct compelling arguments, writers must engage with critical thinking activities, including interpreting and analyzing information, evaluating evidence, and drawing persuasive conclusions (Zare et al., 2025). NPCs in educational games have presented a unique opportunity to develop critical thinking skills in argumentative writing, especially in terms of evaluation, as students need to evaluate viewpoints from multiple characters within specific scenarios (Song & Sparks, 2019a, 2019b). ChatGPT-powered NPCs can take this further by fostering an inquiry-based learning environment, whereby learners are encouraged to actively explore a topic,

continuously pose questions, and seek evidence. This process will enable them to construct arguments and draw conclusions grounded in evidence and reasoning (Lee, 2014; Wale & Bishaw, 2020). Such an iterative process has shown to be useful in refining critical argumentation skills, which are crucial for enhancing argumentative writing (Zare et al., 2025; Zhang et al., 2025). Despite these potential benefits, to the best of our knowledge, no study has yet utilized ChatGPT-powered NPCs to support argumentative writing instruction. This gap underscores a promising and underexplored avenue for future research.

2.5. Theoretical framework: CAMIL

Previous research has demonstrated that storyline-driven game-based learning environments serve as a typical example of immersive learning settings, where NPCs play a crucial role in enhancing immersion through interactions with students (Naul & Liu, 2020). These environments captivate learners by providing interactive and immersive experiences that foster deep affective and cognitive engagement in the learning process. This engagement can lead to a shift in attentional focus, where individuals may temporarily disconnect from their physical surroundings (Beck et al., 2023). In light of this, the present study adopts the CAMIL framework, a theoretical model that explores the learning process within immersive environments (Makransky & Petersen, 2021).

The CAMIL framework is well-established as it builds upon prior learning process research in immersive learning environments (e.g., Petersen et al., 2022) as well as established psychological theories such as the self-determination theory (Deci et al., 1994), the cognitive theory of multimedia learning (Mayer, 1997), and the control value theory of achievement emotions (Pekrun, 2006). This framework highlights the central role of *immersion* and *interactivity* in immersive learning environments (see Fig. 1). When learners are immersed in realistic problem-solving scenarios and engage in dynamic interactions, their sense of presence, i.e., feeling of being physically present in the virtual learning environment, improves (Lee, 2004). Meanwhile, dynamic interactions can enhance student agency, empowering them to actively control their actions and impact the learning environment (Haggard & Chambon, 2012).

Drawing on insights from the literature, storylines can immerse students in realistic learning contexts (Kua et al., 2024). Leveraging ChatGPT to power NPCs within these storylines holds great potential to elevate the interactivity of conventional student-NPC interactions by enabling dynamic and personalized conversations. This approach can also deepen immersion through human-like, engaging interactions (Huang & Huang, 2023; Makridakis et al., 2023; Moon et al., 2024). Unlike static, predefined scripts, ChatGPT-powered NPCs allow students to freely ask questions, express ideas, and make decisions, while providing immediate and tailored responses to their inputs (Chen et al., 2024; Wang et al., 2025). Accordingly, students' sense of presence and agency can be reinforced within the storyline-driven game-based learning environments (Makransky & Petersen, 2021).

The subsequent part of the CAMIL framework proposes that students' cognitive and affective factors identified as being most relevant to learning can be influenced with increased immersion and interactivity. These factors include intrinsic motivation, situational interest, cognitive load, and effort regulation. *Intrinsic motivation* arises when individuals engage in an activity for its inherent satisfaction rather than external rewards (Ryan & Deci, 2000). *Situational interest* refers to the focused attention and emotional reactions triggered by specific environmental stimuli (Hidi & Renninger, 2006). *Cognitive load* denotes the total amount of mental effort being utilized in working memory during learning or task performance (Mayer, 1997). *Effort regulation* refers to one's management and control of their efforts and energy expenditure in pursuing a goal or completing a task (Pintrich et al., 1991). Ultimately, the enhancement of affective and cognitive aspects of learning can lead to improved learning performance. These factors, as outlined in the CAMIL framework, inform the selection of measures for this study.

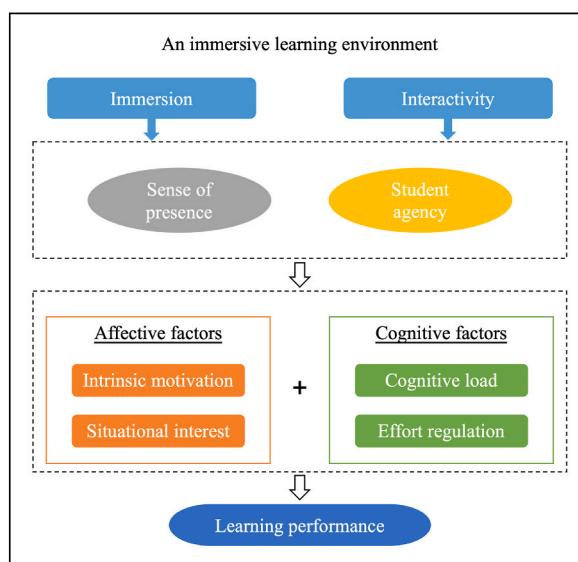


Fig. 1. The CAMIL framework.

2.6. The study

ChatGPT-powered NPCs provide unique advantages by facilitating personalized and dynamic interactions with students, greatly enhancing immersion and interactivity in storyline-driven game-based learning environments. In the context of argumentative writing instruction, these NPCs offer more than just affective appeal; they can play a crucial role in developing critical thinking skills that are essential for improving argumentation abilities. Despite their promising potential, there has been limited research on implementing this approach and experimentally assessing its effectiveness in educational settings. To address this research gap, the present study aims to integrate ChatGPT-powered NPCs to storyline-driven game-based EFL writing learning. Guided by the CAMIL framework, the following four research questions (RQs) were explored.

- RQ1: What are the effects of using ChatGPT-powered NPCs on EFL students' affective factors (intrinsic motivation and situational interest) in storyline-driven game-based writing learning?
- RQ2: What are the effects of using ChatGPT-powered NPCs on EFL students' cognitive factors (cognitive load and effort regulation) in storyline-driven game-based writing learning?
- RQ3: What are the effects of using ChatGPT-powered NPCs on EFL students' writing performance in storyline-driven game-based writing learning?
- RQ4: How do EFL students interact with ChatGPT-powered NPCs in storyline-driven game-based writing learning?

While RQ1, RQ2, and RQ3 explore the impacts of this innovative approach on the affective and cognitive aspects of students' writing learning, as well as their learning outcomes, RQ4 delves into the processes underlying students' engagement with it. The significance of this study lies in its potential to inform educational practices by providing insights into how AI technologies can enhance writing instruction, thereby contributing to more effective and engaging language learning experiences. Additionally, the findings may guide curriculum development and the integration of digital tools in educational contexts.

3. Methods

3.1. Context and participants

This study involved 85 undergraduate students from two classes (42 and 43 students, respectively) at a comprehensive university in China. Among the participants, 38.8 % identified as male, 57.6 % as female, and three students chose not to disclose their gender. The average age of the participants was 19.32 years ($SD = 0.54$). All participants exhibited English proficiency levels ranging from B2 to C1 according to the Common European Framework of Reference for Languages (CEFR). They were from various disciplines, including humanities and social science, management and economics, STEM (science, technology, engineering, and mathematics), and medicine. All students were enrolled in an English writing course taught by the second author, and the research was conducted during a session specifically designed to enhance students' argumentative writing skills. Notably, none of the students had prior experience with the storyline-driven game-based learning approach used in the study. Informed consent was obtained from all participants prior to their involvement in the study, ensuring that they were fully aware of the research purpose and their rights.

3.2. Storyline-driven game-based learning design

3.2.1. Storyline development

To create immersive environments for argumentative writing learning, we designed a detective story where students solved cases and wrote investigation reports. The storyline design focused on four key elements: *setting*, *character*, *plot* and *lesson* (Cohen, 2011). First, *setting* defines the location of the story, enabling the audience to mentally 'walk there'. In this study, we established three primary settings: the detective's office (settings 1 and 2) and a suspect's home (setting 3).

Second, *characters* are individuals who animate the setting and make a situation come to life. In this study, students played the role of the detective assigned to investigate a suicide case of a college student named Sarah Collins. They were to investigate three suspects (NPCs): Sarah's father, Michael Collins; her teacher, Ms. Chow; and her classmate, Alex Johnson.

Third, *plots* refer to the incidents that unfold within the storyline. Our storyline comprised three rounds of investigation, each featuring a central plot. Round 1: introduction of the case and suspects, where the detective gathered corroborative details and distinct perspectives from the three NPCs through inquiries; Round 2: escalation of the suspense with an emergent clue – a photo revealing a dark bruise on Sarah's ankle, which prompted the detective to inquire further; Round 3: revealing of the truth, where the detective uncovered a crucial lead indicating that Sarah's father had a second account, ultimately leading to the resolution of the case.

Fourth, *lesson* represents the educational insight that the storyline wishes to impart. In this study, we sought to highlight the characteristics of cyberbullying and expose its harmful influence on individuals and society as a whole.

Additionally, we incorporated visual elements drawing on the key design principles of the cognitive theory of multimedia learning (Mayer & Moreno, 1998). Specifically, we applied the *multiple representation principle* by combining textual storyline descriptions with visual elements, such as character illustrations, social media screenshots, and photographs, to promote dual-channel processing. To support the *contiguity principle*, we ensured that visual elements were presented spatially and temporally close to the corresponding textual descriptions, facilitating the construction of referential links between the two. Furthermore, we adhered to the *coherence principle* by including only visuals that were directly relevant and complementary to the textual content, avoiding unnecessary

elements that could distract learners. Finally, we followed the *redundancy principle* by ensuring that information was not duplicated across different modes, but rather presented in a way that each mode contributed uniquely to the learning process.

Appendix A presents the details of the four elements, including their definitions, specific designs in this study, and exemplary visual representations.

3.2.2. Using ChatGPT to enhance NPCs

We utilized the ‘Create bot’ function on the Platform for Open Exploration (Poe; <https://poe.com/>) to develop three conversational agents as the NPCs (Sarah’s father, her teacher, and her classmate) in the storyline.

The most critical aspect of bot creation was prompt engineering. Prompt engineering involves providing explicit instructions or queries to the language model, specifically GPT-3.5 in this study, to guide it toward achieving the desired outcomes. Key elements of a prompt include instruction, context, input data, and output indicators (Giray, 2023). *Instruction* refers to a specific task that directs the model’s behavior and guides it toward the intended output. *Context* encompasses external information or background knowledge that helps the model generate more accurate and relevant responses. In this study, all three NPCs were introduced with the overarching context and instruction as follows:

You are one of the main characters in a detective story. The victim is Sarah Collins, who was found dead in her bedroom a week ago. Your name is [...] Your relationship with the victim is [...] A detective has been assigned to investigate the case and will question you. You need to respond to the detective’s inquiries and provide any information you know to move the case forward.

Input data refers to the specific question or prompt that we want the model to process and respond to. This forms the core of the prompt and drives the model’s understanding of the task. *Output indicators* specify the desired format of the response, helping to shape it by defining whether we need a short answer, a paragraph, or another specific format. In this study, the creation of input data and output indicators was facilitated by ChatGPT. We provided ChatGPT with the complete story as source material and requested it to generate potential questions and outline specific key points as answers. We then reviewed, selected, and modified these questions and key points as necessary, which were ultimately provided to the three ChatGPT-powered NPCs as their knowledge bases. They were instructed: “Only if the detective asks about [potential questions], should your response include [specific key points].”

Given that the storyline unfolded gradually with emerging evidence, we established guidelines: “You must only answer the questions asked by the detective. Don’t provide extra information or admit/reveal anything unless the detective brings it up.” To prevent the bots from deviating from the relevant investigation, we instructed them: “If the detective asks any question irrelevant to the case, refuse to answer and redirect the conversation back to the point.”

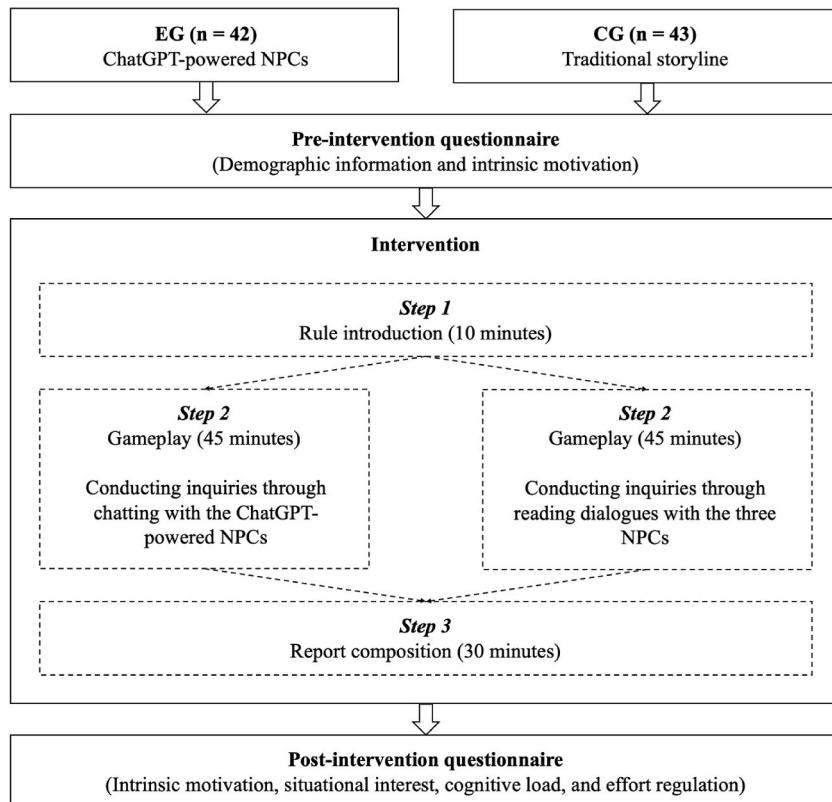


Fig. 2. Research procedures.

The prompt engineering process was iterative. Throughout the creation phase, we constantly tested the bots by (1) asking questions in order to verify the accuracy of their responses and (2) asking questions in random order or about unrelated topics to examine if this disrupted the plot. Once we finalized the process, we piloted the bots with four undergraduate students, and the bots operated effectively.

3.3. Research procedures

The study utilized a quasi-experimental design, assigning one class as the experimental group (EG; $n = 42$) and the other as the control group (CG; $n = 43$). Fig. 2 displays the research procedures. Prior to the intervention, the EG students who interacted with the ChatGPT-powered NPCs were guided to set up their Poe accounts. We also provided them with tips on how to conduct inquiries and demonstrated effective communication with bots on Poe to minimize confusion and offer initial scaffolding. The CG students who used traditional storyline without ChatGPT-powered NPCs were not instructed to use Poe. Additionally, both groups completed a pre-intervention questionnaire that collected demographic information and assessed their initial intrinsic motivation for EFL writing learning.

The intervention lasted approximately 85 min and consisted of three key steps: *rule introduction* (10 min), *gameplay* (45 min), and *report composition* (30 min). In the first step, we introduced the game's background and rules. Specifically, participants assumed the role of detective investigating a college student's suicide case, with the primary task of conducting three rounds of inquiries with three suspects (NPCs). In the second step, participants were assigned digital texts for each round of investigation, each lasting 15 min, during which they were encouraged to take notes on their suspicions regarding the true culprit. The CG students received a re-readable digital text that detailed the progression of the story for each round, presenting the inquiries primarily as dialogues between the detective and the three NPCs. In contrast, the EG students conducted inquiries with the ChatGPT-powered NPCs, engaging interactively rather than simply reading dialogues. In the third step, after the three-round investigation, participants were asked to write a 200-word report articulating their conclusions about the case in 30 min. In their reports (essays), they were required to justify their conclusions and include relevant examples from the investigation.

Of note is that the source information for both groups remained the same during the gameplay, although they interacted with the story in different ways. Specifically, the fixed plots of the storyline, presented through texts and pictures, were consistent for both groups. These plots included the suspects' introduction in round 1, the clue regarding Sarah's injured ankle in round 2, the crucial leads – the login reminder on Michael's phone and Sarah's record of harasser, and the truth – the unread message in the Mr2000 account, in round 3. The only difference lay in how students interacted with the NPCs. For the CG, perspectives and explanations from the NPCs were presented as dialogues in the digital texts for students to 'read' (see Appendix B for the CG's game interface). Hyperlinks were included, allowing students to navigate between pages and read the conversations in any order they chose. In contrast, the EG engaged in inquiries by chatting with the ChatGPT-powered NPCs, which were built using the same source information (see Appendix C for the EG's game interface).

Upon completing the task, both groups were asked to respond to a post-intervention questionnaire aimed at assessing their learning experiences.

3.4. Data collection and analysis

3.4.1. For addressing RQ1 and RQ2

RQ1 and RQ2 focused on students' affective and cognitive factors in the novel learning design. To assess students' affective factors, we evaluated their prior levels of *intrinsic motivation* in writing learning through a pre-intervention questionnaire adapted from the Intrinsic Motivation Inventory (IMI, Deci et al., 1994). The questionnaire contained three dimensions: (1) perceived interest/enjoyment, which reflected the pleasure derived from engaging in writing learning (e.g., "English writing is a pleasant/enjoyable experience to me"); (2) perceived value/usefulness, which referred to the belief that engaging in writing learning can yield beneficial outcomes (e.g., "English writing is important to my improvement"), and (3) perceived autonomy, which referred to the sense of control in writing learning (e.g., "I felt a sense of control while learning English writing"). Students' post-intervention levels of intrinsic motivation were assessed using the same items.

The post-intervention questionnaire also measured students' *situational interest*. This concept was evaluated using items adapted from the Situational Interest Scale (SIS, Chen et al., 1999), which included three dimensions: (1) attention quality, which referred to the level of focus during the writing learning activity (e.g., "I was concentrated in this writing activity"); (2) exploration intention, which referred to the desire to investigate the writing learning activity (e.g., "I want to discover all the clues in this writing activity"), and (3) sense of challenge, which referred to the perception of the writing learning activity as demanding yet achievable (e.g., "This writing activity is a demanding task").

Students' cognitive factors were assessed through *cognitive load* and *effort regulation* using the post-intervention questionnaire. Cognitive load encompasses both intrinsic and extraneous load. Intrinsic load stems from the inherent nature of information (Sweller, 2011), while extraneous load pertains to how information is presented to the learner (Makransky & Petersen, 2021). We measured these aspects using items adapted from the Cognitive Load Scale (CLS, Andersen & Makransky, 2021), including both intrinsic load (e.g., "The vocabulary in the suspects' responses was difficult to understand") and extrinsic load (e.g., "The way to seek information in this writing activity made it harder to learn"). Effort regulation was measured using items adapted from the Motivated Strategies for Learning Questionnaire (MSLQ, Pintrich et al., 1991), e.g., "I tried hard to do well in this writing activity".

These scales, namely the IMI (e.g., Bosch, 2024), SIS (e.g., Cheng et al., 2022), CLS (Andersen & Makransky, 2021), and MSLQ (e.g.,

([Zhou & Wang, 2021](#)), were selected as they have been widely used to measure corresponding concepts in previous studies, demonstrating adequate reliability (Cronbach's alpha >0.7) and validity (Root Mean Square Error of Approximation <0.08, Comparative Fit Index >0.9).

Notably, we did not include perceived situational interest, cognitive load, and effort regulation measures in the pre-intervention questionnaire, as these factors could only be assessed after experiencing the interventions.

Students were asked to indicate the extent to which the statements in the pre- and post-intervention questionnaires reflected their experiences on a seven-point Likert scale, ranging from 1 (very untrue of me) to 7 (very true of me).

The questionnaire achieved a response rate of 87 %, with 7 out of 42 participants in the EG and 4 out of 43 in the CG not responding to the post-intervention questionnaire. Logistic regression analyses indicated that participant dropout from the post-intervention questionnaire was not significantly related to age, gender, or initial intrinsic motivation levels. Therefore, we conclude that the dropout of the 11 participants was relatively random.

An Analysis of Covariance (ANCOVA) was conducted to compare intrinsic motivation between the EG and the CG while controlling for their initial motivational states towards writing. Independent sample *t*-tests were performed to compare the two groups regarding situational interest, cognitive load, and effort regulation reported in the post-intervention questionnaire.

3.4.2. For addressing RQ3

RQ3 examined the impact of the novel learning approach on students' writing performance. To investigate this, we collected participants' reports (argumentative essays) produced during the third step of the intervention. The study achieved an essay submission rate of 84 %, with 8 out of 42 participants in the EG and 6 out of 43 in the CG not submitting their essays. Logistic regression analyses showed that essay non-submission was not significantly related to participants' age, gender, or responses to any of the measured questionnaire items. Consequently, we concluded that the dropout of these 14 participants was relatively random and did not introduce systematic bias.

To evaluate the quality of these essays, we employed the rating rubric in ([Guo, Zhong, Li, and Chu \(2023\)](#)). This rubric was specifically designed to evaluate the quality of arguments in Chinese EFL students' argumentative writing, making it highly relevant to our context. The rubric uses a 5-point scale across seven dimensions of argumentation quality: organization, sufficiency, clarity, elaboration, relevance, persuasiveness, and addressing opposing views (see [Appendix D](#)). According to this framework, higher-quality arguments are characterized by greater organization, sufficiency, clarity, elaboration, relevance, persuasiveness, and the ability to effectively address opposing viewpoints.

To ensure inter-rater reliability, two raters independently scored a randomly selected subset of 15 essays (21 % of the dataset). Reliability for the rubric dimensions was assessed using Weighted Kappa, a method appropriate for ordinal data ([Li et al., 2023](#)). The analysis demonstrated good agreement across the seven dimensions, with the kappa values exceeding 0.6 ([McCluggage et al., 1998](#)): organization (0.667), sufficiency (0.615), clarity (0.722), elaboration (0.672), relevance (0.762), persuasiveness (0.737), and addressing opposing views (0.727). Any discrepancies between raters were resolved through discussion until consensus was reached. Following this process, the remaining 56 essays were scored by a single rater.

We performed independent sample *t*-tests to compare the EG and the CG across the seven dimensions of argumentation quality. Additionally, we assessed their overall writing performance by summing the scores of these seven dimensions.

3.4.3. For addressing RQ4

To answer RQ4 regarding students' interactions with ChatGPT-powered NPCs, the EG students' chat histories with the three NPCs on Poe were collected. Of the 42 students in the EG, 30 provided their chat histories.

Our analysis focused on students' messages to the NPCs, totaling 1171 messages. A 'talk move' served as the unit of analysis, representing a single discrete thought or idea that could be expressed by a single word, a phrase, sentence or even a paragraph to initiate and sustain productive interactions during problem-solving. The coding framework utilized (see [Appendix E](#)) was adapted from ([Hu and Chen \(2024\)](#)). Initially, the first and second authors independently coded the messages of eight randomly selected participants (27 % of the dataset) based on the coding framework ([Hu & Chen, 2024](#)), achieving an inter-coder agreement exceeding 95 %. To ensure consistency, the two coders engaged in detailed discussions to resolve any discrepancies and reached full consensus on the coding criteria. Following this, the second author coded the messages of the remaining 22 participants. Subsequently, a total of 1311 talk moves were identified in the student messages, categorized into 13 types across six themes: *elicitation, confrontation, disclosure, bonding, clarification, and attention* (see [Appendix E](#)).

It is important to note that nonsensical messages were excluded from this analysis, as they did not qualify as talk moves. Additionally, some messages contained multiple talk moves. For instance, Participant 8's statement, "According to Alex, you asked Alex and Sarah to meet you at school the day Sarah died. This implies you were not at home that day, which raises the question: are you lying?" contained two distinct interrogation moves: *disclose information* and *disagree*. First, the student introduced new evidence gathered during the interrogation with Alex (*disclose information*), providing crucial context to the dialogue. Second, this evidence prompted the students to voice their disagreement (*disagree*), questioning the credibility of the NPC's statements.

4. Results

4.1. Effects on affective factors (RQ1)

Students' affective factors in writing learning included intrinsic motivation and situational interest. Their intrinsic motivation was

measured through perceived enjoyment, value, and autonomy. The three dimensions demonstrated adequate reliability in both the pre-intervention (Cronbach's alpha = 0.932, 0.864, 0.774) and post-intervention questionnaires (Cronbach's alpha = 0.956, 0.946, 0.885). Typically, Cronbach's alpha values higher than 0.70 are considered acceptable (Kline, 1999). There were no violations of normal distribution for the three concepts in either questionnaire, as indicated by skewness values within the range of [-2, 2] and kurtosis values within [-7, 7] (Hair et al., 2010). Additionally, the homogeneity of regression slopes was checked, with non-significant interaction terms between pre-intervention enjoyment ($p = 0.114$), value ($p = 0.528$), and autonomy ($p = 0.442$) with the group (Field, 2009).

It was found that there was a significant effect of using ChatGPT-powered NPCs on students' perceived enjoyment of writing learning after controlling for their initial enjoyment level, with a large effect size, $F(1, 72) = 13.981, p < 0.001$, partial $\eta^2 = 0.165$ (Cohen, 1969). Students in the EG (adjusted $M = 6.296$, adjusted $SE = 0.157$) reported a significantly higher level of enjoyment in writing learning than their counterparts in the CG (adjusted $M = 5.484$, adjusted $SE = 0.148$).

Furthermore, a significant effect of using ChatGPT-powered NPCs was found on students' perceived value of writing learning, controlling for their initial level of perceived value. This effect was medium to large in size, $F(1, 72) = 10.214, p = 0.002$, partial $\eta^2 = 0.126$. Students in the EG (adjusted $M = 5.944$, adjusted $SE = 0.152$) indicated a significantly higher level of perceived value of writing learning than their counterparts in the CG (adjusted $M = 5.264$, adjusted $SE = 0.144$).

Additionally, the innovative learning approach also significantly impacted students' perceived autonomy in writing learning after controlling for their initial autonomy level, with a medium to large effect size, $F(1, 72) = 5.580, p = 0.021$, partial $\eta^2 = 0.073$. Students interacting with ChatGPT-powered NPCs (adjusted $M = 5.934$, adjusted $SE = 0.161$) achieved a significantly higher level of autonomy during the learning activity than those in the CG (adjusted $M = 5.409$, adjusted $SE = 0.152$).

Students' situational interest was measured through perceived attention quality, exploration intention, and sense of challenge. The post-intervention questionnaire demonstrated strong reliability across all three dimensions, with Cronbach's alpha values of 0.964, 0.907, and 0.875, respectively. Moreover, no violations of normal distribution were detected for these concepts, as indicated by skewness values within the range of [-2, 2] and kurtosis values within [-7, 7].

Regarding attention quality, the EG ($M = 6.457, SD = 0.767$) significantly outperformed the CG ($M = 6.009, SD = 0.711$), with a medium to large effect size, $t(72) = 2.610, p = 0.011$, 95 % CI = [0.106, 0.791], $d = 0.608$.

For exploration intention, the EG ($M = 6.352, SD = 0.690$) again showed a significant advantage over the CG ($M = 5.923, SD = 0.843$), with a medium to large effect size, $t(72) = 2.380, p = 0.020$, 95 % CI = [0.070, 0.789], $d = 0.554$.

No significant difference was found in their sense of challenge, with $t(72) = -0.162, p = 0.872$, 95 % CI = [-0.604, 0.513], between the EG ($M = 3.390, SD = 1.209$) and the CG ($M = 3.436, SD = 1.197$).

4.2. Effects on cognitive factors (RQ2)

Students' cognitive factors in writing learning included cognitive load and effort regulation. Their cognitive load included both intrinsic and extraneous load. Both of these two dimensions demonstrated adequate reliability (Cronbach's alpha = 0.733, 0.882) in the post-intervention questionnaire. The scale measuring students' effort regulation also demonstrated good reliability (Cronbach's alpha = 0.810). Normal distribution assumptions were met for both cognitive load and effort regulation, as indicated by skewness values within the range of [-2, 2] and kurtosis values within [-7, 7].

No significant difference was found between the EG ($M = 2.133, SD = 1.073$) and the CG ($M = 2.556, SD = 0.938$) in terms of intrinsic load, with $t(72) = -1.806, p = 0.075$, 95 % CI = [-0.888, 0.044]. Similarly, there was no significant difference, with $t(72) = -1.532, p = 0.130$, 95 % CI = [-0.931, 0.122], between the EG ($M = 2.243, SD = 1.109$) and the CG ($M = 2.647, SD = 1.157$) regarding extraneous load.

The results showed that the EG ($M = 6.029, SD = 0.818$) significantly outperformed the CG ($M = 5.573, SD = 0.924$) in effort regulation, with a medium to large effect size, $t(72) = 2.237, p = 0.028$, 95 % CI = [0.050, 0.862], $d = 0.521$.

4.3. Effects on writing performance (RQ3)

Students' writing performance was measured by the quality of arguments in their essays, including seven dimensions: organization, sufficiency, clarity, elaboration, relevance, persuasiveness, and addressing opposing views. There were no violations of normal distribution for all dimensions, as indicated by skewness values within the range of [-2, 2] and kurtosis values within [-7, 7].

The EG significantly outperformed the CG in four aspects, namely, clarity, with a medium to large effect size, $t(68.917) = 2.438, p = 0.017$, 95 % CI = [0.050, 0.503], $d = 0.576$; elaboration, with a large effect size, $t(67.188) = 4.095, p < 0.001$, 95 % CI = [0.244, 0.707], $d = 0.976$; persuasiveness, with a medium to large effect size, $t(69) = 2.343, p = 0.022$, 95 % CI = [0.040, 0.502], $d = 0.577$; addressing opposing views, with a medium to large effect size, $t(68.010) = 3.074, p = 0.003$, 95 % CI = [0.117, 0.552], $d = 0.724$.

There was no significant difference between the EG and the CG in terms of organization, with $t(64.526) = -1.043, p = 0.301$, 95 % CI = [-0.299, 0.094]; sufficiency, with $t(69) = 1.526, p = 0.131$, 95 % CI = [-0.055, 0.416]; and relevance, with $t(65.417) = -1.268, p = 0.209$, 95 % CI = [-0.352, 0.079].

Regarding overall writing performance, the EG exhibited significantly higher results than the CG, with a small to medium effect size, with $t(69) = 2.093, p = 0.040$, 95 % CI = [0.061, 2.535].

Appendix F presents a summary of the quantitative results concerning the effects of the novel learning approach on students' affective factors (RQ1) and cognitive factors (RQ2), as well as writing performance (RQ3).

4.4. Student interactions with ChatGPT-powered NPCs (RQ4)

The analysis of EG students' chat histories with the ChatGPT-powered NPCs revealed notable patterns in the conversational strategies used during their interrogations. These strategies were categorized into six distinct types: *elicitation*, *confrontation*, *disclosure*, *bonding*, *clarification*, and *attention*. These strategies illustrated how students navigated their role as a detective to foster productive dialogues while interacting with the NPC suspects.

4.4.1. Elicitation

Elicitation emerged as the predominant category, with students primarily focusing on eliciting *unknown information*. This talk move accounted for 673 instances (51 %) of all interactions. Students utilized direct questioning techniques to gather essential factual details from the NPCs, exemplified by inquiries such as "Where were you on the weekend of Sarah's death?" (Participant 26). This straightforward approach enabled them to collect vital information, such as alibis, timelines, and personal connections that could clarify the context of the case. Such strategies were essential for establishing a knowledge base necessary for further inquiry, ultimately guiding the direction of the investigation.

Building on the NPCs' initial responses, students prompted them to provide *further details*, which accounted for 178 instances (14 %). For example, by asking, "Can anybody testify for you about your being at home?" (Participant 18), they sought clarification and additional evidence, a key skill for constructing a more comprehensive and nuanced narrative.

Furthermore, students elicited the NPCs' *attitudes and opinions*, as shown by 144 instances (11 %) of inquiries such as "Do you regret?" (Participant 15). These questions encouraged self-reflection among the NPCs and provided valuable insights into their motivations and emotional states, enhancing the depth of the interrogation.

In addition, students prompted the NPCs to articulate the *reasoning* behind their actions by asking 'why' questions, noted in 92 instances (7 %).

In sum, these techniques enriched the interrogation and laid the groundwork for more assertive strategies aimed at challenging the NPCs' narratives and claims.

4.4.2. Confrontation

Confrontation played a crucial role in the interrogation process, enabling students to evaluate the reliability of the information gathered. They *questioned* the NPCs' responses in 29 instances (2 %), using questions like "Are you sure?" (Participant 2) to prompt the NPCs to reassess the credibility of their statements. This often led to more truthful and detailed responses.

Expressing *disagreement* was another tactic used, noted in 23 instances (2 %), where students challenged the NPCs by saying, "You are lying" (Participant 16). This confrontational approach not only prompted the NPCs to defend their statements but also forced them to engage more deeply with the conversation and the detective's line of questioning, creating a dynamic interplay that heightened the stakes of their responses.

Moreover, students' *persistence in questioning*, recorded in 9 instances (0.6 %), compelled the NPCs to confront critical issues. Statements like "Answer it. It is important." (Participant 28) underscored the urgency of the inquiry, reinforcing the need for thorough and honest responses.

Additionally, students engaged in *speculation* about potential plot developments based on the NPCs' responses (8 instances, 0.6 %). For instance, they posed inquiries such as, "You seem to be hiding something" (Participant 4). This form of speculative questioning elicited responses ranging from outright denial to reluctant acknowledgment, potentially uncovering critical information relevant to the unfolding narrative.

4.4.3 Disclosure

Disclosing information, with 102 instances (8 %), served a mediating role in the interrogation by introducing new findings and evidence that facilitated both information elicitation and confrontation. This tactic often accompanied other talk moves. For instance, statements like "I found out that you told Sarah's dad that dying hair is not good" (Participant 8) were typically followed by eliciting further details or reasoning. Additionally, when confronted with evidence such as, "According to Sarah's father, you were not satisfied with her hair color" (Participant 11), the NPCs were prompted to reassess the credibility of their statements. This strategic approach fostered deeper reflection and contributed to a more fruitful information exchange.

4.4.4. Bonding

Establishing and maintaining rapport is essential in interrogation work for eliciting information from suspects. In this study, students created *emotional connections* at various points in the interrogation, accounting for 28 instances (2 %) of their talk moves. For example, the statement "I am sorry for your loss" (Participant 3) helped foster trust with the NPC. This ongoing empathetic approach not only cultivated a supportive atmosphere but also encouraged the NPCs to disclose more information, thereby increasing the likelihood of obtaining critical details.

Additionally, *strategic concessions* were employed in 8 instances (0.6 %). Students recognized NPC refusals to engage on certain topics and allowed them moments of control, as illustrated by phrases like "OK, let's change the topic" (Participant 22). This tactic reduced the defensiveness of the NPCs, giving them a sense of autonomy and paving the way for more in-depth inquiries throughout the interrogation.

4.4.5. Clarification

By employing the technique of *clarifying oneself* (14 instances, 1 %), students ensured that the information elicitation process remained focused and productive. For example, statements like “I am not saying she’s the suspect.” (Participant 5) helped clarify the context, preventing misunderstandings and keeping the dialogue on track.

4.4.6. Attention

Active listening, while less frequent (3 instances, 0.2 %), was evident in phrases such as, “You seem to want to say more” (Participant 30), which demonstrated students’ attentiveness and reinforced that the NPCs’ thoughts were valued.

5. Discussion

5.1. Effects of using ChatGPT-powered NPCs

This study introduced an innovative ChatGPT-powered approach to NPC design within storyline-driven game-based learning, specifically applied to EFL writing instruction. To evaluate the effectiveness of this novel approach, the study examined both the affective and cognitive dimensions of students’ writing learning experiences, as well as their writing performance. In addition, the study analyzed students’ interactions with ChatGPT-powered NPCs by exploring their conversational strategies.

5.1.1. Heightened intrinsic motivation and enhanced situational interest

The results indicated that integrating ChatGPT-powered NPCs into storyline-driven game-based writing learning significantly improved students’ intrinsic motivation compared to traditional NPCs. This suggests that students engaged more deeply in the learning activity due to its inherent satisfaction. Specifically, students interacting with ChatGPT-powered NPCs reported finding the learning process more enjoyable and perceived the activities as more valuable for enhancing their argumentative writing performance. Additionally, they felt a greater sense of autonomy in their learning. Our results align with the findings of Chen et al. (2024), who found that role-play dialogue activities with ChatGPT-powered agents had significant positive influence on student intrinsic motivation in practicing English speaking.

Several factors may contribute to the heightened intrinsic motivation observed in the EG compared with the CG. First, NPCs can enhance the playfulness of learning activities, thus making the learning experiences more enjoyable for students (Yang et al., 2022). The integration of ChatGPT can further improve this playability by enabling dynamic conversations, thereby amplifying enjoyment (Karaosmanoglu et al., 2024; Zhang et al., 2025). Second, interacting with NPCs provided students with opportunities to evaluate the information provided by various NPCs (Song & Sparks, 2019a, 2019b). The incorporation of ChatGPT further enhanced this process by empowering students to ask questions and promptly receive tailored feedback. This facilitated ongoing and iterative inquiry, ultimately enhancing their comprehension and refining their critical argumentation (Zare et al., 2025; Zhang et al., 2025). These enhancements can increase the perceived value of the activity among students. Third, the opportunity for free conversation with ChatGPT-powered NPCs empowered students by allowing them to control their learning pace. As suggested by Chiu et al. (2023), this sense of control in chatting with AI bots can satisfy students’ need for autonomy. The increased perceived enjoyment, value and autonomy collectively signify heightened intrinsic motivation.

Moreover, the results highlighted that the use of ChatGPT-powered NPCs also significantly increased students’ situational interest compared to traditional NPCs. This suggests that these enhanced NPCs fostered greater focus and emotional involvement among students. Participants reported heightened concentration during the writing activities and expressed a stronger desire to explore the learning content. Several factors may explain this heightened situational interest. First, students may have felt a natural curiosity toward new experiences (Malone & Lepper, 1987); this was their first opportunity to interact with ChatGPT-powered NPCs within a storyline-driven game-based learning context, which generated excitement and stimulated students’ high levels of attention and a desire to explore. Second, according to the CAMIL framework (Makransky & Petersen, 2021), increased interactivity can enhance students’ interest in immersive learning environments. The interactivity offered by ChatGPT-powered NPCs engaged students more deeply in both the story and the learning activity, in contrast to conventional approaches where students passively read the narrative. Third, ChatGPT-powered NPCs can enhance students’ sense of immersion, fostering emotional attachment to the learning activity, which may keep them attentive and eager to explore further. Previous studies suggest that NPCs can immerse students by contextualizing learning scenarios (Liew et al., 2024). When students identified with the characters within the game world, it enhanced their perceived social presence (Yuan & Gao, 2024). The integration of ChatGPT further amplified this immersive experience by enabling human-like conversations (Huang & Huang, 2023; Makridakis et al., 2023; Moon et al., 2024).

The results indicated that students did not find the novel learning activity to be more challenging than the tasks associated with traditional storyline-driven game-based learning. This could be attributed to the initial scaffolding provided by the instructor, such as tips on how to conduct inquiries, which may have alleviated the challenges that students in the EG might have otherwise faced.

5.1.2. Greater effort investment without increased cognitive load

In terms of cognitive outcomes, the findings revealed that this innovative learning approach encouraged students to invest more effort in their tasks and overall learning. Several factors may explain this increased investment. First, the interactive nature of ChatGPT-powered NPCs fostered more meaningful interactions. Such engagement often leads to a greater willingness among students to self-regulate their time and effort in learning (Pan et al., 2024), aligning with assumptions of the CAMIL theory (Makransky & Petersen, 2021). Second, the enhanced motivational states—encompassing both intrinsic motivation and situational interest—are

closely linked to improved effort regulation. Specifically, increased feelings of playability, autonomy, curiosity, and immersion, along with the instrumental value derived from meaningful training in critical thinking and argumentation skills can facilitate students' commitment to achieving their goals. The link between affective factors and effort regulation has been the key assumptions of fundamental theories, including the control-value theory (Pekrun et al., 2023) and gamified learning theory (Landers et al., 2018), and has been supported by empirical studies (e.g., Xiao & Hew, 2024).

Importantly, the novel learning design did not impose additional cognitive load compared to traditional methods. This indicates that the integration of ChatGPT-powered NPCs facilitated a more effective and engaging learning experience without overwhelming students. This finding is encouraging, though it does not match the results from Wang et al. (2025), which indicated that ChatGPT-powered NPCs significantly reduced students' intrinsic cognitive load in understanding the course content. The difference may be attributed to variations in the NPCs' roles. In Wang et al. (2025), the NPCs were specifically designed to articulate key concepts and provide direct guidance, thereby significantly easing cognitive load. However, in our study, the NPCs were designed to provide information for students to explore and analyze independently.

Specifically, our study measured both intrinsic and extraneous cognitive load among students in the two groups. It is not surprising that no differences were found in intrinsic cognitive load, as both groups were exposed to the same underlying information—utilizing the same storyline. The distinction between the groups lay in how the information was presented: students in the EG interacted with the ChatGPT-powered NPCs to collect information, while those in the CG gathered information by reading the NPCs' dialogues. The finding that no differences were observed in extraneous cognitive load suggests that the interactive method of information collection did not demand any additional cognitive effort from students. There are some possible explanations for this finding. First, students may have been familiar with both reading dialogues and engaging in conversations, resulting in a similar processing experience. This familiarity likely reduced the cognitive effort required for either method. Previous psychological experiments have suggested that familiarity can modulate the effect of complexity on memory performance, thus reducing cognitive load (Zhang et al., 2020). Second, the interactive approach may have heightened student engagement without introducing additional complexity. When students are more engaged, they tend to be more focused, which can streamline cognitive processing and diminish the perception of cognitive load. Third, the increased motivation and interest fostered by the interactive method may have led students to view the tasks as less burdensome, even if the cognitive demands were comparable.

5.1.3. Superior writing performance and effective interaction with NPCs

The findings indicated that students interacting with ChatGPT-powered NPCs demonstrated significantly higher writing performance, specifically in argumentation quality, compared to those using conventional NPCs. This supports the CAMIL theory which posits that improving the affective and cognitive aspects of learning can ultimately enhance overall learning performance (Makransky & Petersen, 2021). Notably, students utilizing the innovative learning approach demonstrated greater clarity, elaboration, persuasiveness, and the ability to address opposing views in their argumentative essays. Our analysis of students' chat histories with ChatGPT-powered NPCs during gameplay provided insights into the reasons behind their superior performance compared to the control group.

Students employed a range of conversational strategies during their interactions with ChatGPT-powered NPCs, which was fostered by the inquiry-based learning environment embedded in the instructional design (Lee, 2014). Inquiry-based learning involves activities including examining sources, gathering, analyzing, interpreting, and synthesizing information, as well as formulating answers, explanations, and predictions. Such active engagement has been shown to effectively elicit critical thinking skills essential for improving argumentative writing performance (Wale & Bishaw, 2020).

Specifically, elicitation emerged as the most prevalent strategy employed by students, characterized by their active focus on gathering unknown information through direct questioning. Students sought not only factual details but also deeper insights into the NPCs' underlying motivations by asking follow-up questions. This enriched the overall interrogation process and highlighted their proactive approach to learning. By prioritizing the collection of information, students could enhance their comprehension while simultaneously refining their questioning techniques (Abdelghani et al., 2024). The ability to ask targeted and purposeful questions allowed the EG to extract specific details about the suspects' actions, intentions, and justifications. This, in turn, enabled them to build stronger arguments supported by direct quotes from the NPCs, contributing to more detailed and well-elaborated reasoning in their essays (Guo, Zhong, Li, & Chu, 2023).

The confrontation strategy strengthened students' ability to critically evaluate the reliability of the information they gathered. By expressing disagreement, challenging inconsistencies, and seeking clarification, students prompted NPCs to provide more accurate and defensible responses. This approach underscores their capacity for critical thinking, as they actively assessed the credibility of the information rather than passively accepting it. Consequently, the use of confrontation as a strategy contributed to enhanced persuasiveness in their reasoning, as students could craft more convincing arguments based on critically interrogated evidence (Zhang et al., 2025).

A related strategy, disclosure, further enhanced the productivity of these interrogations. Students consistently provided new evidence to guide their questioning and encouraged NPCs to reassess their statements. This led to a more dynamic exchange of information and allowed students to gather more accurate and relevant details. The strategy of disclosure illustrates how students effectively used evidence to drive conversations while reinforcing their ability to formulate arguments using elaborated and persuasive evidence (Zhang et al., 2025).

Meanwhile, when misunderstandings arose during their interactions with ChatGPT-powered NPCs, students consistently clarified their points to ensure the dialogue remained productive and focused. By readjusting and articulating their thoughts during these interactions, students could further develop their communication proficiency. This ability to express their thoughts clearly and

effectively could explain why their essays exhibited greater clarity (Kaur, 2017).

In addition, students actively worked to establish rapport and build emotional connections with the NPCs. Students may feel more personally involved in the investigation while engaging in a highly immersive and interactive environment, thus making empathetic statements in inquiry. To reduce defensiveness and facilitate deep investigations, students strategically allowed NPCs to shift topics when necessary, ensuring the conversations remained comfortable yet productive. Furthermore, while receiving responses from the NPCs, students actively listened, acknowledged their thoughts, and validated their perspectives. This reciprocal engagement could encourage students to consider multiple perspectives and explore alternative viewpoints (Anand, 2023). Consequently, their essays displayed more balanced and empathetic discussions of the NPCs' actions, rather than focusing solely on one-sided blame or assigning culpability.

No significant differences were observed in the organization, sufficiency, and relevance of arguments between students interacting with ChatGPT-powered NPCs and those using conventional NPCs. This is understandable, as key information was explicitly presented in the pre-written NPC conversations for the CG. As a result, students from both groups were able to compose well-organized essays with a sufficient number of arguments directly related to the case.

Overall, these findings underscore the diverse advantages of incorporating ChatGPT-powered NPCs in storyline-driven game-based learning. The interactive nature of the conversations between students and NPCs encouraged deeper exploration and comprehension, resulting in arguments that were clearer, more elaborate, more persuasive, and better able to address opposing viewpoints.

5.2. Implications

The study has important implications. Theoretically, it applied the CAMIL framework to investigate the impact of ChatGPT-powered NPCs on students' writing learning. The findings demonstrate how increased immersion and interactivity influence both the cognitive and affective dimensions of students' writing learning, as well as their writing performance. This validates the applicability of the CAMIL framework in immersive game-based learning environments.

Practically, as one of the first studies to explore the potential of generative AI in enhancing NPC design and development within storyline-driven game-based learning, this study opens new avenues in this field. The positive findings associated with our innovative design provide encouragement for educators to employ it in their classrooms. First, our development of the storyline using four key elements (setting, character, plot, and lesson), combined with ChatGPT-powered NPCs using Poe, offers a versatile model for educators to adapt and customize this design across different educational levels and subject areas. ChatGPT has been recognized as an innovative and helpful tool to enhance and enrich the learning process in both K-12 and higher education (for reviews, see Perera & Lankathilaka, 2023; Zhang & Tur, 2024). Its integration with NPCs is promising in enabling educators to harness the affordances of games toward a new pedagogy of AI-powered game-based learning. This integration can nurture various domain-specific skills as well as higher-order cognitive skills such as critical thinking and problem-solving (Huber et al., 2024).

Second, it is crucial for educators to recognize that ChatGPT-powered pedagogical agents are not a substitute for their own professional roles. To maximize the benefits of this innovative design, educators must adopt a facilitative role (Luo et al., 2023). This includes training students in advance to effectively pose meaningful questions during their interactions with NPCs. Additionally, educators should actively monitor these interactions to ensure that students remain focused and aligned with the intended learning objectives throughout the process.

Last but not least, educators must remain vigilant about the potential ethical concerns associated with employing ChatGPT-powered NPCs (Tai et al., 2023). During the creation phase, educators should critically evaluate the responses generated by ChatGPT to identify and address potential biases present in the input data. During implementation, it is essential to maintain transparency by informing students about the use of ChatGPT, including how their interaction data is collected, stored and utilized.

5.3. Limitations and future research

The study has several limitations. First, the intervention lasted only 85 min, which, although effective in producing immediate effects on students' learning, may not accurately reflect long-term outcomes. Additionally, since it was the participants' first experience interacting with ChatGPT-powered NPCs in a storyline-driven game-based writing learning context, the novelty effect should be taken into account when interpreting and generalizing the findings. Future research should aim to implement this learning approach over a more extended period to evaluate its lasting impact on students' learning.

Second, the study focused exclusively on one narrative type (detective stories), which may limit the generalizability of our findings. Future studies could explore additional story genres and enhance the NPCs within these narratives using ChatGPT. This would provide a more comprehensive understanding of the learning approach's effects.

Finally, the interactions between students and NPCs in this study were limited to text-based communication. Future research could utilize multimodal generative AI, such as GPT-4o and Sora, to facilitate more dynamic interactions, allowing students to engage with NPCs in diverse modalities and creating a more interactive and immersive learning environment.

6. Conclusion

This study introduces, for the first time, an innovative design that leverages ChatGPT to power NPCs in storyline-driven game-based learning for writing education. The goal is to enhance interactivity and immersion in immersive learning environments. Grounded in the assumptions of the CAMIL theory, this design is assumed to improve both the affective and cognitive dimensions of writing

learning, ultimately enhancing writing performance. We experimentally examined its effectiveness in comparison to the conventional approach. The findings revealed that ChatGPT-powered NPCs significantly enhanced students' intrinsic motivation and situational interest, encouraging greater effort investment in learning tasks without increasing cognitive load. Additionally, this innovation resulted in students achieving superior writing performance, particularly in clarity, elaboration, persuasiveness, and their ability to address opposing viewpoints in their arguments. Furthermore, our analysis of students' interactions with ChatGPT-powered NPCs revealed their effective engagement through a range of conversational strategies, which positively influenced their writing performance. This study provides significant insights for future research to bridge generative AI and game-based learning.

CRediT authorship contribution statement

Ya Xiao: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Danling Li:** Writing – original draft, Methodology, Investigation, Formal analysis. **Kai Guo:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used GPT-4o in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

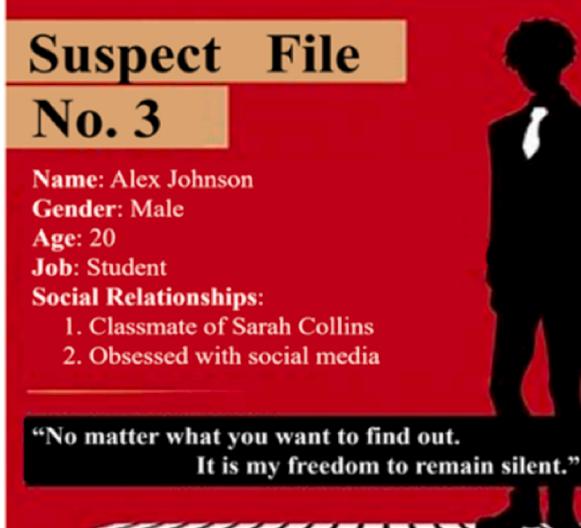
Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of competing interest

None.

Appendix A. Storyline development and illustrations within the game

Elements	Definitions	Design	Exemplary visual representations
Setting	The location of the storyline, enabling the audience to mentally 'walk there'	Settings 1 and 2: The detective's office Setting 3: Michael Collins' home	N/A
Character	Individuals who animate the setting and make a situation come to life	Student player: Detective NPC No. 1: Michael Collins (Sarah's father) NPC No. 2: Ms. Chow (Sarah's teacher) NPC No. 3: Alex Johnson (Sarah's classmate)	

(continued on next page)

(continued)

Elements	Definitions	Design	Exemplary visual representations
Plot	Incidents that take place within the storyline	<p>Round 1 investigation: Introduction of the case and suspects (three NPCs) Common threads in the NPCs' accounts: Sarah dyed her hair pink, posted about it, and subsequently faced online harassment. Distinct perspectives from the three NPCs:</p> <ul style="list-style-type: none"> • Michael: Sarah aspired to be a teacher. She bravely fought against internet trolls, struggling with the rapid spread of harmful content. He supported her online, particularly against critics from her classmate Alex. • Ms. Chow: She addressed Alex's behavior, urging him to apologize. • Alex: He claimed his actions stemmed from discomfort with Sarah's differences, not malice. 	

Round 2 investigation: Escalation of the suspense with an emergent clue

The clue: A photo revealing a dark bruise on Sarah's right ankle
Explanations from the three NPCs:

- Michael: He noted Sarah's injury and was concerned. However, Sarah was reluctant to discuss the injury, but he suspected Alex's involvement.
- Ms. Chow: She confirmed Alex was responsible, revealing that Sarah had requested this information remain hidden after his apology
- Alex: He admitted to pushing Sarah in response to her father's criticism but insisted it did not cause her death.



(continued on next page)

(continued)

Elements	Definitions	Design	Exemplary visual representations
		<p>Round 3 investigation: Revealing of the truth</p> <p>Crucial lead: As the detective updated Michael on the case, a log-in reminder for an account Mr2000 showed on his phone, linked to Sarah's harasser records.</p> <p>Explanations from the three NPCs:</p> <ul style="list-style-type: none"> • Michael: He admitted to creating Mr2000 to remind Sarah of a teacher's appearance after talking with Ms. Chow, intending it as guidance, not criticism. • Ms. Chow: She confirmed having a phone call with Sarah's father, but no clue of this account. • Alex: He didn't know about this account. <p>Truth: The detective urged Michael to log into Mr2000, where he found an unread message from Sarah expressing her frustration with mental illness and disappointment at his secret account.</p>	
Lesson	Educational insight that the storyline wishes to impart	The characteristics of cyberbullying and its harmful influence on individuals as well as the society.	N/A

Note. The above-described plots in three rounds were the same for both the CG and the EG.

Appendix B. The interface of game-based learning activities in the CG

You are an experienced detective, known as Sherlock. You are sitting at your desk, taking a sip of freshly brewed coffee. Suddenly, your boss calls you and gives you a new case:



A 20-year-old college student named Sarah Collins, is found dead in her dorm room a week ago. It seems that Sarah ended her life by herself. However, this case may not be a simple suicide. Now, you are on the case.

There are three suspects. Here are their profiles:

Suspect File No. 1

Name: Michael Collins
Gender: Male
Age: 45
Job: Engineer
Social Relationships:
1. Father of Sarah Collins
2. Widowed at age 32

*I must get justice for my daughter.
No matter what it takes.

Suspect File No. 2

Name: Ms. Chow
Gender: Female
Age: 38
Job: Teacher
Social Relationships:
1. Teacher of Sarah Collins
2. Dedicated to improve students' academic performance

*Students are like my own children.
Everything I do is for the good of them.

Suspect File No. 3

Name: Alex Johnson
Gender: Male
Age: 20
Job: Student
Social Relationships:
1. Classmate of Sarah Collins
2. Obsessed with social media

*No matter what you want to find out.
It is my freedom to remain silent.

You are to investigate **all three suspects** and find out who should be responsible for Sarah's death. Here is your 1st round investigation with:

Michael Collins
Ms. Chow
Alex Johnson



Michael Collins (Sarah's dad):

Good to see you, Detective. Thank you for coming to me. My sweet Sarah was a bright, strong-willed girl. After losing her mother at a young age, Sarah worked tirelessly to get into the best teaching program at HOPE University. This year, she was in her third year at HOPE, having already passed the tests and earned her teaching certificate. Her dream was to become a teacher and help kids. It warmed my heart to see her dream was coming true.

But then, the unthinkable happened. About two months ago, Sarah had dyed her hair pink, which she was really proud of. However, the backlash on social media was harsh. Sarah tried to refute those comments and report the accounts, but there were simply too many anonymous trolls. It was very difficult to track who incited and spread these comments. As her father, I was heartbroken and angry. I did my best to argue with and confront those Internet bullies, but I'm not very familiar with social media.

I found that a classmate of Sarah named Alex had been leaving threatening messages on Sarah's profile, so I confronted him under his posts. But he must have blocked me cause I couldn't no longer find his accounts.

"Now, my precious daughter is gone, taken from me in the worst possible way. I'm counting on you, Detective, to help me find the one responsible and get the justice Sarah deserves. I've always been my daughter's best friend and supported her. I can't let her down." Michael Collins lowered his head, an indescribable light flashed through his eyes...

[Back to Page 1](#)

Note. For the CG, the perspectives and explanations from the NPCs were presented as dialogues in the digital texts for students to 'read'. Students could click on the hyperlinks to read the dialogues in any order they preferred.

Appendix C. The interface of game-based learning activities in the EG

You are an experienced detective, known as Sherlock. You are sitting at your desk, taking a sip of freshly brewed coffee. Suddenly, your boss calls you and gives you a new case:

Suspect File No. 1

Name: Michael Collins
Gender: Male
Age: 35
Job: Engineer
Social Relationships:
1. Best friend of Sarah Collins
2. Widowed at age 32
"I must get justice for my daughter. No matter what it takes."

Suspect File No. 2

Name: Ms. Chow
Gender: Female
Age: 38
Job: Teacher
Social Relationships:
1. Teacher of Michael Collins
2. Dedicated to improve students' academic performance
"Students are like my own children. Everything I do is for the good of them."

Suspect File No. 3

Name: Alex Johnson
Gender: Male
Age: 22
Job: Student
Social Relationships:
1. Classmate of Sarah Collins
2. Obsessed with social media
"No matter what you want to find out. It is my freedom to remain silent."

There are three suspects. Here are their profiles:

You are to talk with all three suspects and find out who should be responsible for Sarah's death in the 1st round of investigation:

Talk with Michael Collins
https://poe.com/Michael_Collins_dad

Talk with Ms. Chow
https://poe.com/Ms.Chow_teacher

Talk with Alex Johnson
https://poe.com/Alex_J_classmate

ChatGPT conversation with Michael Collins' dad:

Michael.Collins_dad Poe
Hi, I'm Michael Collins, how can I assist you?
How close are you with your daughter?

Michael.Collins_dad Poe
Very close. You've always been Sarah's best friend and supported her. Sarah trusted you. You beg the detective to help you find the one responsible and get the justice Sarah deserves. You can't let Sarah down. (An indescribable light flashed through Michael's eyes...)

Michael.Collins_dad Poe
I last saw Sarah at home the night before she... before everything happened.
And what is she doing?

Michael.Collins_dad Poe
Sarah was studying for her upcoming exams and working on some assignments for her teaching program at HOPE Normal University.
Is she has any problem with anyone lately?

Michael.Collins_dad Poe
Sarah had been facing some online harassment recently, which had been really troubling her.
And why is that happened?

Michael.Collins_dad Poe
Sarah dyed her hair pink two months ago, and then she posted it on her social media. Unfortunately, Internet trolls made lots of offensive remarks about her new hair color.
Then what happened to her?

Michael.Collins_dad Poe
Sarah was very hurt by the offensive remarks, but she was tough and determined to fight back.

Note. For the EG, students engaged in inquiries by conversing with the ChatGPT-powered NPCs, which were developed using the same source information. Students could click on the hyperlinks to talk with the NPCs in any order they preferred.

Appendix D. Rating rubric for evaluating argumentation quality (Guo, Zhong, Li, and Chu (2023))

Aspects of argumentation quality	Definition	Score				
		1	2	3	4	5
Organization	The essay is well organized (coherent) and flows well.	The essay lacks coherence and flow.	The essay has some organization, but it is somewhat difficult to follow.	The essay is generally well organized and flows smoothly.	The essay is well organized and flows smoothly.	The essay is exceptionally well organized and flows seamlessly.
Sufficiency	The number of arguments is adequate and sufficient.	There are too few arguments, and they are insufficient to support	There are some arguments, but they are not sufficient to support the thesis.	There are adequate arguments to support the thesis, but some more could have been added.	There are adequate arguments to support the thesis.	There are considerable arguments, all of which are sufficient to support the thesis.
Clarity	The essay is compressible. The arguments are clearly formulated.	The arguments are poorly formulated, and the essay is difficult to understand.	The arguments are somewhat clear, but the essay requires some effort to comprehend.	The arguments are generally clear and easy to understand.	The arguments are clear and well formulated.	The arguments are exceptionally clear and easy to comprehend.
Elaboration	The arguments are well elaborated (e.g., with examples, analogies, and citing authorities).	The arguments are not elaborated or supported by any evidence or examples.	The arguments are somewhat elaborated, but additional evidence or examples could have been provided.	The arguments are adequately elaborated, with some supporting evidence or examples.	The arguments are well elaborated, with convincing supporting evidence or examples.	The arguments are exceptionally well elaborated, with strong and nuanced supporting evidence or examples.
Relevance	The arguments presented are relevant.	The arguments presented are irrelevant to the topic or thesis.	Some of the arguments presented are relevant but do not directly support the thesis.	The majority of the arguments presented are relevant and support the thesis.	All of the arguments presented are relevant and support the thesis.	The arguments presented are not only relevant but also insightful, original, and add significant value to the topic.

(continued on next page)

(continued)

Aspects of argumentation quality	Definition	Score				
		1	2	3	4	5
Persuasiveness	The arguments presented are convincing.	The arguments presented are not at all convincing.	Some of the arguments presented are somewhat convincing, but overall, the essay is unconvincing.	The arguments presented are generally convincing but may lack depth or insight.	The arguments presented are highly convincing, with well-supported evidence and insights.	The arguments presented are exceptionally persuasive, with powerful, nuanced reasoning, and evidence
Addressing the opposing view	The essay addresses the opposing view (s) adequately.	The essay does not address any opposing views.	The essay acknowledges some opposing views, but the response is weak or inadequate.	The essay addresses some opposing views adequately but could have been more thorough.	The essay addresses opposing views comprehensively and provides a thoughtful response.	The essay addresses opposing views exceptionally well, providing a nuanced and insightful response that strengthens the argument.

Appendix E. Coding scheme and results of students' chat histories with ChatGPT-powered NPCs

Theme	Talk move	No. of instances (%)	Definition	Example
Elicitation	Elicit unknown information	673 (51 %)	Gather factual details from the NPCs	<ul style="list-style-type: none"> - Where were you on the weekend of Sarah's death? - Did you have any conflict recently? - Can anybody testify for you about your being at home? - How's her attitude when talking about this with you?
	Elicit further details	178 (14 %)	Prompt the NPCs to elaborate on previous responses to clarify a viewpoint and gather additional evidence	<ul style="list-style-type: none"> - Why? - Why did you hurt Sarah? - Are you sure? - So you have a very good impression on her. Is that correct? - You are lying. - It's not true. - Answer it. It is important. - I have to know everything. I think it is relevant.
	Elicit attitudes and opinions	144 (11 %)	Elicit the NPCs' attitudes and opinions	<ul style="list-style-type: none"> - What do you think of her new hair style? - Do you regret it?
	Elicit reasoning	92 (7 %)	Prompt the NPCs to provide reasoning behind their actions or statements	<ul style="list-style-type: none"> - Why? - Why did you hurt Sarah? - Are you sure? - So you have a very good impression on her. Is that correct? - You are lying. - It's not true. - Answer it. It is important. - I have to know everything. I think it is relevant.
Confrontation	Question	29 (2 %)	Question the credibility of the NPC's statements	<ul style="list-style-type: none"> - I suspect your wife's leaving may cause some emotional damage to your daughter. - You seem to be hiding something. - I found out that you told Sarah's dad that dying hair is not good. - According to Sarah's father, you were not satisfied with her hair color?
	Disagree	23 (2 %)	Express disagreement with the NPCs' statements	<ul style="list-style-type: none"> - I am sorry for your loss. - I understand your concern and you are a great father.
	Persist	9 (0.6 %)	Persist in questioning to force the NPCs to confront the issues	<ul style="list-style-type: none"> - OK, let's change the topic. - I am sorry. Let's talk about the case.
	Speculate	8 (0.6 %)	Conjecture about potential plot developments based on the NPCs' responses	<ul style="list-style-type: none"> - I mean, have you apologized to her? - I am not saying she's the suspect. - You seem to want to say more. - You can say anything to me. Anything will help us to find justice for Sarah.
Disclosure	Disclose information	102 (8 %)	Reveal new findings or evidence to the NPCs	
Bonding	Establish emotional connections	28 (2 %)	Establish trust and emotional connections with the NPCs	
	Make strategic concessions	8 (0.6 %)	Show flexibility and understanding to reduce defensiveness	
Clarification	Clarify oneself	14 (1 %)	Provide clarifications based on the NPCs' responses to improve understanding	
Attention	Active listening	3 (0.2 %)	Demonstrate attentiveness and reinforce that the NPCs' thoughts are valued	

Appendix F. Summary of the quantitative results (RQ1, RQ2, and RQ3)

RQ1 ANCOVA analysis results of intrinsic motivation							
Affective factors	Variables	Groups	M (SD)	Adjusted M (SE)	F (df)	Sig.	partial η^2
	Enjoyment	EG	6.307 (0.791)	6.296 (0.157)	13.981 (72)	<0.001	0.165 (L)
		CG	5.474 (1.016)	5.484 (0.148)			
	Value	EG	6.048 (1.042)	5.944 (0.152)	10.214 (72)	0.002	0.126 (M-L)
		CG	5.171 (0.908)	5.264 (0.144)			
	Autonomy	EG	5.981 (0.963)	5.934 (0.161)	5.580 (72)	0.021	0.073 (M-L)
		CG	5.368 (1.017)	5.409 (0.152)			
Independent sample t-test results of situational interest							
	Variables	Groups	M (SD)		t (df)	Sig.	Cohen's d
	Attention quality	EG	6.457 (0.767)		2.610 (72)	0.011	0.608 (M-L)
		CG	6.009 (0.711)				
	Exploration intention	EG	6.352 (0.690)		2.380 (72)	0.020	0.544 (M-L)
		CG	5.923 (0.843)				
	Sense of challenge	EG	3.390 (1.209)		-0.162 (72)	0.872	/
		CG	3.436 (1.197)				
RQ2	Independent sample t-test results of cognitive load and effort regulation						
Cognitive factors	Variables	Groups	M (SD)		t (df)	Sig.	Cohen's d
	Intrinsic load	EG	2.133 (1.073)		-1.806 (72)	0.075	/
		CG	2.556 (0.938)				
	Extraneous load	EG	2.243 (1.109)		-1.523 (72)	0.130	/
		CG	2.647 (1.157)				
	Effort regulation	EG	6.029 (0.818)		2.237 (72)	0.028	0.521 (M-L)
		CG	5.573 (0.924)				
RQ3	Independent sample t-test results of writing performance						
Writing performance	Variables	Groups	M (SD)		t (df)	Sig.	Cohen's d
	Organization	EG	3.735 (0.448)		-1.043 (64.526)	0.301	/
		CG	3.838 (0.374)				
	Sufficiency	EG	3.559 (0.504)		1.526 (69)	0.131	/
		CG	3.378 (0.492)				
	Clarity	EG	3.735 (0.448)		2.438 (68.917)	0.017	0.576 (M-L)
		CG	3.459 (0.505)				
	Elaboration	EG	3.529 (0.507)		4.095 (67.188)	<0.001	0.976 (L)
		CG	3.054 (0.468)				
	Relevance	EG	4.647 (0.485)		-1.268 (65.417)	0.209	/
		CG	4.784 (0.417)				
	Persuasiveness	EG	3.676 (0.475)		2.343 (69)	0.022	0.557 (M-L)
		CG	3.405 (0.498)				
	Addressing opposing views	EG	2.794 (0.410)		3.074 (68.010)	0.003	0.724 (M-L)
		CG	2.459 (0.505)				
	Overall performance	EG	25.677 (2.727)		2.093 (69)	0.040	0.497 (S-M)

Note. S denotes a small effect size, and M denotes a medium effect size, while L denotes a large effect size. S-M denotes a small to medium effect size, while M-L denotes a medium to large effect size.

Data availability

Data will be made available on request.

References

- Abdelghani, R., Wang, Y. H., Yuan, X., Wang, T., Lucas, P., Sauzéon, H., & Oudeyer, P. Y. (2024). GPT-3-driven pedagogical agents to train children's curious question-asking skills. *International Journal of Artificial Intelligence in Education*, 34(2), 483–518. <https://doi.org/10.1007/s40593-023-00340-7>
- Adams, E. (2013). In *Fundamentals of game design* (3rd ed.) New Riders.
- Ahlquist, J., & Novak, J. (2008). Game development essentials: Game artificial intelligence. Clifton Park, NY: Thomson Del-mar Learning.
- Anand, A. (2023). Exploring the applications and limitations of large language models: A focus on ChatGPT in virtual NPC interactions. Drexel University [unpublished master's thesis].
- Andersen, M. S., & Makransky, G. (2021). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning*, 37(1), 183–196. <https://doi.org/10.1111/jcal.12478>.
- Bado, N. (2022). Game-based learning pedagogy: A review of the literature. *Interactive Learning Environments*, 30(5), 936–948. <https://doi.org/10.1080/10494820.2019.1683587>
- Barz, N., Benick, M., Dörrenbächer-Ulrich, L., & Perels, F. (2024). The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis. *Review of Educational Research*, 94(2), 193–227. <https://doi.org/10.3102/00346543231167795>
- Beck, D., Morgado, L., & O'Shea, P. (2023). Educational practices and strategies with immersive learning environments: Mapping of reviews for using the metaverse. *IEEE Transactions on Learning Technologies*, 17, 319–341. <https://doi.org/10.1109/TLT.2023.3243946>

- Bosch, C. (2024). Assessing the psychometric properties of the intrinsic motivation inventory in blended learning environments. *Journal of Education and e-Learning Research*, 11(2), 263–271. <https://doi.org/10.20448/jeelr.v1i2.5468>
- Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18(1), 1–21.
- Chan, H. Y., Liu, S. W., & Hou, H. T. (2023). Interacting with real-person non-player characters to learn history: Development and playing behavior pattern analysis of a remote scaffolding-based situated educational game. *Interactive Learning Environments*, 32(7), 3892–3912. <https://doi.org/10.1080/10494820.2023.2192745>
- Chen, A., Darst, P. W., & Pangrazi, R. P. (1999). What constitutes situational interest? Validating a construct in physical education. *Measurement in Physical Education and Exercise Science*, 3(3), 157–180.
- Chen, H. L., & Wu, C. T. (2021). A digital role-playing game for learning: Effects on critical thinking and motivation. *Interactive Learning Environments*, 31(5), 3018–3030. <https://doi.org/10.1080/10494820.2021.1916765>
- Chen, Y. C., & Hou, H. T. (2024). A mobile contextualized educational game framework with ChatGPT interactive scaffolding for employee ethics training. *Journal of Educational Computing Research*, 62(7), 1737–1762. <https://doi.org/10.1177/07356331241268505>
- Chen, A., Jia, J., Li, Y., & Fu, L. (2024). Investigating the effect of role-play activity with GenAI agent on EFL students' speaking performance. *Journal of Educational Computing Research*, 1–27. <https://doi.org/10.1177/07356331241299058>
- Cheng, K. H., Lee, S. W. Y., & Hsu, Y. T. (2022). The roles of epistemic curiosity and situational interest in students' attitudinal learning in immersive virtual reality environments. *Journal of Educational Computing Research*, 61(2), 494–519. <https://doi.org/10.1177/07356331221121284>
- Chiù, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments*, 32(7), 3240–3256. <https://doi.org/10.1080/10494820.2023.2172044>
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press, Inc.
- Cohen, S. D. (2011). The art of public narrative: Teaching students how to construct memorable anecdotes. *Communication Teacher*, 25(4), 197–204. <https://doi.org/10.1080/17404622.2011.601726>
- Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*, 56, 815–822. <https://psycnet.apa.org/doi/10.1037/0022-3514.56.5.815>
- Daviault, C. (2012). Does game playing experience have an impact on the player-PNPC relationship? *Bulletin of Science, Technology & Society*, 32(6), 441–446. <https://doi.org/10.1177/0270467612469075>
- de Almeida, J. L. F., & dos Santos Machado, L. (2021). Design requirements for educational serious games with focus on player enjoyment. *Entertainment Computing*, 38, Article 100413. <https://doi.org/10.1016/j.jentcom.2021.100413>
- De Smedt, F., Merchie, E., Barendse, M., Rosseel, Y., De Naeghel, J., & Van Keer, H. (2018). Cognitive and motivational challenges in writing: Studying the relation with writing performance across students' gender and achievement level. *Reading Research Quarterly*, 53(2), 249–272. <https://doi.org/10.1002/rq.193>
- Deci, E. L., Eghbari, H., Patrick, B. C., & Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, 62, 119–142. <https://doi.org/10.1111/j.1467-6494.1994.tb00797.x>
- Dickey, M. D. (2011). Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology*, 42(3), 456–469. <https://doi.org/10.1111/j.1467-8535.2009.01032.x>
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications Ltd.
- Gao, Y., Hu, Z., & Liu, J. (2024). The potential and mechanism of artificial intelligence driven non-player characters in video games for anxiety management. In *2024 international conference on artificial intelligence and communication* (pp. 353–362). Atlantis Press.
- Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*, 51(12), 2629–2633. <https://doi.org/10.1007/s10439-023-03272-4>
- Guo, M., & Liu, T. (2024). ChatGPT & RPG game: A new option for student to study with artificial intelligence. In *2024 3rd international conference on artificial intelligence, internet and digital economy* (pp. 11–22). Atlantis Press.
- Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, 203, Article 104862. <https://doi.org/10.1016/j.compedu.2023.104862>
- Guo, K., Zhong, Y., Zainuddin, Z., & Chu, S. K. W. (2023). Applying game-related methods in the writing classroom: A scoping review. *Education and Information Technologies*, 29(4), 4481–4504. <https://doi.org/10.1007/s10639-023-11998-w>
- Güvendir, E., & Uzun, K. (2023). L2 writing anxiety, working memory, and task complexity in L2 written performance. *Journal of Second Language Writing*, 60, Article 101016. <https://doi.org/10.1016/j.jsw.2023.101016>
- Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, 22(10), 390–392. <https://doi.org/10.1016/j.cub.2012.02.040>
- Hair, J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, New Jersey: Pearson Educational International.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4
- Hilderbrand, R., Romershausen, S., & Lee, N. (2024). *Encyclopedia of computer graphics and games*. Cham: Springer.
- Hu, L., & Chen, G. (2024). A systematic review and meta-analysis of productive peer talk moves. *Journal of Behavioral Education*, 33, 798–830. <https://doi.org/10.1007/s10864-023-09513-9>
- Huang, J. (2024). Generating dynamic and lifelike NPC dialogs in role-playing games using large language model. LUT University [unpublished bachelor's thesis].
- Huang, J., & Huang, K. (2023). ChatGPT in gaming industry. In A. I. Beyond (Ed.), *ChatGPT, Web3, and the business landscape of tomorrow* (pp. 243–269). Cham: Springer Nature Switzerland.
- Hubale, A., Moorin, J., & Khuman, A. S. (2021). Artificial intelligence in FPS games: NPC difficulty effects on gameplay. In J. Carter, F. Chiclana, A. S. Khuman, & T. Chen (Eds.), *Fuzzy logic* (pp. 165–190). Cham: Springer. https://doi.org/10.1007/978-3-03-66474-9_11
- Huber, S. E., Kili, K., Nebel, S., Ryan, R. M., Sailer, M., & Ninaus, M. (2024). Leveraging the potential of large language models in education through playful and game-based learning. *Educational Psychology Review*, 36(25). <https://doi.org/10.1007/s10648-024-09868-z>
- Iarovoii, D., Hebblewhite, R., & Teh, P. L. (2024). AI's influence on non-player character dialogue and gameplay experience. In *Science and information conference* (pp. 76–92). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-62281-6_6
- Iuppa, N., & Borst, T. (2007). *Story and Simulations for Serious Games: Tales from the trenches*. Burlington: Focal Press.
- Jaguš, T., Botički, I., & So, H. J. (2018). Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers & Education*, 125, 444–457. <https://doi.org/10.1016/j.compedu.2018.06.022>
- Karaosmanoglu, S., Fittschen, E. L., Eyalicis, H., Kraus, D., Nickelmann, H., Tomko, A., & Steinicke, F. (2024). Language of Zelda: Facilitating language learning practices using ChatGPT. In *Extended abstracts of the CHI conference on human factors in computing systems* (pp. 1–5). <https://doi.org/10.1145/3613905.3648107>
- Kasapakis, V., Dzardanova, E., & Agelada, A. (2023). Virtual reality in education: The impact of high-fidelity nonverbal cues on the learning experience. *Computers & Education: X Reality*, 2, Article 100020. <https://doi.org/10.1016/j.cexr.2023.100020>
- Kaur, J. (2017). Ambiguity related misunderstanding and clarity enhancing practices in ELF communication. *Intercultural Pragmatics*, 14(1), 25–47. <https://doi.org/10.1515/ip-2017-0002>
- Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London: Routledge.
- Kua, M. Y., Dolo, F. X., Suparmi, N. W., Gelu, A., Dhena, G. V., & Meme, Y. O. (2024). Interactive articulate storyline application based on real world problem and local Ngada wisdom. *Journal of Research in Instructional*, 4(1), 77–90. <https://doi.org/10.30862/jri.v4i1.337>
- Lam, Y. W., Hew, K. F., & Chiu, K. F. (2018). Improving argumentative writing: Effects of a blended learning approach and gamification. *Language, Learning and Technology*, 22(1), 97–118.
- Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming*, 49(3), 315–337. <https://doi.org/10.1177/1046878118774385>
- Lawrence, A. M., & Sherry, M. B. (2021). How feedback from an online video game teaches argument writing for environmental action. *Journal of Literacy Research*, 53(1), 29–52. <https://doi.org/10.1177/1086296X20986598>

- Lee, K. M. (2004). Presence, explicated. *Communication Theory*, 14(1), 27–50. <https://doi.org/10.1111/j.1468-2885.2004.tb00302.x>
- Lee, Y. H. (2014). Inquiry-based teaching in second and foreign language pedagogy. *Journal of Language Teaching and Research*, 5(6), 1236–1244. <https://doi.org/10.4304/jltr.5.6.1236-1244>. Academy Publisher Manufactured in Finland.
- Lee, S. M. (2019). Her Story or their own stories? Digital game-based learning, student creativity, and creative writing. *ReCALL*, 31(3), 238–254. <https://doi.org/10.1017/S0958344019000028>
- Lester, D., Skulmoski, G. J., Fisher, D. P., Mehrotra, V., Lim, I., Lang, A., & Keogh, J. W. (2023). Drivers and barriers to the utilisation of gamification and game-based learning in universities: A systematic review of educators' perspectives. *British Journal of Educational Technology*, 54(6), 1748–1770. <https://doi.org/10.1111/bjet.13311>
- Li, M., Gao, Q., & Yu, T. (2023). Kappa statistic considerations in evaluating inter-rater reliability between two raters: Which, when and context matters. *BMC Cancer*, 23, 1–5. <https://doi.org/10.1186/s12885-023-11325-z>
- Liew, T. W., Siradjy, Y., Tan, S. M., Roedavan, R., Khan, M. T. I., & Pudjoatmodjo, B. (2024). Game-changer NPCs: Leveling-up technology acceptance and flow in a digital learning quest. *International Journal of Human-Computer Interaction*, 1–22. <https://doi.org/10.1080/10447318.2024.2344917>
- Lin, C. J., Hwang, G. J., Fu, Q. K., & Chen, J. F. (2018). A flipped contextual game-based learning approach to enhancing EFL students' English business writing performance and reflective behaviors. *Journal of Educational Technology & Society*, 21(3), 117–131. <https://www.jstor.org/stable/26458512>
- Luo, W., He, H., Liu, J., Berson, I. R., Berson, M. J., Zhou, Y., & Li, H. (2023). Aladdin's genie or Pandora's box for early childhood education? Experts chat on the roles, challenges, and developments of ChatGPT. *Early Education & Development*, 35(1), 96–113. <https://doi.org/10.1080/10409289.2023.2214181>
- Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3), 937–958. <https://doi.org/10.1007/s10648-020-09586-2>
- Makridakis, S., Petropoulos, F., & Kang, Y. (2023). Large language models: Their success and impact. *Forecasting*, 5(3), 536–549. <https://doi.org/10.3390/forecast5030030>
- Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 4, 333–369. [https://doi.org/10.1016/S0364-0213\(81\)80017-1](https://doi.org/10.1016/S0364-0213(81)80017-1)
- Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow, & M. J. Farr (Eds.), *Aptitude, learning, and instruction* (Vol. 3, pp. 223–253). Hillsdale: Lawrence Erlbaum Associates.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1
- Mayer, R. E., & Moreno, R. (1998). A cognitive theory of multimedia learning: Implications for design principles. *Journal of Educational Psychology*, 91(2), 358–368. Retrieved from <https://esoluk.co.uk/calling/pdf/chi.pdf>.
- McCluggage, W. G., Walsh, M. Y., Thornton, C. M., Hamilton, P. W., Date, A., Caughley, L. M., & Bharucha, H. (1998). Inter-and intra-observer variation in the histopathological reporting of cervical squamous in traepithelial lesion susing a modified Bethesda grading system. *BJOG: An International Journal of Obstetrics & Gynaecology*, 105(2), 206–210. <https://doi.org/10.1111/j.1471-0528.1998.tb10054.x>
- Moon, J., Lee, U., Koh, J., Jeong, Y., Lee, Y., Byun, G., & Lim, J. (2024). Generative artificial intelligence in educational game design: Nuanced challenges, design implications, and future research. *Technology, Knowledge, and Learning*. <https://doi.org/10.1007/s10758-024-09756-z>.
- Nagarkar, S. B. (2024). *Improving realism and interactivity in games: A study of AI-integrated non player characters (NPCs)*. Universitat Politècnica de Catalunya [unpublished master's thesis].
- Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. In J. L. Shane, & C. R. Snyder (Eds.), *Handbook of positive psychology* (pp. 195–206). Oxford Handbooks Online. <https://doi.org/10.1093/oxfordhb/9780195187243.013.0018>
- Naul, E., & Liu, M. (2020). Why story matters: A review of narrative in serious games. *Journal of Educational Computing Research*, 58(3), 687–707. <https://doi.org/10.1177/0735633119859904>
- Novak, E. (2015). A critical review of digital storyline-enhanced learning. *Educational Technology Research & Development*, 63(3), 431–453. <https://doi.org/10.1007/s11423-015-9372-y>
- Obedkov, E. (2023). NetEase to add game version of ChatGPT to Justice Online Mobile for dialogue generation and unique reactions. *GAME WORLD OBSERVERS*. <https://gamerworldobserver.com/2023/02/16/netease-chatgpt-justice-online-mobile-dialogue-generation>.
- Pan, M., Lai, C., & Guo, K. (2024). Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. *The Internet and Higher Education*, , Article 100991. <https://doi.org/10.1016/j.iheduc.2024.100991>
- Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, 62(4), 625–633. <https://doi.org/10.1037/0022-3514.62.4.625>
- Pasupuleti, R. S., & Thiyyagura, D. (2024). An empirical evidence on the continuance and recommendation intention of ChatGPT among higher education students in India: An extended technology continuance theory. *Education and Information Technologies*, 29(14), 17965–17985. <https://doi.org/10.1007/s10639-024-12573-7>
- Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences*, 14(5), 2074. <https://doi.org/10.3390/app14052074>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology*, 124(1), 145–178. <https://doi.org/10.1037/pspp0000448>
- Perera, P., & Lankathilaka, M. (2023). AI in higher education: A literature review of ChatGPT and guidelines for responsible implementation. *International Journal of Research and Innovation in Social Science*, VII(VI), 306–314. <https://doi.org/10.47772/IJRIS.2023.7623>
- Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Computers & Education*, 179, Article 104429. <https://doi.org/10.1016/j.compedu.2021.104429>
- Pfau, A., Pollio, C., & Xu, Y. (2023). Exploring the potential of ChatGPT in assessing L2 writing accuracy for research purposes. *Research Methods in Applied Linguistics*, 2 (3), Article 100083. <https://doi.org/10.1016/j.rmal.2023.100083>
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, Michigan: The University of Michigan.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>
- Sánchez, L., & Sunesson, M. (2023). Grasping the effects of storyline complexity, task structure and proficiency in narrative writing performance. *Journal of Second Language Writing*, 60, Article 100986. <https://doi.org/10.1016/j.jslw.2023.100986>
- Solhi, M., Derakhshan, A., Pawlak, M., & Ünsal-Görkemoglu, B. (2024). Exploring the interplay between EFL learners' L2 writing boredom, writing motivation, and boredom coping strategies. *Language Teaching Research*. <https://doi.org/10.1177/1362168824123917>
- Song, Y., & Sparks, J. R. (2019a). Building a game-enhanced formative assessment to gather evidence about middle school students' argumentation skills. *Educational Technology Research & Development*, 67, 1175–1196. <https://doi.org/10.1007/s11423-018-9637-3>
- Song, Y., & Sparks, J. R. (2019b). Measuring argumentation skills through a game-enhanced scenario-based assessment. *Journal of Educational Computing Research*, 56 (8), 1324–1344. <https://doi.org/10.1177/0735633117740605>
- Song, Y., Wu, K., & Ding, J. (2024). Developing an immersive game-based learning platform with generative artificial intelligence and virtual reality technologies – "LearningverseVR". *Computers & Education: X Reality*, 4, Article 100069. <https://doi.org/10.1016/j.cexr.2024.100069>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Stojanov, A., Liu, Q., & Koh, J. H. L. (2024). University students' self-reported reliance on ChatGPT for learning: A latent profile analysis. *Computers & Education: Artificial Intelligence*, 6, Article 100243. <https://doi.org/10.1016/j.caiei.2024.100243>
- Sweller, J. (2011). Cognitive load theory. In J. P. Mestre, & B. H. Ross (Eds.), *The psychology of learning and motivation* (pp. 37–76). Elsevier Academic Press.

- Tai, A. M. Y., Meyer, M., Varidel, M., Prodan, A., Vogel, M., Iorfino, F., & Krausz, R. M. (2023). Exploring the potential and limitations of ChatGPT for academic peer-reviewed writing: Addressing linguistic injustice and ethical concerns. *Journal of Academic Language and Learning*, 17(1), T16–T30.
- Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System*, Article 103528. <https://doi.org/10.1016/j.system.2024.103528>
- Uludaglı, M.Ç., & Oğuz, K. (2023). Non-player character decision-making in computer games. *Artificial Intelligence Review*, 56(12), 14159–14191. <https://doi.org/10.1007/s10462-023-10491-7>
- Wale, B. D., & Bishaw, K. S. (2020). Effects of using inquiry-based learning on EFL students' critical thinking skills. *Asian-Pacific Journal of Second and Foreign Language Education*, 5, 1–14. <https://doi.org/10.1186/s40862-020-00090-2>
- Wang, M., Zhang, D., Zhu, J., & Gu, H. (2025). Effects of incorporating a large language model-based adaptive mechanism into contextual games on students' academic performance, flow experience, cognitive load and behavioral patterns. *Journal of Educational Computing Research*. <https://doi.org/10.1177/07356331251321719>
- Warpefelt, H., & Verhagen, H. (2015). Towards an updated typology of non-player character roles. In *Proceedings of the international conference on game and entertainment technologies* (pp. 1–9). International Association for the Development of the Information Society.
- Wehbe, R. R., Lank, E., & Nacke, L. E. (2017). Left them 4 dead: Perception of humans versus non-player character teammates in cooperative gameplay. In *Proceedings of the 2017 conference on designing interactive systems* (pp. 403–415).
- Xiao, Y., & Hew, K. F. (2024). Personalized gamification versus one-size-fits-all gamification in fully online learning: Effects on student motivational, behavioral and cognitive outcomes. *Learning and Individual Differences*, 113, Article 102470. <https://doi.org/10.1016/j.lindif.2024.102470>
- Yang, Y., Lin, J., Chen, T., Lin, S., Chen, J., Miao, W., Wei, W., Sun, H. C., Sun, J., & Gu, C. (2022). A study on the effects of perceived playfulness and emotional support in interactive learning environments for German language acquisition—A role-playing library system case. *Frontiers in Psychology*, 13, Article 1073985. <https://doi.org/10.3389/fpsyg.2022.1073985>
- Yuan, Q., & Gao, Q. (2024). Being there, and being together: Avatar appearance and peer interaction in VR classrooms for video-based learning. *International Journal of Human-Computer Interaction*, 40(13), 3313–3333. <https://doi.org/10.1080/10447318.2023.2189818>
- Zare, J., Al-Issa, A., & Madiseh, F. R. (2025). Interacting with ChatGPT in essay writing: A study of L2 learners' task motivation. *ReCALL*, 1–18. <https://doi.org/10.1017/S0958344025000035>
- Zhang, J., Liu, X. L., So, M., & Reder, L. M. (2020). Familiarity acts as a reduction in objective complexity. *Memory & Cognition*, 48, 1376–1387. <https://doi.org/10.3758/s13421-020-01055-z>
- Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. *European Journal of Education*, 59(2), Article e12599. <https://doi.org/10.1111/ejed.12599>
- Zhang, R., Zou, D., Cheng, G., & Xie, H. (2025). Flow in ChatGPT-based logic learning and its influences on logic and self-efficacy in English argumentative writing. *Computers in Human Behavior*, 162, Article 108457. <https://doi.org/10.1016/j.chb.2024.108457>
- Zhou, Y., & Wang, J. (2021). Psychometric properties of the MSLQ-B for adult distance education in China. *Frontiers in Psychology*, 12, Article 620564. <https://doi.org/10.3389/fpsyg.2021.620564>
- Zuo, T., Birk, M. V., Van der Spek, E. D., & Hu, J. (2023). The effect of fantasy on learning and recall of declarative knowledge in AR game-based learning. *Entertainment Computing*, 46, Article 100563. <https://doi.org/10.1016/j.entcom.2023.100563>