Do Generative Artificial Intelligence (GenAI) and Science Education Mix?
A Systematic Review of the Literature

Authors
Kason Ka Ching Cheung*
cheungkac@eduhk.hk
Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong, China

Amina Zerouali*
amina.zerouali@tum.de
Department of Educational Sciences, School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

Jenna Koenen
jenna.koenen@tum.de
Department of Educational Sciences, School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

Sibel Erduran
sibel.erduran@education.ox.ac.uk
Department of Education, University of Oxford, Oxford, United Kingdom

*Both authors have equal contribution towards the manuscript.

Declaration
The authors declare that they have no conflicts of interest.

Funding
The work has been supported by ESERA Early Career Researcher Travel Grant 2024 given to Kason Cheung.

Abstract

Generative Artificial Intelligence (GenAI) has been incorporated in different contexts of science education, such as K-12 education and teacher education. However, science educators have raised concerns about the dissonance between what GenAI can offer and what science education aspires to teach to learners. For example, GenAI may undermine social learning goals, learning of science, and epistemic understanding of science. Using a systematic review approach, we examined the variables, data collection tools, and features of teaching interventions in 22 empirical studies to trace how GenAI is integrated into science education. Critically, we identified the extent to which research studies reflect synergy as well as dissonance between GenAI and science education. In so doing, we explored examples of fruitful integrations of GenAI in science education. Our findings showed that variables, data collection tools, and teaching interventions focused on students' understanding of either science or GenAI. We propose a framework that guides the unification of diverse learning goals in relation to disciplinary practices in science education in the context of GenAI. Based on this framework, we propose future research and teaching interventions that harness synergy as well as dissonance to foster students' and teachers' higher-order thinking while using GenAI in learning science or developing their pedagogical competence, respectively.

Keywords GenAI, ChatGPT, epistemic, social learning

1. Introduction

Generative Artificial Intelligence (GenAI) has been described as a "game-changer" (Zhai et al., 2024) in the field of science education. Contrasting with generic artificial intelligence (AI) applications, GenAI applications are capable of creating new artificial content by building on existing digital content, such as images, videos, audio, and texts, through the distribution and learning patterns of such digital content (Hu, 2023; Jovanović, 2023). At the current time, there are systematic reviews on the use of AI in science or STEM (science, technology, engineering, and mathematics) education (Jia, Sun, & Looi, 2024; Xu & Ouyang, 2022). However, these systematic reviews focus merely on generic AI applications and on summarizing pedagogical affordances of various AI technologies in science and STEM education. In fact, it is as if GenAI is considered a discipline, and science is considered another discipline, although they both lend themselves to specialized ways of teaching and learning. The absence of a systematic review in identifying the disciplinary use of GenAI in the teaching and learning of science is notable in the literature.

More importantly, there is growing interest in and concern about the application of GenAI and science education (Avraamidou, 2024a; Cooper & Tang, 2024; Cheung et al., 2025; Nyaaba, 2024; K.-S. Tang & Cooper, 2024). On the one hand, researchers see the potential of GenAI in facilitating students' learning of scientific practices (Erduran and Levrini, 2024), such as analyzing data and model prediction. Some scholars argue that educators should not lose sight of epistemic outcomes in science (Avraamidou, 2024a; Cooper & Tang, 2024). For example, AI tools in educational studies mainly focus on cognitive domains instead of the social or dimensions of the nature of scientific knowledge (Avraamidou, 2024a). To use GenAI appropriately in science learning, students also need to understand the epistemic authority of scientific information created by GenAI in contrast with that of scientists (Tang & Cooper, 2024). GenAI offers both synergy and dissonance with different aspects of science education, including K-12 education (Cheung et al., 2025) and teacher training (e.g., Lee & Zhai, 2024b). Such synergy entails the promotion of students' cognitive and affective outcomes, as well as facilitating scientific practices and giving feedback to students' scientific products. Also, dissonance between science and GenAI disciplines is related to ethics, inaccuracies, and bias of scientific outputs created by GenAI, while researchers and teachers can capitalize on such opportunities to develop students' critical thinking and metacognition. Such synergy and dissonance between GenAI and science education need to be studied through methodological approaches, measurement, and teaching interventions specific to the learning of science.

To date, no systematic review has investigated the extent to which variables, instruments, teaching interventions, and outcomes of teaching interventions can or cannot characterize the relationship between GenAI and the teaching and learning of science. Given concerns raised by science educators regarding the use of GenAI in science education (e.g., Avraamidou, 2024), the objective of this review is to find out the extent to which emerging empirical studies go beyond the simplification of GenAI as a tool to promote learning of science content knowledge. Existing review studies (e.g., Almasri, 2024), which mainly target generic AI applications instead of GenAI, only focus on bibliometric details of empirical papers and do not provide sufficient insights into disciplinary applications to science education. By synthesizing literature at this early stage, we aim to contribute to the body of knowledge at the intersection of science education and GenAI and provide some potential cautionary points in pursuing future research in this area. Such an approach enables us to propose research directions that foster meaningful use of GenAI in science teaching and learning.

The main aims of the paper are to identify (a) how to target variables from the reviewed empirical studies to reflect the synergy and dissonance between GenAI and science education, (b) how data collection methods enable the investigation of the interactions between GenAI and science education, and (c) the characteristics, contributions, and shortcomings of interventions that incorporate GenAI in science education. In order to address these aims, first, we illustrate the different perspectives on the use of GenAI in science education. Subsequently, the notions of synergy and dissonance are problematized in relation to how GenAI is incorporated into science education. Based on the analysis of empirical studies, we aimed to generate a theoretical framework that conceptually guides future research in science education, while components within the framework can be used as an analytical tool in further empirical studies.

2. Literature Review

2.1 Perspectives on the use of GenAI in science education

Generative artificial intelligence (GenAI) is a kind of artificial intelligence that creates new texts, images, audio, and videos (Baidoo-Anu & Ansah, 2023). Extant science education literature (Cooper & Tang, 2024; Feldman-Maggor, Blonder, & Alexandron, 2024; Tang & Cooper, 2024; K. S. Tang, 2024) focuses predominantly on the Generative Pre-trained Transformer (GPT). GPT interprets and creates output of human-like texts in different languages by using publicly available digital content data and natural language processing techniques (Baidoo-Anu & Ansah, 2023). GPT can now be categorized into different versions, with the current version of GPT-o1 that surpasses the accuracy of GPT-4 to understand and reason in the context of different medical scenarios (Xie et al., 2024). Because of the evolution of GenAI in the field of scientific research, science education researchers (e.g., Feldman-Maggor et al., 2024) are currently exploring various perspectives to examine GenAI's potential for supporting science teaching and learning.

One key perspective on GenAI in science education focuses on the development of learners' cognitive outcomes. In educational research, GenAI is often positioned as an agent of learning and assessment of cognitive outcomes (Chiu, 2023; Lo, 2023). GenAI tools like ChatGPT can answer questions and summarise information, as well as facilitate peer collaboration (Lo, 2023). More importantly, ChatGPT can also check students' concepts and provide real-time feedback (Kolade, Owoseni, & Egbetokun, 2024; Lo, 2023). For students' learning in science, evidence suggests that the use of GenAI benefits students' learning of content knowledge (Ng, Tan, & Leung, 2024). In terms of the utility of GenAI for teachers, teachers can use GenAI to plan activities that engage students in learning content knowledge of science (Lee & Zhai, 2024b).

Another perspective on GenAI in science education relates to the epistemic aspects of science education. From an epistemic standpoint, content knowledge should be taught explicitly alongside the nature of scientific knowledge (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Schwartz, Lederman, & Crawford, 2004). In relation to GenAI, there are two aspects regarding the epistemic perspective: *learners' epistemic understanding* and *learners' engagement in epistemic practices* of GenAI in science education. Scientific information in GenAI often inherits bias (Krist & Kubsch, 2023; Cheung et al., 2025), which requires students' and teachers' epistemic understanding of how GenAI creates scientific knowledge. For instance, scientists are often portrayed by GenAI as white males who wear laboratory coats (Cooper & Tang, 2024). Developing students' and teachers' epistemic understanding through GenAI can potentially counteract traditional stereotypes as well as social injustices. Such epistemic understanding can entail aims and values of using GenAI in scientific activities, methods of how GenAI generates scientific knowledge, sources and limitations of scientific knowledge in GenAI, scientific practices influenced by GenAI, as well as socio-institutional understanding of how societal bias influenced GenAI-scientific knowledge (Cheung et al., 2025; Cheung and Zhang, 2025). Such a broad and holistic account of GenAI's contextualization of epistemic understanding is consistent with existing literature on the nature of science (e.g., Erduran & Dagher, 2014). Apart from enhancing epistemic understanding, scholars call for the theorization of epistemic practices (Ho, 2023; Tang, 2024) to clarify how GenAI can address epistemic objectives for science teaching and learning. Epistemic practices are socially accomplished ways that members within a community propose, evaluate, and formalize knowledge claims (Kelly, 2018). Given social injustices that may be inherent in GenAI outputs (Feldman-Maggor et al., 2024; Krist & Kubsch, 2023), teachers and students need to be given explicit opportunities to evaluate and negotiate the legitimacy of claims in GenAI outputs.

Our approach to harnessing the potential of empirical studies is to craft methodological approaches and variables that help us understand the interactions between the features of GenAI and science education. Although an array of theoretical papers argue for cognitive (Zhai & Nehm, 2023) and epistemic perspectives (Cooper, 2023; Cooper & Tang, 2024) on the use of GenAI in science education, empirical studies in this field are still in their infancy. To advance the field of science education, measurement variables, methodological approaches, and intervention methods need to be examined. Furthermore, science education research will benefit from nuanced approaches to understanding how GenAI incorporates disciplinary aspects of science in science education. Hence, our current review seeks to identify the extent to which existing research on GenAI in science education considers disciplinary learning outcomes in science.

2.2 Synergy and dissonance between GenAI and science education

Building on the preceding discussion, in the following sections, we explain the synergy and dissonance between GenAI and science education. For example, in terms of synergy, GenAI tools can help students acquire science content knowledge and develop their intrinsic motivation to learn science (Ng et al., 2024). On the other hand, in terms of dissonance, GenAI tools without tailored prompt training might undermine effective learning of science. First, we explore aspects of synergy, particularly in relation to content and affective aspects of science learning. Subsequently, we highlight some issues of dissonance, namely in terms of epistemic and social injustices that GenAI may present in science education (Krist and Kubsch, 2024). Together, these contrasting accounts present a nuanced articulation of the role that GenAI can play in science education. Ultimately, the goal of science education is to seek the means by which dissonance can be transformed into useful experiences that can enhance the learning of science.

2.2.1 Synergy regarding science content and affective domain

Some researchers argue that GenAI improves students' content knowledge to learn science (Alasadi & Baiz, 2023; Hamid, Zulkifli, Naimat, Yaacob, & Ng, 2023). This is because when students input scientific questions, GenAI provides immediate, detailed answers to the questions. More significantly, when assessment questions are input into GenAI (e.g., GPT-4), GenAI outperformed students in cognitive-demanding tasks such as problem-solving (Zhai, Nyaaba, & Ma, 2024). Thus, GenAI can be an intellectual agent that provides feedback on students' problem-solving. It can also personalize learning materials, provide real-time feedback, and help overcome language barriers (Alasadi & Baiz, 2023). The recent version of OpenAI's GPT-4V has the capability to process and produce texts, images, audio, and sound (Bewersdorff et al., 2024). Such capability aligns with the necessities of teaching and learning science as it involves different modalities (Tang, 2024; 2023). In other words, existing GenAI tools can help students acquire content knowledge from both multimodal and assessment perspectives.

The affordances of GenAI can be linked to improving students' motivation to learn science. GenAI tools such as ChatGPT provide flexibility for students to practice content knowledge questions (Du & Alm, 2024). For instance, human-chatbot interactions and human-robot interactions have the capability to promote students' self-esteem and motivation despite facing challenges in learning content knowledge (Chiu, 2023; Chiu, Moorhouse, Chai, & Ismailov, 2024). Also, these kinds of chatbots are able to boost students' confidence and lower students' anxiety (Chiu, 2023; Yang & Shulruf, 2019). Despite the cognitive and affective affordances of GenAI in science education, the disciplinary-specific potential of GenAI in facilitating students' science learning remains a relatively uncharted research territory.

2.2.2 Dissonance about the epistemic domain and social injustices

Researchers have raised concerns that GenAI will simplify students' learning goals of science, specifically at an epistemic and social learning level (Avraamidou, 2024b; Cheung et al, 2025). Science education entails three learning goals: conceptual, epistemic, and social learning (Duschl, 2008). In the emerging line of GenAI-science education literature, the focus seems to emphasise conceptual learning. Epistemic and social learning goals seem to be underinvestigated so far in the literature. This observation warrants the rationale for investigating the extent of such dissonance in contemporary empirical papers that use GenAI in science education.

From an epistemic perspective, students develop and evaluate scientific knowledge by drawing on their understanding of the nature of science (Duschl, 2008; Erduran and Dagher, 2014). As GenAI tools like ChatGPT provide seemingly authoritative scientific information, this might limit students' opportunities to develop and evaluate scientific claims, as well as understand the nature and source of scientific knowledge in GenAI (Cheung et al, 2024). In contrast, the way that GenAI has been used in scientific research may be less authoritative and more generative, for instance, in playing a role in developing, revising, and validating scientific claims (Cheung et al., 2025). An example of the generative aspect of the GenAI tool AlphaFold, which predicts the structure of proteins and facilitates speeding up the discovery of drug therapeutics (Abriata, 2024). This tool has warranted the 2024 Nobel Prize in Chemistry, which showcases the significance of GenAI tools in helping to drive knowledge generation in science. In the context of science education, mere emphasis on conceptual goals related to GenAI would deprive students and teachers of having such insight about how GenAI is contributing to scientific research and lose sight of understanding science from an epistemic perspective.

In addition to epistemic understanding of science, there have been arguments for the need for students to engage in communication and legitimization of scientific claims from a social perspective (Duschl, 2008). To some extent, GenAI can reproduce social injustices because of the way the tools have been trained. GenAI might reinforce learners' stereotypes regarding the race and gender of scientists (Avraamidou, 2024b; Cooper & Tang, 2024). There may also be ethical aspects related to the use of GenAI tools in science education (Usher, Barak & Erduran, 2025). Such stereotypes are often exacerbated by the fact that GenAI provides personalized student learning and does not afford opportunities for social negotiation. As Longino (2018) argued, the public needs opportunities for discussion, criticism, and tolerance, which then can lead to shared authority and understanding. Owing to a high degree of personalized learning, there is a need for a pedagogical framework and technical modification that fosters social learning during the use of GenAI in science lessons.

2.3 Turning dissonance into synergy

Despite the dissonances between GenAI and science education, existing systematic reviews do not consider the identification of such dissonance specific to science teaching and learning, nor provide suggestions to reconcile it. For example, according to Almasri (2024) AI-powered tools can inherit pedagogical benefits, such as creating quizzes and assessments and predicting students' performance. Although these pedagogical benefits can be applied across all subjects, the systematic review does not reveal a nuanced account of how GenAI and science education may interact in not only positive terms but also in relation to limitations and shortcomings. Jia et al. (2024)'s systematic review focused on categorizing whether AI-facilitated science learning creates a positive or a negative effect on science learning. The authors (2024) characterised learning outcomes as entailing affection, cognition, skills behaviour and correlations. However, they overlooked disciplinary learning goals, including epistemic understanding of science. Furthermore, the review also did not outline the specific application of GenAI in teaching and learning science. Without an in-depth description of the disciplinary strengths and limitations of GenAI, it is difficult to guide future research studies to harness GenAI as an intellectual agent for epistemic and social learning goals of science.

Given that at this point in time, the integration of GenAI in science education is in its infancy, it is pressing to identify the loci of synergy and dissonance between GenAI and the teaching and learning of science so that constructive input can be made in designing future interventions. These synergies and dissonances can be explored in some target variables, data collection tools, content, and measured outcomes of teaching interventions, as described in the next section. With the identification of the potential for synergy, future research studies can capitalize on advancing disciplinary-specific GenAI in science learning. More significantly, following the identification of dissonance, we provide critical suggestions for future research that reconcile the dissonance and synergy of introducing GenAI into the teaching and learning of science.

3. Methodology

3.2 Research questions

The study reported in the paper employed a systematic review that was guided by the following research questions:

RQ1. How do target variables from the reviewed empirical studies reflect the synergy and dissonance between GenAI and science education?

RQ2. How do data collection methods enable the investigation of the interactions between GenAI and science education?

RQ3. What are the characteristics, contributions, and shortcomings of interventions that incorporate GenAI in science education?

3.2 Method

The systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) approach to identify and include eligible articles for analysis (Moher et al., 2015) (**Figure 1**). The search was conducted on 26th June 2024. We restricted the search starting from 2022 as the most popular GenAI tool, ChatGPT, was released on 30th November 2022 (OpenAI, 2024) and sparked discussion among educational researchers around that date.

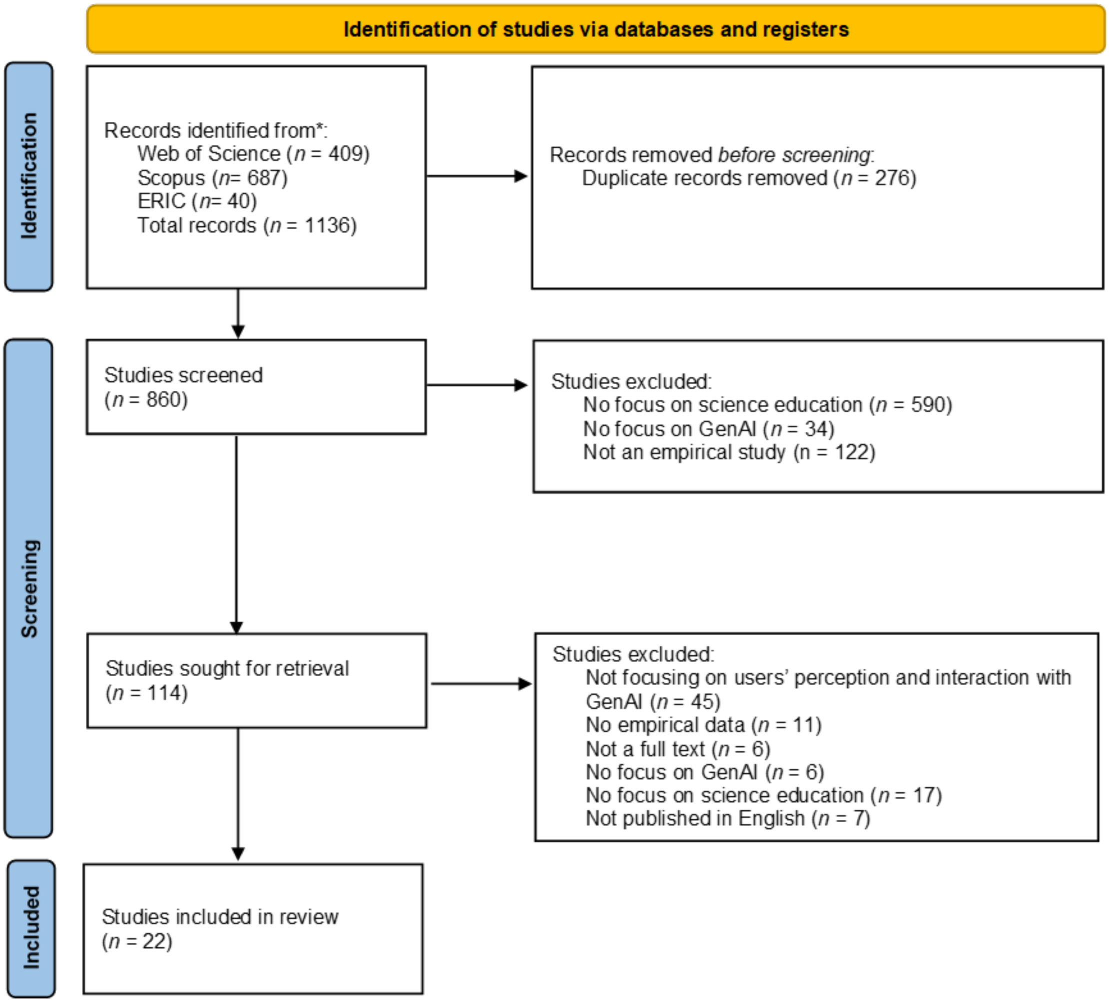

Figure 1. PRISMA flow diagram

3.2 Phase 1: Literature search and screening

Three library databases were used to perform an initial search of the studies, including Web of Science, Scopus, and Eric. Web of Science and Scopus are multidisciplinary databases, and ERIC is a disciplinary database specifically for educational studies. As the aim of this review was to put an exclusive focus on GenAI instead of generic AI in science education, the stipulated keywords need to differentiate GenAI from generic AI applications. We examined the keywords in systematic reviews of GenAI in other fields of education, such as language education (Law, 2024) and general education (Bozkurt, 2023), thus stipulated the first set of keywords on GenAI for searching in Title-Abstract-Keywords (or Title-Description-Subject for the ERIC database) (**Table 1**). In the first set of keywords, we included different written forms of GenAI, examples of GenAI applications, and the mechanisms of GenAI. These keywords include "generative ai*" OR "generative artificial intelligence" OR "gen ai*" OR "gpt" OR "chatgpt" OR "llm" OR "bert" OR "dall*" OR "pre*trained language model" OR "large language model". Three literature databases were not sensitive to small or capital letters; hence, we decided to input these keywords with small letters. However, our initial search (*S1*) in three databases yielded a large amount of literature. Hence, we added the second set of keywords (*S2*), including "science" OR "chemistry" OR "physics" OR "biology" OR "earth science". However, S2 yielded over 8000 search records as well as records that used GenAI in scientific research; we decided to specify the search (*S3*) in relation to the issue of education by adding the keywords "education" OR "school" OR "lesson" OR "class" OR "instruction*" OR “teaching”. As some of the records date back to 20 years ago, when the issue of GenAI in education had not yet emerged, we specified studies that were published on or after 2022, when ChatGPT had been released. A total of 1136 records were downloaded from the three databases and were combined into an Excel file.

Table 1. Search terms in databases and the number of returned hits

Search	Search terms	Number of hits		
		Web of science	Scopus	ERIC
S1	("generative ai*" OR "generative artificial intelligence" OR "gen ai*" OR "gpt" OR "chatgpt" OR "llm" OR "bert" OR "dall*" OR "pre*trained language model" OR "large language model")	78618	75585	629
S2	("generative ai*" OR "generative artificial intelligence" OR "gen ai*" OR "gpt" OR "chatgpt" OR "llm" OR "bert" OR "dall*" OR "pre*trained language model" OR "large language model") AND ("science" OR "chemistry" OR "physics" OR "biology" OR "earth science")	4138	4603	72
S3	("generative ai*" OR "generative artificial intelligence" OR "gen ai*" OR "gpt" OR "chatgpt" OR "llm" OR "bert" OR "dall*" OR "pre*trained language model" OR "large language model") AND ("science" OR "chemistry" OR "physics" OR "biology" OR "earth science") AND ("education" OR "school" OR "lesson" OR "class" OR "instruction*" OR "teaching")	897	973	56
S4	("generative ai*" OR "generative artificial intelligence" OR "gen ai*" OR "gpt" OR "chatgpt" OR "llm" OR "bert" OR "dall*" OR "pre*trained language model" OR "large language model") AND ("science" OR "chemistry" OR "physics" OR "biology" OR "earth science") AND ("education" OR "school" OR "lesson" OR "class" OR "instruction*" OR "teaching") *with specification of studies published on after 2022*	409	687	40

We detected duplicate records manually. These duplicate records were then removed, yielding 860 records in a single Excel file. The abstracts were screened, retaining empirical studies with a focus on GenAI in science education. 590 abstracts were excluded as they did not focus on science education; 34 abstracts were excluded as they did not focus on GenAI; 122 abstracts were excluded as they did not collect empirical data. The first two authors conducted a screening of more than 10% of the abstracts and achieved an intercoder reliability of Cohen’s $\kappa = 0.9$. The remaining abstracts were screened by one of the first two authors.

After screening the abstracts, a total of 114 full texts were further examined. At this stage, 45 studies were excluded because they did not focus on users’ perception or interaction with GenAI in science education. We decided to focus on this aspect, as this review aimed to identify if and how these studies addressed synergy or dissonance between GenAI and science education in terms of data collection methods and variables measured. The excluded studies mostly focused on using students' responses to train assessment (e.g., Cohn, Hutchins, Le, & Biswas, 2024). Although this line of studies was important in developing reliable science assessments by GenAI, these studies only provided technical data of GenAI models and offered less qualitative information regarding designing pedagogical activities that reconcile dissonance in using GenAI in science education. Apart from excluding such studies, we excluded 11 records with no empirical data involved; 6 records with no full texts; 6 records not focusing on GenAI; 17 records not focusing on science education, and 7 records not published in English. Eventually, a total of 22 studies were retained for an in-depth qualitative analysis.

3.4 Phase 2: Developing a coding scheme

Following the screening of articles, the first two authors developed a coding scheme. The coding scheme was then refined through several meetings by all four authors. To address the first research question, *"How do target variables from the reviewed empirical studies reflect the synergy and dissonance between GenAI and science education?"* we analyzed variables by complementing the three learning goals from Duschl (2008) (i.e., cognitive, epistemic, social) with other salient variables for science learning. These variables are (a) cognitive, (b) affective, (c) metacognitive, (d) epistemic, (e) ethical dimensions, and (f) pedagogical competence. Although Duschl (2008) suggested that science education has cognitive, epistemic, and social learning goals, we did not identify any social learning variables in the empirical studies that we investigated. Moreover, affective, metacognitive, and ethical dimensions and pedagogical competence were targeted variables that emerged from the inductive analysis of these empirical studies. For example, Ng (2024) characterised learning goals of GenAI in science classrooms as goal setting, while this entailed the process of thinking about one's thinking (Hartman, 2001). Hence, we characterised this as metacognitive outcomes instead of cognitive outcomes.

Focusing on these variables, we inductively and qualitatively developed sub-categories. We then sorted these variables into those that aligned more with the GenAI discipline, science discipline or characterize the interaction between GenAI and science disciplines. We further classified whether these variables framed GenAI as a positive tool for helping students' learning of science and as inheriting inaccuracies that allow students to develop their critical thinking. Such an approach helped identify whether variables in these empirical studies can help researchers understand the nature of synergy and dissonance between GenAI and science education. Furthermore, it highlighted particular information about how GenAI and science disciplines were covered in research in science education.

To address the second research question, "*How do data collection methods enable the investigation of the interactions between GenAI and science education?"* we analyzed whether these empirical studies use (a) questionnaires, (b) knowledge tests, and (c) interviews. These data collection tools were further named and sorted into alignment with the GenAI, science discipline, or characterizing interaction between GenAI and science. Interestingly, we observed that there were validated and unvalidated scales in these empirical studies. We further classified the knowledge tests and questionnaires into validated and unvalidated instruments. Furthermore, we analyzed whether or not these data collection tools framed GenAI as a useful or ineffective tool for teaching and learning science. Such classification can inform researchers what tools might be available to characterize the synergy or dissonance between GenAI and science education.

To address the third research question, *"What are the characteristics, contributions, and shortcomings of interventions that incorporate GenAI in science education?"* we inductively generated categories to characterize the role of GenAI in interventions, including (a) students' learning of science content knowledge, (b) incorporating GenAI into teaching and learning, and (c) mediating discussion. Furthermore, we analyzed the impact of GenAI on interventions, including the benefits and the limitations. Qualitative findings were documented under these categories, with findings further sorted into their alignment with the GenAI, science discipline, or characterizing interaction between GenAI and science education. Finally, we analyzed whether or not these teaching interventions can embed GenAI in science education or allow learners to critique and evaluate outputs from GenAI in the science learning environments.

10% of empirical studies were double-coded by the first two authors. To ensure coding reliability, Cheung and Tai (2023) suggest that readers could have understood the discrepancies in coding between the two authors and how such discrepancies can be resolved. The first two authors sometimes had disputes about defining codes and examples. In case of disputes, several rounds of online meetings were conducted to clarify the definition of codes. For example, one author thought that the Likert-scale such as *"Using ChatGPT enables me to accomplish task more quickly"* in Bitzenbauer (2023)'s study, was measuring "affective variables", while the other author perceived that these items were measuring "cognitive variables". We then arrived at a consensus that the Likert scale in this context measures affective variables as they measure moods and feelings about the usefulness of ChatGPT in science learning. After ensuring consistency in coding, the remaining studies were coded by the first author and then cross-checked by the other authors.

4. Results and Findings

In this section, we present an overview of the demographic information contained in the studies. Subsequently, we critically analyzed how these studies (*n*=22) demonstrate dissonance and synergy between GenAI and science education regarding targeted variables, methods, and features of teaching interventions.

4.1 Sample description

Empirical studies on GenAI in science education were often led by authors from North America (*n*=10), followed by authors from Asia (*n*=6), Europe (*n*=4), and South America (*n*=2) (see **Table 2**). Similarly, the geographical location of the empirical studies on GenAI in science was North America (*n*=8), followed by Asia (*n*=7), Europe (*n*=5), and South America (*n*=2). It is interesting to note that no empirical studies originated from authors in Australia, New Zealand, or African countries. This observation might be interpreted in different ways, including a potential lack of local large language models or potential restrictions of GenAI in educational policies. For example, Australian public schools banned GenAI when it first emerged (Tang et al., 2024).

Table 2. Demographics and background information of studies

Categories	Codes	Number (Percentages)
Continent of the first author	Asia	6 (27%)
	Central and South America	2 (9%)
	Europe	4 (18%)
	North America	**10 (45%)**
	Asia	7 (32%)

Continent where the study takes place	Central and South America	2 (9%)
	Europe	5 (23%)
	North America	**8 (36%)**
Grade levels	Primary	1 (5%)
	Secondary	6 (27%)
	Tertiary	**8 (36%)**
	Teacher education courses	4 (18%)
	In-service teachers	3 (14%)
Sample size	1 to 20	4 (18%)
	21 to 40	1 (5%)
	41 to 60	3 (14%)
	61 to 80	**8 (36%)**
	over 80	6 (27%)
Intervention duration	"a limited duration"	1 (5%)
	60 minutes	1 (5%)
	< 2 weeks	**7 (32%)**
	2-10 weeks	4 (18%)

Research on GenAI in science education mostly focused on tertiary (*n*=8) education, followed by secondary education (*n*=6). Only one study focused on primary science education. The sample size (*n*=8) mostly ranged from 61 to 80 individuals, while most empirical intervention studies (*n*=7) designed a less than two-week intervention. Future studies could extend to the role of GenAI in primary science education with a longer teaching intervention.

4.2 How do target variables from the reviewed empirical studies reflect the synergy and dissonance between GenAI and science education? (RQ1)

Empirical studies that use GenAI in science education targeted variables that aligned more with the field of GenAI. The study of dissonance between GenAI and science education was minimal in these surveyed studies. These studies targeted mostly synergy, which refers to studying cognitive (*n*=10, 45%) and affective variables (*n*=16, 73%), with epistemic variables examined in only one study (**Table 3**). Although science education needs to consider harmony between cognitive, epistemic, and social goals (Duschl, 2008), current research on GenAI in science education seems to ignore such harmony. This observation is supported historically when science education research tended to emphasise more the cognitive and affective variables rather than social learning or epistemic learning goals (Duschl, 2008). The continued lack of emphasis on epistemic and social learning goals could be because GenAI, such as ChatGPT and Bing, did not have functions tailored for these learning goals. As GenAI dissonates the goals of epistemic practices and social learning, seldom do variables focus on how GenAI changes these important variables as part of learning goals. Students who use GenAI in scientific activities might not tend to develop their epistemic knowledge or social learning, while studies did not track whether these important learning goals have been downplayed. If these variables do not show improvement during an intervention, this might caution how science educators leverage such an emerging technology to promote these learning goals alongside the affordances of GenAI.

Table 3. Type of variables examined in the studies

Type of learning variables	Number of studies	Specific variables examined	Studies
Cognitive	10 (45%)	*<u>Science discipline:</u>*	
		Scientific knowledge (including knowledge about lab reports)	(Chen & Chang, 2024; Cheng et al., 2024; Ghazali, Zaki, Ali, & Harous, 2024; Ng et al., 2024)
		Knowledge about the structure of lab reports	(Ruff et al., 2024)
		<u>GenAI discipline:</u>	
		Knowledge about GenAI	(Tassoti, 2024)
		Writing effective prompting strategies	(Guo & Lee, 2023; Tassoti, 2024)
		Knowledge about strengths and weaknesses of large language models	(Ruff et al., 2024)
		<u>Interaction between GenAI and science disciplines:</u>	
		Knowledge about accuracies and linguistic quality of GenAI in scientific outputs	(M. N. Dahlkemper, S. Z. Lahme, & P. Klein, 2023)
		<u>Others:</u>	
		Cognitive load and learning behaviour	(Chen & Chang, 2024)
Affective	16 (73%)	*<u>Science discipline:</u>*	
		Intrinsic motivation to learn science	(Cheng et al., 2024; Ng et al., 2024)
		<u>GenAI discipline:</u>	
		Trust in GenAI's information	(Lu Ding, Tong Li, Shiyan Jiang, & Albert Gapud, 2023; Lee & Zhai, 2024a)
		Perception of inaccuracies/bias in GenAI	(M. N. Dahlkemper et al., 2023; Hamid et al., 2023; Lee & Zhai, 2024a; Taani & Alabidi, 2024; West et al., 2023)
		Perception of GenAI as a positive development in educational technology	(Bitzenbauer, 2023; Garofalo & Farenga, 2024; Lee & Zhai, 2024a; Rezende Junior & López-Simó, 2024; West et al., 2023)

		Perception of GenAI in affecting assessment	(Garofalo & Farenga, 2024; Lee & Zhai, 2024a; Monteiro et al., 2024; Rezende Junior & López-Simó, 2024; Taani & Alabidi, 2024)
		Perception of critical thinking competence and future use of GenAI	(Guo & Lee, 2023; Lee & Zhai, 2024a)
		User experiences and acceptance towards GenAI	(Lieb & Goel, 2024)
		<u>*Interaction between GenAI and science disciplines:*</u>	
		Perception of GenAI as a useful tool lab report writing	(Ruff et al., 2024; West et al., 2023)
		<u>*General discipline:*</u>	
		Engagement in problem-based learning	(Hamid et al., 2023)
		Intrinsic motivation to play learning game and cultural engagement	(Chen & Chang, 2024)
		<u>*Other disciplines:*</u>	
		Attitudes towards environment and climate change	(Cheng et al., 2024)
Metacognitive	5 (23%)	<u>*General discipline:*</u>	
		Goal setting, environmental structuring, task strategies, time management, helping-seeking and self-evaluation	(Ng et al., 2024)
		Verdict on their prompting strategies (set the scene, be specific, simplify language, structure output, shared feedback)	(Tassoti, 2024)
		<u>*Interaction between GenAI and science disciplines:*</u>	
		Distinguish and evaluate whether lab report drafts are written by GenAI	(Ruff et al., 2024)
		Compare lab reports written by ChatGPT and students	(Clark, Anderson, Dickson-Karn, Soltanirad, & Tafini, 2023)
		Validate information from multiple sources	(Guo & Lee, 2023)
Epistemic	1 (5%)	<u>*General discipline:*</u>	
		The effect of GenAI in changing the ways of thinking and knowledge generation	(Garofalo & Farenga, 2024)
Ethical dimensions	3 (14%)	<u>*General discipline:*</u>	
		Perception of GenAI as a cheating tool	(Monteiro et al., 2024; West et al., 2023)
		Privacy concerns	(Garofalo & Farenga, 2024)

Pedagogical competence	4 (18%)	*General application:*	
		GenAI facilitates prospective teachers to develop text-based assessment tasks	(Kuechemann et al., 2023)
		GenAI supports instructional tasks such as lesson planning and recommending lesson resources	(Clark, Fhaner, Stoltzfus, & Queen, 2024)
		Pre-service teachers incorporate ChatGPT in their lesson planning by drawing on TPACK rubric	(Lee & Zhai, 2024a)
		In-service teachers' self-reported teaching practices of using ChatGPT	(Taani & Alabidi, 2024)

Cognitive variables Regarding cognitive variables (n=10), the variables were often inclined to either GenAI or science, while only one study examined variables that entail the interaction between GenAI and science. For example, despite focusing on applying GenAI in science classrooms, four studies targeted scientific content knowledge (e.g., Chen & Chang, 2024), and one study focused on knowledge about the structure of lab reports (Ruff et al., 2024). Regarding the discipline of GenAI, various variables included knowledge about GenAI (Tassoti, 2024), writing effective prompting strategies (Guo & Lee, 2023; Tassoti, 2024), and the strengths and weaknesses of large language models (Guo & Lee, 2023; Tassoti, 2024). Interestingly, the cognitive learning goals targeted by studies only focus on either science or GenAI, while studies examining cognitive learning goals from both disciplines are rare among these empirical studies.

Affective variables For affective variables (n=16), empirical studies were more inclined to examine GenAI-specific affection, though these studies were situated within an issue of science education. These affective variables included trust in GenAI in information (e.g., Lu Ding et al., 2023), perception of GenAI inaccuracies (e.g., Merten Nikolay Dahlkemper, Simon Zacharias Lahme and Klein, 2023), and perception of GenAI as a positive development in educational technology (e.g., Bitzenbauer, 2023). By contrast, the affective variable studied was limited only to intrinsic motivation in learning science (Cheng et al., 2024; Ng et al., 2024). Cheng et al. (2024) studied how students' motivation to learn science changes after engaging with a culturally relevant augmented reality application powered by GPT-4; Ng et al. (2024) also explored the effectiveness of a ChatGPT-powered chatbot in fostering students' motivation to learn science. Affective variables in science education entail enjoyment of science, the value of science, the usefulness of science, science self-concept, instrumental motivation, and future-oriented motivation (Ozel, Caglak, & Erdogan, 2013). Yet, existing empirical studies have not addressed the effectiveness of GenAI on these science-specific affective variables. We envisage that future studies would theorize and measure affective variables in different discipline-specific situations when using GenAI. For example, Ruff et al. (2024) reported that upper-division students perceived GenAI as a useful tool for revising their writing of lab reports. These affective variables or learning goals could be extended to wider scenarios, such as learning motivation to learn scientific languages and learning motivation to engage in scientific practices.

Metacognitive variables Metacognitive learning goals in these studies were either unrelated to a discipline or characterising the interaction between GenAI and science. For example, Ng et al. (2024) examined the effectiveness of GenAI-powered chatbots on goal setting and task strategies. These learning goals helped students understand the thought processes behind using GenAI in science learning. Yet, this study did not theorize disciplinary-specific thought processes behind GenAI in learning science, which warrants further empirical studies. In the study by Ruff et al. (2024), students were asked to compare lab reports written by ChatGPT with those they had written themselves. This process encouraged them to reflect critically on their own writing, while the GenAI-generated reports they had prompted provided opportunities for self-questioning and awareness of writing strengths and weaknesses. GenAI outputs in science lessons can facilitate self-questioning and identifying learning gaps.

Epistemic variables Only one study targeted epistemic variables as the learning goals. Garofalo and Farenga (2024) studied 24 secondary science educators' epistemic cognition. In their study, the authors prompted the educators' understanding of how GenAI changed their thinking and human-generated knowledge. The study demonstrates that emerging studies began to examine how GenAI further fosters the tentativeness of knowledge. However, questions related to epistemic cognition did not focus on how GenAI changed their scientific thinking and scientists-generated knowledge.

Ethical dimensions Three studies perceived the ethical dimension as an important variable to be studied. For example, in Monteiro et al. (2024)'s study it was reported that both teachers in private and public schools were neutral about whether ChatGPT-use entailed plagiarism. West et al. (2023) also examined whether students perceived ChatGPT as a cheating tool. They found that students did not remain neutral about this issue. On the other hand, Garofalo and Farenga (2024) also examined science educators' ethical and privacy concerns regarding the use of AI in education. Yet, the ethical dimension remained a variable to be studied in questionnaire surveys among science educators, teachers, and students. This had not been transformed into a learning goal of a teaching intervention that used GenAI in science education.

Pedagogical competencies Four studies examined the pedagogical competence of science teachers in the era of GenAI. However, the identified pedagogical competence remained unrelated to a discipline and does not shed light on how GenAI can specifically help students learn science. Kuechemann et al. (2023) examined if prospective teachers can develop physics questions, but they used generic indicators to measure teachers' pedagogical competence. These generic indicators include specificity, clarity, correctness, adequate difficulty, and context (Kuechemann et al., 2023). On the other hand, Clark et al. (2024) used a self-study approach to examine if GenAI can facilitate lesson plans, recommending resources, describing calculations and offering calculations. Taani and Alabidi (2024) identified teachers' self-reported frequency, effectiveness, educational outcomes offered by ChatGPT. For example, some teachers reported that ChatGPT can facilitate critical thinking and learning motivation. Yet, these empirical studies have not developed a conceptual framework that characterizes the pedagogical competence of using GenAI in teaching science.

A more theoretically-informed study draws on GenAI-Technological Pedagogical Content Knowledge (GenAI-TPACK) to examine teachers' lesson planning of science (Lee & Zhai, 2024a). According to the study, GenAI-TPACK comprises four dimensions: curriculum goals and ChatGPT, instructional methods/strategies and ChatGPT, ChatGPT function selections, and "Fit" (Lee & Zhai, 2024a). The GenAI-TPACK theoretical framework can be further enhanced by incorporating disciplinary-specific features of teaching and learning science, such as considering students' difficulties in translation across chemical representations (Talanquer, 2022).

From **Table 3**, a few learning variables reflect synergy and dissonance between the disciplines of GenAI and science, while others tend to focus on outcomes related to either GenAI or science. For instance, synergy entails GenAI as a useful tool for writing laboratory reports (Ruff et al., 2024; West et al., 2023). For dissonance, scientific outputs created by GenAI were deemed inaccurate, while students' higher-order thinking about these outputs can be elicited by inaccurate GenAI-created scientific information. Among all cognitive variables, some studies measured students' knowledge about the accuracy and linguistic quality of GenAI in scientific outputs (Dahlkemper, Lahme and Klein, 2023), positioning students as epistemic agents to interact with GenAI in learning science. More importantly, for metacognitive outcomes, students critically evaluate GenAI-created lab reports and compare student-generated lab reports to those created by GenAI.

For *affective* variables, intrinsic motivation to learn science was measured as well as learners' perceptions of inaccuracies and bias in GenAI. This indicates that GenAI raised students' intrinsic motivation to learn science while simultaneously eliciting students' perceptions of inaccuracies and biases in learning science. For the ethical dimension, the learning variables examined focused mainly on dissonance, such as learners' perception of GenAI as a cheating tool and privacy concerns. Studies in the current science education landscape assumed that students might violate ethics while using GenAI. More recent research has now begun to examine what role ethics could play in the context of GenAI and science education, including from students' points of view (Usher et al., 2025).

4.3 How do data collection methods enable investigating the interactions between GenAI and science education? (RQ2)

When studies aimed to investigate the use of GenAI in science education (**Table 4**), the dominant data collection tools used were questionnaires (*n*=17), followed by knowledge tests (*n*=5) and interviews (*n*=4). Table 4 shows a lack of validated scales that characterize relationships between GenAI and science learning. The instruments or scales focus on either learning outcomes targeted in technological education or science education. For example, some instruments can ask students to evaluate GenAI-created scientific products (e.g., scientific models, scientific information) (Cheung et al., 2024). Questions can target students' appreciation of the strengths of GenAI as well as evaluating its inaccuracies.

Questionnaire The validated scales in these empirical studies focused only on technology, science, or general learning outcomes. The technology scale used was the System Usability Scale (SUS) and the Technology Acceptance Model (TAM2) (Kuechemann et al., 2023), which surveyed users' perception of technology in general instead of its application in science education. Also, the Students' Motivation Toward Science Learning (SMTSL) questionnaire was used in the study by Cheng et al. (2024), although the questions did not specifically address how GenAI influenced students' motivation towards learning science. On the other hand, some scales non-specific to a particular discipline, neither characterizing learning of GenAI nor learning of science, were used in some studies. For example, in the study by Chen and Chang (2024) students' intrinsic motivation and cognitive load were measured using a validated questionnaire.

Table 4. Data collection tools of the study

Data collection tool	Exact data collection tool	Studies
Questionnaire (*n* = 17)	*Validated scale:*	
	System Usability Scale (SUS)	(Kuechemann et al., 2023)
	Technology Acceptance Model 2 (TAM2)	(Kuechemann et al., 2023)
	Students' Motivation Toward Science Learning (SMTSL) questionnaire	(Cheng et al., 2024)
	Giggle Gauge	(Cheng et al., 2024)
	Intrinsic motivation and cognitive load	(Chen & Chang, 2024)
	Self-regulated learning and motivation	(Ng et al., 2024)
	Unvalidated scales:	
	Knowledge and perception about ChatGPT	(Bitzenbauer, 2023; Guo & Lee, 2023; Ruff et al., 2024;

		Tassoti, 2024; West et al., 2023)
	Impact of ChatGPT on quality of student education and assessment	(Monteiro et al., 2024)
	Frequency and purpose of using ChatGPT, perceived effectiveness of using ChatGPT, student engagement, benefits and challenges	(Taani & Alabidi, 2024)
	The role of ChatGPT in facilitating engagement in problem-based learning and resolution	(Hamid et al., 2023)
	Perception of ChatGPT responses in relation to physics questions	(Dahlkemper et al., 2023)
	Students' confidence in applying critical thinking skills	(Guo & Lee, 2023)
	Attitudes towards AI in Education	(Garofalo & Farenga, 2024)
	Perception of usefulness, ease and intention to use AI	(Cheng et al., 2024; Ding, Li & Gapud, 2023)
	Behavioural change	(Cheng et al., 2024)
	Cultural references	(Cheng et al., 2024)
	Epistemic cognition	(Garofalo & Farenga, 2024)
	Qualitative survey determining teaching methods	(Lee & Zhai, 2024a)
Knowledge test (*n* = 5)	*Test on physics:*	
	Force Concept Inventory	(Kuechemann et al., 2023)
	Physics concept test	(Chen & Chang, 2024)
	Science knowledge related to force and motion	(Ng et al., 2024)
	Test on other disciplines:	
	Knowing, applying and reasoning of chemical compounds	(Ghazali et al., 2024)
	Knowledge about climate change, astronomy and environmental science	(Cheng et al., 2024)
Interview (*n* = 4)	Perception of students' interaction with ChatGPT	(Chen & Chang, 2024; Ng et al., 2024; Rezende Junior & López-Simó, 2024)
	Interpret survey questions (e.g. epistemic cognition and attitude towards AI in education)	(Garofalo & Farenga, 2024)

Among the unvalidated scales in the reviewed empirical studies were studies focusing on users' general perceptions and experiences with GenAI tools. For example, studies examined users' knowledge and perception of ChatGPT (e.g., Bitzenbauer, 2023) and their general attitudes toward AI in education (Garofalo & Farenga, 2024). This reflects the need for a validated questionnaire for measuring students' affective, epistemic, ethical, and pedagogical outcomes that characterize the interaction between GenAI and science learning. For example, Garofalo and Farenga (2024) asked teacher educators about their attitudes toward AI in education. According to Erduran and Levrini (2024) AI is already impacting scientific practices, such as the case of machine learning algorithms being used to analyze medical data. Such items in the questionnaire can be modified to entail their teacher educators' attitudes towards the use of AI in science disciplinary practices.

Knowledge test Among the five studies, the knowledge tests primarily measured the physics-specific knowledge. Most studies on GenAI in science education measured learners' physics knowledge, namely the Force Concept Inventory (Kuechemann et al., 2023), Physics concept test (Chen & Chang, 2024), and scientific knowledge related to force and motion (Ng et al., 2024). Although there was one study in the domain of chemistry (Ghazali et al., 2024) and another in environmental science (Cheng et al., 2024), no study examined the effectiveness of GenAI in improving students' biology knowledge.

Interview There were four studies that use interview techniques to qualitatively analyze learners' *static* perception and interaction with GenAI in teaching and learning science. Chen and Chang (2024) used post-activity interviews to recall students' learning experiences of playing a game powered by large language models. Similarly, Ng et al. (2024) conducted post-interviews on the limitations of the rule-based design of GenAI-powered chatbots. Using the focus group technique, Brazilian physics teachers were asked about doubts and questions about ChatGPT in class (Rezende Junior & López-Simó, 2024). These interview questions were declarative in nature, without characterizing their in-the-moment decision-making while interacting with GenAI. To advance the field of science education, there could be interview approaches that characterize individuals' *situational*, *dynamic* interaction with GenAI while at the same time engaging them in scientific activities. Such interview approaches might be fruitful in eliciting synergy and dissonance between GenAI and learning science.

More importantly, the interview questions were not related to a specific discipline and did not capture students' and teachers' perceptions of GenAI in various scientific activities, such as planning a scientific investigation. For example, in the study by Garofalo and Farenga (2024), science teacher educators were asked, "Do you think students should be taught about artificial intelligence as part of their education? If so, at what age and in what capacity?". These interview questions did not prompt teachers about the disciplinary applications of GenAI in learning science.

4.4 What are the characteristics, contributions, and shortcomings of interventions that incorporate GenAI in science education? (RQ3)

4.4.1 Designing interventions

Most teaching interventions focused on students' acquisition of content knowledge (**Table 5**), while GenAI's facilitation of scientific and engineering practices is minimal. In the study by Lieb and Goel (2024), students used GenAI to facilitate their completion of electromagnetism questions. Apart from content knowledge, undergraduate pharmacy students used ChatGPT to seek clarification of scientific terms (Hamid et al., 2023). More significantly, Cheng et al. (2024) embedded GPT-4-powered chatbots into a VR-based game. The chatbot facilitated interaction between a fictional character and students, hence improving students' learning of science in astronomy and environmental science. Despite the potential of GenAI tools, teaching interventions did not consider how large language models facilitate disciplinary practices (NRC, 2012), such as asking a question and making an observation. Without meaningful engagement in these practices, students cannot develop their epistemic understanding of how scientific knowledge is generated, validated, and revised (Cheung et al., 2025).

Table 5. Role of GenAI in science education interventions

Categories	Features of intervention	Studies

Students' learning of science content knowledge	Prompt ChatGPT to answer the questions and solve problems	(Ding et al., 2023; Lieb & Goel, 2024; Ng et al., 2024; Tassoti, 2024)
	Use ChatGPT to assist searching data and clarification of terms	(Hamid et al., 2023)
	Replace teachers and students and seek clarification and ask questions in relation to lesson objectives	(Ghazali et al., 2024)
	Chatbot incorporated in an application (e.g., VR/game)	(Chen & Chang, 2024; Cheng et al., 2024; Saindane, Prajapati, & Das, 2023)
Incorporate ChatGPT into teacher training	A workshop was held for teachers and teachers used ChatGPT actively in their science classrooms	(Taani & Alabidi, 2024)
	Being incorporated as a part of lesson planning	(Lee & Zhai, 2024a)
ChatGPT's outputs mediate discussion	ChatGPT creates guiding questions in focal group discussions and analyse teachers' responses	(Garofalo & Farenga, 2024)
	Review and discuss collaboratively scientific outputs generated by ChatGPT	(Bitzenbauer, 2023)

Two studies reported that science teachers incorporated ChatGPT into their teaching and learning. Taani and Alabidi (2024) designed a workshop to help students actively utilize GenAI tools. On the other side, Lee and Zhai (2024a) asked teachers to plan science lessons that involve ChatGPT-based activities. The teacher educators familiarized participants with the functions of ChatGPT and discussed the benefits of introducing ChatGPT into the teaching and learning of science (Lee & Zhai, 2024a). Yet, teachers could be further trained on how GenAI creates synergy and dissonance in generating scientific knowledge. To capitalise synergy, teachers can be trained how to use GenAI to facilitate scientific practices or give real-time feedback to students' scientific claims; to harness dissonance, teachers can learn how to engage students to critically evaluate GenAI-created scientific products during professional development program.

GenAI tools can also facilitate science education research. Two studies reported that ChatGPT outputs could mediate the discussion. For example, ChatGPT can create guiding focal group interview questions and qualitatively code teachers' responses (Garofalo & Farenga, 2024). Importantly, ChatGPT outputs can elicit participants' discussion (Bitzenbauer, 2023). However, the GenAI outputs could provide specific classroom scenarios for using ChatGPT in science classrooms so that teachers' discussions about science education can be mediated.

4.4.2 Benefits and limitations

Following an analysis of the features of teaching interventions, **Table 6** shows that the benefits offered by GenAI tools outweigh the harms they cause to science education. For example, GenAI can improve students' understanding of content knowledge (e.g., Taani & Alabidi, 2024), motivation (e.g., Hamid et al., 2023), perception of usability (e.g., Bitzenbauer, 2023), and satisfaction with AI-generated answers (e.g., Tassoti, 2024). In addition to these positive effects on cognitive and affective outcomes, GenAI also improved outcomes such as self-regulated strategies (Ng et al., 2024). Yet, the affordance of GenAI in epistemic outcomes, disciplinary practices, and disciplinary core ideas emphasized in science education reform has not yet been researched.

Table 6. Effectiveness of GenAI in promoting science learning

Categories	Outcomes of intervention	Studies
Benefits	Teaching students to reflect on their prompting strategies can increase their satisfaction of AI-generated answers	(Tassoti, 2024)
	Enhance students' understanding of content knowledge	(Ng et al., 2024; Taani & Alabidi, 2024)
	Increase students' engagement and motivation	(Chen & Chang, 2024; Cheng et al., 2024; Hamid et al., 2023; Ng et al., 2024; Taani & Alabidi, 2024)
	Increase perception of benefits/usability offered by GenAI	(Bitzenbauer, 2023; Lieb & Goel, 2024; Saindane et al., 2023)
	Reduces learners' cognitive load	(Chen & Chang, 2024)
	Enhances learners' self-regulated strategies	
Limitations	Stronger perception of inaccuracy of information	(Hamid et al., 2023; Taani & Alabidi, 2024)
	Mistrust scientific information created by GenAI	(Ding et al., 2023)
	Stronger perception of algorithm bias	(Taani & Alabidi, 2024)
	Disengagement in class	(Ghazali et al., 2024)

Apart from the benefits, GenAI-based interventions highlighted limitations such as a stronger perception of inaccuracy (Hamid et al., 2023) and bias (Taani & Alabidi, 2024) of information. However, these studies did not provide specific descriptions of how teaching intervention fosters such perceptions of inaccuracy and bias. These teaching interventions can also create disengagement in class (Ghazali et al., 2024). Yet, owing to unvalidated measures that were not specific to a particular discipline, the findings remain preliminary and prevent us from drawing firm conclusions about whether GenAI-supported teaching interventions bring more benefits than limitations to teaching and learning science.

5. Discussion

Given the contemporary interest in GenAI in science education, the study reported in the paper provided a timely review of the emerging empirical studies in this line of research. Firstly, this review addresses the extent to which the targeted variables of empirical studies reflect the synergy and dissonance between GenAI and science education. Secondly, the current review also showed to what degree data collection methods can examine the interaction between science education and GenAI. Thirdly, this review also synthesized the characteristics and impacts of GenAI intervention in science education and critically analyzed whether these teaching interventions can capture the synergy and dissonance between GenAI and science education.

5.1 GenAI conceptualizes new learning goals in science education

Before the emergence of GenAI, common learning goals in science education emphasized the harmony among conceptual, epistemic, and social learning aspects (Duschl, 2008) (**Figure 2**). In this review, however, we analyzed a broader range of learning goals related to the use of GenAI in science education, including conceptual epistemic, affective, meta-cognitive, ethical, and pedagogical competencies (**Table 3**). The learning goals of GenAI in science education did not focus on the synergy and dissonance between GenAI and science education. For example, affective variables in these empirical studies focused more on GenAI itself than on the interaction between GenAI and science disciplines. These GenAI-specific affective variables included trust in GenAI's information and perception of inaccuracies or biases in GenAI (Dahlkemper et al., 2023; Hamid et al., 2023; Lee & Zhai, 2024a; Taani & Alabidi, 2024; West et al., 2023). However, no theoretical construct currently characterizes the discipline-specific affective dimensions of using GenAI in science education.

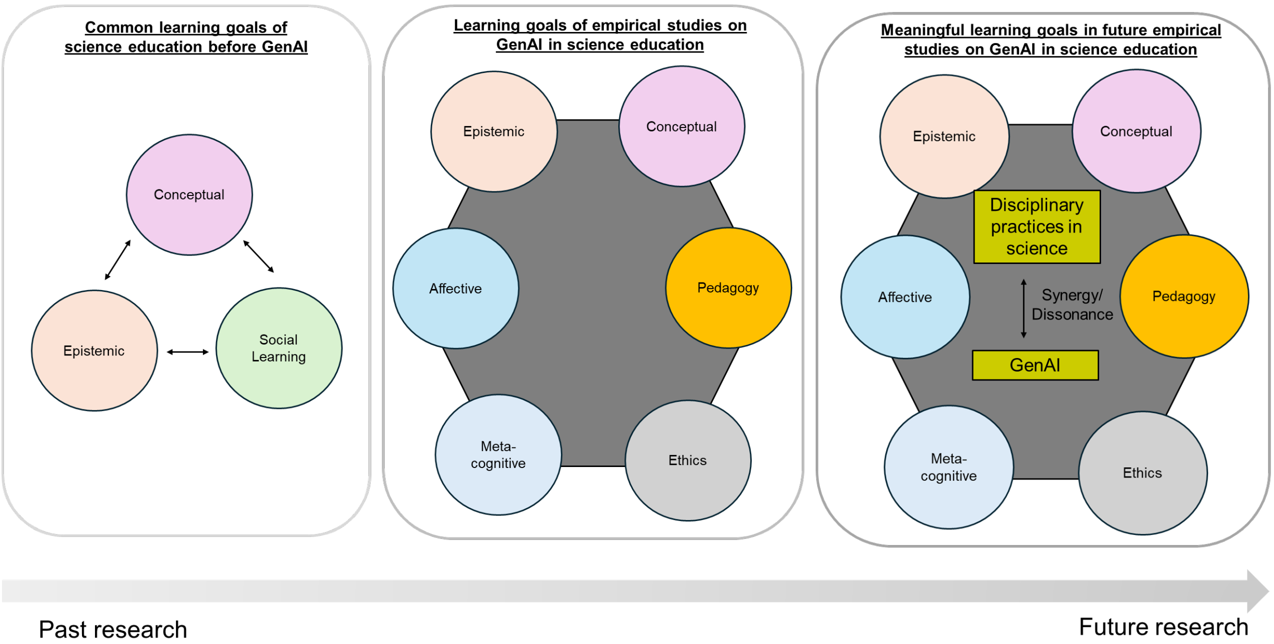

Figure 2. Past learning goals in science education, existing and future learning goals of GenAI in science education

Importantly, as Cheung et al. (2025) argued, GenAI has various disciplinary applications, such as facilitating scientific practices as identified by NRC (2012). However, our review demonstrated that the examined variables, data collection tools, and teaching interventions did not feature such disciplinary applications. Hence, we propose that the learning goals of GenAI in science education should further unpack the *synergy* and *dissonance* between disciplinary practices in science education and GenAI. To capture *synergy*, studies could design variables or instruments that examine learners' motivation to engage in scientific practices using GenAI. Also, GenAI can give personalised feedback to students' scientific claims (Mills et al., 2025). Apart from affective variables, scholars have been arguing for redefining epistemic aspects of science under the influence of GenAI (Tekin, 2025; Cheung et al., 2025). Studies can also theorize new constructs, such as the *"nature of GenAI-influenced science"* and develop instruments to measure students' and teachers' understanding of the nature of GenAI-influenced science. For *dissonance*, learners can engage, for example, by critiquing images generated by DALL.E3 and comparing them with those generated by scientists (Pun et al., 2025). Teachers could then prompt students to explicitly reflect on how different GenAI applications create images on a science topic. For example, DALL.E 3 uses diffusion models to create images, rather than relying on empirical observations and photographs of a scientific phenomenon. This activity would also guide students to reflect on the affordances and challenges of using GenAI tools for creating scientific visuals. Leveraging both *synergy* and *dissonance* could potentially enhance students' understanding and use of GenAI in science education.

Another important learning goal, *social learning*, was largely absent from the existing empirical literature that had been investigated. The use of GenAI could undermine the social nature of scientific knowledge and the development of learners' socio-emotions (Avraamidou, 2024b). This may be because GenAI adopts a more personalized approach to teaching and learning (Bozkurt, 2023) without providing sufficient opportunities for human-human interaction. To address this dissonance, future research could design GenAI-powered chatbots that foster student collaboration and communication, particularly in generating, validating, and revising scientific claims. Using large multi-agent language models can allow multiple decision-making agents to interact with each other to reconcile conflicting goals (Zheng, 2024), which can mimic the disagreement among scientists during scientific investigations. Another potential approach is to engage students in group activities with GenAI, where they act as scientists using GenAI as a tool to support scientific practices. Following that, they peer review others' scientific work. Such an approach shifts the focus from general social learning goals to *disciplinary* social learning goals.

5.2 Current state of research and future research directions

The systematic review identifies several research gaps in existing empirical studies. Current research has only focused on tertiary contexts and typically involved interventions lasting less than 2 weeks. Such results regarding the length of teaching intervention are consistent with previous systematic reviews on the use of AI in science education (Jia et al., 2024). However, although generic AI technologies are commonly used in middle or secondary schools (Jia et al., 2024), the use of GenAI in these educational levels remains far less common than in tertiary science education. This may be because many educators express concerns about the bias and ethical issues related to GenAI (Cooper & Tang, 2024; Feldman-Maggor et al., 2024). We envision future studies that incorporate sensible use of GenAI in science education across a wider age range and with longer teaching interventions.

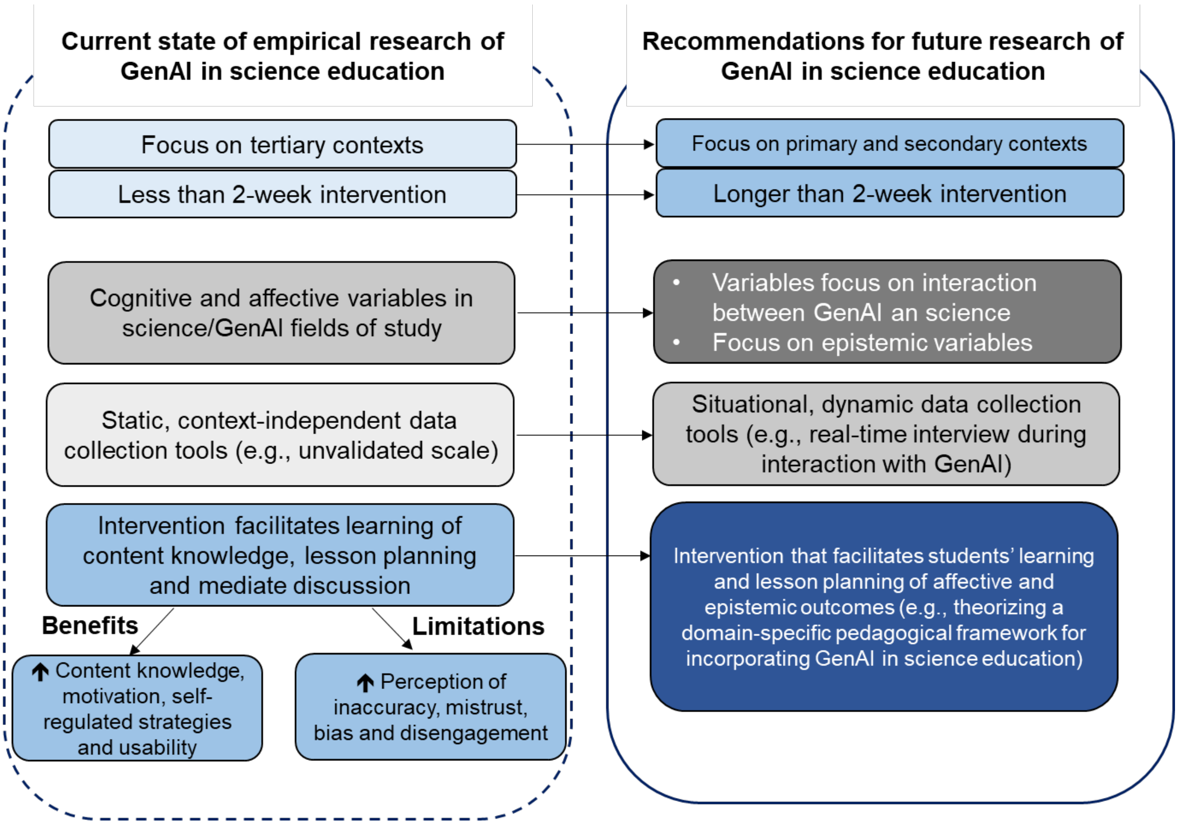

Figure 3. The current state of research on GenAI in science education and recommendations for future research.

Current empirical studies predominantly examine cognitive and affective variables, which are largely based on technological and science education literature (Figure 3). Future research studies could extend this focus to include epistemic or social learning variables. More importantly, the data collection tools employed in existing studies remain static and context-independent. However, learning science is dependent on contexts (Sikorski, 2019; Watkins & Elby, 2013). Also, epistemic variables, such as students' understanding of the nature of science, vary according to familiarity with the contexts (Khishfe, 2019; Khishfe & Lederman, 2007). Following this argument, students' understanding of the nature of GenAI-influenced science could also vary according to contexts. Moreover, their interaction with GenAI is also situational, which involves in-the-moment decision-making. Static data collection instruments, such as questionnaires (Kuechemann et al., 2023), might not be able to capture students' in-the-moment disciplinary decision-making during their interaction with GenAI.

Existing interventions are primarily aimed at developing students' conceptual and affective domains. These interventions commonly reported that GenAI improved students' learning of scientific conceptual knowledge and enhanced their motivation and self-regulated strategies. However, these studies also revealed limitations, such as increasing users' perception of inaccuracy, bias, and mistrust towards GenAI-generated scientific information. In light of these findings, we propose that future studies should develop a pedagogical framework that considers the disciplinary affordances of GenAI in teaching and learning science, as well as incorporating teaching interventions that entail disciplinary practices in science. Such a pedagogical framework needs to consider the synergy between the technical design of the GenAI application and teachers' instruction. For instance, some exploratory studies have designed chatbots as a dialogic partner for students (Tang & Putra, 2025); however, further research is needed to explore how teachers can orchestrate classroom instruction around the technological design of GenAI chatbots. Also, some important learning goals in science classrooms, such as epistemic practices and science identity, seem to contradict the incorporation of GenAI in science classrooms, as GenAI chatbots encourage personalised feedback.

Overall, examining the existing literature provides an evidence-based approach that illustrates the current state of GenAI applications in science education and highlights the gaps that need to be addressed in future research. In other words, the evidence from this systematic review not only informs the state of contemporary research but also provides a rationale for why future studies will need to pay closer attention to certain aspects of science education, such as epistemic reasoning and social learning, that are considered more broadly to be significant for science learning.

5.3 Limitations

We are aware of the fact that GenAI is developing at a fast pace within the field of science education, and this static review may not keep track of updated literature in a timely manner. Nevertheless, this review provides important insights for science education researchers by encouraging the diversification of learning goals and promoting reflection on the synergy and dissonance between GenAI and science prior to researching and incorporating GenAI in science classrooms. Importantly, our review focused solely on empirical studies published in English. There might be some empirical studies that were published in other languages. However, this systematic review conducted an extensive search across three major databases and employed multiple different search strings to ensure comprehensive coverage. This provides a solid foundation and highlights important directions for future research on GenAI in science education.

6. Conclusion

The features identified in this systematic review highlight key research gaps in the variables, data collection tools, and teaching interventions. These variables, data collection tools, and teaching interventions have yet to adequately consider the synergy and dissonance between GenAI and science education. Recognizing this interplay of *synergy* and *dissonance* in science education research is important for examining how GenAI can improve students' understanding of disciplinary practices in science (Ford, 2008).

The systematic review provides a comprehensive overview of the current state of research on the potential of GenAI in science education. It identifies an overemphasis on cognitive and affective outcomes, while other outcomes, such as epistemic and social learning outcomes, are minimally addressed in empirical studies. Importantly, the notion of GenAI diversifies learning goals in science education, for example, by including ethics and meta-cognition. Though ethics and meta-cognition have long been a part of learning goals in science education, our review illustrates that these variables play a particularly pivotal role in GenAI-facilitated science learning. To further advance this line of research, we proposed a research framework (Figure 2) that conceptualizes the connections among diversified learning goals, all centered on disciplinary scientific practices. This framework serves as a foundation for guiding future empirical studies toward a more holistic understanding of how GenAI can be meaningfully integrated into science education.

References

Abriata, L. A. (2024). The Nobel Prize in Chemistry: past, present, and future of AI in biology. *Communications Biology, 7*(1), 1409.

Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education, 100*(8), 2965-2971.

Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education, 54*(5), 977-997.

Avraamidou, L. (2024a). Can we disrupt the momentum of the AI colonization of science education? *Journal of research in science teaching, n/a*(n/a). doi:https://doi.org/10.1002/tea.21961

Avraamidou, L. (2024b). Can we disrupt the momentum of the AI colonization of science education? *Journal of Research in Science Teaching*.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52-62.

Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., . . . Nerdel, C. (2024). Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *arXiv preprint arXiv:2401.00832*.

*Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology, 15*(3). doi:10.30935/cedtech/13176

Bozkurt, A. (2023). Unleashing the potential of generative AI, conversational agents and chatbots in educational praxis: A systematic review and bibliometric analysis of GenAI in education. In (Vol. 15, pp. 261-270): International Council for Open and Distance Education Oslo, Norway.

*Chen, C. H., & Chang, C. L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. *Education and Information Technologies*. doi:10.1007/s10639-024-12553-x

*Cheng, A. Y., Guo, M., Ran, M., Ranasaria, A., Sharma, A., Xie, A., . . . Landay, J. A. (2024). *Scientific and Fantastical: Creating Immersive, Culturally-Relevant Learning Experiences with Augmented Reality and Large Language Models.* Paper presented at the Conference on Human Factors in Computing Systems - Proceedings.

Cheung, K. K. C., Long, Y., Liu, Q., & Chan, H. Y. (2025). Unpacking epistemic insights of artificial intelligence (AI) in science education: A systematic review. *Science & Education*, *34*(2), 747-777.

Cheung, K. K. C., Pun, J. K., & Li, W. (2024). Students' holistic reading of socio-scientific texts on climate change in a ChatGPT scenario. *Research in Science Education*, *54*(5), 957-976.

Cheung, K. K. C., & Tai, K. W. (2023). The use of intercoder reliability in qualitative interview data analysis in science education. *Research in Science & Technological Education*, *41*(3), 1155-1175.

Cheung, K. K. C., & Zhang, W. (2025). Undergraduate Students' Perceptions of the Nature of Generative Artificial Intelligence (GenAI)-Influenced Science: A Structural Equation Modelling Approach. *Journal of Science Education and Technology*, 1-13.

Chiu, T. K. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1-17.

Chiu, T. K., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2024). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments, 32*(7), 3240-3256.

*Clark, T. M., Anderson, E., Dickson-Karn, N. M., Soltanirad, C., & Tafini, N. (2023). Comparing the Performance of College Chemistry Students with ChatGPT for

Calculations Involving Acids and Bases. *Journal of Chemical Education, 100*(10), 3934-3944. doi:10.1021/acs.jchemed.3c00500

*Clark, T. M., Fhaner, M., Stoltzfus, M., & Queen, M. S. (2024). Using ChatGPT to Support Lesson Planning for the Historical Experiments of Thomson, Millikan, and Rutherford. *Journal of Chemical Education, 101*(5), 1992-1999. doi:10.1021/acs.jchemed.4c00200

Cohn, C., Hutchins, N., Le, T., & Biswas, G. (2024). *A Chain-of-Thought Prompting Approach with LLMs for Evaluating Students' Formative Assessment Responses in Science.* Paper presented at the Proceedings of the AAAI Conference on Artificial Intelligence.

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444-452.

Cooper, G., & Tang, K.-S. (2024). Pixels and Pedagogy: Examining Science Education Imagery by Generative Artificial Intelligence. *Journal of Science Education and Technology*, 1-13.

Dahlkemper, M. N., Lahme, S. Z., & Klein, P. (2023). How do physics students evaluate artificial intelligence responses on comprehension questions A study on the perceived scientific accuracy and linguistic quality of ChatGPT. *Physical Review Physics Education Research, 19*(1). doi:10.1103/PhysRevPhysEducRes.19.010142

*Dahlkemper, M. N., Lahme, S. Z., & Klein, P. (2023). How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT. *Physical Review Physics Education Research, 19*(1). doi:10.1103/PhysRevPhysEducRes.19.010142

*Ding, L., Li, T., Jiang, S., & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education, 20*(1). doi:10.1186/s41239-023-00434-1

Du, J., & Alm, A. (2024). The impact of ChatGPT on English for academic purposes (EAP) students' language learning experience: A self-determination theory perspective. *Education Sciences, 14*(7), 726.

Erduran, S., & Dagher, Z. R. (2014). *Reconceptualizing nature of science for science education.* Dordrecht: Springer Netherlands.

Erduran, S., & Levrini, O. (2024). The impact of artificial intelligence on scientific practices: an emergent area of research for science education. *International Journal of Science Education*, *46*(18), 1982-1989.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of research in education, 32*(1), 268-291.

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*, 1-12.

Ford, M. (2008). 'Grasp of practice'as a reasoning resource for inquiry and nature of science understanding. *Science & Education, 17*, 147-177.

*Garofalo, S. G., & Farenga, S. J. (2024). Science Teacher Perceptions of the State of Knowledge and Education at the Advent of Generative Artificial Intelligence Popularity. *Science & Education*. doi:10.1007/s11191-024-00534-y

*Ghazali, S. A., Zaki, N., Ali, L., & Harous, S. (2024). Exploring the Potential of ChatGPT as a Substitute Teacher: A Case Study. *International Journal of Information and Education Technology, 14*(2), 271-278. doi:10.18178/ijiet.2024.14.2.2048

*Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for Enhancing Critical Thinking Skills. *Journal of Chemical Education, 100*(12), 4876-4883. doi:10.1021/acs.jchemed.3c00505

*Hamid, H., Zulkifli, K., Naimat, F., Yaacob, N. L. C., & Ng, K. W. (2023). Exploratory study on student perception on the use of chat AI in process-driven problem-based learning. *Currents in Pharmacy Teaching and Learning, 15*(12), 1017-1025. doi:10.1016/j.cptl.2023.10.001

Ho, C. W.-L. (2023). Generative AI and the foregrounding of epistemic injustice in bioethics. *The American Journal of Bioethics, 23*(10), 99-102.

Hu, L. (2023). Generative AI and Future. Retrieved from https://pub.towardsai.net/generativeai-and-future-c3b1695876f2

Jia, F., Sun, D., & Looi, C.-k. (2024). Artificial intelligence in science education (2013–2023): Research trends in ten years. *Journal of Science Education and Technology, 33*(1), 94-117.

Jovanović, M. (2023). Generative Artificial Intelligence: Trends. Retrieved from https://www.computer.org/csdl/magazine/co/2022/10/09903869/1H0G6xvtREk.0.1109/MC.2022.3192720.

Kelly, G. J. (2018). Developing epistemic aims and supports for engaging students in scientific practices. In (Vol. 27, pp. 245-246): Springer.

Khishfe, R. (2019). The transfer of nature of science understandings: a question of similarity and familiarity of contexts. *International Journal of Science Education, 41*(9), 1159-1180.

Khishfe, R., & Lederman, N. (2007). Relationship between instructional context and views of nature of science. *International Journal of Science Education, 29*(8), 939-961.

Kolade, O., Owoseni, A., & Egbetokun, A. (2024). Is AI changing learning and assessment as we know it? Evidence from a ChatGPT experiment and a conceptual framework. *Heliyon, 10*(4).

Krist, C., & Kubsch, M. (2023). Bias, bias everywhere: A response to Li et al. and Zhai and Nehm. *Journal of research in science teaching, 60*, 2395-2399.

*Kuechemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E., & Kuhn, J. (2023). Can ChatGPT support prospective teachers in physics task development? *Physical Review Physics Education Research, 19*(2). doi:10.1103/PhysRevPhysEducRes.19.020128

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 100174.

Lederman, N. G., Abd‐El‐Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6), 497-521.

Lee, G.-G., & Zhai, X. (2024a). Using ChatGPT for Science Learning: A Study on Pre-service Teachers' Lesson Planning. *Ieee Transactions on Learning Technologies, 17*, 1683-1700. doi:10.1109/tlt.2024.3401457

*Lee, G.-G., & Zhai, X. (2024b). Using ChatGPT for Science Learning: A Study on Pre-service Teachers' Lesson Planning. *IEEE Transactions on Learning Technologies*.

*Lieb, A., & Goel, T. (2024). *Student Interaction with NewtBot: An LLM-as-tutor Chatbot for Secondary Physics Education.* Paper presented at the Conference on Human Factors in Computing Systems - Proceedings.

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410.

Longino, H. E. (2018). The fate of knowledge.

Mills, E., Mizouri, A., & Peach, A. (2025). Prompting Better Feedback: A Study of Custom GPT for Formative Assessment in Undergraduate Physics. *Education Sciences*, *15*(8), 1058.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Group, P.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews, 4*, 1-9.

*Monteiro, F. F., Souza, P. V. S., da Silva, M. C., Maia, J. R., da Silva, W. F., & Girardi, D. (2024). ChatGPT in Brazilian K-12 science education. *Frontiers in Education, 9*. doi:10.3389/feduc.2024.1321547

*Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*. doi:10.1111/bjet.13454

National Research Council [NRC] (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* . Washington, DC: The National Academies Press.

Nyaaba, M. (2024). *Generative AI Conception of the Nature of Science.* Paper presented at the Society for Information Technology & Teacher Education International Conference.

OpenAI. (2024). ChatGPT — Release Notes: A changelog of the latest updates for ChatGPT. Retrieved from https://help.openai.com/en/articles/6825453-chatgpt-release-notes

Ozel, M., Caglak, S., & Erdogan, M. (2013). Are affective factors a good predictor of science achievement? Examining the role of affective factors based on PISA 2006. *Learning and Individual Differences, 24*, 73-82.

Pun, J., Cheung, K. K. C., & Kenneth-Li, W. (2025). Secondary Students' Reading of Socio-Scientific Image-Texts on Climate Change in a GPT-4 Scenario. *Research in Science Education*, 1-20.

*Rezende Junior, M. F., & López-Simó, V. (2024). *What are the perceptions of physics teachers in Brazil about ChatGPT in school activities?* Paper presented at the Journal of Physics: Conference Series.

*Ruff, E. F., Engen, M. A., Franz, J. L., Mauser, J. F., West, J. K., & Zemke, J. M. O. (2024). ChatGPT Writing Assistance and Evaluation Assignments Across the Chemistry Curriculum. *Journal of Chemical Education, 101*(6), 2483-2492. doi:10.1021/acs.jchemed.4c00248

*Saindane, D., Prajapati, S. P., & Das, S. (2023). *Converting Physical Textbooks into Interactive and Immersive 'Phygital' Textbooks: A Proposed System Architecture Design for Textbook Companion Apps.* Paper presented at the CEUR Workshop Proceedings.

Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education, 88*(4), 610-645.

Sikorski, T.-R. (2019). Context-dependent "upper anchors" for learning progressions. *Science & Education, 28*(8), 957-981.

Spunt, R., & Adolphs, R. (2017). A new look at domain specificity: insights from social neuroscience. *Nat Rev Neurosci* **18**, 559–567. https://doi.org/10.1038/nrn.2017.76

*Taani, O., & Alabidi, S. (2024). ChatGPT in education: benefits and challenges of ChatGPT for mathematics and science teaching practices. *International Journal of Mathematical Education in Science and Technology*. doi:10.1080/0020739x.2024.2357341

Talanquer, V. (2022). The complexity of reasoning about and with chemical representations. *JACS Au, 2*(12), 2658-2669.

Tang, K.-S., & Cooper, G. (2024). The role of materiality in an era of generative artificial intelligence. *Science & Education*, 1-16.

Tang, K. S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning & assessment with generative AI in secondary education: A proof of concept. *Pedagogies: An International Journal*, *19*(3), 493-503.

Tang, K. S. (2023). The characteristics of diagrams in scientific explanations: Multimodal integration of written and visual modes of representation in junior high school textbooks. *Science education, 107*(3), 741-772.

Tang, K. S. (2024). Informing research on generative artificial intelligence from a language and literacy perspective: A meta-synthesis of studies in science education. *Science Education*. doi:10.1002/sce.21875

Tang, K. S., & Putra, G. B. S. (2025). Generative AI as a Dialogic Partner: Enhancing Multiple Perspectives, Reasoning, and Argumentation in Science Education with Customized Chatbots. *Journal of Science Education and Technology*, 1-13.

*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. doi:10.1021/acs.jchemed.4c00212

Tekin, N. (2025). Using Artificial Intelligence to Evaluate the Nature of Science Myths: A Comparison of Pre-Service Science Teachers and ChatGPT. In *AI Applications and Strategies in Teacher Education* (pp. 253-274): IGI Global.

Usher, M., Barak, M. & Erduran, S. (2025). What role should higher education institutions play in fostering AI ethics? Insights from science and engineering graduate students. *IJ STEM Ed*12, 51. https://doi.org/10.1186/s40594-025-00567-x

Watkins, J., & Elby, A. (2013). Context dependence of students' views about the role of equations in understanding biology. *CBE—Life Sciences Education, 12*(2), 274-286.

*West, J. K., Franz, J. L., Hein, S. M., Leverentz-Culp, H. R., Mauser, J. F., Ruff, E. F., & Zemke, J. M. (2023). An Analysis of AI-Generated Laboratory Reports across the Chemistry Curriculum and Student Perceptions of ChatGPT. *Journal of Chemical Education, 100*(11), 4351-4359. doi:10.1021/acs.jchemed.3c00581

Xie, Y., Wu, J., Tu, H., Yang, S., Zhao, B., Zong, Y., . . . Zhou, Y. (2024). A Preliminary Study of o1 in Medicine: Are We Closer to an AI Doctor? *arXiv preprint arXiv:2409.15277*.

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: a systematic review from 2011 to 2021. *International Journal of STEM Education, 9*(1), 59.

Yang, Y.-Y., & Shulruf, B. (2019). An expert-led and artificial intelligence system-assisted tutoring course to improve the confidence of Chinese medical interns in suturing and ligature skills: a prospective pilot study. *Journal of Educational Evaluation for Health Professions, 16*.

Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching, 60*(6), 1390-1398.

Zhai, X., Nyaaba, M., & Ma, W. (2024). Can generative AI and ChatGPT outperform humans on cognitive-demanding problem-solving tasks in science? *Science & Education*, 1-22.

Zhai, X., Tai., R. H., Lee, G.-G., Latif, E., Feldman-Maggor, Y., & Bewersdorff, A. (2024). Special Issue Call: The Game-Changer: Generative Artificial Intelligence for Science Education and Research. *International Journal of Science Education*.

Zheng, X. (2024). Exploring Multi-Agent Dynamics for Generative AI and Large Language Models in Mobile Edge Networks. *Authorea Preprints*.