education sciences

MDPI

Article

Designing Flipped-Interaction Prompts: A Framework for Generative AI as an Intelligent Tutor in Higher Education

Stefanus Johannes Scheepers [1,*] and Angela Elisabeth Stott [2]

[1] Senior and FET Phase Studies, Akademie Reformatoriese Opleiding en Studies (Aros), Pretoria 0182, South Africa
[2] Department of Mathematics, Natural Sciences and Technology Education, University of the Free State, Bloemfontein 9300, South Africa; stottae@ufs.ac.za
* Correspondence: stefan.scheepers@aros.ac.za

Abstract

Students' use of generative artificial intelligence (GAI) to avoid engaging in generative processing can undermine the validity of higher education. In contrast, Flipped-Interaction Intelligent Tutoring Systems (FIITSs) may promote active engagement by leading a personalised dialogue. The underutilisation of FIITS may stem from the lack of a framework to guide prompt creation and from a dearth of published FIITS prompt examples. This article presents the Flipped-Interaction Prompt (FIP) Framework, abstracted from two validated prompts. To achieve this validation, 26 preservice science education students at a South African university engaged with either prompt in a free GAI five times over ten weeks. The resulting 114 engagements, each involving at least 10 flipped-interaction dialogue exchanges, were analysed for implementation fidelity and for students' engagement in generative processing. Findings were triangulated against questionnaire and group interview responses, as well as written reflections. The technical implementation was closely aligned with the prompt instruction, with minor deviations noted for not providing answers outright. Additionally, students demonstrated moderate to high levels of generative processing. Findings support the efficacy of the abstracted FIP Framework in guiding the creation of FIITS prompts. Investigating instantiations for additional subject domains would further strengthen confidence in this framework.

Keywords: flipped-interaction intelligent tutoring systems; generative artificial intelligence; generative processing; prompting; Global South; intelligent tutoring systems; student engagement

Check for updates

Academic Editor: Sergio Miranda

Received: 2 February 2026
Revised: 20 March 2026
Accepted: 23 March 2026
Published: 3 April 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) license.

1. Introduction

The underutilisation of free generative artificial intelligence (GAI) as a Flipped-Interaction Intelligent Tutoring System (FIITS) may be due to a lack of a prompt framework and exemplars. In a flipped-interaction (FI), GAI asks, rather than answers, questions (White et al., 2023). An FI that is customised to the user's needs is an intelligent tutoring system (ITS) (Zheng et al., 2025). Underutilisation of free GAI to lead, rather than respond to, a dialogical exchange, is evident in its absence in reports on students' GAI use, for example, by South African preservice teachers (Ramnarain et al., 2024) and biology students across nine universities in the United States of America (Coelho et al., 2025). Other information and communication technologies (ICTs), such as search engines and websites, can provide information. Therefore, using GAI to answer users' questions involves substitution or augmentation relative to these other ICTs (Franco, 2019). In contrast,

whereas search engines and websites cannot, GAI can ask users questions and respond with personalised feedback. Therefore, prompting GAI as a FIITS can be seen as potentially redefining ICT-mediated learning by providing functionality absent in other ICTs.

Consistent with the argument that GAI's unique capabilities are underutilised in education, a meta-analysis of GAI-related research revealed that GAI continues to perpetuate, rather than redefine, education paradigms (Lee et al., 2025). Despite calls to pioneer innovative integration of free GAI into education (Zhai, 2024), few concrete guidelines, such as template-like frameworks and illustrative examples, have been published for FIITS. Possibly, the paucity of prompt templates and examples of this kind results from a preference for more user-friendly chatbots and custom-built intelligent tutoring systems over lengthy prompts that need to be copied from a repository and pasted into GAI. However, these more user-friendly alternatives are prohibitively expensive for many higher institutions, particularly in the developing world (Avcı et al., 2025). Also, obtaining licences for such alternatives may be difficult due to institutional bureaucracy.

Prompt frameworks and examples do, however, abound in the literature for the more common form of using GAI, i.e., to provide or evaluate information. These include frameworks for prompting GAI to formatively assess scientific argumentation (F. M. Watts et al., 2025), pedagogical content knowledge (Blonder et al., 2024), and pedagogical reflection (Hofmann et al., 2025). This article extends the literature on prompt examples and frameworks into the domain of the flipped-interaction intelligent tutor, as it answers the overarching question:

> How can free generative artificial intelligence be prompted to serve as a flipped-interaction intelligent tutor?

To answer this question, the technical and pedagogical efficacy of two FIITS prompts is first investigated, guided by the following research questions:

(1) To what extent do free generative artificial intelligence (GAI) platforms execute the prompts with technical fidelity?
(2) To what extent do the Flipped-Interaction Intelligent Tutoring Systems (FIITSs) resulting from these prompts encourage student generative processing?

Positive responses to these questions provide evidence for abstracting an easily customisable generic framework from these two investigated prompts. This framework is called the Flipped-Interaction Prompt (FIP) Framework. The FIP Framework could guide educators in creating free FIITS tailored to their specifications, without bureaucratic oversight.

2. Theoretical Frameworks

Generative processing serves as a theoretical framework for interpreting students' engagement with GAI. Cognitive Load Theory, in conjunction with the Cognitive Theory of Multimedia Learning, holds that generative processing is necessary for meaningful learning to occur (Mayer, 2024). Generative processing involves selecting relevant information, forming connections between prior and new knowledge, and reflecting on feedback to construct a deeper understanding (Fiorella & Mayer, 2021). Generative processing occurs within the limited space of working memory, which can be expanded through skilful collaboration with peers, instructors, or even artificial tools serving these purposes (Kirschner et al., 2018). When used pedagogically appropriately, GAI enhances the working memory space and facilitates and stimulates students' generative processing (Katsenou et al., 2025). When used pedagogically inappropriately, however, GAI promotes gaming the system (K. J. Watts, 2025), i.e., circumventing generative processing. Concerns about GAI misuse in higher education arise from GAI's ability to game traditional assignments by generating

human-like answers while requiring little generative processing from students (Hughes et al., 2025). This concern is irrelevant to a FIITS, given its role of asking students questions.

Against this background, the FIITS prompt examples and abstracted framework presented in this article were pedagogically informed by Laurillard's (2013) Conversational Framework. Dialogue likely to promote generative processing is conceptualised as students expressing their current conception, receiving feedback that challenges or extends that conception, applying revised ideas in practice, and reflecting on the outcomes.

While the questioning dialogue of a FIITS is more likely to promote generative processing than is more passive information access (Katsenou et al., 2025), it is not immune to attempts to game the system. Research into integrating ICTs into education over several decades has revealed various methods for detecting attempts to game the system (Baker et al., 2013). These include avoidance of more difficult questions, over-reliance on hints, and guessing. Guessing may be evidenced by high error rates and repeated selection of a single multiple-choice option. In contrast, behaviour that suggests generative processing includes active selection of relevant information, connection-building, and concept modification (Fiorella & Mayer, 2021).

3. Materials and Methods

3.1. Research Design

This is a pragmatically conducted mixed-methods case study (Creswell & Cresswell, 2003) aimed at evaluating the technical and pedagogical efficacy of two example FIITS prompts to validate a generic FIT framework. The two prompts and their creation are first described, followed by an explanation of the data collection and analysis process.

3.2. The Prompts, Their Creation, and Output Examples

The two investigated prompts are available at https://bit.ly/IntelligentTutorPrompt. When pasted into free GAI, they result in FIITS aimed at developing subject-matter knowledge among preservice high school natural or physical science teachers. These two FIITS prompts were created through a series of iterative interactions with ChatGPT Plus (version 4). These initial prompts were refined through approximately five iterations of testing, guided by the pedagogical intuition of two experienced educators. Each iteration included testing the prompts on three free GAI platforms (ChatGPT, Gemini, and CoPilot). During these tests, the user experience, consistency, and fidelity across the interactions and simulations were assessed, and improvements were implemented by adding content, pedagogical elements, student control, and administrative details as the need for these became evident. The prompts were then input into ChatGPT Plus (version 4) and the free version of Gemini to run 50 simulated dialogues, each consisting of ten original question-answer exchanges. The simulators were prompted to answer some questions incorrectly, resulting in additional follow-up question-answer exchanges. The simulated dialogues were then analysed by ChatGPT Plus (version 4). This analysis suggested that the GAI platforms implemented the prompts with high fidelity, and, therefore, the prompts were finalised, enabling initiation of the implementation and investigation stages of this study.

Table 1 illustrates the setup sequence that results when the physical science prompt is pasted into Gemini. The FIITS provides choices for topic, grade, sub-topic, and difficulty level. Although the grade level refers to school-level physical science, the prompts were used with university students preparing to teach school physical science.

Table 1. An example of a setup sequence resulting from the physical science prompt.

Agent	Text
Gemini	Alright, let's dive into the world of Physical Sciences! Welcome to your interactive learning simulation. I'm here to guide you through the exciting concepts of your curriculum. To start, please choose a main topic from the following: 1. Matter and Materials 2. Chemical Systems 3. Chemical Change 4. Mechanics 5. Waves, Sound, and Light 6. Electricity and Magnetism Which topic would you like to explore first?
Student	4
Gemini	Excellent choice! Mechanics is a fundamental and fascinating area of Physical Sciences. Now, let's narrow it down further. Please select a grade level: 1. Grade 10 2. Grade 11 3. Grade 12
Student	3
Gemini	Great, Grade 12 Mechanics! You're ready to tackle some advanced concepts. Here are the sub-topics we'll be focusing on: 1. Vertical Projectile Motion (1D) 2. Momentum and Impulse (1D & 2D) 3. Work, Energy, and Power Which sub-topic would you like to explore?
Student	2
Gemini	Perfect, Momentum and Impulse! A key area for understanding collisions and forces. Now, let's select a difficulty level: 1. Easy 2. Medium 3. Hard 4. Challenging Which level would you like to start with?
Student	3
Gemini	Alright, you've chosen the "Hard" level for Momentum and Impulse. Let's get started. This will involve multiple-answer questions and fill-in-the-blanks, so be prepared to think critically.

Corresponding to the student's setup instructions, Table 2 shows the first of ten dialogue exchanges in the student's engagement with the resulting FIITS. This FIITS aimed to develop the student's momentum and impulse knowledge, aligned with the South African Grade 12 mechanics curriculum at the hard level. The FIITS provides a scenario with a multi-select question about the truth of five statements. Even though the student correctly selects two of these, the FIITS engages in a dialogue to ascertain the student's understanding, guiding them to perform the necessary calculations and reason through the concepts to evaluate each of the five statements.

Table 2. Example of a hard-level Flipped-Interaction Intelligent Tutoring System (FIITS).

Agent	Text
Gemini	Question 1: A 2 kg trolley moving east at 4 m/s collides with a 3 kg trolley moving west at 2 m/s. After the collision, the 2 kg trolley moves west at 1 m/s. Which of the following statements are correct? (Select all that apply) A. The total momentum of the system is conserved. B. The collision is perfectly elastic. C. The final velocity of the 3 kg trolley is 3 m/s east. D. The impulse experienced by the 2 kg trolley is 10 N.s west. E. The kinetic energy of the system is increased. Take your time, and carefully consider each statement. Remember to show your working where applicable.
Student	A & D
Gemini	Okay, you've selected options A and D. Let's break down each part and see how we can approach this problem. A. The total momentum of the system is conserved. • This is a fundamental principle in physics. In a closed system, the total momentum before a collision is equal to the total momentum after the collision. Do you recall the formula for momentum? • Momentum (p) = mass (m) × velocity (v). • Let's calculate the initial and final momentum of the system to verify this statement. D. The impulse experienced by the 2 kg trolley is 10 N.s west. • Impulse is the change in momentum. Do you remember how to calculate impulse? • Impulse (J) = change in momentum (Δp) = m(vf − vi), where vf is the final velocity and vi is the initial velocity. • Let's calculate the impulse experienced by the 2 kg trolley. Let's begin by calculating the initial momentum of each trolley and the total initial momentum. • Initial momentum of the 2 kg trolley: (2 kg)(4 m/s east) = 8 kg.m/s east. • Initial momentum of the 3 kg trolley: (3 kg)(−2 m/s west) = −6 kg.m/s (or 6 kg.m/s west). • Total initial momentum: 8 kg.m/s east + (−6 kg.m/s west) = 2 kg.m/s east. Now, calculate the final momentum of the 2 kg trolley. • Final momentum of the 2 kg trolley: (2 kg)(−1 m/s west) = −2 kg.m/s (or 2 kg.m/s west). What information do we need to calculate the final momentum of the 3 kg trolley? And after that, what information will we gain?
Student	To calculate the final momentum of the 3 kg trolley we need the mass and final velocity of the 3 kg trolley and with this information we will find the total momentum of the 3 kg trolley
Gemini	Exactly! And to find the final velocity of the 3 kg trolley, we will use the principle of conservation of momentum. Let's do that now. We know that the total initial momentum is 2 kg.m/s east. Let the final velocity of the 3 kg trolley be 'v'. Therefore, the total final momentum is: • (−2 kg.m/s) + (3 kg × v) = 2 kg.m/s (east) Now, solve for 'v'. What do you get?

Table 2. *Cont.*

Agent	Text
Student	V = 4/3 m/s east
Gemini	Excellent! So, the final velocity of the 3 kg trolley is 4/3 m/s east, or approximately 1.33 m/s east. Now, let's revisit your answers and the remaining options: A. The total momentum of the system is conserved. • You've correctly identified this. The total initial momentum (2 kg.m/s east) is equal to the total final momentum. D. The impulse experienced by the 2 kg trolley is 10 N.s west. • Let's calculate the impulse: ○ Impulse (J) = m(vf − vi) ○ J = 2 kg (−1 m/s − 4 m/s) ○ J = 2 kg (−5 m/s) ○ J = −10 kg.m/s or 10 N.s west. • You are correct. Now, let's analyze the other options: B. The collision is perfectly elastic. • In a perfectly elastic collision, kinetic energy is conserved. Let's calculate the initial and final kinetic energies to determine if this is true. ○ Kinetic energy (KE) = 1/2 mv^2. C. The final velocity of the 3 kg trolley is 3 m/s east. • You've already calculated the final velocity of the 3 kg trolley. It is 4/3 m/s east. E. The kinetic energy of the system is increased. • We need to calculate the initial and final kinetic energies to determine if this is true. Would you like to calculate the kinetic energies and verify options B, C, and E?
Student	yes
Gemini	Great, let's calculate the kinetic energies. Initial Kinetic Energy: • KE (2 kg trolley) = 1/2 × (2 kg) × $(4\ m/s)^2$ = 16 J • KE (3 kg trolley) = 1/2 × (3 kg) × $(-2\ m/s)^2$ = 6 J • Total Initial KE = 16 J + 6 J = 22 J Final Kinetic Energy: • KE (2 kg trolley) = 1/2 × (2 kg) × $(-1\ m/s)^2$ = 1 J • KE (3 kg trolley) = 1/2 × (3 kg) × $(4/3\ m/s)^2$ = 8/3 J or approximately 2.67 J • Total Final KE = 1 J + 2.67 J = 3.67 J Now, let's analyze the remaining options: B. The collision is perfectly elastic. • Is the total kinetic energy conserved? C. The final velocity of the 3 kg trolley is 3 m/s east. Is this correct? E. The kinetic energy of the system is increased. • Is the Final KE greater than the initial KE? What are your conclusions based on the calculated kinetic energies and the previously calculated final velocity of the 3 kg trolley?
Student	The kinetic energy is not conserved, the final velocity of the 3 kg is not 3 m/s east but 3.67 m/s east. therefore the collision is not elastic

Table 2. *Cont.*

Agent	Text
Gemini	You've hit the nail on the head! Let's clarify: • Kinetic Energy: ○ As you've calculated, the initial total kinetic energy was 22 J, and the final total kinetic energy was approximately 3.67 J. This means kinetic energy was not conserved. • Elasticity: ○ Because kinetic energy was not conserved, the collision is not perfectly elastic. It's an inelastic collision, where some kinetic energy is transformed into other forms of energy (like heat or sound). • Final Velocity: ○ You correctly noted that the final velocity of the 3 kg trolley was 4/3 m/s (approximately 1.33 m/s) east, not 3 m/s east. Therefore, the correct answers are: • A. The total momentum of the system is conserved. • D. The impulse experienced by the 2 kg trolley is 10 N.s west. And the incorrect options are explained as follows: • B. The collision is perfectly elastic. Incorrect, because kinetic energy is not conserved. • C. The final velocity of the 3 kg trolley is 3 m/s east. Incorrect, the final velocity is 4/3 m/s east. • E. The kinetic energy of the system is increased. Incorrect, the total kinetic energy decreased. You're making great progress! You've successfully navigated a challenging question involving momentum, impulse, and kinetic energy. Are you ready to continue with another question at this level of difficulty, or would you like to try something different?
Student	next question
Gemini	Alright, let's move on to the next "Hard" level question.

3.3. Implementation, Data Collection, Sample, and Ethics

The two FIITS prompts were implemented as part of the Bachelor of Education degree at a South African University. Using population sampling, data were collected from 11 third- and 15 fourth-year BEd students who engaged with the physical science and natural science prompts, respectively, every other week during a 10-week intervention. The students were all Black South Africans (16 male and 10 female) under the age of 25, who had been schooled in South Africa and were majoring in middle and high school natural and physical sciences and mathematics education at the university where this research was conducted. The students copied the prompt from their module's learning management system, pasted it into a free GAI of their choice, selected the topic, grade, and difficulty level, and interacted with the resulting FIITS. This meant that ideally, each of the 26 students submitted five engagements, yielding a total of 130. However, a few students failed to submit all expected engagements, resulting in a data corpus of 114 engagements. Each engagement consists of the 10 questions that the FIITS was prompted to pose, along with any follow-up questions when the student provided an incorrect answer. Additionally, the students completed a reflective questionnaire immediately after each engagement, wrote a reflective essay, and participated in an audio-recorded group interview at the end of the 10-week intervention. Ethical clearance for the study was obtained from the relevant regulatory body (reference number deleted for review). The students were provided with written informed consent

prior to the intervention. In this process, they were informed of the study's purpose, their right for their responses not to be used within the study without any penalty.

Given the use of prompt-based interactions, particular care was taken to ensure that no personally identifiable information was included in the data. Data were stored securely and reported in aggregate or anonymised form to protect student confidentiality.

3.4. Data Analysis

The FIITS student–engagements formed the primary data source. In correspondence to the research questions, this data was analysed for: (1) the fidelity with which the resulting FIITS implemented the prompt, by calculating the percentage of engagements that abided by the prompt instructions, per prompt aspect; (2) evidence for and against generative cognitive processing during the interaction. The engagements were analysed quantitatively for evidence of gaming the system, and the students' post-activity reflections were analysed qualitatively for evidence of generative processing. Qualitative detection of generative processing was guided by the purpose-built rubric given in Table 3. Generative processing is operationalised across three dimensions: E (engagement), B (building on feedback), and C (constructing conceptual understanding). Each student's written reflection was coded on a 0–4 scale for each dimension. Analysis was supported by ChatGPT Plus through an iterative process of exploratory prompting, rubric-guided manual evaluation, and consistency checking, which enhanced analytic efficiency, reliability and transparency. Quantitative summaries were generated using Microsoft Excel. To strengthen validity, findings from the reflection analysis were triangulated against group interview data, allowing for the identification of convergent and divergent patterns across data sources. The group interviews were also used to member check emergent themes.

Table 3. The rubric used to quantify generative processing behaviour from the students' written reflections.

Score	E: Engagement	B: Building on Feedback	C: Constructing Understanding
0	No mention of tutor use/engagement	No reference to feedback influencing learning	No reasoning or understanding described
1	Mentions tutor use or enjoyment (vague)	Mentions feedback/hints helped (vague)	States "I understood more" (no why/how)
2	Describes interaction features (pace, convenience, motivation)	Describes correcting mistakes or identifying gaps	Describes conceptual clarification in general terms
3	Clear engagement narrative (active participation, persistence)	Clear linkage: feedback → change in approach/answers	Explains learning mechanism or reasoning development
4	Rich engagement + evaluative detail (confidence, commitment, sustained use)	Rich, specific adjustment process (misconception repair, step-by-step progression, strategy change)	Strong conceptual account + metacognitive insight (explains why understanding changed; may note limits/accuracy issues)

4. Results

Findings reveal that when the prompts were pasted into free GAI, the resulting FIITS demonstrated high technical fidelity. Also, students employed moderate to high levels of generative processing as they engaged with the FIITS.

4.1. High Fidelity

Table 4 shows the overall fidelity during the implementation of the FIITS prompt, indicating a high to perfect success rate. Three of the five structural elements analysed showed perfect enactment of the prompt, while the remaining two showed minor implementation fidelity errors. These errors are indicated in Table 4 as structural errors (not staying within the chosen difficulty level) and answer violations (providing answers outright after an incorrect answer).

Table 4. Implementation fidelity.

Prompt Aspect	Implementation Fidelity (%) and Error Identification
Options provided as per the prompt.	100
Implementation of the chosen difficulty level aligns with the descriptions in the prompt (a).	~98.2% Structural error
The correct answer should never be given outright. Rather, guiding questions and hints should be provided (b).	~91% Answer violation
All content should be aligned to the specified curriculum.	100
Ten questions are given before summarising progress and asking how to proceed.	100

Table 5 shows that fidelity errors were unevenly distributed across platforms. The most popular platform, ChatGPT, accounted for the two structural errors and 74 of the 76 answer violations, with Gemini accounting for the remaining 2. Regarding the structural errors, the FIITS prompted progression to a higher difficulty level without explicitly offering the option to remain at the previously chosen level. Although the students could have instructed the FIITS to remain at their current difficulty level, the prompt's suggestive nature appeared to advance the students by default. These two cases of structural error, therefore, reflect reduced interface clarity rather than a breakdown in prompt logic.

Table 5. Breakdown of fidelity error rates.

Platform	Total Engagements	Structurally Sound	Structural Error	Structural Error Rate %	Number of Questions	Answer Violations	Answer Violation Rate %
ChatGPT	96	94	2	2.13%	960	74	7.71%
Gemini	15	15	0	0.00%	150	2	1.33%
Copilot	2	2	0	0.00%	20	0	0.00%
DeepSeek	1	1	0	0.00%	10	0	0.00%

Regarding answer violation errors, Tables 6 and 7 illustrate two ways in which ChatGPT violated the instruction not to provide answers outright in response to a student's incorrect answer. Table 6 shows the FIITS providing the correct answer with reasons on the student's first incorrect attempt.

In contrast to the overt violation of the prompt not to provide correct answers, Table 7 shows closer adherence to it. In this example, the FIITS did provide a guiding follow-up question in response to the student's incorrect answer, as per the prompt, but then provided the correct answer directly when the student continued to answer incorrectly.

Table 6. An outright answer provision violation by a ChatGPT FIITS.

Agent	Text
Gemini	Question 7: When a magnetic field through a coil increases, what happens to the induced current in the coil (according to Lenz's Law)? (A) The induced current increases the magnetic field, resisting the change. (B) The induced current weakens the magnetic field to allow more flux to pass through. (C) The induced current does not change. (D) The induced current causes a drop in the voltage across the coil. 👉 Type the letter of your answer (A, B, C, or D).
Student	B
ChatGPT	✗ Incorrect. Let's review the question carefully! Lenz's Law tells us that the induced current will always oppose the change that caused it. In this case, if the magnetic field through the coil increases, the induced current will generate its own magnetic field that opposes the increase. It will try to reduce the magnetic flux rather than allow it to increase. The correct answer is (A) (A) The induced current increases the magnetic field, resisting the change. Explanation: If the magnetic flux increases, the induced current creates a magnetic field in the opposite direction to reduce the flux. This is why Lenz's Law exists: it ensures the conservation of energy by opposing any change in the system. Would you like to try another question or revisit a topic from this session?
Student	CONTINUE

Table 7. An answer provision violation for an incorrectly answered follow-up question.

Agent	Text
Gemini	Next Question: When we inhale, the diaphragm plays a key role in pulling air into the lungs. What happens to the diaphragm when you breathe in? (A) It relaxes and moves up (B) It contracts and moves down (C) It stays in the same position (D) It expands sideways 👉 Type your answer (A, B, C, or D) and explain your choice! 😊
Student	D
ChatGPT	✗ Not quite! But you're thinking in the right direction! Let's break it down. Hint: The diaphragm is a dome-shaped muscle at the bottom of your ribcage. When you breathe in, your lungs expand, and they need more space. 👉 Try again! What happens to the diaphragm when you inhale? (A) It relaxes and moves up (B) It contracts and moves down (C) It stays in the same position (D) It expands sideways

Table 7. *Cont.*

Agent	Text
Student	A
ChatGPT	✗ Not quite! But I love that you're thinking through the options! Let's go step by step. Correct Answer: (B) It contracts and moves down. Explanation: When you inhale, the diaphragm contracts and moves downward, increasing the space in the chest cavity. This causes the lungs to expand, allowing air to rush in. When you exhale, the diaphragm relaxes and moves upward, decreasing space in the chest cavity and pushing air out of the lungs.
ChatGPT	Try This Quick Exercise: Take a deep breath and place your hand on your stomach. What do you feel when you breathe in? (A) Your stomach moves out as the diaphragm moves down. (B) Your stomach moves in as the diaphragm moves up. 👉 Type A or B based on what you feel! 😊
Student	A
ChatGPT	☑ Correct! Well done! 🎉 Explanation: When you inhale, your diaphragm contracts and moves down, pushing your abdominal organs slightly forward. This makes your stomach move outward as your lungs expand and fill with air. When you exhale, your diaphragm relaxes and moves up, pushing air out of your lungs, and your stomach moves inward."

4.2. Moderate to High Levels of Generative Processing

In general, moderate levels, with some evidence of high levels of generative processing, were deduced. This deduction is based on the observations that the students: voluntarily chose medium to harder difficulty levels for a considerable fraction of their engagement; displayed a reasonable correct-incorrect response rate; only occasionally avoided follow-up questions; did not noticeably resort to random or repetitive button pressing activity; and scored moderately to highly in their written reflection using a rubric (Table 3) designed to detect self-reported engagement in generative processing, as previously explained.

4.2.1. Difficulty Levels

Figure 1 indicates the number of students who chose each difficulty level within the FIITS for each of the five fortnightly engagements. This choice was voluntary. Although students were verbally encouraged to progress to more difficult levels over time, this was not incentivised. The 11 cases where students chose the easy option after week 1 could be evidence of avoidance of generative processing, as suggested by a student's confession in a group interview: "There were times that it was hard to focus because maybe the thing was too easy for me, or maybe sometimes I was just rushing to finish the questions to submit". In contrast, some students reported intentionally progressing through the levels, for example, "Choosing difficulty levels was fun too—I started with Easy questions to understand the basics, then moved up to Hard and Challenging, which made me apply my knowledge to solve more complex problems. This was a great learning experience for me." Consistent with this student's statement, Figure 2 shows migration towards the hard level in the fifth and ninth weeks, and to a lesser extent, towards the challenging level in the seventh week. Inconsistent with this student's statement, however, the challenging level was largely avoided, and it appears that some students regressed from hard to medium between weeks 5 and 7.

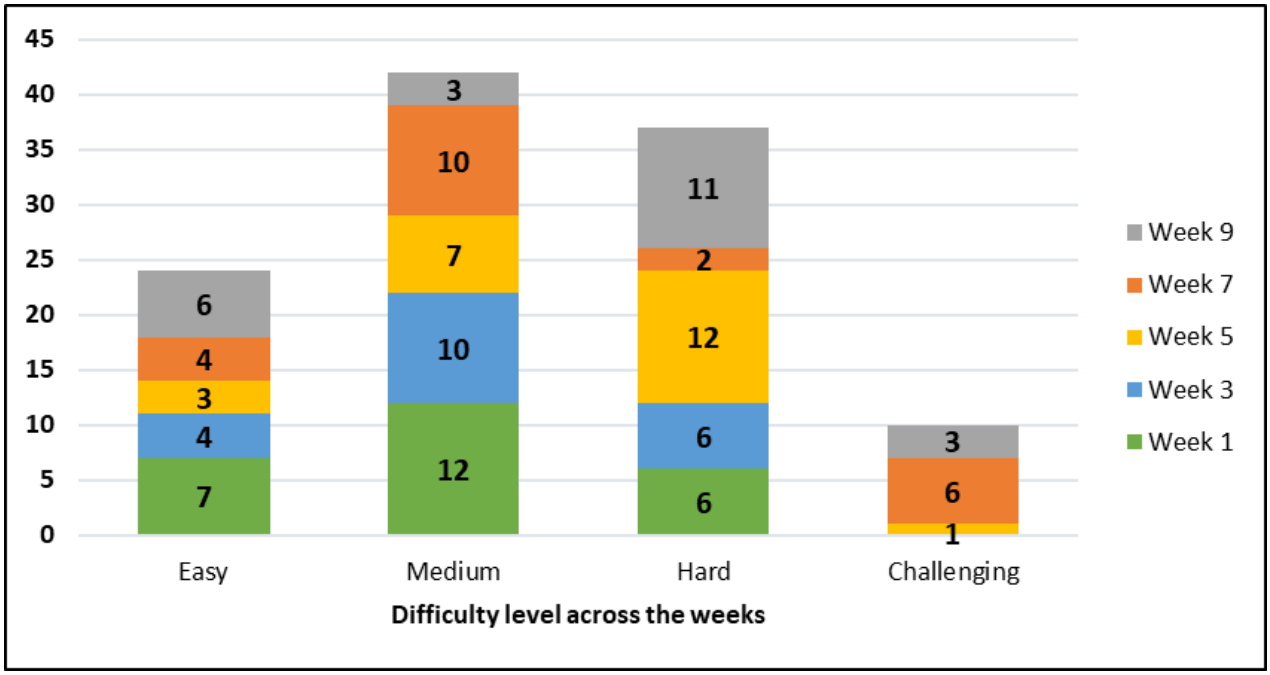


Figure 1. Number of students who chose each difficulty level across the 10-week intervention period (n = 114).

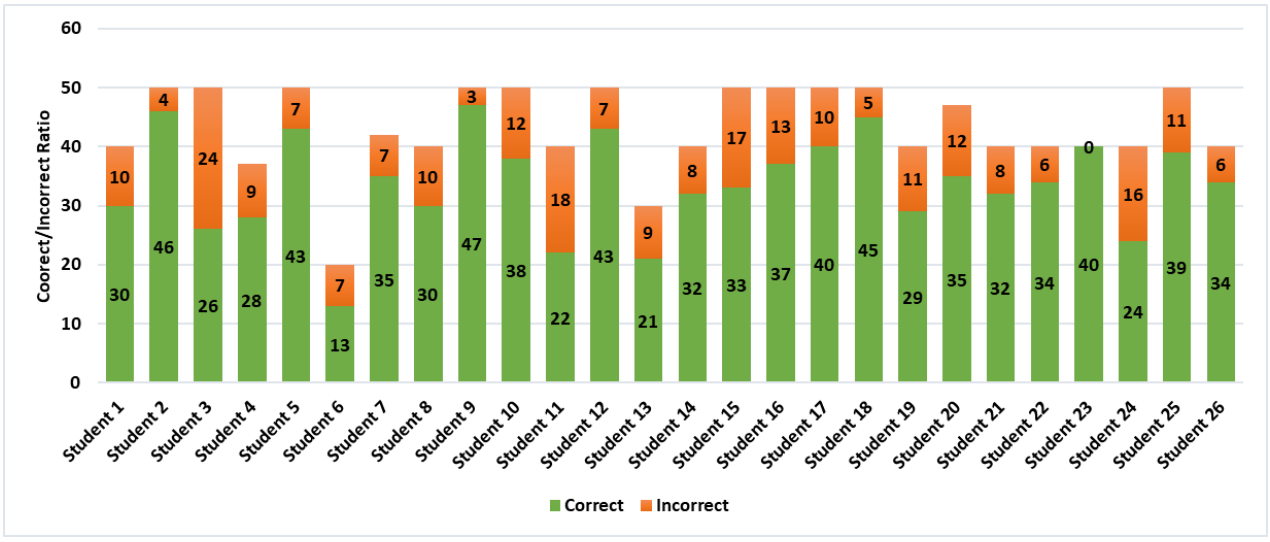


Figure 2. Number of sets of engagements that showed each possible correct and incorrect score (/10).

4.2.2. Correct–Incorrect Response Rates

Figure 2 displays the frequency with which each possible score was attained across the sets of engagement. Since each engagement set consisted of 10 questions, the maximum score per set was 10, and the maximum total score across the five weeks was 50. Figure 2 shows a high frequency of perfect scores. Similarly, a high mean score (35/50) was calculated. It is possible that some of this success arose from gaming the system by using an external GAI platform to obtain correct answers. During the focus group discussions, one student admitted to doing this when pressed for time. However, these high success rates appeared more largely attributable to genuine engagement in generative processing, as evidenced by the following statement:

> Using ChatGPT as an intelligent tutor made me feel as though I always had a personal advisor available. Whether I was stuck on a perplexing concept or simply wanted to explore an idea further, ChatGPT offered straightforward explanations, relatable examples, and immediate feedback. It adapted to my questions and learning pace, making studying feel less formal and more conversational. For example, if I didn't understand a mathematical formula or the context of a historical event, I could ask clarifying questions without feeling judged or rushed. This created a learning process that felt more natural and less intimidating, especially for topics I typically struggled with.
>
> (Student 6—Reflections)

Additional explanations for the high scores are the revisionary nature of the FIITS and students' misunderstanding of the incentive system. Since the FIITS were aligned to a school-level curriculum, these university students may have experienced them as a revision of previously mastered work. Moreover, the focus group discussion revealed that some students had misunderstood the assessment expectations: "I thought we would lose marks if we got answers wrong, so I redid it until I got full marks" (Focus group 3).

4.2.3. Avoiding Follow-Up Questions

Table 8 illustrates a student avoiding engagement with the FIITS's optional follow-up questions. Since answering these follow-up questions was not incentivised by mark allocation, a student who nevertheless answered them was likely self-directed, suggesting generative processing. In contrast, a student who avoided these follow-up questions may have been attempting to minimise generative processing by doing only the minimum required to secure a mark. Even if this avoidance of follow-up questions indicates an attempt to game the system, its occurrence was rare, affecting only one or two students each week.

Table 8. An example of a student ignoring further engagement. Student input is highlighted.

Agent	Text
Student	Substance sodium and gas hydrogen
Gemini	Excellent! You've made a great adjustment. Let's analyze your new answer: • Sodium: ○ Sodium is indeed a solid at room temperature. ○ It has a shiny, metallic appearance when freshly cut. ○ Critically, sodium reacts very vigorously with water, producing heat and a gas. This aligns perfectly with the observation. ○ The resulting solution is sodium hydroxide (NaOH), which is a strong base. And strong bases will turn red litmus paper blue. • Hydrogen: ○ When sodium reacts with water, it produces hydrogen gas (H_2). This gas is flammable and the reaction is exothermic, producing heat. Therefore, your selection of sodium as the substance and hydrogen as the gas is correct. Now, to make this even more relevant to a South African context, consider this: Imagine a scenario where a group of learners are conducting an experiment with different metals in a school lab. If they accidentally introduce a small piece of sodium into water, it could lead to a dangerous reaction. This highlights the importance of safety precautions when handling reactive substances in a lab setting. Do you understand why Sodium and Hydrogen are the correct answers? Do you have any further questions?
Student	Next question

4.2.4. Button-Pressing Possibilities

Random or repetitive button pressing is a possible way to game the system, which may be encouraged if students are aware that marking does not incentivise giving the correct answer. Possibly conducive to random button-pressing behaviour, several students said they prioritised completing the activity quickly rather than answering carefully, particularly under time constraints. However, the pattern displayed in Figure 3 does not suggest widespread random responding or an overreliance on any single option. Instead, the spread of A–C responses appears relatively balanced. Fewer D responses are evident, corresponding to the observation that the FIITS rarely positioned the correct option at D.

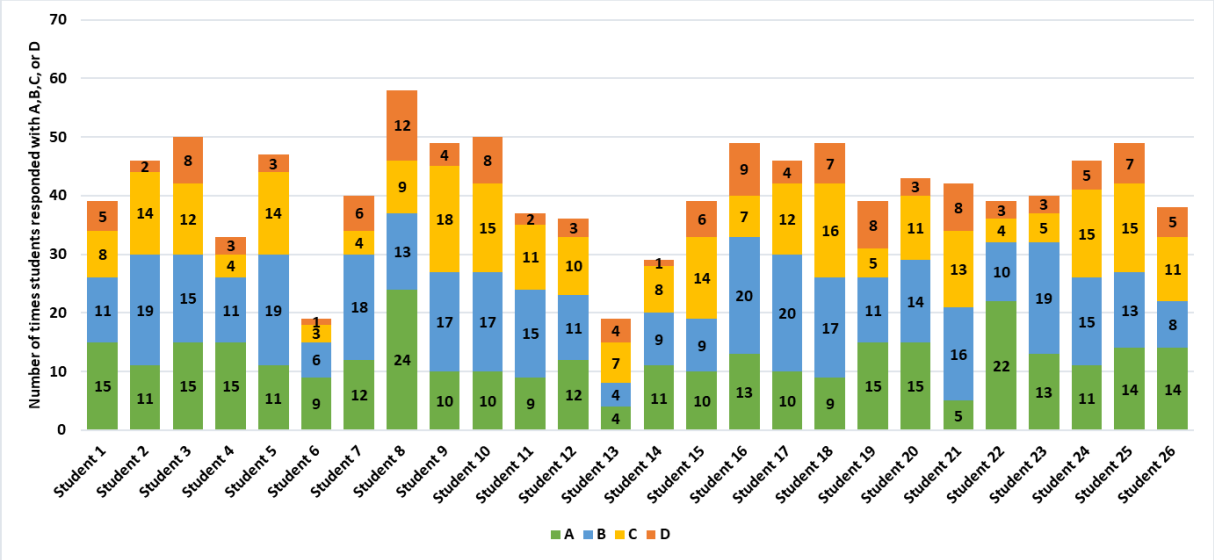


Figure 3. The number of times each student chose each multiple-choice option.

Figure 3 shows variations in total responses across the students. Totals under 50 resulted from students failing to complete all engagements (e.g., Student 6) or engaging in more challenging levels, which deviated from the multiple-choice format (e.g., Student 13). Totals over 50 resulted from the student providing incorrect responses to one or more of the 10 initial questions per set, prompting the FIITS to ask follow-up questions to guide them to the correct answers.

4.2.5. Generative Processing

Figure 4 represents the analysis of the reflection-based generative processing scores using the 0–4 scale rubric for each of the dimensions of engagement, building, and constructing (see Table 3). Engagement scores were strongly concentrated at the upper end of the scale, with all but one student's reflection demonstrating rich (n = 15) or clear (n = 8) engagement, i.e., engagement at levels 4 and 3, respectively. No students scored below level 2 on this dimension, indicating consistently high levels of reported interaction, motivation, and sustained participation when using the intelligent tutor.

A similar but more pronounced pattern was observed for building. Most students (n = 23) achieved the maximum score of 4, with only one student scoring 3 and no scores recorded below level 3. This indicates widespread evidence of feedback-driven adjustment, including the use of hints, explanations, and corrective guidance to refine understanding.

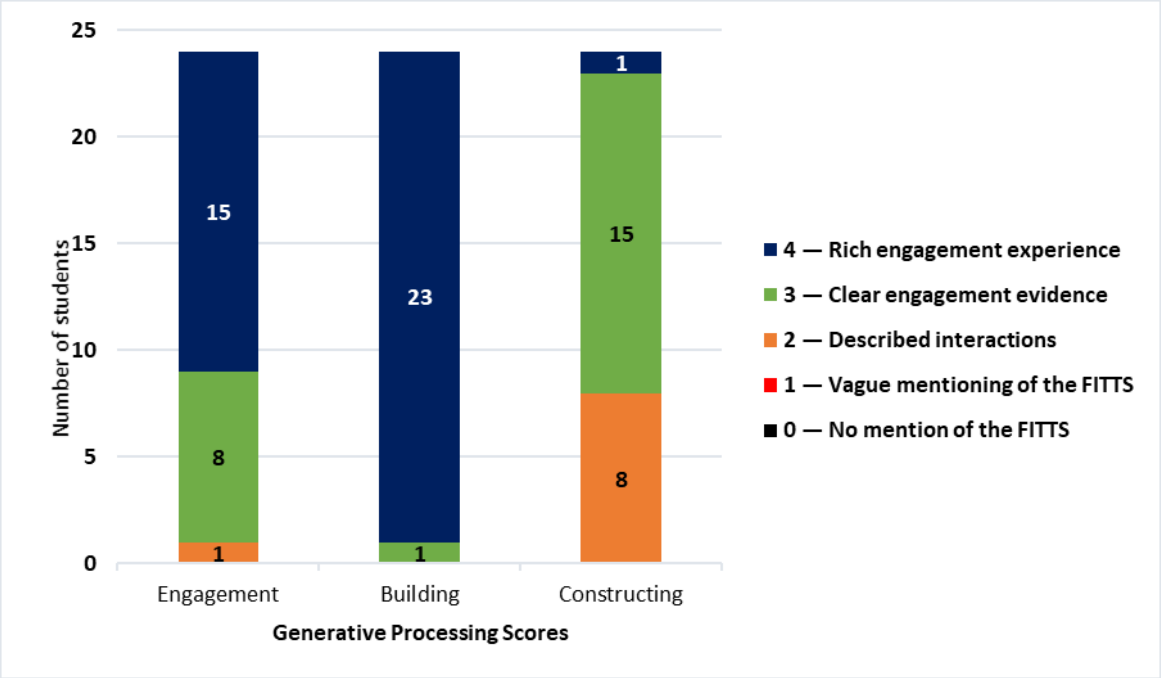


Figure 4. Reflective-based generative processing scores.

In contrast, Constructing scores showed greater spread and moderation. While most students scored at level 3 (n = 15), fewer reached level 4 (n = 1), and a notable number remained at level 2 (n = 8). This suggests that although students frequently reported improved understanding, fewer articulated fully reconstructed or transferable conceptual reasoning in their reflections.

Student 18 is an example of a student who demonstrated sustained engagement, feedback-driven adjustment, and conceptual construction when engaging with the FIITS. Their written reflection was evaluated at the maximum level (4) for each aspect of generative processing. Below is an extract from this reflection:

> When I struggled to determine whether a species is reduced or oxidised, the tutor broke the redox process into less intimidating steps: determining oxidation numbers, determining electron transfer, and writing half-reactions. It immediately corrected me if I had gotten it wrong and explained why the solution was wrong.

This extract evidences B—building through iterative feedback and adjustment, while the explicit articulation of redox mechanisms and half-reactions reflects C—constructing at a high level. The student not only reports improvement but also explains how conceptual understanding was reorganised, aligning with their maximum C score.

In contrast, Student 17 (assigned the scores: engagement = 3, building = 4, and construction = 2) reflected strong engagement and reliance on feedback, but with more limited conceptual articulation:

> Having to have immediate feedback, made the learning process more enacting as If I was unclear about something, the interactive AI would give me the correct answer. Unlike the typical system of having to search or wait to come in contact with a teacher to get clarification. The AI gave it immediately and even went further on justifying its answers where I was correct. Where the answer was wrong it also gave me correct answer immediately and even asked if I needed further explanation.

The above quote emphasises interaction, immediacy, and clarification, with little reference to the student providing further explanation; it focuses only on correct answers. Thus, the student primarily frames learning as receiving confirmation of correct answers. This pattern supports the lower constructing score, as conceptual reconstruction appears to be externally driven by feedback rather than internally articulated by the student.

Across reflections, a consistent pattern emerged: high engagement and responsiveness to feedback do not automatically translate into articulated conceptual construction. Students who scored highest on C explicitly described disciplinary mechanisms, causal relationships, or conceptual reorganisation. By contrast, students with moderate C scores emphasised immediacy, correctness, and reassurance, without detailing how knowledge structures changed.

This pattern helps explain the quantitative distribution observed in the findings, where E and B scores cluster at the upper end, while C scores show greater spread and moderation. The reflections suggest that the FIITS supported engagement and feedback-driven learning, but either deeper generative processing was limited, or the prompt's emphasis on multiple-choice questions, for ease of answering, obscured the verbalisation of underlying conceptual change.

5. Discussion

The findings demonstrate high to perfect technical implementation fidelity and moderate to high levels of student generative processing for the two FIITSs investigated. These findings contribute to a growing body of evidence that student–GAI dialogue can result in deep learning through generative processing for the same reasons that student–lecturer dialogue can (Hofmann et al., 2025; Katsenou et al., 2025). Dialogue enables the sharing of cognitive resources, thereby decreasing cognitive load (Kirschner et al., 2018). Additionally, pedagogically skilful dialogue can direct attention towards relevant items for selection, prompt concept linkage, and provide formative feedback to guide concept construction (Fiorella, 2023). The finding that even free GAI can be prompted to engage in such pedagogically skilled dialogue is consistent with findings from other studies (Ling & Jan, 2025; Yan et al., 2024; Zeng, 2025).

The positive findings reported above justified our use of the two investigated prompts as a starting point, from which we abstracted a generic prompt framework. We present the Flipped-Interaction Prompt (FIP) Framework in Figure 5. The FIP Framework is a template for creating customised prompts for GAI to behave as a FIITS. Square brackets indicate placeholders to guide educators in applying their pedagogical and curricular expertise, without requiring technical expertise such as coding, to create a suitable FIITS prompt for their context. The FIP Framework has contextual and pedagogical framing sections, which are described and justified below.

The context framing sections of the prompt require customised completion of the relevant square brackets to specify the following: subject domain (e.g., history, nursing, or economics); level of instruction (e.g., Grade 12, first-year undergraduate, or postgraduate); mode of interaction (e.g., multiple-choice simulation and dialogue, case-based); list of topics to choose from; and difficulty levels to choose from. Providing difficulty level choices enables differentiation and cognitive load management (Fiorella & Mayer, 2021).

Please create an interactive, gamified, [*Insert Format Type: e.g., multiple-choice, case-based*] dialogical learning simulation for students in [*Insert Subject*] at [*Insert Level or Grade*], following the [*Insert Curriculum or Outcomes*]. I will pretend to be such a student.
Begin by presenting a list of main topics for me to choose from, such as:

- [*Insert Topic 1*]
- [*Insert Topic* 2]
- [*Insert Topic 3*]

Once I select a topic, ask me to choose a sub-topic and a difficulty level:

- Easy – [*Insert Type of Question e.g., basic identification or definition*]
- Medium – [*Insert Type e.g., application questions*]
- Hard – [*Insert Type e.g., multi-step reasoning or multiple-answer formats*]
- Challenging – [*Insert Type e.g., dual-layered decision-making*]

The simulation should never give me the correct answer outright. If I make a mistake, provide detailed, interactive guidance, using probing questions, hints, or reasoning steps to help me reach the answer myself.
Ensure all content is aligned with [Insert Curriculum/Framework], and contextualised using real-life or local examples, such as:

- [*Insert Example 1*]
- [*Insert Example 2*]

Guide me through practical scenarios, such as:

- [*Insert Simulation or Problem Context Here*]

After ten questions, ask if I'd like to change topics, increase difficulty, or revisit concepts. Provide feedback and summarise my progress at the end.
Goal: Help me learn through engagement, reflection, and conceptual reasoning, not memorisation.

Figure 5. The Flipped-Interaction Prompt Framework (FIP Framework).

The pedagogical framing sections of the prompt are guided by Laurillard's (2013) Conversational Framework, which aims to facilitate dialogue that promotes articulation, feedback, practice, and conceptual refinement. Articulation is promoted by asking questions. The students' answers to these questions are followed by personalised feedback. Practice is provided through multiple questions asked on each chosen topic. Conceptual refinement is encouraged when a student chooses an incorrect answer, and a Socratic dialogue follows to guide them to analyse alternative answers. Socratic dialogue may help students revise their misconceptions and develop metacognition (Elder & Paul, 2019), as well as refine their understanding iteratively (Shute, 2008). The framework emphasises conceptual understanding over memorisation since this is more durable and promotes higher-order thinking skills (Bransford et al., 2000).

The FIP Framework aligns with UNESCO's (2024) AI competency framework for teachers. The flipped, interactive nature of the FIITS generated using this framework encourages active learning. This human-centric, ethical use of GAI to promote learning contrasts with students' passive, unethical dependence on GAI to complete tasks with minimal generative processing. The FIP Framework's guidance on using free GAI platforms enhances

equitable access to AI for pedagogical and professional development. Furthermore, the FIP Framework is intended to enhance the prompt proficiency of students, lecturers, and teachers, thereby advancing their competence in using GAI in pedagogically powerful, ethical, and human-centric ways.

6. Conclusions

This article presented the Flipped-Interaction Prompt (FIP) Framework to address this underutilisation. This framework was abstracted from two prompts that were found to be implemented with high to perfect fidelity when pasted into free GAI and to prompt moderate to high generative processing, engaging students. Lecturers and students could benefit from entering their module details into the FIP Framework template. The resulting prompt can be pasted into free GAI to turn it into a FIITS or application in various contexts and disciplines. In this way, the FIP Framework could expand educators' GAI pedagogical repertoires and enhance students' generative processing. Future research could replicate this study in other contexts and with larger samples to evaluate the influence of these variables on limiting the generalisability of this study. Additionally, the technical and pedagogical efficacy of instantiations across other discipline domains and study levels, and with various GAI platforms, could be investigated. Other areas of interest include evaluating how students' engagement in FIITS can be assessed for university credit, as well as the role of frameworks, such as this one, in developing prompt literacy among students and educators. The FIP Framework is offered as a tool to increase the uptake of free GAI prompting as Flipped-Interaction Intelligent Tutoring Systems (FIITSs). Such an uptake of FIITS prompting should promote student thinking through GAI use at a time when fears that GAI undermines higher education are abundant.

Author Contributions: Conceptualization, S.J.S. and A.E.S.; methodology, S.J.S. and A.E.S.; software, S.J.S.; validation, S.J.S.; formal analysis, S.J.S.; investigation, S.J.S. and A.E.S.; resources, S.J.S. and A.E.S.; data curation, S.J.S. and A.E.S.; writing—original draft preparation, S.J.S. and A.E.S.; writing—review and editing, A.E.S.; visualization, S.J.S.; supervision, A.E.S.; project administration, S.J.S. and A.E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the General/Human Research Ethics Committee (GHREC) of the University of the Free State (protocol code: UFS-HSD2024/2116; date of approval: 20 February 2025).

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are not publicly available due to ethical restrictions related to participant privacy and institutional research ethics approval.

Acknowledgments: The authors declare that GAI was used in a limited and supportive capacity during the preparation of this manuscript. Specifically, ChatGPT-4 was used to proofread selected sections of the manuscript and to enhance language clarity and readability. GAI was not used to generate substantive scholarly content or to write the manuscript in full. ChatGPT-5 was used to guide analytical processes, including structuring analyses and assisting with certain data analysis tasks. All analytical decisions, interpretations, and conclusions remain the responsibility of the authors. The authors have reviewed and approved the final version of the manuscript and take full responsibility for its content.

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

Abbreviations

The following abbreviations are used in this manuscript:

GAI	Generative Artificial Intelligence
FIITS	Flipped-Interaction Intelligent Tutoring System
ICTs	Information and Communication Technologies

References

Avcı, G., Owusu-Boateng, O., Vampa, M., & Khalid, A.-M. (2025). *Intelligent tutoring systems* (pp. 46–57). Deep Science Publishing. [CrossRef]

Baker, R. S. J., Corbett, A. T., Roll, I., Koedinger, K. R., Aleven, V., Cocea, M., Hershkovitz, A., de Caravalho, A., Mitrovic, A., & Mathews, M. (2013). Modeling and studying gaming the system with educational data mining. In *International handbook of metacognition and learning technologies* (pp. 97–115). Springer.

Blonder, R., Feldman-Maggor, Y., & Rap, S. (2024). Are they ready to teach? Generative AI as a means to uncover pre-service science teachers' PCK and enhance their preparation program. *Journal of Science Education and Technology*, *34*(6), 1301–1310. [CrossRef]

Bransford, J. D., Brown, A. L., & Cocking, R. (2000). *How people learn*. National Academy Press.

Coelho, R., Bjune, A. E., Ellingsen, S., Solheim, B. M., Thormodsæter, R., Wasson, B., & Cotner, S. (2025). A call for clarity: Biology students advocate for guidelines for the use of generative AI in higher education. *Journal of Science Education and Technology*, *34*(4), 853–865. [CrossRef]

Creswell, J. W., & Cresswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd ed.). SAGE Publications.

Elder, L., & Paul, R. (2019). *The art of asking essential questions: Based on critical thinking concepts and Socratic principles*. Bloomsbury Publishing PLC.

Fiorella, L. (2023). Making sense of generative learning. In *Educational psychology review* (Vol. 35, Issue 2). Springer. [CrossRef]

Fiorella, L., & Mayer, R. E. (2021). The generative activity principle in multimedia learning. In *The cambridge handbook of multimedia learning* (pp. 339–350). Cambridge University Press. [CrossRef]

Franco, M. (2019). SAMR and technology integration. In *Technology and the curriculum: Summer 2019*. Power Learning Solutions.

Hofmann, F., Daunicht, T.-M., Plößl, L., & Gläser-Zikuda, M. (2025). Promoting reflection skills of pre-service teachers—The power of AI-generated feedback. *Education Sciences*, *15*(10), 1315. [CrossRef]

Hughes, L., Malik, T., Dettmer, S., Al-Busaidi, A. S., & Dwivedi, Y. K. (2025). *Reimagining higher education: Navigating the challenges of generative AI adoption*. Information Systems Frontiers. [CrossRef]

Katsenou, R., Kotsidis, K., Papadopoulou, A., Anastasiadis, P., & Deliyannis, I. (2025). Beyond assistance: Embracing AI as a collaborative co-agent in education. *Education Sciences*, *15*(8), 1006. [CrossRef]

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano, J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 213–233. [CrossRef] [PubMed]

Laurillard, D. (2013). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge. [CrossRef]

Lee, G., Yun, M., Zhai, X., & Crippen, K. (2025). Artificial intelligence in science education research: Current states and challenges. *Journal of Science Education and Technology*, *35*, 110–127. [CrossRef]

Ling, Y., & Jan, J. M. (2025). Voices from the Flip: Teacher Perspectives on Integrating AI Chatbots in Flipped English Classrooms. *Education Sciences*, *15*(9), 1219. [CrossRef]

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, *36*(1), 8. [CrossRef]

Ramnarain, U., Ogegbo, A. A., Penn, M., Ojetunde, S., & Mdlalose, N. (2024). Pre-service science teachers' intention to use generative artificial intelligence in inquiry-based teaching. *Journal of Science Education and Technology*, *34*(6), 1272–1285. [CrossRef]

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. [CrossRef]

UNESCO. (2024). AI competency framework for teachers. In *AI competency framework for teachers*. UNESCO. [CrossRef]

Watts, F. M., Liu, L., Ober, T. M., Song, Y., Jusino-Del Valle, E., Zhai, X., Wang, Y., & Liu, N. (2025). A framework for designing an ai chatbot to support scientific argumentation. *Education Sciences*, *15*(11), 1507. [CrossRef]

Watts, K. J. (2025). Paying the cognitive debt: An experiential learning framework for integrating AI in social work education. *Education Sciences*, *15*(10), 1304. [CrossRef]

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. *arXiv*, arXiv:2302.11382. Available online: http://arxiv.org/abs/2302.11382.

Yan, W., Nakajima, T., & Sawada, R. (2024). Benefits and challenges of collaboration between students and conversational generative artificial intelligence in programming learning: An empirical case study. *Education Sciences*, *14*(4), 433. [CrossRef]

Zeng, S. A. (2025). Dialogues in play: Conversational AI and early mathematical thinking. *Education Sciences*, *15*(11), 1516. [CrossRef]
Zhai, X. (2024). Transforming teachers' roles and agencies in the era of generative AI: Perceptions, acceptance, knowledge, and practices. *Journal of Science Education and Technology*, *34*, 1323–1333. [CrossRef]
Zheng, L., Shi, Z., Fu, Z., & Liu, S. (2025). The impacts of intelligent feedback on learning achievements and learning perceptions in inquiry-based science learning: A meta-analysis of studies from 2013 to 2023. *Journal of Science Education and Technology*, *34*(4), 737–756. [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.