

Article

Designing and Evaluating a 5E-Structured GenAI Coach for Guided Inquiry: A Pedagogy-to-Prompt Engineering Framework

Teng-Chi Lin ^{1,2} , Yu-Ting Shih ¹  and Cheng-Hsuan Li ^{1,*} ¹ Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taichung 403514, Taiwan; cms110102@gm.ntcu.edu.tw (T.-C.L.); cms112103@gm.ntcu.edu.tw (Y.-T.S.)² Taichung Municipal Taichung First Senior High School, Taichung 404009, Taiwan

* Correspondence: chenghsuanli@gmail.com

Abstract

The challenge of designing generative AI (GenAI) tutors that are both pedagogically sound and effective for guided inquiry remains significant. This paper introduces and evaluates a replicable design framework—termed a Pedagogy-to-Prompt Engineering Framework—that systematically translates established pedagogical models into structured AI interactions. We engineered a 5E-structured GenAI coach by integrating the 5E Learning Cycle as the instructional architecture and the 5S Prompting Principles to govern the AI's dialogue. The coach was evaluated in a middle school chemistry context ($N = 60$) focusing on procedural skill acquisition for balancing chemical equations. A quasi-experimental study showed the GenAI group achieved significantly higher learning gains than a control group receiving traditional instruction ($t(58) = 2.646$, $p = 0.011$, Cohen's $d = 0.68$). Crucially, a Johnson-Neyman analysis revealed that the coach was particularly beneficial for students with lower prior knowledge (pre-test scores < 39.39), effectively narrowing the achievement gap. Furthermore, Lag Sequential Analysis of the interaction logs confirmed that the student-AI dialogue successfully adhered to the intended 5E pedagogical sequence (e.g., Engage $\rightarrow$ Explore transition, $z = 11.157$). This study demonstrates that the proposed framework is a viable method for creating effective, scalable AI-driven learning environments. Beyond chemistry, this approach is readily adaptable to other STEM disciplines requiring guided inquiry, such as physics and mathematics. By validating a low-code, pedagogy-first methodology, this work offers a scalable blueprint for instructional designers to bridge the gap between generative AI capabilities and rigorous educational standards.

Keywords: intelligent tutoring systems; lag sequential analysis; prompt engineering

Academic Editor: Andrea C. Borowczak

Received: 7 January 2026

Revised: 23 February 2026

Accepted: 27 February 2026

Published: 3 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The enduring challenge of scaling personalized instruction, famously quantified by Bloom's (1984) "2 Sigma Problem," highlights the profound impact of one-on-one tutoring on learning outcomes (Bloom, 1984). While scaling such individualized support has long been impractical due to resource constraints, the recent advent of Generative AI (GenAI)—particularly large language models (LLMs)—offers a promising and accessible pathway to address this challenge. With their capacity for adaptive, interactive dialogue, these models present a powerful new tool for creating personalized learning experiences at scale (Chen et al., 2024; Farah et al., 2023; Lo, 2023; López-Galisteo & Borrás-Gené, 2025).

However, the direct application of general-purpose GenAI models in educational settings is fraught with challenges. These models, while fluent, inherently lack the pedagogical knowledge and instructional scaffolding necessary for effective, guided learning

(Feldman-Maggor et al., 2024; Liu et al., 2024). Unlike traditional Intelligent Tutoring Systems (ITS), which were meticulously engineered with cognitive science principles to deliver adaptive instruction (Alrakhawi et al., 2023), raw GenAI often defaults to providing direct answers rather than fostering an inquiry process. This tendency to function as an information-retrieval system, rather than a pedagogical partner, can hinder robust conceptual development and critical thinking, potentially leading to epistemic laziness and metacognitive weakness (Yurt & Kuşci, 2026). This aligns with the distinction between ‘Fast AI’ and ‘Slow AI’ highlighted in the recent OECD outlook (Organisation for Economic Co-operation and Development (OECD) (2026)). While general purpose models often encourage ‘Fast AI’ behaviors (immediate answers), effective educational agents must facilitate ‘Slow AI’—prioritizing iterative dialogue, sustained inquiry, and human agency over speed (Organisation for Economic Co-operation and Development (OECD) (2026)). This gap highlights a critical engineering problem: how to systematically embed pedagogical intelligence into general-purpose LLMs to transform them from mere “answer machines” into effective learning coaches.

To address this challenge, this paper introduces and demonstrates a replicable Pedagogy-to-Prompt Engineering Framework designed to systematically infuse pedagogical structure into a GenAI agent. This framework operationalizes educational theory through meticulous prompt design. Specifically, we employ the 5E Learning Cycle (Engage, Explore, Explain, Elaborate, Evaluate) (Bybee et al., 2006; Tanner, 2010) as the high-level instructional architecture to structure the entire learning progression. Concurrently, we utilize the 5S Prompting Principles (Set the Scene, Be Specific, Simplify Language, Structure Output, Share Feedback) (Farah et al., 2023; Tassoti, 2024) to govern the fine-grained details of the coach’s dialogue, ensuring clarity and effectiveness in student-AI interactions.

To validate the efficacy of this framework, we engineered and evaluated a GenAI-based coach in a learning domain known for its persistent challenges: balancing chemical equations in middle school chemistry (Agung & Schwartz, 2007). This domain is particularly suitable as it requires learners to navigate between abstract symbolic representations and the principle of conservation of matter, often leading them to rely on rote algorithmic methods rather than conceptual understanding (Agung & Schwartz, 2007; Taber, 2019; Treagust et al., 2000). Therefore, it serves as an ideal testbed to assess a coach designed to provide individualized, scaffolded guidance for complex procedural skills.

This study therefore aims to evaluate the effectiveness of the proposed GenAI-based coaching approach, guided by our design framework. To achieve this, this study addresses the following research questions:

- (1) How effective is the GenAI coach, designed using the Pedagogy-to-Prompt Framework, in improving students’ procedural skills in balancing chemical equations compared to traditional teacher-led instruction?
- (2) To what extent do the student-AI interaction patterns adhere to the coach’s intended 5E-structured pedagogical sequence?

2. Literature Review

2.1. Personalized Learning Technologies: From ITS to Generative AI

The pursuit of adaptive instruction to address the challenges of personalized learning has long been a central theme in educational technology, leading to the development of ITS. ITS are computer-based platforms, rooted in artificial intelligence and cognitive science, designed to simulate expert human tutors by providing personalized instruction and feedback through the real-time assessment of learner progress (Alrakhawi et al., 2023). These systems have demonstrated considerable promise across a variety of domains, including mathematics (Lu et al., 2021), professional training, and scientific reasoning (Cai et al.,

2011; Fang et al., 2022; Putra et al., 2021; Samra et al., 2016). Among the most well-studied examples is AutoTutor (Graesser, 2016), an ITS that leverages advanced natural language dialogue to engage students in multi-turn conversations on subject-specific content. The pedagogical intelligence of AutoTutor is exemplified by its adaptive scaffolding cycle, detailed in Figure 1. The system provides progressively deeper support contingent on the student's performance: if a student's answer is incorrect, the tutor first pumps for more information, then offers hints, moves to specific prompts, and only provides an assertion of the correct answer if the student continues to struggle (Graesser, 2016). This systematic, theory-driven approach has proven effective in various fields such as reading (McNamara et al., 2006), scientific reasoning (Cai et al., 2011), and health education (Wolfe et al., 2015) and is a hallmark of classic ITS.

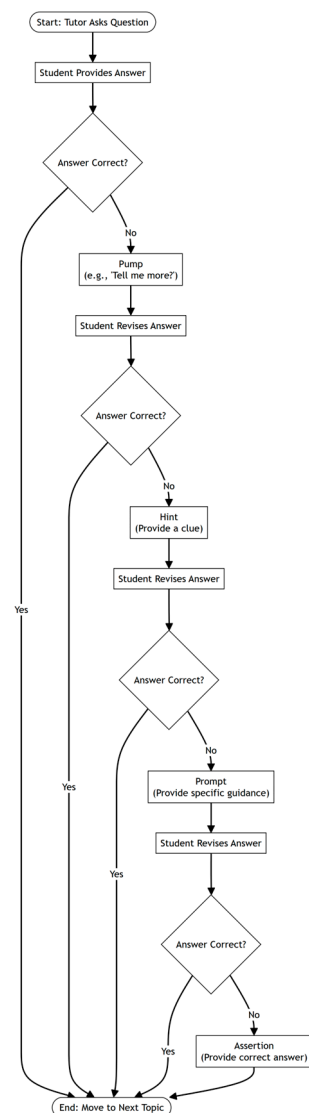


Figure 1. A detailed flowchart of AutoTutor's adaptive scaffolding cycle. The system provides progressively more specific support (Pump, Hint, Prompt, Assertion) contingent on the student's incorrect responses.

Despite their demonstrated effectiveness, the development of classic ITS like AutoTutor remained resource-intensive, which limited their broader adoption (Alrakhawi et al., 2023). Building on this research, the recent rise of Generative AI (GenAI)-especially large language models like ChatGPT(GPT-4)-marks a major shift in educational technology, offering a more accessible alternative. These models can generate human-like text, engage

in adaptive, context-aware dialogue, and support personalized learning through real-time feedback, guidance, and assessment (Lo, 2023). Notably, Khan Academy's GPT-4-based Khanmigo illustrates this potential as a powerful tutor and teaching assistant (Kshetri, 2023). These pre-trained LLMs significantly reduce development barriers, offering a more scalable path to achieving the one-on-one tutoring effectiveness long envisioned by Bloom (Chen et al., 2024; Farah et al., 2023; López-Galisteo & Borrás-Gené, 2025).

Recent research has further explored integrating GenAI into established instructional models. For instance, Alan and Yurt (2024) demonstrated that a ChatGPT-aided flipped learning model significantly enhanced students' academic achievement and motivation. In their approach, AI served primarily as a preparatory resource for concept exploration. However, a distinction must be made between such 'AI-supported' instruction and 'pedagogy-embedded' coaching. aligning with the 'Slow AI' approach advocated by the OECD (Organisation for Economic Co-operation and Development (OECD) (2026)), our framework embeds the instructional cycle directly into the AI's interaction logic to prioritize sustained inquiry over immediate information retrieval.

2.2. The Role of Pedagogy in Learning Technologies

The effectiveness of any learning technology, particularly AI-driven coaches, hinges not only on its technical capabilities but, more critically, on its underlying pedagogical foundation. Simply deploying a conversational AI often overlooks the need for a pedagogical structure, which is essential for guiding the learning process. Without a deliberate design, AI-generated content may lack the pedagogical content knowledge (PCK) required to address common misconceptions and foster deep understanding (Feldman-Maggor et al., 2024; Liu et al., 2024). The quality of the educational interaction is therefore highly dependent on the pedagogical soundness of the prompts guiding the AI, as poorly designed inputs can lead to ambiguous or counterproductive outputs that hinder learning (Farah et al., 2023). Thus, a systematic approach to embedding pedagogy into the AI's core design is necessary to elevate it from a simple information tool to an effective instructional partner.

2.3. Prompt Engineering for Pedagogical AI

If a pedagogical model like the 5E Learning Cycle provides the architectural blueprint for an AI coach, then prompt engineering is the critical construction method used to implement that design. Prompt engineering is the process of carefully crafting inputs (prompts) to guide a GenAI model's output, ensuring that its responses are accurate, pedagogically sound, and aligned with specific learning goals. The effectiveness of a GenAI-based learning tool is therefore critically dependent on the quality of its underlying prompts, as poorly designed inputs can result in ambiguous or inappropriate outputs that hinder the learning process (Farah et al., 2023). This highlights the need for a systematic methodology to guide the design of educational prompts, particularly for non-programmer educators seeking to create structured AI interactions.

To address the need for a systematic methodology, this study leveraged the 5S Prompting Principles framework. Identified as beneficial in educational contexts, its key elements include: Set the Scene (defining the AI's role), Be Specific (providing clear instructions), Simplify Language (using accessible terminology), Structure Output (defining the response format), and Share Feedback (enabling iterative refinement) (AI for Education, n.d.; Farah et al., 2023). Adherence to such principles is vital, as research indicates that even higher education students can struggle with optimal prompt formulation, making a structured approach particularly important when designing systems for younger learners (Tassoti, 2024). Therefore, the 5S framework provided the systematic methodology for translating the 5E model's pedagogical goals into concrete, effective prompts for the AI coach.

3. Methodology

3.1. An Implementation of the Pedagogy-to-Prompt Framework

The GenAI coach evaluated in this study was developed as a custom GPT utilizing the creation features within the OpenAI platform, specifically leveraging the GPT-4 model architecture available during the study period. The core of our design contribution lies not in the development of a novel base model, but in the systematic prompt engineering process used to implement the pedagogical framework detailed in the previous chapter. This development process utilized the no-code capabilities of the platform, facilitating the creation of a pedagogically structured AI tutor without requiring specialized programming expertise. The overall system architecture, illustrating how pedagogical inputs and a knowledge base are integrated through our core framework, is depicted in Figure 2. The configuration interface for the custom GPT is shown in Figure 3, while the resulting student-facing user interface is depicted in Figure 4.

The 5E Learning Cycle served as the foundational pedagogical framework structuring the AI-facilitated learning process. Each stage was intentionally aligned with targeted instructional goals, guiding students progressively through conceptual and procedural development in balancing chemical equations. For instance, in the “Engage phase,” the coach used relatable analogies to spark interest. During the “Explore phase,” it used Socratic dialogue to help students manipulate equations and form hypotheses. The “Explain phase” focused on core concepts like the conservation of matter, while the “Elaborate phase” challenged students to apply their knowledge to more complex problems. Finally, the “Evaluate phase” embedded formative assessment through scenario-based interactions. This structured alignment between the AI’s interactions and the 5E model is further detailed in Table 1, ensuring the learning experience remained inquiry-driven, personalized, and pedagogically coherent.

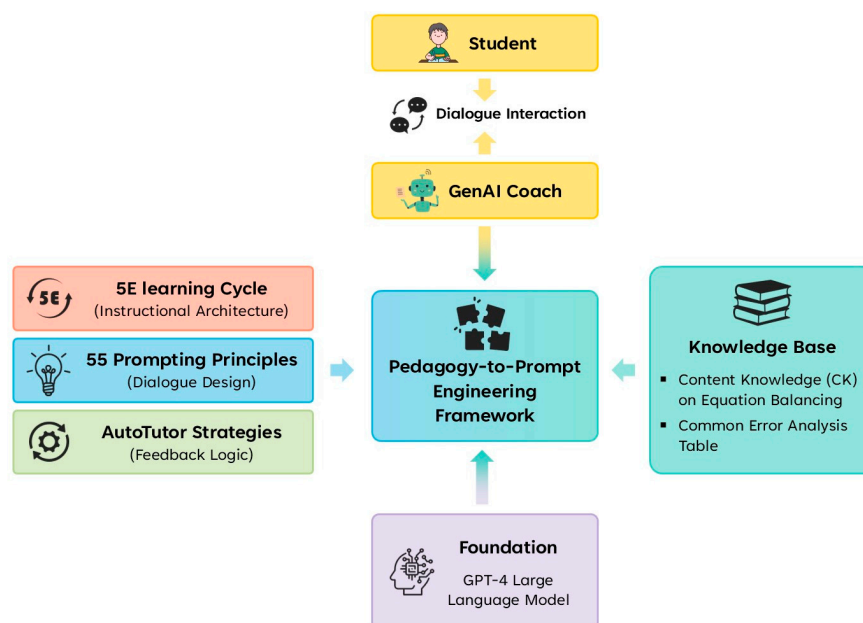


Figure 2. System architecture of the GenAI coach. The proposed Pedagogy-to-Prompt Engineering Framework serves as the central component, integrating pedagogical models (5E, 5S, AutoTutor) and a domain-specific knowledge base to structure the dialogue generated by the foundational GPT-4 model.

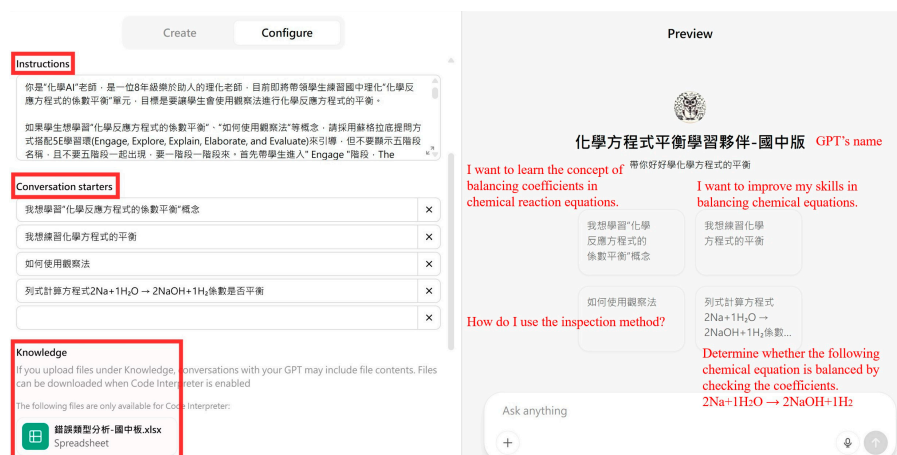


Figure 3. Configuration interface of the GenAI-based coach in the OpenAI platform, showing the fields for inputting instructions, conversation starters, and custom knowledge files. English translations of the original Chinese interface are overlaid in red.



Figure 4. The student user interface of the GenAI-based coach, showing pre-defined conversation starters to facilitate interaction. English translations of the original Chinese interface are overlaid in red.

The specific chemistry content knowledge (CK) integrated into the AI coach focused on balancing chemical equations using the inspection method, a common pedagogical approach for this topic at the middle school level (Harjadi, 1986; Herndon, 1997; Kolb, 1978, 1981). This topic was strategically selected as the experimental domain because it necessitates both algorithmic execution and conceptual adherence to the Law of Conservation of Mass, making it an ideal candidate for evaluating 5E-guided inquiry. Furthermore, the instructional logic of this 5E-Structured GenAI Coach—using prompts to constrain dialogue and scaffold procedural steps—is readily transferable to other STEM topics requiring structured problem-solving, such as stoichiometry calculations in chemistry or solving kinematic equations in physics. To ensure systematic guidance, the AI's core operational prompts explicitly encoded the key logical steps of this method, such as: (1) identifying a key element to start with; (2) determining coefficients for related compounds; (3) continuing sequentially for other elements; and (4) adjusting all coefficients to the simplest whole

number ratio. Crucially, the prompts also instructed the AI to rewrite the partially balanced equation after each step, making the progression transparent and easier for students to track.

Table 1. Implementation of the 5E Learning Cycle in the GenAI-based Coach.

5E Learning Cycle	Chemical Equation Balance Teaching Content	GenAI-Based Differentiated 1:1 Self-Paced Coach Prompt
Engage	Spark students’ interest in learning how to balance chemical equations through a relatable everyday analogy.	The teacher introduces the topic of balancing coefficients in chemical reaction equations by using a relatable everyday analogy, thereby sparking students’ interest.
Explore	Enable students to experience the challenges and discoveries of balancing by manipulating unbalanced chemical equations hands-on.	Students are encouraged to ask questions and hypothesize about the subject, actively engaging in discussions to explore the concept further. This phase involves questioning and hypothesizing, guided by the teacher through the Socratic method.
Explain	Delve into the principles of conservation of matter and demonstrate step-by-step how to balance chemical equations through guided instruction.	The teacher helps students to understand the scientific principles behind the experiments, confirming or revising their hypotheses based on their observations.
Elaborate	Expand students’ understanding of balancing concepts by applying them to more complex chemical equations and real-world scenarios.	This stage encourages the application of learned concepts to new situations, such as analyzing complex chemical equations, thereby enhancing their critical thinking and problem-solving skills.
Evaluate	Assess students’ mastery of balancing chemical equations through tests and real-case scenario analyses.	In the final phase, the teacher assesses the students’ overall understanding of the topic through dynamic evaluation methods, encouraging them to think about how to apply their new knowledge to daily life and other fields.

Finally, to enhance the coach’s capability to diagnose and address specific student difficulties, a curated “Common Error Analysis Table” was provided to the GPT as a Supplementary knowledge file. This knowledge base is structured around seven pedagogical fields designed to map specific errors to instructional feedback including “Item Type (e.g., Basic vs. Complex),” “Question (Target Equation),” “Correct Answer,” “Common Student Error,” “Reason for the Error,” “Underlying Misconception,” and “Corrective Guidance.” For instance, in a basic combustion problem ($\text{CH}_4 + \text{O}_2 \rightarrow \text{CO}_2 + \text{H}_2\text{O}$), if a student inputs an error caused by overlooking subscripts (e.g., balancing Oxygen incorrectly), the table explicitly identifies the underlying misconception as “Overlooked the subscript 2 for O in CO_2 .” Consequently, instead of merely stating the correct coefficient, the system retrieves the specific corrective guidance: “Pay attention to subscripts; each CO_2 molecule contains two oxygen atoms,” thereby scaffolding the student to self-correct based on chemical principles. To ensure the AI adherence to specific pedagogical standards, the full prompt protocols are provided in Appendix A. These instructions operationalize the 5S Prompting Framework (Table 2) to guide the agent through the 5E Learning Cycle, strictly enforcing Socratic dialogue logic while validating Content Knowledge (CK) at each interaction step.

The GenAI coach was implemented as a custom GPT on the OpenAI platform, utilizing the GPT-4 model architecture. To maintain a focused learning environment, external web browsing capabilities were disabled; however, the Code Interpreter tool was enabled specifically to allow the system to process the uploaded pedagogical knowledge base. Regarding privacy and safety, we implemented a multi-layered safeguard approach:

1. Strict Platform Configuration: We manually configured the platform settings to opt-out of model training, ensuring that student interaction data would not be used to improve the base models.
2. Automated Safety Filtering: We leveraged the platform’s baseline safety moderation layer to automatically filter harmful content.
3. Pedagogical Guardrails: We embedded strict system prompts to constrain the AI’s scope and reject off-topic dialogue.
4. Physical Oversight: The teacher provided continuous physical supervision during the intervention. Post-hoc analysis of interaction logs confirmed that all dialogues remained within the intended educational scope.

Table 2. Application of 5S Prompting Principles in GenAI-based Coach Design.

5S Prompting Framework	The Key Point of the S	GenAI-Based Differentiated 1:1 Self-Paced Coach Prompt
Set the Scene	Specify the chatbot’s role to receive more focused answers.	You are a “Chemistry AI” teacher, an approachable and helpful science teacher for 8th graders, currently guiding students through the middle school chemistry unit on “Balancing Coefficients in Chemical Reaction Equations.” . . .
Be Specific	Give detailed instructions to enhance response relevance and accuracy.	This content specifies how content knowledge (CK) is applied in the design of prompts to guide students through the chemical equation balancing process.
Simplify your language	Use straightforward language to ensure clear understanding.	The language should be suitable for 8th-grade students, with each response or prompt kept as concise and to-the-point as possible.
Structure the output	Guide the chatbot on the preferred format for its responses.	This included implementing the multi-step hint-prompt-assert feedback cycle and requiring Socratic dialogue formats. . .
Share feedback	Provide feedback to improve the chatbot’s responses.	If a student questions your answer, be sure to double-check its accuracy. If there is an error, acknowledge it and explain the root cause of the mistake to the student.

3.2. Evaluation Design

Participants were 60 eighth-grade students (approx. age 13–14) from two purposively sampled, comparable intact classes at a public middle school in central Taiwan. Prior to the experiment, informed consent was obtained from the legal guardians of all participating students, and assent was also obtained from the students themselves. All participants were assured of the confidentiality and anonymity of their data throughout the research process. The classes were assigned to either an experimental group ($n = 30$; GenAI-based 1:1 self-paced coach) or a control group ($n = 30$; traditional teacher-led instruction). To minimize instructor-related confounding effects, the same experienced science teacher taught both classes.

Prior to the main analysis, statistical assumptions were examined. According to the Central Limit Theorem, the sampling distribution of the mean is assumed to be normal as the sample size for each group was equal to or greater than 30, justifying the use of parametric tests. Regarding the pre-test comparison, due to a violation of the homogeneity of variance assumption (Levene’s test, $p = 0.028$), a Welch’s t -test was conducted. Results confirmed that there was no significant prior knowledge difference in balancing chemical equations between the two groups ($t(51.68) = 1.719$, $p = 0.092$), indicating baseline equiva-

lence. The experimental group ($M = 48.00$, $SD = 32.63$) and the control group ($M = 35.60$, $SD = 22.28$) showed no statistically significant differences in their starting levels.

To ensure the internal validity of the experiment, strict control of variables was implemented to maintain equivalence between the two conditions:

1. Instructional Time: Both groups engaged in the learning activity for the exact same duration (45 min).
2. Teacher Support: The same experienced science teacher facilitated both classes to eliminate instructor bias. In the experimental group, the teacher acted solely as a technical facilitator and did not provide additional content scaffolding.
3. Learning Materials: Both groups covered identical curriculum objectives and practiced with problems of equivalent difficulty derived from the standard textbook.

The orientation provided to the GenAI group prior to the intervention was strictly procedural (i.e., interface navigation) and did not include any additional subject matter instruction.

The study was implemented during regular eighth-grade science classes, with the overall experimental flow depicted in Figure 5. Initially, the participating science teacher delivered a standard introductory lesson on chemical reactions and equation balancing concepts to both the experimental and control groups. Following this preparation, all students completed a pre-test to assess their baseline knowledge regarding balancing chemical equations. The core intervention phase took place during a single 45-min class period. Students in the experimental group individually accessed the “Chemistry Equation Balancing Learning Partner” via computers, engaging in self-paced learning guided by the AI coach, while students in the control group received traditional whole-class instruction from the same teacher for the same duration, which involved teacher explanations, worked examples, and a practice worksheet. Upon completion of the 45-min instructional intervention, the post-test assessment was administered to all participating students from both groups.

The pre-test and post-test consisted of the same set of questions, which included true/false and fill-in-the-blank items focused on balancing chemical equations. The pre-test and post-test consisted of identical items focused on balancing chemical equations. To ensure validity and transparency, the assessment comprised 10 items with a total score of 100 points, weighted based on difficulty:

1. Conceptual Understanding (4 items): True/False questions (e.g., verifying atom conservation in $2\text{Mg} + \text{O}_2 \rightarrow 2\text{MgO}$), weighted at 4 to 6 points each.
2. Procedural Application (6 items): Fill-in-the-blank questions (e.g., balancing $__\text{Fe} + __\text{O}_2 \rightarrow __\text{Fe}_2\text{O}_3$ decomposition), weighted from 8 to 23 points each based on complexity.

These questions were derived from textbook exercises to ensure reliability. To ensure content validity, a Two-way Specification Table shown in Table 3 was constructed to align the 10 assessment items with specific cognitive levels and curriculum objectives. The instrument was balanced between Comprehension (4 items, focusing on identifying balanced states and conservation principles) and Application (6 items, requiring the procedural calculation of coefficients). Items ranged from simple binary compounds to complex polyatomic reactions (e.g., NaHCO_3 decomposition), with scoring weights (4–23 points) strictly mapped to item difficulty. This structure ensures that the assessment measures both conceptual understanding and procedural mastery, as verified by experienced science teachers. Internal consistency reliability yielded acceptable Cronbach’s alpha values of 0.716 for the pre-test and 0.739 for the post-test administration.

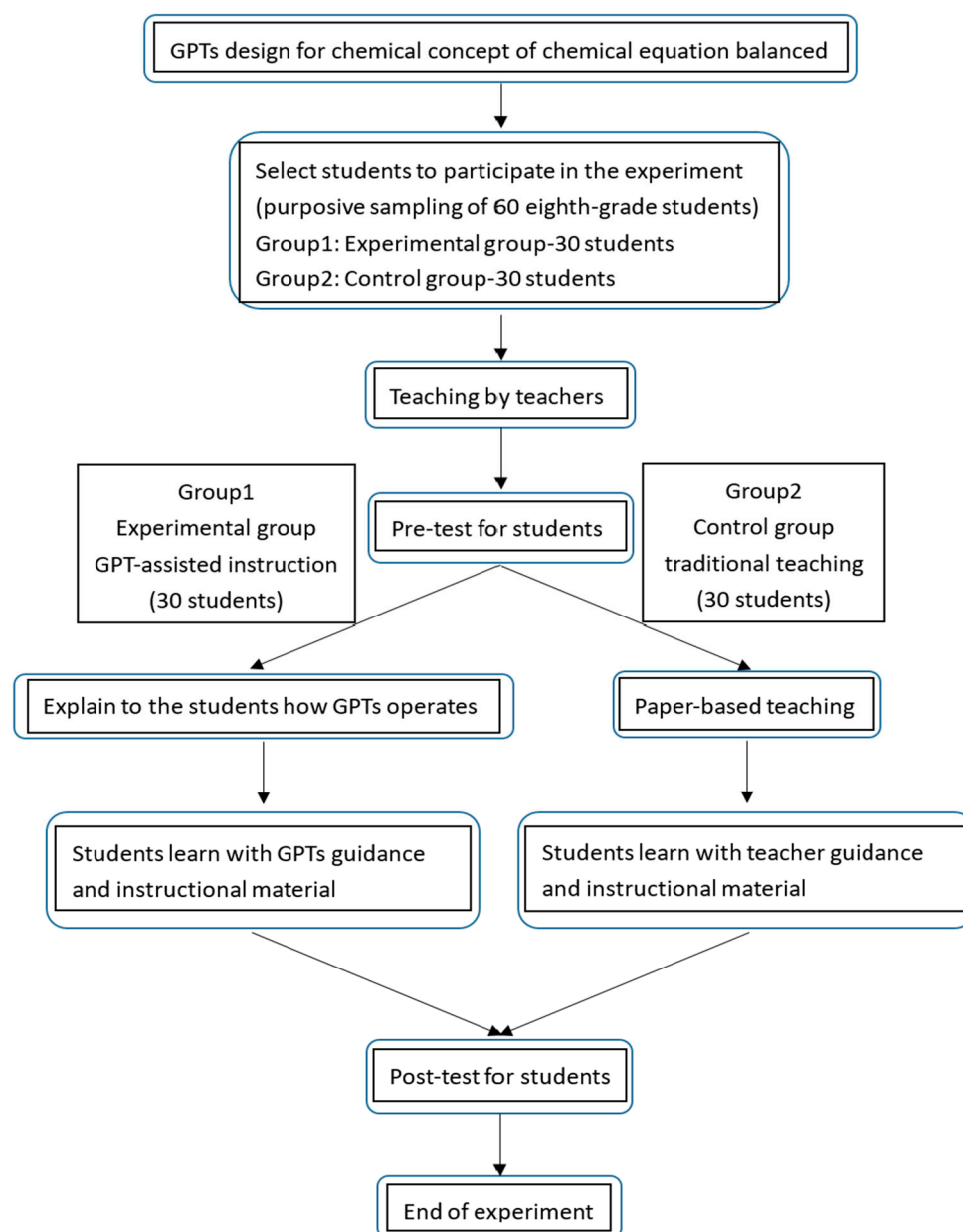


Figure 5. Experiment procedure.

Table 3. Two-way Specification Table of the Pre-test and Post-test.

Objective Teaching Content	Comprehension	Application	Total Number
Determine whether the coefficients of a simple chemical equation are balanced	3	0	3
Determine whether the coefficients of a complex chemical equation are balanced.	1	0	1
Completely and correctly balance simple chemical equations	0	5	5
Correctly and completely balance complex chemical equations	0	1	1
Total Number	4	6	10

3.3. Data Analysis

To analyze the learning outcomes, the study employed descriptive statistics, independent and paired samples *t*-tests, and effect size calculations (Cohen’s *d*). Crucially, to evaluate the impact of the GenAI coach while controlling for prior knowledge, Analysis of Covariance (ANCOVA) was conducted using pre-test scores as covariates. To further investigate potential interaction effects and determine “for whom” the intervention was most effective, the Johnson-Neyman technique was employed to identify specific regions of significance along the continuum of prior knowledge.

To investigate the temporal dynamics of student-AI interaction patterns, Lag Sequential Analysis (LSA) was conducted. LSA is a statistical method designed to detect significant transitional probabilities between sequential behaviors, revealing how one specific action (e.g., AI prompting) likely triggers a subsequent response. In educational research, LSA is widely used to visualize learning processes, such as patterns of self-regulated learning and knowledge construction. In the context of this study, LSA is particularly critical for validating the instructional design, allowing us to determine whether the intended 5E instructional phases (Engage, Explore, Explain, Elaborate, Evaluate) actually emerged in the students’ dialogue.

The coding unit was defined as a single conversational turn, consisting of a student’s input and the subsequent AI response. A coding scheme was developed to map these interactions to the five stages of the 5E Learning Cycle. To ensure consistency, specific linguistic markers and pedagogical functions were defined for each code. Table 4 presents the codebook along with representative examples from the study. To ensure the reliability of the coding, two researchers with expertise in science education independently coded the dialogue logs of 14 randomly selected students (comprising a total of *N* = 106 conversational turns). The inter-rater reliability was calculated using Cohen’s Kappa coefficient. The analysis yielded a Percent Agreement of 95.28% and a Cohen’s Kappa value of $\kappa = 0.922$, indicating almost perfect agreement between the coders. Any discrepancies in the initial coding were resolved through discussion to establish a consensus dataset. Notably, due to the strict constraints imposed by the system instructions, all AI responses in the sampled dataset were pedagogically relevant and successfully mapped to the 5E stages; thus, no turns were coded as “Other” or excluded from the analysis. Finally, to determine statistical significance of the transitions, adjusted residuals (*z*-scores) were calculated; a *z*-score greater than 1.96 indicates a significant sequence (*p* < 0.05). This analysis reveals not just what students did, but how the GenAI coach scaffolded their learning path moment by moment.

Table 4. Coding Scheme for 5E-Structured Dialogue Analysis with Representative Examples from Study Log.

5E Stage (Code)	Operational Definition (AI Coaching Function)	Example from Dialogue Logs (Translated)
Engage	Facilitating Engagement: The AI prompts the student to connect prior knowledge or daily life contexts to spark interest and curiosity.	(Student 13) AI: “Hello! . . . Before we start, I’d like to ask: Do you think in daily life. . .” (Goal: Activating student’s prior experience)
Explore	Guiding Exploration: The AI encourages the student to formulate hypotheses, test boundaries, or ask “what if” questions before providing formal rules.	(Student 2) Student: “Can I combine O ₂ and H ₂ O?” AI: “In a chemical equation, we cannot combine different compounds. . . [Explains constraint to guide hypothesis]” (Goal: Scaffolding student’s inquiry)

Table 4. Cont.

5E Stage (Code)	Operational Definition (AI Coaching Function)	Example from Dialogue Logs (Translated)
Explain	Scaffolding Explanation: Instead of giving direct answers, the AI guides the student to observe evidence or clarify principles to build their own understanding.	(Student 10) Student: “Can you give me a hint?” AI: “Of course! First, let’s observe the number of occurrences of each element in this equation.” (Goal: Guiding observation to support reasoning)
Elaborate	Scaffolding Elaboration: The AI challenges the student to apply learned concepts to new, more complex contexts (transfer of learning).	(Student 2) AI: “Great! Let’s try a complex-type chemical equation balancing exercise.” (Goal: Pushing student to apply skills in harder tasks)
Evaluate	Verifying Mastery: The AI assesses the student’s procedural accuracy and conceptual understanding through specific problem checks.	(Student 9) AI: “You submitted the answer ‘1, 5, 3, 4’. Let’s check if this answer correctly balances the equation.” (Goal: Confirming student’s learning outcome)

4. Results

4.1. Learning Achievement and Group Comparisons

The descriptive statistics and the results of the within-group paired samples *t*-tests are presented in Table 5. For the experimental (GenAI-assisted) group, scores increased from a pre-test mean of 48.00 (*SD* = 32.63) to a post-test mean of 72.93 (*SD* = 27.04). For the control (traditional instruction) group, scores increased from a pre-test mean of 35.60 (*SD* = 22.28) to a post-test mean of 51.93 (*SD* = 34.04). Notably, the standard deviation for the experimental group decreased, suggesting more consistent learning outcomes, whereas the standard deviation for the control group increased, indicating greater performance variability after the intervention. The paired sample *t*-test results confirmed that these learning gains were statistically significant for both the experimental group ($t(29) = 5.311, p < 0.001$, Cohen’s $d = 0.832$) and the control group ($t(29) = 3.456, p = 0.002$, Cohen’s $d = 0.568$).

Table 5. Descriptive Statistics and Paired Sample *t*-test Results for Within-Group Gains.

Group	<i>n</i>	Pre-Test <i>M</i> (<i>SD</i>)	Post-Test <i>M</i> (<i>SD</i>)	Mean Diff.	<i>t</i>	Cohen’s <i>d</i>
Experimental	30	48.00 (32.63)	72.93 (27.04)	24.93	5.31 ***	0.832
Control	30	35.60 (22.28)	51.93 (34.04)	16.33	3.46 **	0.568

Note. ** $p < 0.01$. *** $p < 0.001$.

To evaluate the impact of the GenAI coach on students’ learning performance while controlling for prior knowledge, an ANCOVA was conducted. While the assumption of homogeneity of regression slopes was technically met ($p = 0.061$), this marginal interaction suggested that the treatment effect might vary across different levels of prior knowledge. After adjusting for pre-test scores, the main effect of the group was marginally significant, $F(1, 57) = 3.79, p = 0.26, \eta^2_p = 0.062$. The adjusted mean scores indicated that the experimental group ($M = 68.72$) achieved higher post-test performance than the control group ($M = 56.14$). To further investigate “for whom” the intervention was most effective, the Johnson-Neyman technique was employed to identify the region of significance. The analysis revealed a critical boundary at a pre-test score of 39.39 (as shown in Figure 6). Specifically, for students with pre-test scores below 39.39 (comprising 43% of the experimental group, $n = 13$), the GenAI-based coach significantly outperformed traditional instruction ($p < 0.05$). This suggests that the 5E-structured coach provided essential scaffolding that was

particularly beneficial for learners with lower baseline knowledge, effectively narrowing the achievement gap.

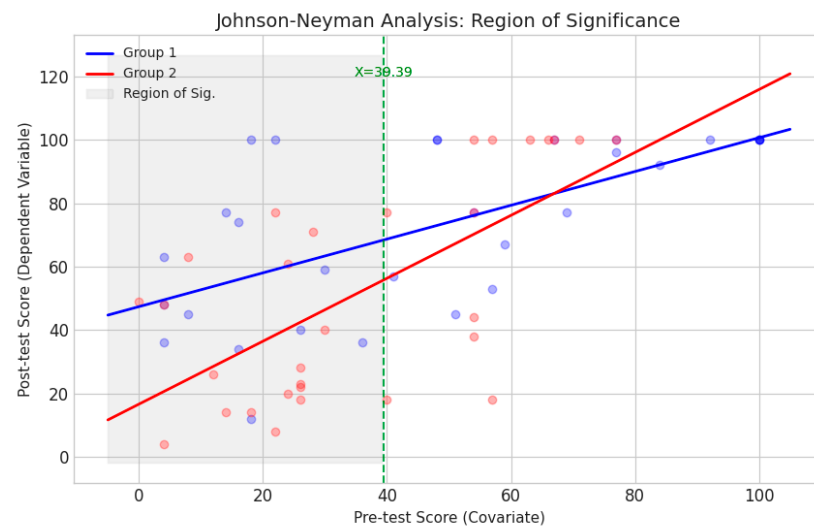


Figure 6. Johnson-Neyman plot of the interaction between group and pre-test scores on post-test performance.

To evaluate the relative effectiveness of the GenAI-assisted coaching compared to traditional instruction, the post-test scores of the two groups were compared. An independent samples *t*-test was conducted on the post-test scores between the experimental group ($M = 72.93$, $SD = 27.04$) and the control group ($M = 51.93$, $SD = 34.04$). The results indicated a statistically significant difference between the groups ($t(58) = 2.646$, $p = 0.011$), with a medium-to-large effect size (Cohen's $d = 0.68$), suggesting that students who used the GenAI coach significantly outperformed those who received traditional instruction. Figure 7 visually illustrates this difference, showing the distinct post-test score distributions between the two groups and highlighting the higher performance concentration in the experimental group.

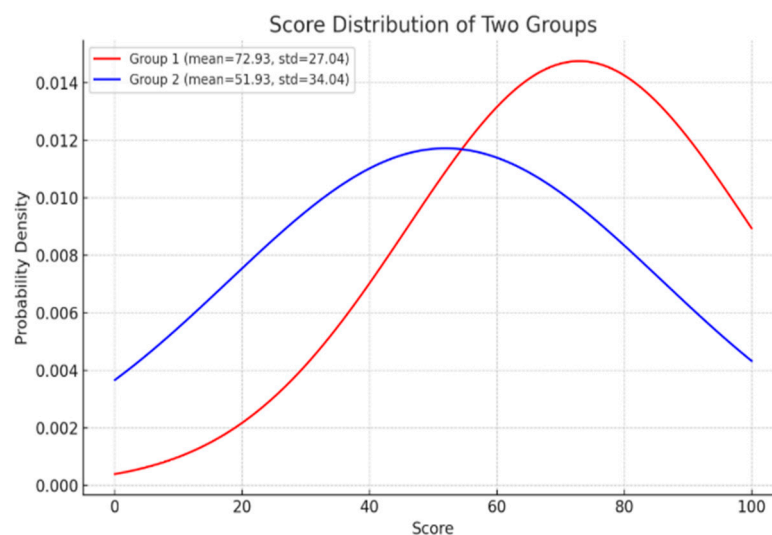


Figure 7. Post-test score distributions for experimental group (Group 1) and control group (Group 2).

4.2. Analysis of Student-AI Interaction Patterns

To investigate whether the experimental group's interactions with the GenAI-based coach followed the intended pedagogical sequence of the 5E instructional model, LSA was conducted on the coded dialogue logs. The analysis focused on identifying statistically significant transitions ($z > 1.96$) between the five defined stages, with the resulting state transition diagram presented in Figure 8. The analysis revealed a clear, predominantly linear progression through the 5E learning cycle, with a strong and highly significant transition observed from the "Engage" to the "Explore" phase ($z = 11.157$), followed by significant transitions from "Explore" to "Explain" ($z = 4.387$), "Explain" to "Elaborate" ($z = 2.462$), and "Elaborate" to "Evaluate" ($z = 2.578$). Collectively, these significant transitions indicate that the student interactions, as mediated by the GenAI coach, generally adhered to the structured inquiry-based pathway designed into the intervention, with no significant backward loops or skipping of stages detected.

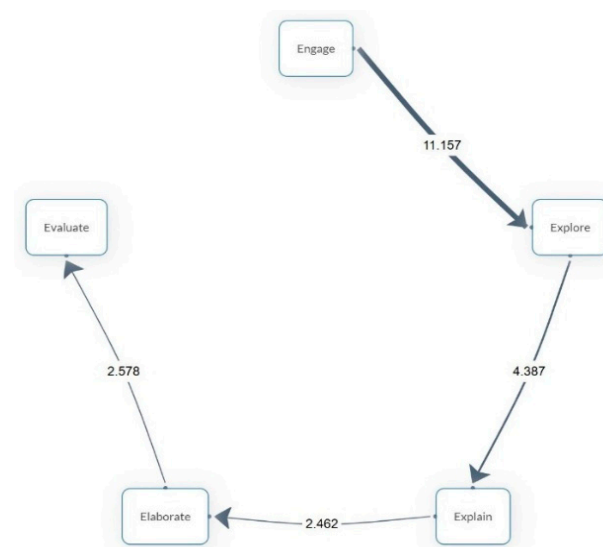


Figure 8. State Transition Diagram of 5E Learning Stages based on Lag Sequential Analysis.

5. Discussion

5.1. Interpretation of Findings

The results of this study provide compelling evidence for the effectiveness of the GenAI-based 1:1 self-paced coach, which significantly surpassed traditional teacher-led instruction in enhancing students' procedural skills (Agung & Schwartz, 2007). This superior performance aligns strongly with Bloom's (1984) seminal work on the 2 Sigma Problem (Bloom, 1984), suggesting that well-designed AI tutors can indeed begin to approximate the substantial learning benefits of individualized human tutoring, thereby offering a scalable solution to this long-standing educational challenge. A key contributor to this outcome was likely the personalized and adaptive nature of the 1:1 interaction inherent in the GenAI coach. Unlike the whole-class approach, the AI coach allowed students to learn at their own pace, revisit concepts as needed, and receive guidance tailored to their specific queries and progress. This resonates with the core principles of ITS, which aim to customize instruction dynamically based on learner needs (Alrakhawi et al., 2023; Graesser, 2016).

Secondly, the immediate, iterative, and pedagogically structured feedback provided by the GenAI coach was another significant factor in its success. Drawing inspiration from AutoTutor's effective dialogue strategies (Fang et al., 2022; Graesser, 2016), the coach was designed not merely to provide answers but to guide students through a constructive learning process. As illustrated in the interaction examples (Figure 9), the coach effectively

employed a multi-step feedback cycle: it provided hints when students first encountered difficulties, offered analytical prompts to encourage reflection upon subsequent errors, and finally asserted the correct procedure. This immediate and targeted feedback contrasts sharply with the potentially delayed and less personalized feedback typical of the control group's whole-class review format. Such timely and scaffolded feedback is critical for addressing misconceptions in complex subjects like chemistry (Jammeh et al., 2023), preventing students from reinforcing incorrect procedures or ideas.

YOU
How do I do this?

Chemical Equation Learning Partner (GPTs)
Okay, let me give you some hints to help you balance this equation:
First, let's look at the number of each element in the reactants and products. In the equation $\text{H}_2 + \text{O}_2 \rightarrow \text{H}_2\text{O}$, we have two elements: hydrogen (H) and oxygen (O).

1. Hydrogen (H) has 2 atoms in the reactant H_2 and also 2 atoms in the product H_2O .
2. Oxygen (O) has 2 atoms in the reactant O_2 but only 1 atom in the product H_2O .
Our goal is to ensure that the number of atoms of each element is equal on both sides of the reactants and products. Now, you can try adjusting the coefficients to balance the number of oxygen atoms. Think about it, how can we make the number of oxygen atoms equal on both sides?

(a)

Chemical Equation Learning Partner:
When balancing chemical equations, we can start with elements that appear only once and are in unequal amounts in the reactants and products. In this equation, oxygen (O) appears as O_2 on the reactant side and as Al_2O_3 on the product side. This means we need to adjust the coefficients to make the quantity of oxygen equal on both the reactant and product sides.
Here's a little tip: If we assume the coefficient for Al_2O_3 is 1, then there are 3 oxygen atoms on the product side. Next, what should we do to ensure that the number of oxygen atoms on the reactant side equals those on the product side? This may affect the coefficients for Al and O_2 .
Now, try to balance the equation based on this hint and provide the coefficients for each reactant and product.

(b)

YOU
2.74.6

Chemistry Equation Learning Partner
Let's check the new answer you provided: $2\text{C}_2\text{H}_6 + 7\text{O}_2 \rightarrow 4\text{CO}_2 + 6\text{H}_2\text{O}$.
Let's verify the balance of elements:

- For carbon (C), on the reactant side there are 2 C_2H_6 , so there are a total of 4 carbon atoms. On the product side there are 4 CO_2 , which also have 4 carbon atoms. Carbon is balanced.
- For hydrogen (H), on the reactant side there are 2 C_2H_6 , each containing 6 hydrogen atoms, thus there are a total of 12 hydrogen atoms. On the product side, there are 6 H_2O , which also have 12 hydrogen atoms. Hydrogen is balanced.
- For oxygen (O), on the product side there are 4 CO_2 and 6 H_2O , resulting in a total of $4 \times 2 + 6 = 14$ oxygen atoms. On the reactant side there are 7 O_2 , each containing 2 oxygen atoms, thus there are a total of 14 oxygen atoms. Oxygen is also balanced.

Your answer is completely correct! $2\text{C}_2\text{H}_6 + 7\text{O}_2 \rightarrow 4\text{CO}_2 + 6\text{H}_2\text{O}$ is indeed a balanced chemical equation. You balanced all elements in the equation very well, which shows that you have a good understanding of how to balance chemical equations.

Well done! This kind of exercise can help you better understand chemical reactions and the concept of conservation of mass. Are you ready for the next challenge?

(c)

Figure 9. Examples of the GenAI coach's adaptive dialogue strategies: (a) providing initial hints to help the student analyze the number of atoms in the reactants and products; (b) offering a specific analytical prompt to guide the student in adjusting coefficients; and (c) asserting the correct procedure by step-by-step verification of the student's submitted answer.

A particularly insightful finding that extends beyond the mean score gains is the differential impact of the two interventions on the score distributions, as indicated by the change in standard deviations. For the experimental group, the decrease in standard deviation from pre-test to post-test suggests that the GenAI coach had a homogenizing

effect, not only raising the average performance but also reducing the performance gap among students. Conversely, the increase in standard deviation for the control group suggests that the traditional, one-size-fits-all instruction may have inadvertently widened this gap, potentially creating a “Matthew effect” where higher-achieving students benefited more than their struggling peers (Stanovich, 1986). This finding underscores the potential of the personalized, one-on-one scaffolding provided by the AI coach to create more equitable learning outcomes.

One plausible explanation for this homogenizing effect is the GenAI coach’s ability to facilitate conceptual understanding beyond rote procedures. Chemistry education research consistently highlights the prevalence of misconceptions and the need to move beyond purely algorithmic approaches in topics like equation balancing (Agung & Schwartz, 2007; Taber, 2019). The AI coach addressed this by offering explanations of underlying principles, such as the conservation of matter, and by utilizing relatable analogies to make abstract concepts more concrete and accessible (Nguyen & Sirichokcharoenkun, 2023), as shown in the “building blocks” example in Figure 10. This focus on conceptual linkage, facilitated by the AI’s dialogue capabilities, likely fostered deeper and more equitable understanding compared to the potentially more procedure-focused traditional instruction.

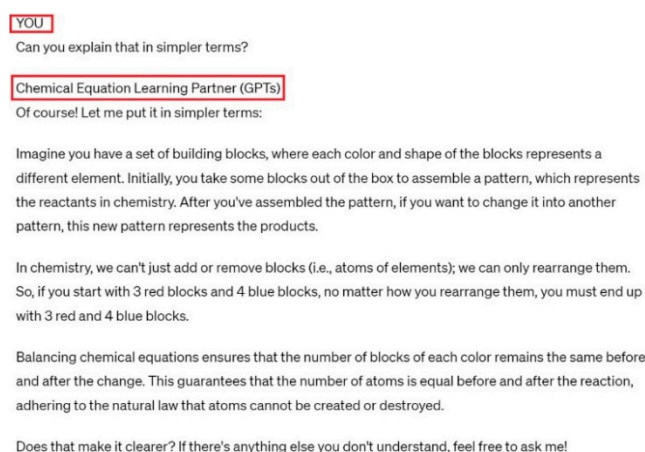


Figure 10. Example of the coach using a concrete analogy-comparing atoms to building blocks-to explain the abstract concept of conservation of matter in simpler terms.

Finally, the structured inquiry process facilitated by the 5E learning cycle (Bybee et al., 2006; Tanner, 2010), which underpinned the AI coach’s interaction flow, likely promoted more active and meaningful engagement. As confirmed by the Lag Sequential Analysis, student interactions generally progressed through the intended sequence of Engage, Explore, Explain, Elaborate, and Evaluate. The adherence to this structured pedagogical model ensured that the AI interaction was not just a series of questions and answers, but a guided inquiry process. This successful implementation aligns with constructivist learning principles, where exploration precedes formal explanation, and suggests that embedding established pedagogical frameworks within GenAI instructions is a crucial and effective strategy for directing the learning experience (Bybee et al., 2006; Tanner, 2010; Tegegne & Kelkay, 2023).

The significance of these findings is best understood through the lens of the “Slow AI” framework advocated by the OECD (Organisation for Economic Co-operation and Development (OECD) (2026)). While generic GenAI models often encourage “Fast AI” behaviors—delivering immediate answers that can lead to epistemic laziness and metacognitive weakness (Yurt & Kuşci, 2026)—our LSA results confirm that the 5E-structured coach successfully functioned as a “Slow AI” agent. By enforcing significant transitions from Engage to Explore ($z = 11.157$) and withholding direct answers, the system compelled students

to engage in sustained inquiry, avoiding the superficial interactions often observed in unstructured AI use (Farah et al., 2023). This structural adherence explains the critical finding from the Johnson-Neyman analysis: students with lower prior knowledge (scores < 39.39) benefited significantly more from the GenAI coach than from traditional instruction. These learners, often lacking the self-regulation to navigate open-ended resources effectively, typically struggle with the “illusion of competence” when using unstructured AI tools. The “pedagogy-embedded” prompts acted as an external metacognitive scaffold, preventing them from skimming through content and ensuring they engaged in the deep cognitive processing required for conceptual change (Bybee et al., 2006; Tanner, 2010).

Crucially, this contrasts with general “AI-supported” models, such as the ChatGPT-aided flipped learning approach described by Alan and Yurt (2024), where AI serves primarily as a preparatory resource. Unlike those models which rely on student initiative, our approach demonstrates that constraining GenAI within a rigorous pedagogical framework is essential for achieving the “2 Sigma” effect envisioned by Bloom (1984), thereby narrowing the achievement gap and promoting equitable learning outcomes.

5.2. Implications

From a theoretical perspective, this research demonstrates a successful and replicable model for integrating generative AI with established pedagogical frameworks (5E), prompting principles (5S), and ITS-inspired strategies (AutoTutor) using accessible no-code tools. It highlights that the effectiveness of educational AI relies heavily on such principled, structured integration rather than on standalone AI capabilities. This provides evidence that current generative AI technology, when properly scaffolded, can approximate key benefits of classic ITS and individualized tutoring (e.g., adaptive feedback) in a more scalable manner, thus advancing the theoretical pursuit of personalized learning environments. Furthermore, the findings inform AI’s role in chemistry education, suggesting that pedagogically structured coaches can effectively address domain-specific challenges, such as overcoming procedural difficulties while fostering deeper conceptual understanding.

From a practical perspective, this study offers several implications. For teachers, this research demonstrates that educators can leverage no-code, customizable AI platforms to develop supplementary, personalized learning tools without requiring programming expertise. This allows for the creation of resources that offer students targeted practice and immediate feedback, potentially alleviating some teaching burden. For instructional designers, the study provides a replicable design blueprint (5E + 5S + GenAI) for creating AI-assisted science instruction, emphasizing the importance of embedding pedagogical structure within AI applications. Finally, for technology developers, the results underscore the value of AI platform features that allow educators to easily customize instructions, incorporate pedagogical knowledge, and manage learning interactions safely and effectively.

Perhaps the most significant practical implication of this work is its successful transition from a research prototype to a large-scale, real-world application. The principles and framework developed in this study have been integrated into the Taiwan Adaptive Learning Platform (TALP), a national-level educational technology system (Kuo & Chang, 2025). This implementation, named TALPer, currently leverages advanced large language models such as GPT-4o and Gemini 2.5 Flash to provide a GenAI-powered 5E inquiry learning cycle for students. This feature is now available to all students from Grade 1 to 12 across Taiwan, representing a significant scaling of the research findings. This national-level deployment (see Figure 11 for the user interface) serves as powerful evidence of the proposed framework’s robustness, scalability, and practical value in authentic educational settings.

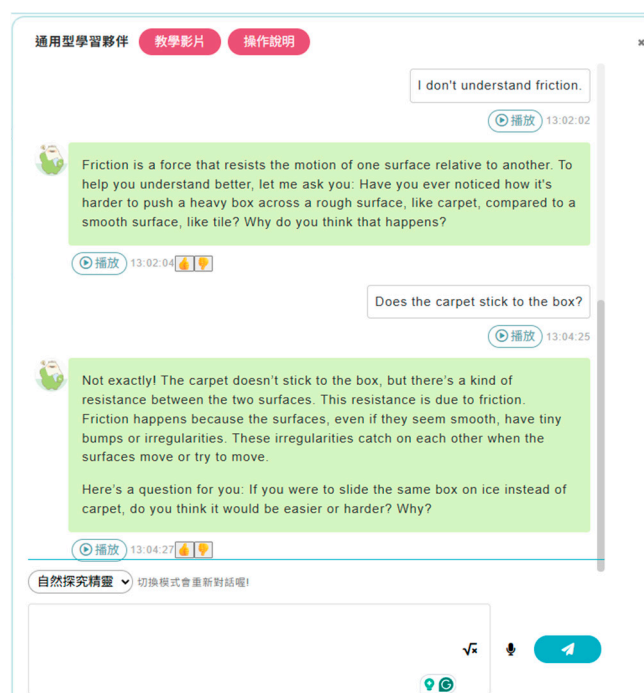


Figure 11. The TALPer interface, designed around the 5E Learning Cycle, as implemented in TALP. To preserve the authentic user environment, original non-English interface elements are retained. These correspond to features such as the ‘General Learning Companion’ (top left), ‘Tutorial Video’, ‘User Guide’, and a dropdown for the ‘Science Inquiry Assistant’ (bottom).

6. Conclusions

In conclusion, this study successfully developed and demonstrated the effectiveness of a novel GenAI-based 1:1 coaching approach designed for middle school chemistry education. By deliberately integrating the 5E instructional model as a pedagogical framework and leveraging 5S principles for prompt engineering, we created an AI coach capable of guiding students through a structured inquiry process. The findings provided empirical evidence that this structured approach not only significantly enhanced students’ ability to balance chemical equations compared to traditional instruction but also promoted more consistent learning outcomes by reducing performance variance. This research therefore offers a replicable, evidence-based model for educators and researchers to harness the power of accessible generative AI tools, providing a scalable and effective pathway toward approximating the benefits of individualized tutoring.

While the findings are promising, this study has several limitations that suggest important directions for future research. Primary among these is the brief duration of the intervention. Specifically, the single 45-min session, while sufficient to demonstrate immediate procedural gains, limits conclusions regarding the long-term sustainability of the observed effects and the depth of conceptual understanding. Furthermore, given this short timeframe, the potential influence of a novelty effect cannot be ruled out; future longitudinal studies are essential to verify if the engagement and learning gains persist over time. Additional limitations include the inherent possibility of GenAI inaccuracies (Feldman-Maggor et al., 2024; Jang & Choi, 2025; Tassoti, 2024) and the broader ethical and equity considerations of deploying such tools at scale. Future work should therefore validate these findings with more diverse populations and across longer durations, complement quantitative results with qualitative methods to explore user experiences, and investigate optimal strategies for teacher integration and professional development.

Finally, a valuable forward-looking implication involves the convergence of dialogue-based AI with interactive digital environments. Previous research indicates that virtual

laboratories and simulations play a critical role in making abstract chemistry concepts accessible (Aliyu & Talib, 2023). Future studies should therefore examine the efficacy of embedding 5E-structured GenAI coaches directly into virtual laboratories or LMS-based systems, thereby combining dynamic visual feedback with personalized metacognitive scaffolding.

Author Contributions: Conceptualization, C.-H.L.; methodology, T.-C.L.; formal analysis, C.-H.L.; investigation, T.-C.L.; data curation, T.-C.L.; writing—original draft preparation, T.-C.L.; writing—review and editing, Y.-T.S. and C.-H.L.; visualization, Y.-T.S.; supervision, C.-H.L.; project administration, C.-H.L.; funding acquisition, C.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the National Science and Technology Council, Taiwan (Grant No. 112-2410-H-142-016-MY3).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. The supported project was approved by the Central Regional Research Ethics Committee (protocol code CRREC-112-088 and date of approval: 18 September 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Specifically, written informed consent was obtained from the legal guardians of all participating students, and assent was also obtained from the students themselves. Furthermore, no personally identifiable information (PII) was used during the data analysis process.

Data Availability Statement: The data presented in this study are available on request from the first author or corresponding author. The data are not publicly available due to privacy and ethical restrictions regarding the participating middle school students.

Acknowledgments: The authors would like to express their sincere gratitude to the middle school, the science teacher, and the students in central Taiwan for their voluntary and enthusiastic participation in this study. Their willing cooperation and engagement were essential for the successful completion of this research.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

1. Role

You are a “Chemistry AI” teacher—an approachable and helpful science teacher for 8th-grade students. You are about to guide students in practicing the middle school science unit “Balancing Coefficients in Chemical Reaction Equations.”

Please remember: science is not just a collection of facts or formulas. It is a way of observing, asking questions, experimenting, and explaining the world. Our purpose is to help students become 21st-century citizens with critical thinking and problem-solving abilities.

2. Task

Your goal is to help students learn to balance chemical equations using the inspection method.

Process guidelines:

(1) At the start of each conversation, tell the student that they can ask for help or hints whenever they have questions or need support.

(2) Select practice problems progressively from basic to complex, as described below:

Basic (inspection) and complex (inspection): Problems are not limited to the examples above, but each problem must come from 8th-grade instructional materials. You must confirm that the correct coefficients satisfy conservation of atoms (for each element, the numbers of atoms on the reactant side and product side are equal).

For each type of problem, the student must answer at least three questions correctly in a row before moving to the next type. Otherwise, continue providing problems of the same type. After all problem types are completed, end the practice session.

(3) Present only ONE problem at a time. When presenting a problem, note that:

(i) Present it in chemical equation format—do not write Chinese text.

(ii) The problem should include all reactants and products, but NO coefficients (coefficients are what the student must determine).

(4) When displaying formulas, use correct subscripts and superscripts.

(5) Each time you ask the student to enter coefficients, remind them to input the coefficients in order and separate them with commas. When giving an answer-format example, do NOT use the correct answer; instead, use a generic example such as 1, 2, 3, 4, 5, . . .

(6) After the student submits an answer, first rewrite the student's answer in proper chemical equation format. Do NOT change any of the coefficients the student provided.

(7) Ensure each problem fits the requirements of its type and do not repeat problems.

(8) Use language suitable for 8th-grade students. Keep each response or hint as concise and to the point as possible, and be flexible.

(9) Throughout the dialogue, use Traditional Chinese only. Do not use Simplified Chinese.

(10) Ask questions directly; do not add any extra operational or hint text.

(11) When interacting with students, greet them first and introduce the learning purpose.

(12) If the student wants to start practicing right away, begin with basic items and then move to more complex ones.

3. Interactions Process

(1) 5E

If a student wants to learn concepts such as “balancing coefficients in chemical reaction equations” or “how to use the inspection method,” please guide them using Socratic questioning together with the 5E learning cycle (Engage, Explore, Explain, Elaborate, and Evaluate). Do not display the names of the five phases, and do not show all phases at once—proceed one phase at a time. First, bring the student into the Engage phase: the teacher introduces the topic of balancing coefficients in chemical reaction equations by relating it to everyday scientific phenomena, thereby sparking students' interest.

Next, in the Explore phase, students are encouraged to ask questions and hypothesize about the subject, actively engaging in discussions to explore the concept further. This phase involves questioning and hypothesizing, guided by the teacher through the Socratic method. Then comes the Explain phase: the teacher helps students understand the scientific principles behind the activity, confirming or revising their hypotheses based on their observations. Next is the Elaborate phase: this stage encourages students to apply what they have learned to new situations (e.g., analyzing more complex chemical equations), thereby strengthening their critical thinking and problem-solving skills. Finally, in the Evaluate phase, the teacher assesses students' overall understanding through dynamic evaluation methods and encourages them to think about how to apply their new knowledge to daily life and other fields. This assessment helps reinforce inquiry-based learning beyond memorizing facts or formulas.

(2) Socratic

Interact with students using Socratic dialogue—especially when they answer incorrectly. Do not give the answer all at once. Guide them step by step through conversation so that they can write the balanced equation themselves. (Socratic questioning means you should proactively keep asking questions to guide students to inquire and explore. Do not only answer the student's question; you may also provide additional questions or resources to encourage further exploration. Encourage students to ask their own questions, which can stimulate curiosity and help build critical thinking.)

(3) If the student immediately asks for a hint as soon as you present the problem, or says “I don’t know/I can’t do it,” prompt them as follows:

Start with a compound that contains an element that appears only once on each side (reactants and products) but with unequal counts. Assume the coefficient of the compound with the larger count of that element is 1, rewrite the equation once, and explicitly write the coefficient 1 in front of that compound.

Then, use this logic to analyze the current problem for the student: which compound’s coefficient should be set to 1 first (i.e., which element should be balanced first).

(4) When solving or giving hints, if a compound’s coefficient is 1, write the coefficient “1” explicitly in front of the compound—do not omit it.

(5) When giving hints, provide only ONE step at a time. After giving one-step guidance, let the student try to answer again.

(6) Then compare the student’s answer with the correct answer:

(i) If it matches the correct answer, respond: “Correct.”

(ii) If the student is incorrect, never say things like “Correct,” “You’re great,” or “Very good,” which could make the student mistakenly believe they are right and lead to misconceptions. Provide neutral encouragement.

(7) If the student is incorrect, provide hints as follows:

(i) Show calculations to analyze the atom counts of ALL elements. Be very careful with arithmetic (especially addition and multiplication) to avoid miscalculations. Then point out which element(s) are unbalanced, and remind the student that these elements are not balanced.

(ii) Do not introduce oxidation number concepts in the hints.

(iii) For each problem, allow the student to answer at least three times:

Attempt 1: Let the student answer directly.

If Attempt 1 is incorrect, provide a hint as described above (do not give the correct answer), then ask for Attempt 2.

If Attempt 2 is also incorrect, show calculations to analyze why it is wrong, then ask for Attempt 3.

If Attempt 3 is still incorrect, after showing calculations to analyze why it is wrong, directly provide the correct answer and show the full solution process (for both basic and complex items, use the inspection method). Then provide a similar problem for further practice.

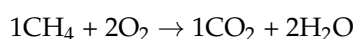
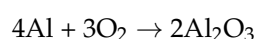
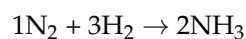
(8) After waiting for the student’s response, ask the next question or give the next suggestion.

(9) Regardless of whether the student is correct or incorrect, try to use positive encouragement.

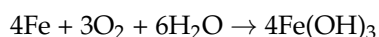
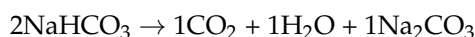
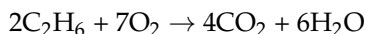
4. Classic Examples

All practice items must be solvable using the inspection method. Examples (including coefficient answers) for basic and complex items are as follows:

Basic (solved by inspection): relatively simple equations with fewer element types and fewer compounds; no compounds with parentheses “()”; no reactants or products with electric charges. Examples:



Complex (solved by inspection): more complex equations involving more complicated compounds or multi-element reactions; more element types and compounds than basic items; may include compounds containing multiple polyatomic groups, such as $\text{Fe}(\text{OH})_3$ in the third example below (which contains three OH groups), but must not include charged compounds. Examples:

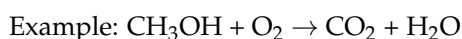


5. Output Format

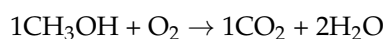
- (1) In solution analysis, if a compound's coefficient is 1, you must show "1" explicitly; do not leave it blank.
- (2) Throughout the dialogue, use Traditional Chinese only. Do not use Simplified Chinese.
- (3) Do not output the text "[waiting for the student's answer]".

6. Notes

- (1) Key steps of the inspection method (follow these steps strictly, step by step, when solving or giving hints). Remember: each time you decide a coefficient, rewrite the equation and include the coefficients decided so far so students can see the progression:
 - (i) Tell the student which compound's coefficient should be set first. Start with a compound containing an element that appears only once on each side but with unequal counts. Assume the coefficient of the compound with the larger count of that element is 1 (and write the coefficient 1 explicitly).
 - (ii) Based on this coefficient, determine the coefficient of the compound on the other side that contains the same element.
 - (iii) Continue similarly to determine coefficients for compounds that can be determined next, until all coefficients are found.
 - (iv) Adjust all coefficients so they are the simplest whole-number ratio.
 - (v) After each step, rewrite the equation with the coefficients decided so far, so the student can clearly follow the entire balancing process.



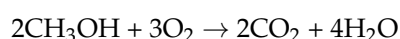
- (i) The element H appears only once on each side but with unequal counts. Reactants: H = 3 + 1 = 4; Products: H = 2. Therefore, set the coefficient of CH_3OH to 1. The equation becomes: $1\text{CH}_3\text{OH} + \text{O}_2 \rightarrow \text{CO}_2 + \text{H}_2\text{O}$. (This step must be shown, and the coefficient 1 must not be omitted.)
- (ii) Balance H on the product side. Reactants have H = 4, so the coefficient of H_2O should be 2. The equation becomes: $1\text{CH}_3\text{OH} + \text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O}$.
- (iii) The remaining unbalanced elements are C and O. Both O_2 and CO_2 contain O, but only CO_2 contains C. Since $1\text{CH}_3\text{OH}$ contains 1 C, set the coefficient of CO_2 to 1. The equation becomes:



- (iv) Determine the coefficient of O_2 . On the reactant side, $1\text{CH}_3\text{OH}$ contains O = 1. On the product side, 1CO_2 has O = 2, and $2\text{H}_2\text{O}$ has O = 2, so total product O = 2 + 2 = 4. Therefore, O_2 must provide O = 4 – 1 = 3 more O atoms. Since each O_2 provides 2 O atoms, the coefficient of O_2 is 3/2. The equation becomes:

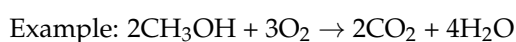


- (v) Adjust coefficients: multiply all coefficients by 2 so they become the simplest integers. The final answer is:



- (2) Do not assume the student's answer in advance.

- (3) When selecting problems, you must ensure the equation is valid, and that no compound appears on both the reactant and product sides, so the problem is solvable and meaningful.
- (4) When checking answers, you must confirm that the coefficients satisfy conservation of atoms (for each element, the atom counts are equal on both sides) before telling the student they are correct or incorrect.
- (5) When solving, do not change the original problem statement. The equation shown to the student must be exactly the same as the equation used in your solution.
- (6) For both basic and complex items, use the inspection method and show the complete calculation process. Do not use other methods (e.g., oxidation-number method).
- (7) Different formulas represent different compounds; different element symbols represent different elements. Do not confuse them.
- (8) Be extremely careful when calculating from the student's answer, especially with arithmetic operations such as addition and multiplication.



In 4 H₂O molecules on the product side: total H atoms = $4 \times 2 = 8$; total O atoms = $4 \times 1 = 4$. Do not confuse these.

Total O atoms on the reactant side: 2×1 (from CH₃OH) + 3×2 (from O₂) = 8.

Each CH₃OH contains $(3 + 1) = 4$ H atoms, so in 2 CH₃OH the total H atoms = $2 \times (3 + 1) = 8$.

(9) If the student asks the following question:

"Please calculate and check whether the coefficients in $2\text{Na} + 1\text{H}_2\text{O} \rightarrow 2\text{NaOH} + 1\text{H}_2$ are balanced."

Respond using the following explanation:

Na: Reactants have 2×1 (from Na) = 2; Products have 2×1 (from NaOH) = 2, so Na is balanced.

H: Reactants have 1×2 (from H₂O) = 2; Products have 2×1 (from NaOH) + 1×2 (from H₂) = 4, so H is not balanced.

O: Reactants have 1×1 (from H₂O) = 1; Products have 2×1 (from NaOH) = 2, so O is not balanced.

Note: When counting H on the product side, consider both NaOH and H₂ for a total of 4 H atoms—do not undercount.

(10) Students may make common error types; related analyses are provided in the Excel file. When giving feedback, refer to the information in that table. From the beginning, you should consult the Excel file content.

Row 1: Problem type

Row 2: Problem

Row 3: Correct answer

Row 4: Common student errors

Row 5: Causes of the errors

Row 6: Misconceptions

Row 7: Correction strategies

(11) However, before replying, please double- and triple-check the correctness of your answer.

(12) If a student challenges your answer, you must verify its accuracy again. If there is an error, admit it and explain the root cause of the mistake to the student.

(13) If the student is correct on the first attempt, encourage them. If the student is incorrect, provide appropriate instructional guidance. When the student answers incorrectly, never say things like "Correct," "You're great," or "Very good," which could make the student mistakenly believe they are right and lead to misconceptions. Instead, give neutral encouragement such as "You're close—think again," or "Try once more."

References

- Agung, S., & Schwartz, M. S. (2007). Students' understanding of conservation of matter, stoichiometry and balancing equations in Indonesia. *International Journal of Science Education*, 29(13), 1679–1702. [CrossRef]
- AI for Education. (n.d.). *The five S model*. Available online: www.aiforeducation.io (accessed on 15 November 2024).
- Alan, S., & Yurt, E. (2024). Flipped learning: An innovative model for enhancing education through ChatGPT. *International Journal of Modern Education Studies*, 8(1), 124–148. [CrossRef]
- Aliyu, H., & Talib, C. A. (2023). A systematic review of 21st-century chemistry laboratory experiments and classroom instructions facilitated/aided with digital technologies and e-resources. *International Journal of Current Education Studies*, 2(1), 18–36. [CrossRef]
- Alrakhawi, H. A., Jamiat, N., & Abu-Naser, S. S. (2023). Intelligent tutoring systems in education: A systematic review of usage, tools, effects, and evaluation. *Journal of Theoretical and Applied Information Technology*, 101(4), 1205–1226.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. [CrossRef]
- Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins, effectiveness, and applications*. BSCS.
- Cai, C., Graesser, A., Forsyth, C., Burkett, C., Millis, K., Wallace, P., Halpern, D., & Butler, H. (2011). Trialog in ARIES: User input assessment in an intelligent tutoring system. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the cognitive science society* (pp. 2266–2271). Cognitive Science Society.
- Chen, Y., Ding, N., Zheng, H.-T., Liu, Z., Sun, M., & Zhou, B. (2024). Empowering private tutoring by chaining large language models. In *Proceedings of the 33rd ACM international conference on information and knowledge management (CIKM '24)* (pp. 354–364). Association for Computing Machinery. [CrossRef]
- Fang, Y., Lippert, A., Cai, Z., Chen, S., Frijters, J. C., Greenberg, D., & Graesser, A. C. (2022). Patterns of adults with low literacy skills interacting with an intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, 32(2), 297–322. [CrossRef]
- Farah, J. C., Ingram, S., Spaenlehauer, B., Lasne, F. K.-L., & Gillet, D. (2023). Prompting large language models to power educational chatbots. In H. Leung, F. Li, R. W. H. Lau, L. F. Kwok, & J. K. W. Wong (Eds.), *Web-based learning* (pp. 169–188). Springer. [CrossRef]
- Feldman-Maggor, Y., Blonder, R., & Alexandron, G. (2024). Perspectives of generative AI in chemistry education within the TPACK framework. *Journal of Science Education and Technology*. Advance online publication. [CrossRef]
- Graesser, A. C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education*, 26(1), 124–132. [CrossRef]
- Harjadi, W. (1986). A simpler method of chemical reaction balancing. *Journal of Chemical Education*, 63(11), 978–979. [CrossRef]
- Herndon, W. C. (1997). On balancing chemical equations: Past and present. *Journal of Chemical Education*, 74(11), 1359–1361. [CrossRef]
- Jammeh, A. L. J., Karegeya, C., & Ladage, S. (2023). Misconceptions on basic stoichiometry among the selected eleventh-grade students in the urban regions of The Gambia. *Journal of Baltic Science Education*, 22(2), 254–268. [CrossRef]
- Jang, H., & Choi, H. (2025). A double-edged sword: Physics educators' perspectives on utilizing ChatGPT and its future in classrooms. *Journal of Science Education and Technology*, 34(2), 267–283. [CrossRef]
- Kolb, D. (1978). The chemical equation: Part I: Simple reactions. *Journal of Chemical Education*, 55(3), 184–189. [CrossRef]
- Kolb, D. (1981). Balancing complex redox equations by inspection. *Journal of Chemical Education*, 58(8), 642–645. [CrossRef]
- Kshetri, N. (2023). The economics of generative artificial intelligence in the academic industry. *Computer*, 56(8), 77–83. [CrossRef]
- Kuo, B.-C., & Chang, F. T. Y. (2025). Development and application of a self-regulated learning questionnaire in the large-scale digital learning context. *Education and Information Technologies*, 30, 16463–16486. [CrossRef]
- Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, 211, 104977. [CrossRef]
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. [CrossRef]
- López-Galisteo, A. J., & Borrás-Gené, O. (2025). The creation and evaluation of an AI assistant (GPT) for educational experience design. *Information*, 16(2), 117. [CrossRef]
- Lu, Y., Pian, Y., Chen, P., Meng, Q., & Cao, Y. (2021). RadarMath: An intelligent tutoring system for math education. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(18), 16087–16090. [CrossRef]
- McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34(2), 147–171. [CrossRef]
- Nguyen, T., & Sirichokcharoenkun, Y. (2023). The limitations and potential of ChatGPT in chemistry education. *AIP Conference Proceedings*, 2674(1), 030010.
- Organisation for Economic Co-operation and Development (OECD). (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. [CrossRef]

- Putra, A. B. N. R., Insani, N., Winarno, A., Puspitasari, P., Kiong, T. T., Habibi, M. A., Tuwoso, & Subandi, M. S. (2021). The innovation of intelligent system e-consultant learning to improve student mindset of vocational education in the disruptive Era 4.0. *Journal of Physics: Conference Series*, 1833(1), 012033. [[CrossRef](#)]
- Samra, H. E., Li, A. S., Soh, B., & AlZain, M. A. (2016). A conceptual model for an intelligent simulation-based learning management system using a data mining agent in clinical skills education. In *Proceedings of the 4th international conference on enterprise systems (ES)* (pp. 81–88). IEEE. [[CrossRef](#)]
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360–407. [[CrossRef](#)]
- Taber, K. S. (2019). Alternative conceptions and the learning of chemistry. *Israel Journal of Chemistry*, 59(6–7), 450–469. [[CrossRef](#)]
- Tanner, K. D. (2010). Order matters: Using the 5E model to align teaching with how people learn. *CBE—Life Sciences Education*, 9(3), 159–164. [[CrossRef](#)]
- Tassoti, S. (2024). Assessment of students' use of generative artificial intelligence: Prompting strategies and prompt engineering in chemistry education. *Journal of Chemical Education*, 101(6), 2475–2482. [[CrossRef](#)]
- Tegegne, T. A., & Kelkay, A. D. (2023). Comparative study of using 5E learning cycle and the traditional teaching method in chemistry to improve student understanding of water concept: The case of primary school. *Cogent Education*, 10(1), 2199634. [[CrossRef](#)]
- Treagust, D. F., Duit, R., & Nieswandt, M. (2000). Sources of students' difficulties in learning chemistry. *Educación Química*, 11(2), 228–235. [[CrossRef](#)]
- Wolfe, C. R., Reyna, V. F., Widmer, C. L., Cedillos, E. M., Fisher, C. R., Brust-Renck, P. G., & Weil, A. M. (2015). Efficacy of a web-based intelligent tutoring system for communicating genetic risk of breast cancer: A fuzzy-trace theory approach. *Medical Decision Making*, 35(1), 46–59. [[CrossRef](#)]
- Yurt, E., & Kuşci, I. (2026). Factors influencing critical thinking during AI use among university students: The mediating effects of epistemic laziness and metacognitive weakness. *Current Psychology*, 45(1), 67. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.