

RESEARCH ARTICLE

Open Access



# Empowering ChatGPT with guidance mechanism in blended learning: effect of self-regulated learning, higher-order thinking skills, and knowledge construction

Hsin-Yu Lee<sup>1</sup> , Pei-Hua Chen<sup>1</sup>, Wei-Sheng Wang<sup>1</sup> , Yueh-Min Huang<sup>1\*</sup>  and Ting-Ting Wu<sup>2</sup> 

\*Correspondence:  
huang@mail.ncku.edu.tw

<sup>1</sup> Department of Engineering  
Science, National Cheng Kung  
University, No. 1, University Road,  
Tainan 701, Taiwan, ROC

<sup>2</sup> Graduate School  
of Technological and Vocational  
Education, National Yunlin  
University of Science  
and Technology, Yunlin 640,  
Taiwan, ROC

## Abstract

In the evolving landscape of higher education, challenges such as the COVID-19 pandemic have underscored the necessity for innovative teaching methodologies. These challenges have catalyzed the integration of technology into education, particularly in blended learning environments, to bolster self-regulated learning (SRL) and higher-order thinking skills (HOTS). However, increased autonomy in blended learning can lead to learning disruptions if issues are not promptly addressed. In this context, OpenAI's ChatGPT, known for its extensive knowledge base and immediate feedback capability, emerges as a significant educational resource. Nonetheless, there are concerns that students might become excessively dependent on such tools, potentially hindering their development of HOTS. To address these concerns, this study introduces the Guidance-based ChatGPT-assisted Learning Aid (GCLA). This approach modifies the use of ChatGPT in educational settings by encouraging students to attempt problem-solving independently before seeking ChatGPT assistance. When engaged, the GCLA provides guidance through hints rather than direct answers, fostering an environment conducive to the development of SRL and HOTS. A randomized controlled trial (RCT) was employed to examine the impact of the GCLA compared to traditional ChatGPT use in a foundational chemistry course within a blended learning setting. This study involved 61 undergraduate students from a university in Taiwan. The findings reveal that the GCLA enhances SRL, HOTS, and knowledge construction compared to traditional ChatGPT use. These results directly align with the research objective to improve learning outcomes through providing guidance rather than answers by ChatGPT. In conclusion, the introduction of the GCLA has not only facilitated more effective learning experiences in blended learning environments but also ensured that students engage more actively in their educational journey. The implications of this study highlight the potential of ChatGPT-based tools in enhancing the quality of higher education, particularly in fostering essential skills such as self-regulation and HOTS. Furthermore, this research offers insights regarding the more effective use of ChatGPT in education.

**Keywords:** Higher education, Blended learning, Self-regulated learning, ChatGPT, Higher-order thinking skills, Knowledge construction

## Introduction

The onset of COVID-19 accelerated the adoption of blended learning in higher education, merging traditional and online teaching methodologies. This approach has been crucial in meeting the complex demands of university courses, ensuring educational continuity, facilitating access to resources, enabling virtual collaboration, and combining online and in-person problem-solving strategies (Mali & Lim, 2021; Menon & Azam, 2021). Moreover, blended learning emphasizes the importance of self-regulated learning (SRL), as proposed in model by Zimmerman (1990), essential for students in complex academic environments. Higher-order thinking skills (HOTS) are also vital, fostering critical analysis, idea synthesis, and innovative problem-solving, skills imperative in academia and professional practice (Hwang et al., 2019). However, the autonomous nature of blended learning can pose challenges in maintaining consistent teacher support, which may impact the development of students' SRL and HOTS (Rasheed et al., 2020).

The introduction of ChatGPT by OpenAI in late 2022 has transformed the blended learning landscape. Incorporating ChatGPT offers both opportunities and challenges (Labadze et al., 2023). It provides immediate feedback and a plethora of information, enhancing learning efficiency and allowing for personalized educational paths (Stojanov, 2023; Wu et al., 2023a). Nevertheless, it could also foster a dependency that might impede the development of HOTS. The ease of accessing information and problem-solving assistance through tools like ChatGPT could discourage students from engaging in thorough thinking or independent problem-solving (Chan & Hu, 2023; Ding et al., 2023). White and Gunstone (2014) underscored the importance of prediction in knowledge acquisition, suggesting that learners should initially hypothesize and develop their own solutions and viewpoints, then validate their assumptions through observation and analysis.

Although ChatGPT's capabilities have revolutionized pedagogical methods and learning dynamics at various educational levels, including higher education, the absence of structured guidelines may limit the development of HOTS, potentially affecting the efficacy of blended learning. To address this, this study introduces the guidance-based ChatGPT-assisted learning aid (GCLA), designed specifically for higher education. The GCLA requires students to articulate their initial thoughts and perspectives before consulting ChatGPT. Then, they iteratively refine their responses using ChatGPT's feedback until a well-substantiated answer is developed. This approach promotes deeper engagement with the subject matter, thus enhancing knowledge construction and educational outcomes in higher education blended learning environments.

This study aims to tackle the critical issue of students' excessive reliance on ChatGPT in higher education blended learning settings, and its impact on their SRL, HOTS, and knowledge construction. It also evaluates the effectiveness of the GCLA, as compared to traditional ChatGPT usage, in enhancing these aspects among learners. The study seeks to answer the following research questions:

1. How does the GCLA, compared to traditional ChatGPT use, affect the SRL of higher education students in blended learning environments?
2. How does the GCLA, compared to traditional ChatGPT use, influence the development of HOTS in these students?

3. How does the GCLA, compared to traditional ChatGPT use, impact knowledge construction in higher education students in these environments?

## Literature review

### Self-regulated learning in blended learning

Self-regulated learning (SRL) is a framework that empowers learners to autonomously steer their educational paths. Originally delineated by Zimmerman (1990), the SRL model is segmented into three pivotal phases: forethought, performance, and self-reflection. The first phase is the forethought phase, which is the earliest and most important phase in the learning process. Learners first analyze the learning task, set goals and develop methods to achieve them. During this process, learners often need to adjust their learning motivation to ensure they have enough motivation to complete the learning task. The second phase is the performance phase, which is the phase where learners actually engage in learning activities. Learners need to actively participate in learning activities, monitor and adjust their own learning behavior to ensure that they achieve the expected learning goals. The final phase is the self-reflection phase, which is the last phase in the learning process. Learners need to review and evaluate their own effectiveness and learning outcomes to gain more learning experience and knowledge (Pintrich, 2000; Zimmerman, 2008).

In blended learning environments, which integrate face-to-face and online instruction, learners are granted the autonomy to tailor their learning objectives and pace (Rasheed et al., 2020; Snodin, 2013). These environments are well-aligned with SRL, particularly facilitating the performance phase by providing opportunities for learners to practice and apply their strategies. Students adeptly navigate their learning journeys, employing the most suitable strategies and resources for their needs (Wu et al., 2023a). Within such settings, as noted by Rasheed et al. (2020), SRL necessitates significant self-discipline and initiative, especially online where learners choose their engagement levels—a dynamic essential for the self-reflection phase, as discussed by Hood et al. (2015) and grounded in Zimmerman (1990)'s work.

Motivation, deeply intertwined with the forethought phase, drives individuals to engage fully with the SRL process and reach their goals. This link between motivation and goal-setting is supported by a robust body of research highlighting motivation's central role in self-regulation (Wu et al., 2023a; Zhu et al., 2020). Engagement, critical in the performance phase, stems from active involvement in learning tasks and is closely tied to the strategic application of learning techniques and interaction with educational content and peers. Research has shown that such engagement is vital in reinforcing SRL and educational dedication (Hershcovits et al., 2020; Wu et al., 2023a). Finally, self-efficacy, particularly relevant in the self-reflection phase, reflects one's confidence in successfully completing learning tasks and is strongly correlated with self-regulatory skills and academic outcomes. This sense of self-efficacy fosters the ability to embrace challenges and navigate the self-regulatory process effectively, a stance that is widely recognized in academic literature (Salah Dogham et al., 2022; Wu et al., 2023a).

In summary, motivation, engagement, and self-efficacy are vital pillars of SRL, corresponding to its phases and collectively nurturing learner autonomy and proactive behavior. These factors not only amplify the learners' intrinsic drive but also solidify their

learning proficiency, ensuring adept management of learning experiences within the rich contexts of blended learning environments.

### Higher-order thinking skills

In recent years, higher-order thinking skills (HOTS) have increasingly become a focal point in higher education on a global scale (Lu et al., 2021a, 2021b). The discourse in this sector has evolved to consider HOTS indispensable for navigating the complexities of modern society (Lu et al., 2021a, 2021b). These skills, which go beyond basic memory and comprehension, include advanced cognitive processes such as critical thinking, problem-solving, and creativity (Hwang et al., 2018).

In the realm of higher education, critical thinking is essential for students to objectively analyze and evaluate information, leading to informed decisions (Brookhart, 2010; Lu et al., 2021a, 2021b). This level of scrutiny encourages learners to not just passively accept information but to actively engage in questioning and appraising its validity and utility (Krathwohl, 2002).

Problem-solving, a key component of HOTS, is particularly relevant in higher education as it involves identifying complex issues, gathering and scrutinizing data, proposing potential solutions, and selecting the most effective ones (Hwang & Lai, 2017; Lu et al., 2021a, 2021b). In an evolving global landscape, the ability to address new and unprecedented challenges is paramount, equipping students for real-life and workplace scenarios where standard solutions may not apply.

Creativity, integral to HOTS, is championed in higher education as a means of thinking outside the norm and generating innovative, impactful ideas (Hwang & Lai, 2017; Lu et al., 2021a, 2021b). Encouraging creativity allows students to surpass conventional thinking, leading to flexible strategies in problem-solving and pioneering advances (Sternberg, 2003).

When these elements are integrated into the higher education experience, a rich and varied learning environment is created (Cheng et al., 2020). This not only nurtures the development of HOTS through educator and peer interaction but also promotes autonomous and personalized learning pathways (Chen et al., 2023; Jansen & Möller, 2022). Such integration is vital for preparing learners for the twenty-first century, emphasizing the importance of applying HOTS in everyday life and professional pursuits. Consequently, developing appropriate pedagogical approaches and tools to enhance higher-order thinking within the context of blended learning in higher education remains an area of significant interest and activity.

### ChatGPT in education

In recent years, the rapid evolution of large language models has profoundly influenced the domain of natural language processing (NLP). Foremost among these innovations, the Transformer model has surpassed earlier frameworks such as LSTM and RNN in prominence and efficacy (Vaswani et al., 2017). This pivotal shift catalyzed the development of notable pre-trained models like BERT (Ettinger, 2020) and GPT (Radford et al., 2018). In particular, GPT-3, with its unprecedented scale, set benchmarks by leveraging billions of parameters to capture intricate language nuances (Dale, 2021; Zhang & Li, 2021). Building on this foundation, GPT-3.5 introduced directive learning and

Reinforcement Learning from Human Feedback (RLHF) to further optimize model performance (Abramski et al., 2023; Wu et al., 2023a).

Built upon GPT-3.5, ChatGPT has been increasingly adopted within the educational sector. Investigations by Kasneci et al. (2023) elucidate how ChatGPT engenders engagement and fosters a more interactive learning paradigm. Moreover, Jeon and Lee (2023) scrutinized the synergy between ChatGPT and teachers, emphasizing their mutual complementarity in the educational arena and probing how ChatGPT might support teachers in classroom facilitation. Furthermore, the utility of ChatGPT extends to linguistic instruction, as illustrated by Kohnke et al. (2023). Its efficacy as a supplementary educational tool has been further underscored through self-assessment studies (Stojanov, 2023). Innovations like CILA, tailored for blended learning environments, harness the capabilities of ChatGPT to furnish students with precise, on-demand answers—a clear departure from the vast yet unspecific information of traditional search engines like Google (Wu et al., 2023a).

However, the unparalleled potential of ChatGPT in education is not without challenges (Adeshola & Adepoju, 2023). Prominent among these is the risk of students becoming overly reliant on its responses, which could potentially stifle their critical thinking and independent problem-solving acumen (Cooper, 2023). Such dependence could dilute the depth and richness of their educational experiences and compromise their acquisition of vital problem-solving techniques (Montenegro-Rueda et al., 2023). Consequently, it becomes imperative for educators to exercise prudence in integrating ChatGPT into curricula. Implementing comprehensive and pragmatic guidelines for its usage is essential to ensure that digital learning continues to evolve responsibly. With judicious management, Researchers can harness the full potential of ChatGPT while curtailing its potential pitfalls, fostering a robust, effective, and accountable educational landscape.

### **The design of guidance-based ChatGPT-assisted learning aid (GCLA)**

This study introduces GCLA, an innovative educational tool designed for blended learning environments, highlighting its potential applications and benefits. Birthed from a groundbreaking collaboration, it fuses ChatGPT's vast response capabilities with the unparalleled connectivity of Apple's Shortcuts. GCLA harnesses ChatGPT's comprehensive knowledge to proficiently address a broad array of student inquiries. This is enhanced by its smooth integration with Apple's Shortcuts, providing fluid interaction across all Apple devices and ensuring prompt, personalized responses. A key aspect of GCLA is its 'learning log file,' which records past inquiries to support reflective learning. Rather than giving direct answers, GCLA prompts students to formulate their own, offering insightful hints to aid problem-solving. This cultivates a deep engagement with the material, fostering HOTS. Figure 1 depicts the intricate workings of GCLA, showcasing its various components and their interconnections.

Figure 2 visually depicts the GCLA workflow, illustrating the interaction of learners with the tool. This study's objective is to underscore the unique attributes and advantages of GCLA as a revolutionary tool in blended learning environments. It provides comprehensive insights into its functionalities and the potential to revolutionize educational paradigms.

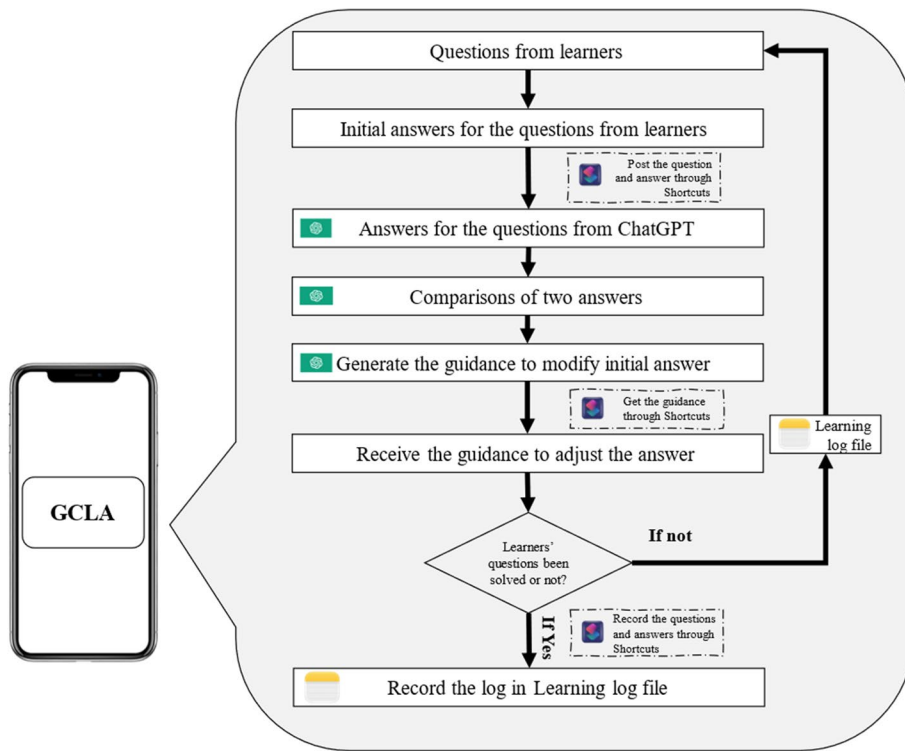


Fig. 1 The procedure of GCLA

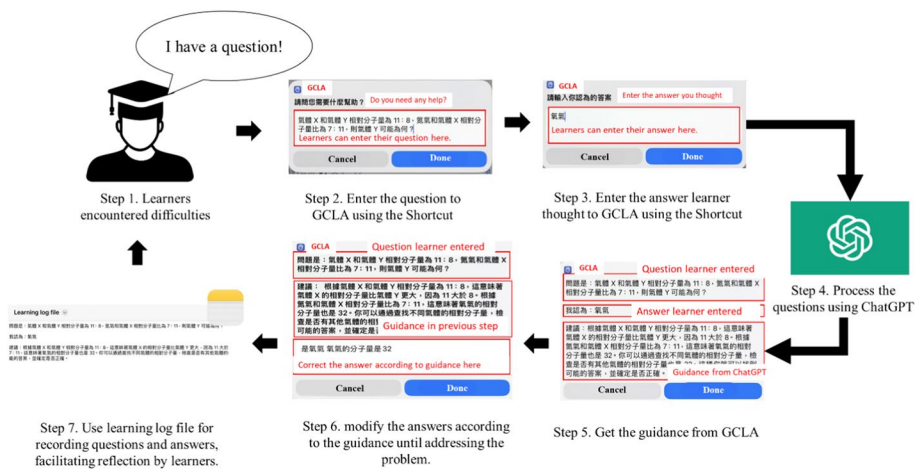
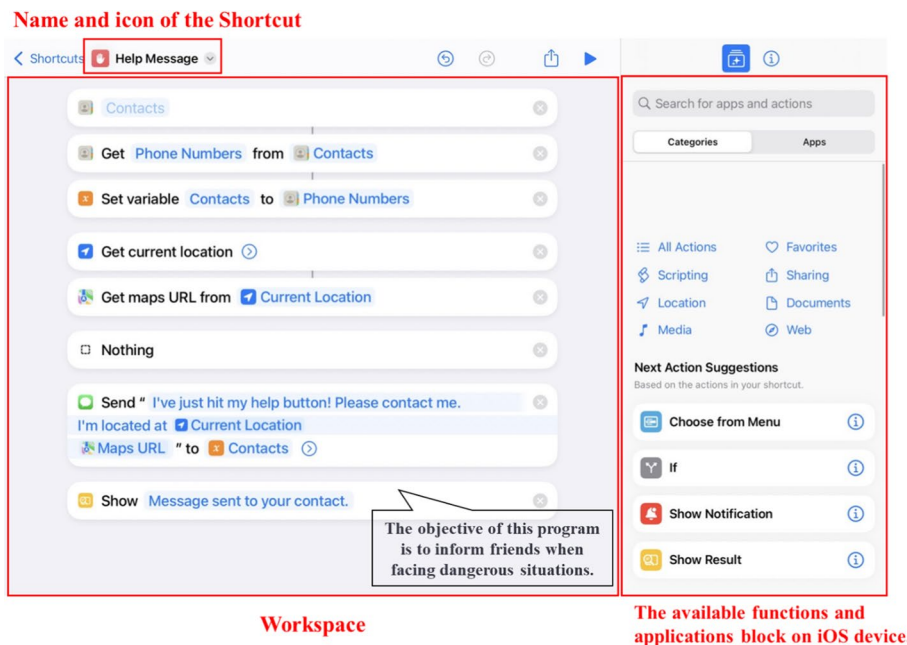


Fig. 2 The workflow of GCLA

### The implementation of GCLA

The GCLA represents an innovative step in educational assistance, fostering a more interactive relationship between learners and technology. Unlike traditional tools that merely provide answers, GCLA prompts users to first articulate their own solutions through Apple’s Shortcuts. This method ensures active engagement with the subject matter, fostering deeper comprehension and enhancing HOTS. The GCLA service,



**Fig. 3** The development interface of Shortcuts

**Table 1** The parameter of ChatGPT in GCLA

Parameter	Model	Max_tokens	Temperature	Presence_penalty
Value	gpt-3.5-turbo-16 k	4000	0.6	0.2

designed to run on Apple devices with iOS 12 or above, streamlines the learning process with automated workflows and personalized interactions.

Central to GCLA is Apple’s Shortcuts, a user-friendly platform that enables custom automation of tasks. The Shortcuts interface, illustrated in Fig. 3, allows learners to easily create workflows for a variety of functions, such as messaging, weather updates, or voice commands through Siri—Apple’s virtual assistant—all without needing extensive programming knowledge.

GCLA also incorporates the ChatGPT engine, specifically the "gpt-3.5-turbo-16 k" model, to manage queries. This study has set precise parameters for ChatGPT, detailed in Table 1. These settings include a 4000 max\_tokens limit for detailed responses, a temperature of 0.6 for varied replies, and a presence\_penalty of 0.2 to reduce repetition and promote novelty.

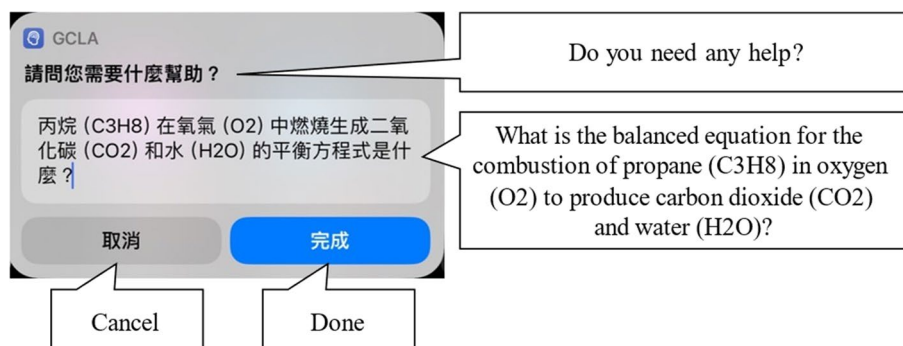
Learners wishing to use GCLA’s ChatGPT feature must obtain an authentication key from OpenAI’s official website. After registration, they can submit questions and their own proposed answers through the GCLA interface. This step is crucial as it compels the learner to engage with the content before receiving feedback. The input is processed by ChatGPT, which, once authenticated, provides customized guidance according to the preset parameters.

Additionally, GCLA includes a learning log file to record all questions and answers, supporting self-regulation and reflection. This log acts as a useful tool for revisiting material, aiding in concept recall and memory retention, and establishing a basis for post-class reflection. In essence, GCLA is an integrated learning solution that combines the adaptability of Apple's Shortcuts with ChatGPT's intelligent query processing, delivering a tool that not only demystifies complex learning but also considerably enriches the educational journey.

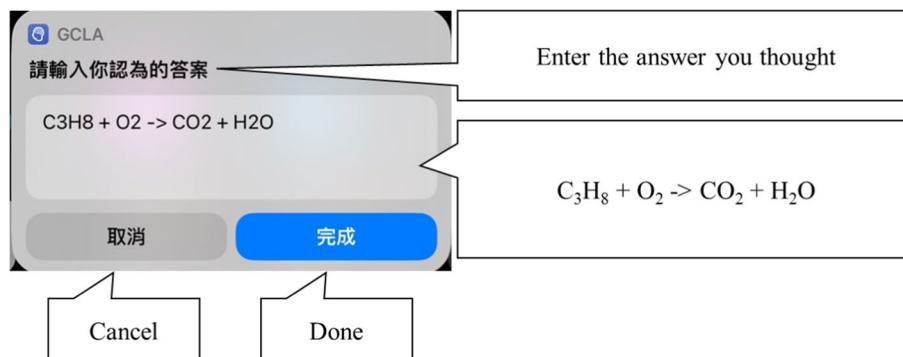
### The example of using GCLA

To demonstrate the practical application of GCLA (Guided Chemical Learning Assistant) and its response to reviewers' inquiries, this example showcases a student utilizing GCLA in their study of chemical reactions. Confronted with an unsolvable problem, the student employed GCLA for guidance and to deepen their understanding. The problem was: "What is the balanced equation for the combustion of propane ( $C_3H_8$ ) in oxygen ( $O_2$ ) to produce carbon dioxide ( $CO_2$ ) and water ( $H_2O$ )?" (refer to Fig. 4).

Upon accessing the GCLA application on their iPad, the student entered the problem and attempted an initial solution, drawing on prior knowledge. Their response was: " $C_3H_8 + O_2 \rightarrow CO_2 + H_2O$ " (see Fig. 5).



**Fig. 4** Inputting the problem into GCLA



**Fig. 5** Submitting an initial answer in GCLA



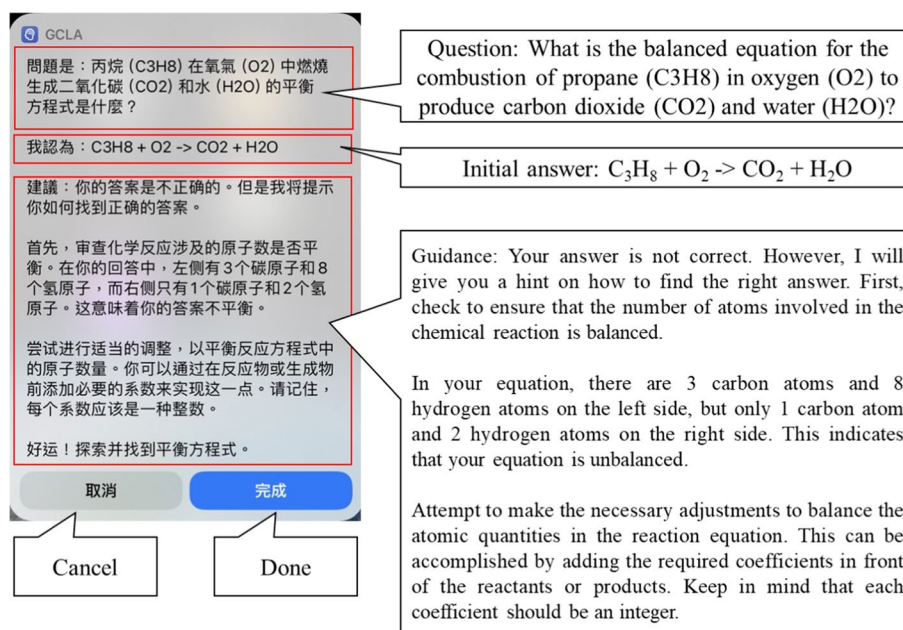


Fig. 6 The guidance from GCLA

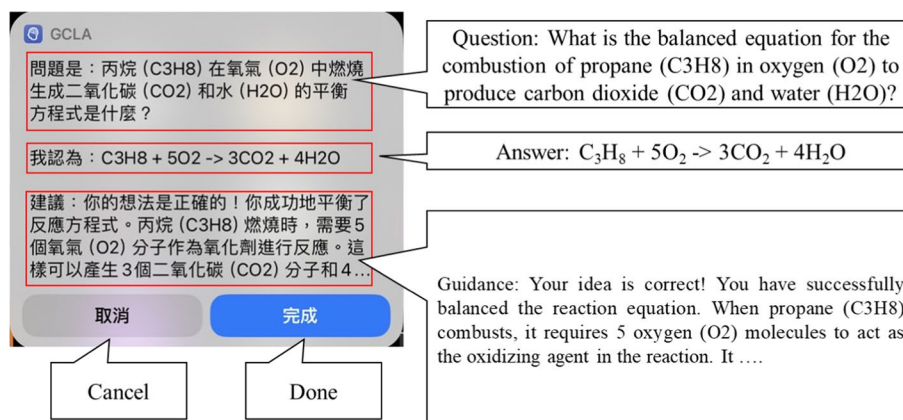


Fig. 7 Refining the answer with GCLA guidance until correct

Following this, GCLA processed the input and interfaced with the ChatGPT engine, subsequently presenting a hint: “Your answer is not correct. However, I will give you a hint on how...” (illustrated in Fig. 6).

Upon reviewing this hint, the student recognized the necessity of atom counting. They identified three carbon atoms, eight hydrogen atoms, and two oxygen atoms on the left, contrasted with one carbon atom, two hydrogen atoms, and three oxygen atoms on the right. Observing the disparity, the student revised their answer to: “ $C_3H_8 + 5O_2 \rightarrow 3CO_2 + 4H_2O$ ”.

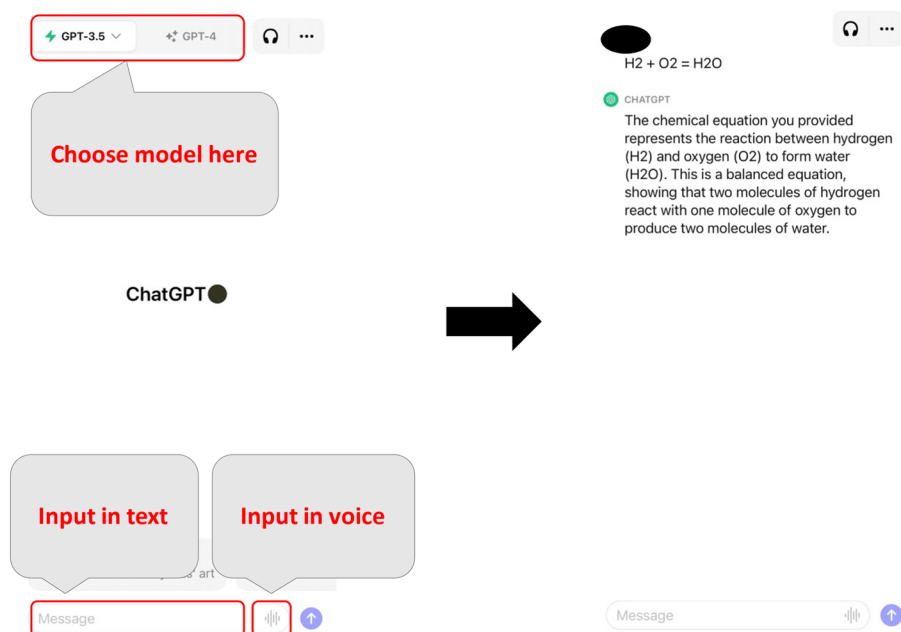
Re-evaluating the oxygen atoms, the student noted an equal count of ten on both sides, affirming the accuracy of their revised answer. This was corroborated by the app, which displayed: “Your idea is correct! You have...” (refer to Fig. 7).

Through the feedback and hints from the GCLA app, the student not only resolved the problem but also engaged in reflective learning. By reviewing the learning log, which documented the problem, responses, and received guidance, the student assessed their learning process, pinpointed strengths and weaknesses, and set future objectives, such as tackling more complex problems and expanding their knowledge of various chemical reactions.

### The comparison of ChatGPT on iOS

In May 2023, OpenAI launched a ChatGPT mobile application for both Android and iOS platforms, aiming to provide a seamless user experience for applying ChatGPT in addressing everyday challenges. However, due to the limited availability of GCLA (generative conversational language assistant) within the iOS ecosystem, this study will focus exclusively on the integration of ChatGPT with iOS. The user interface of this integration is illustrated in Fig. 8.

A notable feature presented at the top of the application allows users to toggle between the GPT-3.5 and GPT-4 models—the latter being accessible to those with an active subscription. Drawing a parallel to GCLA, this application endows learners with the flexibility to relay their inquiries either through textual or vocal input. It's imperative to note that, due to the predominant use of GPT-3.5 among the masses and its utilization in GCLA's developmental phase, this study predominantly leverages the GPT-3.5 iteration of ChatGPT on iOS. Upon query submission, the selected GPT model proffers an immediate response. This interactive mechanism mirrors that of GCLA, albeit with a distinction: GCLA mandates learners to offer their preliminary solutions post-question submission.



**Fig. 8** The interface of ChatGPT on iOS

## **Methodology**

### **Research design**

A randomized controlled trial (RCT) is a scientific study design that involves allocating participants into different groups using randomization (Stanley, 2007). This method is widely regarded as the 'gold standard' for evaluating the efficacy of new interventions or treatments. In an RCT, participants are randomly assigned to either a treatment group, which receives the intervention, or a control group, which receives a standard treatment or a placebo. This random allocation helps to minimize biases and ensures that the groups are comparable at the start of the study. The outcomes of these groups are then compared to determine the effectiveness of the intervention (Stolberg et al., 2004).

In this study, we employed an RCT to assess the impact of the GCLA on student performance in a blended learning environment. The study was conducted in a first-year university chemistry course in southern Taiwan, involving 61 students, comprising 31 males and 30 females. These participants were randomly divided into two groups: the Treatment Group (TG), consisting of 16 males and 15 females, and the Control Group (CG), with 15 males and 15 females. The TG interacted with the GCLA, a tool developed for this study to enhance blended learning. Unlike the traditional ChatGPT, the GCLA does not provide direct answers but rather encourages students to propose their solutions first. It then offers guidance to promote critical thinking and the iterative improvement of their answers. In contrast, the CG used the standard ChatGPT application on iOS. To ensure consistency in terms of hardware, all students were provided with iPads for the duration of the experiment. Both groups were taught by the same instructor, used identical teaching materials, and were in the same learning environment. The only variable was the use of either GCLA or traditional ChatGPT. This experiment was seamlessly integrated into the existing chemistry curriculum without requiring additional courses or incentives. All participants were fully informed about the study's objectives, methods, potential risks, and benefits through detailed consent forms. They were also assured that they could withdraw from the study at any time without any penalty or negative consequences.

### **Population**

The population of this study consisted of first-year undergraduate students enrolled in a chemistry course at a university in southern Taiwan. The course was a compulsory general education course for college of science students, regardless of their major. The course covered basic concepts and principles of chemistry, such as atomic structure, periodic table, chemical bonding, chemical reactions, and stoichiometry.

### **Sample size and sampling technique**

The sample size of this study was 61 students, which was determined by the availability of the participants and the feasibility of the experiment. The sampling technique used in this study was convenience sampling, which is a non-probability sampling method that selects participants based on their accessibility and willingness to participate (Creswell & Creswell, 2017). Convenience sampling was chosen because it was

the most practical and economical way to recruit participants for this study, given the time and resource constraints.

### Measurement

In this study, pre-tests, post-tests, and delayed tests were utilized to measure participants' advancement in chemistry knowledge. Researchers utilized a multiple-choice questionnaire consisting of 20 items, each valued at 5 points. This questionnaire was developed collaboratively by two expert chemistry educators, both possessing over a decade of teaching experience. 'Knowledge Construction'—as defined in this study—is based on the premise that individuals progressively enhance their understanding of a subject (van Kesteren & Meeter, 2020), evolving from simple information acquisition to in-depth comprehension through active cognitive engagement, leading to new insights that integrate with existing cognitive structures (Gan et al., 2020). The delayed test was particularly revealing regarding the long-term retention of knowledge. By administering pre-, post-, and delayed tests, Researchers can more accurately assess how learners acquire and retain chemistry knowledge, thus mapping the progression of their knowledge construction more clearly.

The higher-order thinking skills scale developed by Hwang et al. (2018) was utilized for a comprehensive assessment of higher-order thinking skills. This scale includes eleven items that correspond to three core dimensions: critical thinking, problem-solving, and creativity. Critical thinking involves reflective thinking and informed judgment-making, as defined by Hwang et al. (2018). Problem-solving focuses on the thorough gathering and analysis of information to overcome challenges effectively. The third dimension, creativity, highlights the ability to generate and develop new ideas. The importance of this instrument, built upon the framework by Hwang et al. (2018), is significant. It equips both researchers and educators with a powerful tool to explore and enhance individual competencies in these crucial domains.

The method proposed by Wu et al., (2023a) was applied to measure SRL, focusing on motivation, engagement, and self-efficacy—each aligning with the forethought, performance, and self-reflection phases of SRL, respectively. Wu et al., (2023a) underline the importance of sustaining motivation in the forethought phase (Pintrich, 2000; Zimmerman, 2000, 2008), actively engaging in the performance phase (Bernardo et al., 2022; Doo & Bonk, 2020), and enhancing self-efficacy during self-reflection (Rabin et al., 2020; Stephen et al., 2020). The Situational Motivation Scale (SIMS) by Guay et al. (2000) was employed, which differentiates motivation into intrinsic motivation, identified regulation, external regulation, and amotivation. Intrinsic motivation is driven by personal interest, identified regulation by a recognition of relevance, external regulation by rewards or pressures, and amotivation reflects a lack of motivation. For engagement, the Math and Science Engagement Scales proposed by Wang et al. (2016) was adapted, dividing engagement into cognitive, behavioral, and emotional aspects. Cognitive engagement involves self-regulation and strategy use, behavioral engagement involves participation and positive conduct, and emotional engagement involves positive feelings towards the educational environment. The New General Self-Efficacy Scale proposed by Chen et al. (2001) was adapted, defining self-efficacy as confidence in mobilizing resources to meet situational demands,

particularly in an academic context. The SRL scale, comprising three primary dimensions and nine sub-dimensions, uses a five-point Likert scale for responses. Its reliability and validity were confirmed in prior studies (Chen et al., 2001; Guay et al., 2000; Wang et al., 2016).

**Reliability of the measurement**

The reliability of the instruments used in this study were established through various methods. For the chemistry knowledge questionnaire, the content validity was ensured by the expert judgment of two experienced chemistry educators, who reviewed the items and provided feedback on their clarity, relevance, and difficulty. The reliability was assessed by calculating the Cronbach’s alpha coefficient, which yielded a value of 0.79, indicating a significant level of internal consistency.

For the higher-order thinking skills scale, the content validity was based on the theoretical framework proposed by Hwang et al. (2018), which identified three core dimensions of higher-order thinking skills: critical thinking, problem-solving, and creativity. Following the translation into Chinese, a reliability analysis detailed was performed in Table 2. The reliability was measured by computing the Cronbach’s alpha coefficients for each dimension, which ranged from 0.72 to 0.81, demonstrating a significant level of internal consistency.

For the self-regulated learning scale, the content validity was derived from the model of SRL proposed by Wu et al., (2023a), which aligned the three phases of SRL with three key factors: motivation, engagement, and self-efficacy. Following the translation into Chinese, a reliability analysis detailed was performed in Table 3. The reliability was evaluated by calculating the Cronbach’s alpha coefficients for each factor, which ranged from 0.75 to 0.88, indicating a significant level of internal consistency.

**Table 2** Reliability analysis of higher-order thinking scale

	Critical thinking	Problem-solving	Creativity
Original reliability	0.84	0.85	0.80
Revised reliability	0.81	0.78	0.72

**Table 3** The reliability of the self-regulated learning scale

Dimension	Sub-dimension	Reliability
Motivation (Forethought phase)	Intrinsic motivation	0.86
	Identified regulation	0.88
	External regulation	0.79
	Amotivation	0.80
Engagement (Performance phase)	Cognitive engagement	0.85
	Emotional engagement	0.81
	Behavioral engagement	0.75
	Social engagement	0.77
Self-efficacy (Self-reflection phase)	Self-efficacy	0.79

### Details of pre-test of and post-test

To assess the impact of GCLA on students' knowledge construction, self-regulated learning (SRL), and higher-order thinking skills (HOTS), this study employed a series of tests: pre-tests, post-tests, and delayed tests. These tests served as evaluative instruments. The pre-tests, conducted before the intervention, and the post-tests, administered immediately after the intervention, each had a duration of one hour. The delayed tests, taking place two weeks following the intervention, were 20 min long. All tests were conducted online using Google Forms.

The tests evaluating knowledge construction consisted of 20 multiple-choice questions, each worth 5 points, focusing on fundamental concepts and principles of chemistry, resulting in a total possible score of 100 points for each test. The format of the questions in the pre-tests, post-tests, and delayed tests remained consistent, with variations only in the questions, options, and numerical values in the questions.

The SRL assessment comprised 46 items across three domains: motivation (20 items), engagement (18 items), and self-efficacy (8 items). These items were adapted from Wu et al. (2023a) framework, integrating scales developed by Guay et al. (2000), Wang et al. (2016), and Chen et al. (2001). Responses were recorded on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

The HOTS assessment included 11 items, aligned with three dimensions: critical thinking, problem-solving, and creativity, adapted from the scale developed by Hwang et al. (2018). Responses were also recorded using a five-point Likert scale.

### Instruction methodology

The instruction methodology was based on the self-regulated learning (SRL) model proposed by Zimmerman (1990), which consists of three phases: forethought, performance, and self-reflection. The course involved learners in three-hour weekly classes for ten weeks, following the SRL framework as shown in Fig. 5.

- Week 1: Forethought phase

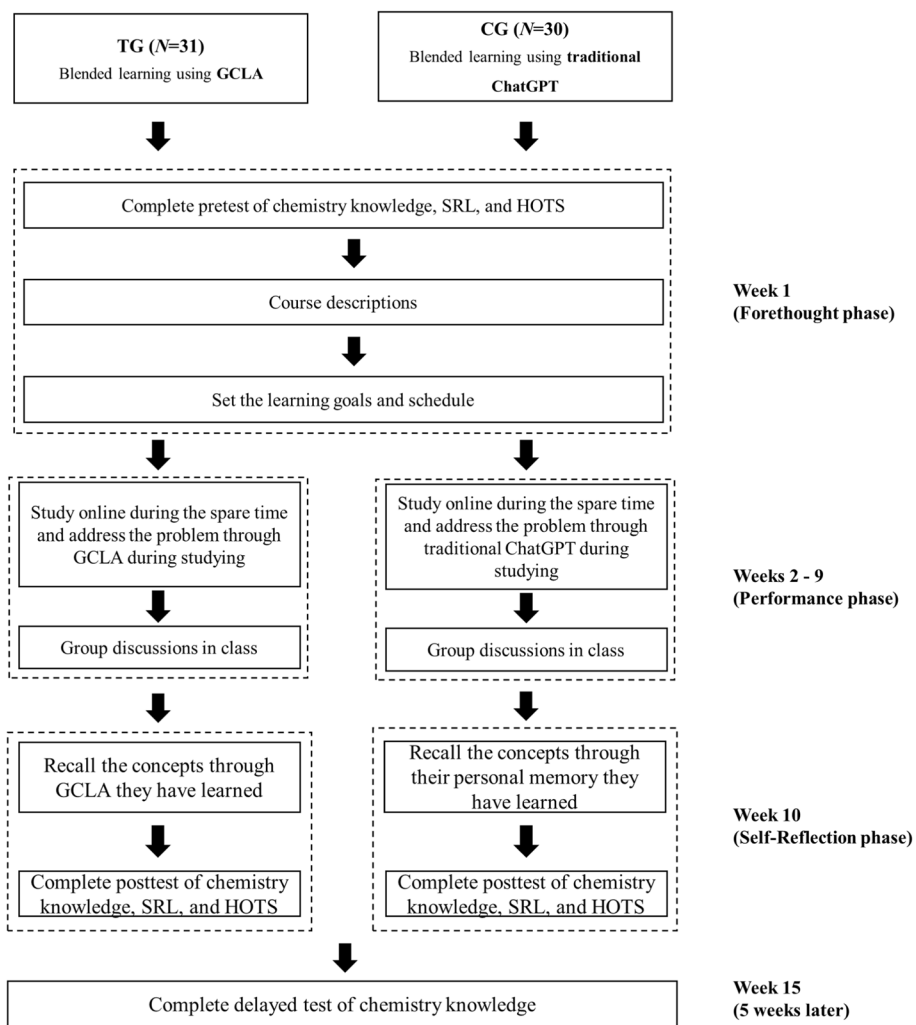
The instructor covered core chemistry concepts and conducted a pre-test to evaluate learners' preliminary understanding. Learners established goals like mastering the periodic table and chemical bonding, coordinating these with their other duties.

- Weeks 2 to 9: Performance phase

Learners dedicated these weeks to achieving their academic objectives, utilizing online tools and resources. Concurrently, weekly three-hour in-person classes promoted teamwork in problem-solving and permitted real-time learning strategy modifications.

- Week 10: Self-reflection phase

Learners appraised the knowledge gained throughout the course. A post-test assessed their comprehension, alongside a survey examining their SRL, HOTS, and chemistry knowledge. A delayed test in Week 15 was designed to evaluate the persistence of their acquired knowledge. The combined results of the pre-test, post-test,



**Fig. 9** Experimental procedure

**Table 4** Differences between groups in SRL at different phase

SRL phase	TG description	CG description
Forethought phase	Before diving into the study of chemistry, students should set daily learning objectives and monitor their achievements	Before diving into the study of chemistry, students should set daily learning objectives and monitor their achievements
Performance phase	When faced with challenges in their studies, students can turn to GCLA to address these difficulties	When facing challenges in their studies, students can turn to ChatGPT on iOS for assistance
Self-reflection phase	Examining the archived entries in the learning log file allows for reflection on past lessons and aids in the retention of concepts	Encourage students to reflect on their classroom experiences and recall concepts by drawing upon their personal memories

and delayed test will provide a comprehensive measure of the learners’ knowledge construction. Figure 9 details the entire experimental protocol.

The TG used GCLA to navigate the performance phase’s challenges, with learning logs aiding Week 10’s reflective activities. Conversely, the CG addressed problems

using ChatGPT on iOS during the same phase and relied on memory for reflective tasks in the final week. Table 4 delineates the variations in SRL dynamics between the TG and CG throughout the study’s duration.

**Variables**

The variables in this study included the independent variable, the dependent variables, and the covariates. The independent variable was the type of learning tool used by the students: GCLA or ChatGPT. The dependent variables were the students’ scores on the SRL, HOTS, and chemistry knowledge tests. The covariates were the students’ scores on the pre-tests, which were used to control for the initial differences between the groups.

**Methods of analysis and statistical tools**

The methods of analysis and statistical tools used in this study were as follows:

- Descriptive statistics: To describe the sample characteristics and the scores of the dependent variables for each group.
- Analysis of covariance (ANCOVA): To compare the mean scores of the dependent variables between the groups, adjusting for the effects of the covariates.
- Effect size: To measure the magnitude of the difference between the groups, using partial eta-squared ( $\eta^2$ ) as the index.
- Statistical software: To perform the data analysis, using JAMOVI version 2.4 (project, 2023).

**Results**

**The impact of GCLA on self-regulated learning (SRL)**

To investigate the potential impact of introducing GCLA on students’ SRL in a blended learning setting, Analysis of Covariance (ANCOVA) was employed in this study. The SRL scores from the pre-tests as covariates and considered the post-test SRL scores as the dependent variables. Prior to the main analysis, The assumption of homogeneity of variances using Levene’s test was confirmed. The results are outlined in Table 5. As seen in Table 5, all sub-dimensions of SRL have p-values exceeding 0.05. These findings underline the validity of the variance equality assumption and confirm the appropriateness of using ANCOVA for the analysis.

**Table 5** The Levene’s test results for SRL

Variable	Levene’s test	
	F	p
Intrinsic motivation	0.104	0.749
Identified regulation	1.82	0.182
External regulation	2.43	0.124
Amotivation	0.075	0.785
Cognitive engagement	0.323	0.572
Behavioral engagement	0.124	0.726
Emotional engagement	2.21	0.143
Self-efficacy	0.029	0.866



Tables 6 and 7 respectively present the descriptive statistics and ANCOVA analysis results for SRL. As evidenced in Table 7, statistically significant differences were observed in intrinsic motivation, amotivation, cognitive engagement, behavioral engagement, and self-efficacy. From the descriptive statistics in Table 6, it can be discerned that the CG notably outperformed the TG in terms of intrinsic motivation ( $M_{EG} = 18.9 > M_{CG} = 16.1$ ) and amotivation ( $M_{EG} = 8.53 < M_{CG} = 12.0$ , this being an adverse indicator). Conversely, the TG surpassed the CG in cognitive engagement ( $M_{EG} = 34.5 > M_{CG} = 27.4$ ), behavioral engagement ( $M_{EG} = 20.1 > M_{CG} = 17.4$ ), and self-efficacy ( $M_{EG} = 31.2 > M_{CG} = 25.8$ ). Thus, in addressing Research Question 1, while the introduction of GCLA may not bolster students' motivational aspect in SRL (Forethought phase), it does indeed foster positive effects on their engagement level (Performance phase) and self-efficacy (Self-reflection phase).

**Table 6** Descriptive results for SRL

	TG (N = 31)				CG (N = 30)			
	Pre-test		Post-test		Pre-test		Post-test	
	M	SD	M	SD	M	SD	M	SD
Intrinsic motivation	16.2	1.88	18.9	2.82	16.0	2.46	16.1	2.55
Identified regulation	14.8	2.32	19.1	3.50	14.7	1.45	17.8	2.53
External regulation	15.4	2.01	18.1	1.53	15.8	2.26	17.3	2.26
Amotivation	11.2	1.64	8.53	2.73	11.6	1.71	12.0	2.73
Cognitive engagement	25.3	3.49	34.5	2.58	24.9	3.35	27.4	2.72
Behavioral engagement	17.0	1.28	20.1	1.71	16.9	2.16	17.4	1.33
Emotional engagement	12.6	1.60	13.5	1.21	12.5	2.32	12.8	1.53
Self-efficacy	25.8	2.53	31.2	3.90	25.7	2.79	25.8	4.16

**Table 7** ANCOVA results for SRL

Variable	SS	df	Mean square	F	p	Partial $\eta^2$
Intrinsic motivation	120.4	1	120.41	17.13	<b>&lt;0.001***</b>	0.228
Identified regulation	24.6	1	24.6	2.94	0.092	0.048
External regulation	11.8	1	11.8	3.33	0.073	0.054
Amotivation	185.98	1	185.98	24.62	<b>&lt;0.001***</b>	0.298
Cognitive engagement	778.4	1	778.4	48.60	<b>&lt;0.001***</b>	0.369
Behavioral engagement	133	1	132.88	41.4	<b>&lt;0.001***</b>	0.333
Emotional engagement	5.98	1	5.98	3.12	0.082	0.051
Self-efficacy	437.06	1	437.06	26.46	<b>&lt;0.001***</b>	0.313

Bold values represent significant difference

**Table 8** The Levene's test results for HOTS

Variable	Levene's test	
	F	p
Critical thinking	1.22	0.273
Problem-solving	0.13	0.719
Creativity	0.261	0.611

**Table 9** Descriptive results for HOTS

	TG (N = 31)				CG (N = 30)			
	Pre-test		Post-test		Pre-test		Post-test	
	M	SD	M	SD	M	SD	M	SD
Critical thinking	13.0	1.47	17.3	2.19	13.0	1.52	13.8	2.00
Problem-solving	12.3	1.05	15.9	2.01	12.0	2.01	13.1	1.96
Creativity	9.58	1.39	11.0	1.18	9.47	1.11	10.1	1.07

**Table 10** ANCOVA results for HOTS

Variable	SS	df	Mean Square	F	p	Partial $\eta^2$
Critical thinking	182.17	1	182.17	41.73	<b>&lt;0.001</b> ***	0.418
Problem-solving	130.6	1	130.6	38.7	<b>&lt;0.001</b> ***	0.400
Creativity	10.63	1	10.63	8.98	<b>0.004</b> **	0.134

Bold values represent significant difference

**The impact of GCLA on higher-order thinking skills (HOTS)**

To investigate the potential impact of introducing GCLA on students’ HOTS in a blended learning setting, ANCOVA was employed in this study. The HOTS scores from the pre-tests as covariates and considered the post-test HOTS scores as the dependent variables. Prior to the main analysis, The assumption of homogeneity of variances using Levene’s test was confirmed. The results are outlined in Table 8. As seen in Table 8, all dimensions of HOTS have p-values exceeding 0.05. These findings underline the validity of the variance equality assumption and confirm the appropriateness of using ANCOVA for the analysis.

Tables 9 and 10 provide the descriptive statistics and ANCOVA analysis results for HOTS, respectively. As seen in Table 10, there were statistically significant differences in the dimension of critical thinking, problem-solving, and creativity. From the descriptive statistics in Table 9, it can be discerned that the TG notably outperformed the CG in terms of critical thinking ( $M_{EG} = 17.3 > M_{CG} = 13.8$ ), problem-solving ( $M_{EG} = 15.9 > M_{CG} = 13.1$ ), and creativity ( $M_{EG} = 11.0 > M_{CG} = 10.1$ ). Thus, in addressing the research question 2, the introduction of GCLA can be deduced to enhance the HOTS of learners, manifesting in significant improvements in critical thinking, problem-solving, and creativity.

**The impact of GCLA on knowledge construction**

To gauge learners’ knowledge construction within a blended learning environment, chemistry comprehension was assessed at three distinct intervals: a pre-test (prior to experimental activity), a post-test (following the experimental activity), and a delayed test (two weeks post-experimental activity). Two ANCOVAs were employed to discern variations in knowledge over these time frames, thereby evaluating students’ knowledge construction. The initial ANCOVA treated post-test scores as the dependent variable, with pre-test scores serving as the covariate. This was to discern

**Table 11** The Levene’s test results for knowledge construction

Variable	Levene’s test	
	F	p
Post-test for chemistry knowledge	1.22	0.273
Delayed test for chemistry knowledge	0.13	0.719

the disparity in chemistry understanding between the TG and CG subsequent to the experimental intervention. The subsequent ANCOVA utilized delayed test scores as the dependent variable and post-test scores as the covariate, aiming to gauge the retention of knowledge after a two-week span. Prior to the ANCOVA implementation, the Levene’s test was executed to confirm the homogeneity of variances, with the results delineated in Table 11. As evident from Table 11, both post-test and delayed test yielded p-values surpassing 0.05, reinforcing the validity of the equal variance assumption and solidifying the justification for ANCOVA in the assessment.

To ensure that there were no initial differences in learners’ prior knowledge, the study initially conducted an ANOVA on the pre-test scores for chemistry knowledge. The results indicated no significant relationship between the TG and the CG in the pre-test scores ( $F=0.231$ ,  $p=0.632 > 0.05$ ). This suggests that there were no significant differences in prior knowledge between the two groups at the onset of the study. Such a finding is crucial as it establishes a baseline equivalence between the TG and CG, thereby allowing for a more accurate assessment of the GCLA intervention’s impact on chemistry comprehension and knowledge retention.

Tables 12 and 13 present the descriptive statistics for the pre-test, post-test, and delayed post-test, along with the outcomes of two distinct ANCOVAs. As elucidated in Table 13, noteworthy statistical discrepancies exist between the TG and the CG concerning both post-test and delayed test scores. A closer analysis of Table 12 distinctly showcases that the scores attained by the TG in both the post-test ( $M_{EG} = 79.7 > M_{CG} = 74.7$ )

**Table 12** Descriptive results for knowledge construction

	TG (N = 31)		CG (N = 30)	
	M	SD	M	SD
Pre-test for chemistry knowledge	63.3	11.8	62.0	7.80
Post-test for chemistry knowledge	79.7	6.77	74.7	8.55
Delayed test for chemistry knowledge	75.9	6.36	69.3	6.94

**Table 13** ANCOVA results for knowledge construction

Variable	SS	df	Mean Square	F	p	Partial $\eta^2$
Post-test for chemistry knowledge	357	1	357	6.10	<b>0.017*</b>	0.100
Delayed test for chemistry knowledge	280	1	280	8.20	<b>0.006**</b>	0.124

Bold values represent significant difference

and delayed test ( $M_{EG}=75.9>M_{CG}=69.3$ ) substantially surpass those of the CG. Addressing the research question 3, the GCLA intervention markedly enhances learners' comprehension of chemistry and the subsequent retention of this knowledge. Consequently, it's plausible to deduce that GCLA plays a pivotal role in augmenting learners' knowledge construction.

## Discussion

### The impact of GCLA on self-regulated learning (SRL)

SRL, fundamental to blended learning, plays a crucial role in higher education, where hybrid models are increasingly prevalent. Zimmerman (1990) articulated SRL's phases as Forethought, Performance, and Self-Reflection. Wu et al. (2023a) explored these stages within the context of higher education, evaluating their impact on motivation, engagement, and self-efficacy. The investigation, aligning with these scholarly works, specifically examines the differential effects of GCLA versus traditional ChatGPT on SRL among higher education students, as Table 6 and 7 illustrate. The data shows that those utilizing GCLA exhibited greater cognitive and behavioral engagement, alongside heightened self-efficacy, compared to peers using ChatGPT.

In the realm of higher education, where independent critical thinking is paramount, GCLA's design, which necessitates that students engage thoughtfully with content prior to receiving answers, proved advantageous. This strategy encourages the deployment of advanced cognitive strategies and the swift integration of complex concepts, key in fostering cognitive engagement at this level. This methodology is supported by the "Probing understanding" concept advocated by White and Gunstone (2014) and the inquiry-based framework proposed by Pedaste et al. (2015). Such an approach is particularly effective in higher education, as confirmed by Al Mamun and Lawrie (2023) and Mamun et al. (2020), and aligns with GCLA's problem-solving orientation.

Enhanced behavioral engagement with GCLA is also critical in higher education settings. The platform's design requires more intensive student interaction, reflecting the "learning-by-doing" pedagogy that is vital in higher learning environments. Students in these settings must not only understand theoretical concepts but also apply them in practical scenarios (Lee et al., 2023a, 2023b; Lee et al., 2023a, 2023b; Wu et al., 2023b). This is consistent with the findings of Dellatola et al. (2020) and Kuo et al. (2020), who reported that active engagement practices, such as collaborative problem-solving and peer feedback, improved students' learning outcomes and satisfaction in blended learning courses. Therefore, GCLA can be seen as a valuable tool to facilitate such practices and enhance students' behavioral engagement.

Furthermore, GCLA's record-keeping feature improves self-efficacy among higher education learners by enabling them to monitor their queries and reflect on their learning journey. This reflective practice is vital for students to develop confidence in their academic abilities, which is especially important in higher education where learners are expected to take greater ownership of their learning (Hsia & Hwang, 2020; Menon & Azam, 2021).

However, in terms of intrinsic motivation and amotivation, traditional ChatGPT holds a slight edge, particularly in higher education contexts. Its ability to provide immediate feedback seems to foster motivation and stave off feelings of helplessness, underscoring

the impact of timely support on student motivation—a critical factor for student success in higher education, as emphasized by Wu et al. (2023a).

#### **The impact of GCLA on higher-order thinking skills (HOTS)**

The increasing acknowledgment of HOTS as indispensable for success in the twenty-first century has been a prominent theme in the field of higher education research (Conklin, 2011). In stark contrast to the passive nature of traditional lecture-based instruction prevalent in higher education, blended learning paradigms afford students a high level of autonomy (Snodin, 2013). This shift necessitates a more deliberate focus on cultivating advanced cognitive skills that enable students to critically analyze educational content, appraise their learning journeys, and craft tailored learning plans. Evidence from Tables 9 and 10 reinforces that GCLA developed in this study notably advances HOTS within higher education's blended learning frameworks, specifically targeting critical thinking, problem-solving, and creativity.

In terms of critical thinking within the higher education context, GCLA systematically provides hints rather than outright solutions, prompting students to iteratively refine their reasoning to reach the correct conclusion. This approach aligns with the dialogic and inquiry-based models that are becoming increasingly prevalent in higher education pedagogy (Al Mamun & Lawrie, 2023; Kuo et al., 2020). It encourages a deeper engagement with material, thus leading students in the TG to develop more robust critical thinking skills than their counterparts in the CG. These findings dovetail with multiple studies that emphasize the importance of discussion-based learning in fostering critical thinking within higher education settings (Al-Husban, 2020; Giacomo & Savenye, 2020; O'Riordan et al., 2021).

When addressing problem-solving abilities, GCLA's structured hinting mechanism compels students to actively engage with problems, reflecting on and revising their approaches. This is particularly valuable in the higher education landscape, where problem-solving is a key learning outcome. The technique employed by GCLA mimics the iterative process fundamental to Problem-Based Learning (PBL), a method well-established and valued in higher education for its efficacy in cultivating problem-solving prowess (Aslan, 2021; Phungsuk et al., 2017; Valentine et al., 2017).

Lastly, regarding creativity, a skill increasingly sought after in higher education graduates, GCLA's methodology nurtures this by offering hints that encourage a multiplicity of perspectives. This contrasts with the more deterministic approach of traditional ChatGPT, fostering a learning environment in higher education where students are prompted to think divergently and conceive innovative solutions. The potential for GCLA to stimulate more creative outcomes is in line with research advocating for the role of brainstorming and creative thinking in higher education (Göçmen & Coşkun, 2022; Gong et al., 2022).

#### **The impact of GCLA on knowledge construction**

Knowledge construction is fundamental in higher education, where evaluating students' learning outcomes is paramount. This study employed a multi-stage assessment approach comprising pre-tests, post-tests, and delayed tests to track students'

knowledge acquisition in higher education settings. As delineated in Tables 12 and 13, higher education learners using the GCLA displayed superior performance in both post-test and delayed test evaluations when compared to peers using conventional ChatGPT tools.

The GCLA framework insists that learners in higher education contexts propose their own responses before receiving any assistance, offering hints instead of outright answers. This process promotes active engagement and allows students to develop, assess, and hone their problem-solving techniques, leading to more effective learning and review. This is a practical application of metacognitive strategies which are particularly relevant in the context of higher education (Schraw & Moshman, 1995). Moreover, by expressing their initial reasoning and iteratively refining their understanding, students experience a form of inquiry-based learning that is pivotal for higher education (Pedaste et al., 2015; White & Gunstone, 2014). The enhancement in knowledge construction and retention within the higher education cohort using GCLA is further substantiated by existing research on the benefits of metacognitive and inquiry-based learning strategies (Carvalho & Santos, 2022; Tawfik et al., 2020; Zhou & Lam, 2019).

## **Implications**

This study has several implications for both theory and practice in the field of education.

### **Theoretical implications**

This study contributes to the existing literature on the use of ChatGPT in education, especially in blended learning environments. It introduces the GCLA, a novel approach that modifies the use of ChatGPT by providing guidance rather than direct answers, and evaluates its impact on students' SRL, HOTS, and knowledge construction. The findings reveal that the GCLA enhances these aspects compared to traditional ChatGPT use, demonstrating the potential of ChatGPT-based tools to foster essential skills such as self-regulation and higher-order thinking. Furthermore, this study offers insights into the more effective use of ChatGPT in education, highlighting the importance of encouraging students to attempt problem-solving independently before seeking ChatGPT assistance, and providing feedback through hints rather than solutions. These insights can inform the design and development of future ChatGPT-based educational tools and interventions.

### **Practical implications**

This study also has practical implications for educators and learners in higher education, particularly in blended learning settings. It suggests that the GCLA can be a valuable tool to supplement blended learning, as it can provide timely and personalized guidance to students, enhance their engagement and self-efficacy, and improve their learning outcomes. The GCLA can also support educators in facilitating blended learning, as it can reduce their workload in providing feedback and assistance, and allow them to monitor

students' progress and performance through the learning log file. Moreover, the GCLA can be easily integrated into existing blended learning platforms and curricula, as it is compatible with Apple devices and can be customized according to different learning objectives and contexts.

### **Conclusion**

This study aimed to address the issue of students' excessive reliance on ChatGPT in higher education blended learning settings, and its impact on their SRL, HOTS, and knowledge construction. It also evaluated the effectiveness of the GCLA, a guidance-based ChatGPT-assisted learning aid, compared to traditional ChatGPT use, in enhancing these aspects among learners. The study involved 61 undergraduate students from a university in Taiwan, who were randomly assigned to either the treatment group (TG) or the control group (CG). The TG used the GCLA, while the CG used the traditional ChatGPT application on iOS, to assist their learning in a foundational chemistry course within a blended learning setting. The study employed a Randomized Controlled Trial (RCT) and used pre-tests, post-tests, delayed tests, and surveys to measure the impact of the GCLA on students' SRL, HOTS, and knowledge construction.

The results showed that the GCLA had a significant positive effect on students' cognitive and behavioral engagement, self-efficacy, critical thinking, problem-solving, creativity, and knowledge construction, compared to traditional ChatGPT use. However, the GCLA did not have a significant effect on students' intrinsic motivation and amotivation, which were higher in the CG than in the EG. These results suggest that the GCLA can effectively enhance students' learning experiences and outcomes in blended learning environments, by providing guidance rather than answers, and fostering an environment conducive to the development of SRL and HOTS. The results also indicate that the GCLA can help students overcome the challenges of blended learning, such as maintaining consistent teacher support, managing learning autonomy, and engaging in complex academic tasks.

### **Limitations**

This study has several limitations that should be acknowledged and addressed in future research. First, the sample size of this study was relatively small, and the participants were from a single university in Taiwan. Therefore, the generalizability of the findings may be limited, and further studies with larger and more diverse samples are needed to validate the results. Second, the study only focused on one subject area, namely chemistry, and one blended learning setting. Thus, the applicability of the GCLA to other subject areas and learning settings may vary, and more studies are needed to explore the effects of the GCLA in different domains and contexts. Third, the study only measured the short-term and medium-term effects of the GCLA on students' learning, using post-tests and delayed tests. The long-term effects of the GCLA on students' learning, such as retention, transfer, and application of knowledge and skills, were not assessed, and future studies should include more longitudinal measures to evaluate the lasting impact of the GCLA.

### Future directions

Based on the findings and limitations of this study, several directions for future research can be suggested. First, future studies can extend the scope of this study by investigating the effects of the GCLA on other aspects of students' learning, such as motivation, interest, satisfaction, and attitude. These aspects are also important for students' learning success and well-being, and can provide a more comprehensive picture of the impact of the GCLA. Second, future studies can explore the underlying mechanisms and factors that mediate or moderate the effects of the GCLA on students' learning. For example, how does the GCLA influence students' cognitive processes, metacognitive strategies, and affective states? How do students' prior knowledge, learning styles, and preferences affect their use and perception of the GCLA? These questions can help to explain the reasons and conditions for the effectiveness of the GCLA, and provide more insights for its improvement and optimization. Third, future studies can compare the GCLA with other ChatGPT-based tools or interventions, such as those that provide different types or levels of feedback, scaffolding, or personalization. These comparisons can help to identify the strengths and weaknesses of the GCLA, and provide more evidence for its relative advantages and disadvantages.

### Acknowledgements

We are extremely grateful to the research assistants and students who participated in this study.

### Author contributions

H-YL is the leader of this research, he is in charge of the research design, conducting teaching and learning experiment, data analysis. P-HC is responsible for assisting in the conduct of experiments and surveying related literature. W-SW is responsible for assisting in the conduct of experiments. Y-MH is responsible for designing research experiments, providing fundamental education theories and comments to this research, and he is also responsible for revising the manuscript. All authors spent more than 2 months to discuss and analyze the data. The author(s) read and approved the final manuscript. T-TW is responsible for assisting in the conduct of experiments and surveying related literature and proofreading the manuscript.

### Funding

This project was funded by the National Science and Technology Council (NSTC), of the Republic of China under Contract numbers NSTC 110-2511-H-006-008-MY3, NSTC 112-2410-H-006-053-MY3.

### Availability of data and materials

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

### Declarations

#### Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 8 November 2023 Accepted: 9 February 2024

Published online: 04 March 2024

### References

- Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive network science reveals bias in GPT-3, GPT-3.5 Turbo, and GPT-4 mirroring math anxiety in high-school students. *Big Data and Cognitive Computing*, 7(3), 124. <https://doi.org/10.3390/bdcc7030124>
- Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2023.2253858>
- Al-Husban, N. A. (2020). Critical thinking skills in asynchronous discussion forums: A case study. *International Journal of Technology in Education*, 3(2), 82–91.
- Al Mamun, M. A., & Lawrie, G. (2023). Student-content interactions: Exploring behavioural engagement with self-regulated inquiry-based online learning modules. *Smart Learning Environments*, 10(1), 1. <https://doi.org/10.1186/s40561-022-00221-x>



- Aslan, A. (2021). Problem-based learning in live online classes: Learning achievement, problem-solving skill, communication skill, and interaction. *Computers & Education*, 171, 104237. <https://doi.org/10.1016/j.compedu.2021.104237>
- Bernardo, A. B., Galve-González, C., Núñez, J. C., & Almeida, L. S. (2022). A path model of university dropout predictors: The role of satisfaction, the use of self-regulation learning strategies and students' engagement. *Sustainability*. <https://doi.org/10.3390/su14031057>
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Ascd.
- Carvalho, A. R., & Santos, C. (2022). Developing peer mentors' collaborative and metacognitive skills with a technology-enhanced peer learning program. *Computers and Education Open*, 3, 100070. <https://doi.org/10.1016/j.caeo.2021.100070>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, 4(1), 62–83. <https://doi.org/10.1177/109442810141004>
- Chen, T., Luo, H., Feng, Q., & Li, G. (2023). Effect of technology acceptance on blended learning satisfaction: The serial mediation of emotional experience, social belonging, and higher-order thinking. *International Journal of Environmental Research and Public Health*, 20(5), 4442.
- Cheng, S.-C., Hwang, G.-J., & Lai, C.-L. (2020). Effects of the group leadership promotion approach on students' higher order thinking awareness and online interactive behavioral patterns in a blended learning environment. *Interactive Learning Environments*, 28(2), 246–263. <https://doi.org/10.1080/10494820.2019.1636075>
- Conklin, W. (2011). *Higher-order thinking skills to develop 21st century learners*. Teacher Created Materials.
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, 27(1), 113–118.
- Dellatola, E., Daradoumis, T., & Dimitriadis, Y. (2020). Exploring students' engagement within a collaborative inquiry-based language learning activity in a blended environment. In S. Yu, M. Ally, & A. Tsinakos (Eds.), *Emerging technologies and pedagogies in the curriculum* (pp. 355–375). Springer Singapore. [https://doi.org/10.1007/978-981-15-0618-5\\_21](https://doi.org/10.1007/978-981-15-0618-5_21)
- Ding, L., Li, T., Jiang, S., & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, 20(1), 63. <https://doi.org/10.1186/s41239-023-00434-1>
- Doo, M. Y., & Bonk, C. J. (2020). The effects of self-efficacy, self-regulation and social presence on learning engagement in a large university class using flipped Learning. *Journal of Computer Assisted Learning*, 36(6), 997–1010. <https://doi.org/10.1111/jcal.12455>
- Ettinger, A. (2020). What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8, 34–48. [https://doi.org/10.1162/tacl\\_a\\_00298](https://doi.org/10.1162/tacl_a_00298)
- Gan, W., Sun, Y., Peng, X., & Sun, Y. (2020). Modeling learner's dynamic knowledge construction procedure and cognitive item difficulty for knowledge tracing. *Applied Intelligence*, 50(11), 3894–3912. <https://doi.org/10.1007/s10489-020-01756-7>
- Giacumo, L. A., & Savenye, W. (2020). Asynchronous discussion forum design to support cognition: Effects of rubrics and instructor prompts on learner's critical thinking, achievement, and satisfaction. *Educational Technology Research and Development*, 68(1), 37–66. <https://doi.org/10.1007/s11423-019-09664-5>
- Göçmen, Ö., & Coşkun, H. (2022). Do De Bono's green hat and green-red combination increase creativity in brainstorming on individuals and dyads? *Thinking Skills and Creativity*, 46, 101185. <https://doi.org/10.1016/j.tsc.2022.101185>
- Gong, Z., Lee, L.-H., Soomro, S. A., Nanjappan, V., & Georgiev, G. V. (2022). A systematic review of virtual brainstorming from the perspective of creativity: Affordances, framework, and outlook. *Digital Creativity*, 33(2), 96–127. <https://doi.org/10.1080/14626268.2022.2064879>
- Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24(3), 175–213. <https://doi.org/10.1023/A:1005614228250>
- Hershcovits, H., Vilenchik, D., & Gal, K. (2020). Modeling engagement in self-directed learning systems using principal component analysis. *IEEE Transactions on Learning Technologies*, 13(1), 164–171. <https://doi.org/10.1109/TLT.2019.2922902>
- Hood, N., Littlejohn, A., & Milligan, C. (2015). Context counts: How learners' contexts influence learning in a MOOC. *Computers & Education*, 91, 83–91. <https://doi.org/10.1016/j.compedu.2015.10.019>
- Hsia, L.-H., & Hwang, G.-J. (2020). From reflective thinking to learning engagement awareness: A reflective thinking promoting approach to improve students' dance performance, self-efficacy and task load in flipped learning. *British Journal of Educational Technology*, 51(6), 2461–2477. <https://doi.org/10.1111/bjet.12911>
- Hwang, G.-J., & Lai, C.-L. (2017). Facilitating and bridging out-of-class and in-class learning: An interactive e-book-based flipped learning approach for math courses. *Journal of Educational Technology & Society*, 20(1), 184–197.
- Hwang, G.-J., Lai, C.-L., Liang, J.-C., Chu, H.-C., & Tsai, C.-C. (2018). A long-term experiment to investigate the relationships between high school students' perceptions of mobile learning and peer interaction and higher-order thinking tendencies. *Educational Technology Research and Development*, 66(1), 75–93. <https://doi.org/10.1007/s11423-017-9540-3>
- Hwang, G.-J., Yin, C., & Chu, H.-C. (2019). The era of flipped learning: Promoting active learning and higher order thinking with innovative flipped learning strategies and supporting systems. *Interactive Learning Environments*, 27(8), 991–994. <https://doi.org/10.1080/10494820.2019.1667150>
- Jansen, T., & Möller, J. (2022). Teacher judgments in school exams: Influences of students' lower-order-thinking skills on the assessment of students' higher-order-thinking skills. *Teaching and Teacher Education*, 111, 103616. <https://doi.org/10.1016/j.tate.2021.103616>

- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-11834-1>
- Kasneji, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Kasneji, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212–218.
- Kuo, Y.-R., Tuan, H.-L., & Chin, C.-C. (2020). The influence of inquiry-based teaching on male and female students' motivation and engagement. *Research in Science Education*, 50(2), 549–572. <https://doi.org/10.1007/s11165-018-9701-3>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Lee, H.-Y., Cheng, Y.-P., Wang, W.-S., Lin, C.-J., & Huang, Y.-M. (2023a). Exploring the learning process and effectiveness of STEM education via learning behavior analysis and the interactive-constructive-active-passive framework. *Journal of Educational Computing Research*, 61(5), 951–976. <https://doi.org/10.1177/07356331221136888>
- Lee, H.-Y., Lin, C.-J., Wang, W.-S., Chang, W.-C., & Huang, Y.-M. (2023b). Precision education via timely intervention in K-12 computer programming course to enhance programming skill and affective-domain learning objectives. *International Journal of STEM Education*, 10(1), 52. <https://doi.org/10.1186/s40594-023-00444-5>
- Lu, K., Pang, F., & Shadiev, R. (2021a). Understanding the mediating effect of learning approach between learning factors and higher order thinking skills in collaborative inquiry-based learning. *Educational Technology Research and Development*, 69(5), 2475–2492. <https://doi.org/10.1007/s11423-021-10025-4>
- Lu, K., Yang, H. H., Shi, Y., & Wang, X. (2021b). Examining the key influencing factors on college students' higher-order thinking skills in the smart classroom environment. *International Journal of Educational Technology in Higher Education*, 18(1), 1. <https://doi.org/10.1186/s41239-020-00238-7>
- Mali, D., & Lim, H. (2021). How do students perceive face-to-face/blended learning as a result of the Covid-19 pandemic? *The International Journal of Management Education*, 19(3), 100552. <https://doi.org/10.1016/j.ijme.2021.100552>
- Mamun, M. A. A., Lawrie, G., & Wright, T. (2020). Instructional design of scaffolded online learning modules for self-directed and inquiry-based learning environments. *Computers & Education*, 144, 103695. <https://doi.org/10.1016/j.compedu.2019.103695>
- Menon, D., & Azam, S. (2021). Investigating preservice teachers' science teaching self-efficacy: An analysis of reflective practices. *International Journal of Science and Mathematics Education*, 19(8), 1587–1607. <https://doi.org/10.1007/s10763-020-10131-4>
- Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers*, 12(8), 153.
- O'Riordan, T., Millard, D. E., & Schulz, J. (2021). Is critical thinking happening? Testing content analysis schemes applied to MOOC discussion forums. *Computer Applications in Engineering Education*, 29(4), 690–709. <https://doi.org/10.1002/cae.22314>
- Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. <https://doi.org/10.1016/j.edurev.2015.02.003>
- Phungsuk, R., Viriyavejakul, C., & Ratanaolarn, T. (2017). Development of a problem-based learning model via a virtual learning environment. *Kasetsart Journal of Social Sciences*, 38(3), 297–306. <https://doi.org/10.1016/j.kjss.2017.01.001>
- Pintrich, P. R. (2000). Chapter 14—The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press.
- project, T. J. (2023). *jamovi*. In <https://www.jamovi.org/>
- Rabin, E., Henderikx, M., Yoram, M. K., & Kalz, M. (2020). What are the barriers to learners' satisfaction in MOOCs and what predicts them? The role of age, intention, self-regulation, self-efficacy and motivation. *Australasian Journal of Educational Technology*, 36(3), 119–131.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf)
- Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A systematic review. *Computers & Education*, 144, 103701. <https://doi.org/10.1016/j.compedu.2019.103701>
- Salah Dogham, R., Elcokany, N. M., Saber Ghaly, A., Dawood, T. M. A., Aldakheel, F. M., Llaguno, M. B. B., & Mohsen, D. M. (2022). Self-directed learning readiness and online learning self-efficacy among undergraduate nursing students. *International Journal of Africa Nursing Sciences*, 17, 100490. <https://doi.org/10.1016/j.ijans.2022.100490>
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351–371. <https://doi.org/10.1007/BF02212307>
- Snodin, N. S. (2013). The effects of blended learning with a CMS on the development of autonomous learning: A case study of different degrees of autonomy achieved by individual learners. *Computers & Education*, 61, 209–216. <https://doi.org/10.1016/j.compedu.2012.10.004>
- Stanley, K. (2007). Design of randomized controlled trials. *Circulation*, 115(9), 1164–1169.
- Stephen, J. S., Rockinson-Szapkiw, A. J., & Dubay, C. (2020). Persistence model of non-traditional online learners: Self-efficacy, self-regulation, and self-direction. *American Journal of Distance Education*, 34(4), 306–321. <https://doi.org/10.1080/08923647.2020.1745619>
- Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. Cambridge University Press.
- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1), 35. <https://doi.org/10.1186/s41239-023-00404-7>

- Stolberg, H. O., Norman, G., & Trop, I. (2004). Randomized controlled trials. *AJR American Journal of Roentgenology*, 183(6), 1539–1544.
- Tawfik, A. A., Graesser, A., Gatewood, J., & Gishbaugher, J. (2020). Role of questions in inquiry-based instruction: Towards a design taxonomy for question-asking and implications for design. *Educational Technology Research and Development*, 68(2), 653–678. <https://doi.org/10.1007/s11423-020-09738-9>
- Valentine, A., Belski, I., & Hamilton, M. (2017). Developing creativity and problem-solving skills of engineering students: A comparison of web- and pen-and-paper-based approaches. *European Journal of Engineering Education*, 42(6), 1309–1329. <https://doi.org/10.1080/03043797.2017.1291584>
- van Kesteren, M. T. R., & Meeter, M. (2020). How to optimize knowledge construction in the brain. *NPJ Science of Learning*, 5(1), 5. <https://doi.org/10.1038/s41539-020-0064-y>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 5998–6008.
- Wang, M.-T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction*, 43, 16–26. <https://doi.org/10.1016/j.learninstruc.2016.01.008>
- White, R., & Gunstone, R. (2014). *Probing understanding*. Routledge.
- Wu, T.-T., Lee, H.-Y., Li, P.-H., Huang, C.-N., & Huang, Y.-M. (2023a). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. *Journal of Educational Computing Research*. <https://doi.org/10.1177/07356331231191125>
- Wu, T.-T., Lee, H.-Y., Wang, W.-S., Lin, C.-J., & Huang, Y.-M. (2023b). Leveraging computer vision for adaptive learning in STEM education: Effect of engagement and self-efficacy. *International Journal of Educational Technology in Higher Education*, 20(1), 53. <https://doi.org/10.1186/s41239-023-00422-5>
- Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT Technology Review 2021. *Fundamental Research*, 1(6), 831–833. <https://doi.org/10.1016/j.fmre.2021.11.011>
- Zhou, M., & Lam, K. K. L. (2019). Metacognitive scaffolding for online information search in K-12 and higher education settings: A systematic review. *Educational Technology Research and Development*, 67(6), 1353–1384. <https://doi.org/10.1007/s11423-019-09646-7>
- Zhu, M., Bonk, C. J., & Doo, M. Y. (2020). Self-directed learning in MOOCs: Exploring the relationships among motivation, self-monitoring, and self-management. *Educational Technology Research and Development*, 68(5), 2073–2093. <https://doi.org/10.1007/s11423-020-09747-8>
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17.
- Zimmerman, B. J. (2000). Chapter 2—Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166–183. <https://doi.org/10.3102/0002831207312909>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Hsin-Yu Lee** received his B.S. degree in the Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan, R.O.C., in 2019, and received the M.S. degree in the Department of Engineering Science, National Cheng-Kung University, Taiwan, R.O.C., in 2021. He is currently pursuing his Ph.D. degree in engineering science at National Cheng Kung University, Taiwan, R.O.C. His research interests include educational technology, learning analytics, computer vision and artificial intelligence.

**Pei-Hua Chen** received her B.S. degrees from the Department of Cultural and Creative Industries, National Chin-Yi University of Technology, Taiwan, R.O.C., in 2023. She is currently pursuing her M.S. degree in engineering science at National Cheng Kung University, Taiwan. Her research interests include educational technology, design thinking, UI/UX design and artificial intelligence.

**Wei-Sheng Wang** received her B.S. and M.S. degrees from the Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taiwan, R.O.C., in 2018 and 2020, respectively. He is currently pursuing his Ph.D. degree in engineering science at National Cheng Kung University, Taiwan.

**Yueh-Min Huang** received the M.S. and Ph.D. degrees in Electrical Engineering from the University of Arizona in 1988 and 1991 respectively. He is a Chair Professor in Department of Engineering Science, National Cheng-Kung University, Taiwan. He has co-authored 3 books and has published more than 250 refereed journal research papers. His research interests include e-Learning, multimedia communications, wireless networks, and artificial intelligence. He is in the editorial board of several international journals in the area of educational technology, computer communications, and web intelligence, including 3 SSCI-indexed e-Learning ones. In e-Learning area, except of publishing more than 100 SSCI-indexed journal

papers and editing 3 special issues in SSCI-indexed journals, he was also serving as the directors of Disciplines of Applied Science Education and Innovative Engineering Education in Taiwan's Ministry of Science and Technology. Dr. Huang is a senior member of the IEEE and became Fellow of British Computer Society in 2011. Dr. Huang has received many research awards, such as Taiwan's National Outstanding Research Award in 2011 and 2014, given to Taiwan's top 100 scholars. According to a paper published in BJET, he is ranked no. 3 in the world on terms of the number of educational technology papers published in the period 2012 to 2017.

**Ting-Ting Wu** is a distinguished professor and director of the Graduate School of Technological and Vocational Education, National Yunlin University of Science and Technology, Taiwan. Her research covers such areas as technology-enhanced language learning and applications of advanced technology to support nursing education.