Education and Information Technologies
https://doi.org/10.1007/s10639-024-12754-4

Improving ethical dilemma learning: Featuring thinking aloud pair problem solving (TAPPS) and AI-assisted virtual learning companion

Yung-Hsiang Hu[1]

Received: 6 September 2023 / Accepted: 29 April 2024

Abstract

In this study, a generative artificial intelligence (AI)-assisted Think-Aloud Pair Problem-Solving (TAPPS) learning strategy was introduced to support ethical dilemma-related problem-solving learning activities. Then, an interactive virtual learning companion system was developed and tested in a business ethics course to evaluate the efficacy of the proposed method. A total of 135 students from a technological university in central Taiwan participated in this experiment. The experimental group employed the generative AI chatbot virtual learning companion that assisted with the TAPPS approach for ethical dilemma learning. Two control groups used, respectively, the conventional TAPPS approach involving paired learners and the conventional individual thinking-aloud problem-solving method. The proposed method not only enhanced the students' problem-solving ability but also fostered their learning motivation and ethical reasoning ability.

Keywords Virtual learning companion · Thinking aloud pair problem solving · AI · Ethical dilemma · Chatbot

1 Introduction

Enhancing problem-solving abilities is pivotal for ethical decision-making (Kavathatzopoulos, 1994). This is particularly challenging when addressing real-life moral issues (Knight et al., 2019). A key component of this process is ethical reasoning (Lewis et al., 2019), which involves the capacity to judge and make rational decisions in the face of moral dilemmas (Kamali et al., 2019). Previous research has shown that school education plays a role in developing these reasoning skills

✉ Yung-Hsiang Hu
hsianghu@go.edu.tw

[1] Department of Education, College of Education, University of Taipei, Taipei 100234, Taiwan

Published online: 18 May 2024

(Hartwell, 1995). However, traditional teacher-centered approaches, such as lectures, have been found to be somewhat lacking (Khaghanizade et al., 2012). Hence, there is a continuous effort in universities to find effective methods to enhance students' problem-solving skills in ethical scenarios, preparing them for future real-world decision-making.

To further elucidate the connection between problem-solving and ethical reasoning, it is imperative to consider the integral role of problem-solving in navigating ethical dilemmas. Effective problem-solving provides a foundation for individuals to explore, evaluate, and select the most ethical courses of action in complex situations, thereby directly influencing their capacity for ethical reasoning (Veatch et al., 2017). Moreover, the development of problem-solving skills is closely tied to enhancing students' ability to engage with and resolve ethical challenges, laying a groundwork for informed and thoughtful ethical decision-making (Chowning et al., 2012).

The thinking-aloud teaching method is used for problem-solving and is a student-centered active learning approach (Millis, 2012) with the unique characteristic of guiding students to address questions without straightforward answers (Van Someren et al., 1994), making it suitable for scenario-based ethical dilemma learning. Research across various disciplines, like biology (Newman et al., 2016), computer science (Porter et al., 2019), and physics (McGinness & Savage, 2016), confirmed the benefits of thinking-aloud in enhancing problem-solving and reasoning skills (Reinhart et al., 2022; Tedesco-Schneck, 2018). An advanced variant, Thinking-Aloud Pair Problem-Solving (TAPPS), introduced dialogic two-person problem-solving method (Whimbey & Lochhead, 1999) which has been shown to further enhance cognitive process organization and problem-solving performance (Schraw, 1998). TAPPS' distinct feature is its collaborative approach, where participants alternate roles as 'listener' and 'problem solver', providing feedback and aiding in understanding reasoning processes (DiNapoli & Miller, 2022; Reinhart et al., 2022Mahyar & Dani, 2021). This contrasts with the conventional method, where problem-solving is pursued individually. TAPPS' introduction of a learning companion (the 'listener') offers the advantage of addressing potential reasoning blind spots that may arise in individual thinking, with listeners supporting problem solvers through feedback and promoting self-reflection (Gourgey, 1998). Additionally, the role of learning motivation in the context of problem-solving and ethical reasoning cannot be understated. Motivated students are more likely to engage deeply with problem-solving tasks(Amerstorfer & Freiin von Münster-Kistner, 2021), showing persistence and a willingness to tackle complex ethical dilemmas.

Numerous studies showed that TAPPS enhanced learning outcomes in areas such as attitude, interest, self-concept, values, and morality (Braier-Lorimer & Warren-Miell, 2022; Noh et al., 2005; Setiawan et al., 2021; Sultan & Alasif, 2021). However, the specific impacts of TAPPS on students' moral reasoning, learning motivation, and problem-solving abilities in ethical dilemma discussions were not yet fully understood. Limited research, like the study by Audie (2019), has explored these aspects, particularly in the context of integrating a learning companion in ethical dilemma learning.

Recent advancements in chatbot technology have spurred innovation in human–machine learning models and interactive dialogue systems for various educational purposes, such as instructional and learning guidance (Iku-Silan et al., 2023), interactive companionship (Liu et al., 2022), and self-assessment (Durall & Kapros, 2020). Chatbots, serving as instructional tools, enhance interactive learning experiences through collaborative scripting and adaptive support (Vogel et al., 2017; Wambsganss et al., 2021). The emergence of artificial intelligence (AI) technologies, like ChatGPT, has underscored their potential in facilitating virtual learning companion interactions in text-based dialogue systems, promoting collaborative problem-solving in learning tasks (Kasneci et al., 2023a, 2023b; Lukpat, 2023; Ratten & Jones, 2023). This study aims to explore the transformative effects of generative AI-assisted virtual learning companions on student learning outcomes, a topic that is both timely and critical (Lim et al., 2023). Although the literature on AI-based chatbots as virtual learning companions is still developing, this research contributes to understanding their role in business ethics education.

Recent studies have started to uncover the potential of generative AI-powered chatbots as virtual learning companions in human–machine interactive learning. For instance, Su et al. (2023) explored their use in argumentative writing tasks, where students engaged with ChatGPT for various writing processes. Similarly, Yilmaz and Yilmaz (2023) demonstrated that such interactions could enhance computational thinking, programming self-efficacy, and motivation in complex programming tasks. Despite these promising findings, the need for more comprehensive empirical research remains to fully understand the impact of AI-assisted virtual learning companions on learning outcomes, particularly in the context of business ethics education (Chaturvedi et al., 2023; Gill et al., 2024; Kasneci et al., 2023a, 2023b). The current literature is still evolving, emphasizing the importance of further exploration in this area.

To address this gap, this study constructed a generative AI-assisted virtual learning companion interaction system and examined the efficacy of human–machine interactions in a type of TAPPS learning applied to ethical dilemma problem-solving activities. Further, an experiment was conducted in a university business ethics course to assess the effects of the proposed method by attempting to answer the following research questions:

(1) Can the virtual learning companion, facilitated by generative AI-assisted TAPPS, enhance students' problem-solving abilities when compared with both conventional TAPPS and individual thinking-aloud approaches?
(2) Can the virtual learning companion, facilitated by generative AI-assisted TAPPS, enhance students' learning motivation when compared with both conventional TAPPS and individual thinking-aloud approaches?
(3) Can the virtual learning companion, facilitated by generative AI-assisted TAPPS, enhance students' ethical reasoning ability when compared with both conventional TAPPS and individual thinking-aloud approaches?

This paper begins by exploring the theoretical background and application of TAPPS, followed by an introduction to the design and functionality of AI-assisted virtual learning companions. The methodology section details the experimental design conducted, while the results and discussion sections present and analyze the experimental data. Finally, the conclusion summarizes the research findings and offers suggestions for future research.

2 Literature review

This section reviews the pertinent literature to explore how virtual learning companions transform traditional learning interaction models and how they enhance learning effectiveness by providing personalized support and real-time feedback. This laid the theoretical and empirical groundwork for the exploration of the application of virtual learning companions in ethical dilemma discussions and the TAPPS (Thinking-Aloud Pair Problem Solving) learning strategy, as discussed in later chapters.

2.1 Applications of virtual learning companions in classroom interactions

"Virtual learning companions" is a general term for human-like (Gnewuch et al., 2017), computer-simulated, nonauthoritative, intelligent tutoring systems that serve as pedagogical agents. They are alternately referred to as co-learners, simulated students, virtual peers, or artificial students (Chou et al., 2003). According to Clarebout and Heidig (2012), a virtual learning companion acts as a lifelike proxy, designed to enhance learning by guiding students. Such a companion changes the learning mode from interpersonal learning interactions to human–machine collaborations, enabling students to accomplish learning tasks or exercises with the assistance of virtual learning companions. Early research indicated that educators utilized virtual learning companions to assist student interactions in various learning contexts (Nkambou et al., 2010). Beyond merely supporting interactive learning tasks or exercises, in-class virtual learning companions can provide students with real-time and tailored feedback (Bowman, 2012) and lead to positive learning outcomes such as increased motivation (Lin et al., 2020) and efficiency in performing learning tasks (Barange et al., 2017; Kim et al., 2006).

In business ethics education, ethical dilemmas are a critical instructional component that is applied to foster ethical reasoning and business problem-solving abilities, and they often are discussed in the context of case studies (Rombout et al., 2022). In conventional lesson units focusing on ethical dilemmas, educators guide students through thinking-aloud processes to practice ethical reasoning. Classroom discussions are effective in promoting dialogue that facilitates ethical reasoning (Rescher, 2014). However, educators face challenges in leading an entire class in thinking-aloud activities, especially when it comes to supporting the students' metacognitive monitoring related to ethical reasoning. A lack of effective metacognitive monitoring can adversely influence problem-solving (Callan et al., 2021).

The role of virtual learning companions in higher education has grown, especially their role as pedagogical agents assisting in highly interactive learning tasks

(Sikström et al., 2022), and their inclusion in the classroom enables students to engage in human–machine interactions that are human-like and simulate genuine human dialogues for learning. Studies elaborated on the applications of virtual learning companions, discussing their merits and implementation and the subsequent influence thereof on educators creating interactive learning environments. For example, Yilmaz and Yilmaz (2019) conducted an experiment utilizing a virtual learning companion to offer metacognitive support to students, and the companion enhanced the students' motivation and metacognitive awareness. Corbett and Anderson (2001) performed comparative research to examine the differences in feedback when virtual learning companions interact with students and found that providing immediate feedback to students accelerated the learning rates of students compared with providing feedback upon request.

Virtual learning companions are perceived as supportive tools for simulating interactive learning and as enabling personalized instruction and fostering student-centered learning approaches, but several scholars have found that the effects of such pedagogical agents on student motivation and learning are not apparent (Heidig & Clarebout, 2011; Schroeder & Gotch, 2015; Schroeder et al., 2017). Further, prior to the advent of generative AI, some experts remained skeptical about the capacity of pedagogical agents to engage in interactive dialogue with students (Kopp & Krämer, 2021). Edwards et al. (2018) cautioned that as students engage with the technology that enlivens virtual learning companions, their outcomes might not align with the initial expectations of educators and students.

Learning motivation and development of students' ethical reasoning and problem-solving ability are crucial for successful ethical dilemma activities. Therefore, this study developed a generative AI-assisted virtual learning companion system to aid students in paired thinking-aloud processes during ethical dilemma activities and to offer them opportunities to receive personalized immediate feedback.

2.2 TAPPS

"Thinking aloud" has been categorized as a self-regulatory tool that promotes active learning (Zimmerman & Schunk, 2008). Conventional thinking aloud methods typically involve instructors demonstrating the process, followed by students implementing it themselves (Kucan & Beck, 1997). When students are asked to think aloud when performing ethical reasoning regarding moral dilemmas, they often struggle to identify their blind spots in ethical reasoning during the process. TAPPS is a learning activity that incorporates both a thinker and a listener (Jeon et al., 2005). Students can better solve problems by undergoing extensive deliberation highlighting various issues (Wardana & Sagoro, 2019). TAPPS had been applied in various discussion-based learning scenarios, and had yielded benefits in numerous areas such as analytical thinking (Whimbey & Lochhead, 1999), problem-solving (Irham & Zainuri, 2017), and fostering learner autonomy (Herbert & Williams, 2023).

The application of TAPPS saw rapid growth across various educational settings and disciplines in the past (Salmi, 2022), and promoted problem-solving learning (Kotsopoulus, 2010). Although the development of a virtual learning companion for implementing TAPPS showed early promise, research into the integration of a virtual learning companion with TAPPS needed to continue. Only Ramachandran et al. (2018) have developed a tutoring robot to support students in TAPPS, demonstrating that a tutoring robot can effectively support students in employing the thinking aloud strategy during cognitively complex problem-solving tasks, leading to a notable improvement in learning performance. How to effectively use virtual learning companion interactive systems to support problem-solving learning activities is thus a pressing research question.

Consequently, this research aimed to enhance university students' problem-solving abilities, learning motivation, and ethical reasoning capabilities in moral dilemma education by employing a virtual learning companion chatbot based on generative AI-assisted TAPPS.

2.3 System structure: Virtual learning companion for TAPPS

This study developed a chatbot virtual learning companion system based on generative AI to facilitate TAPPS. The system is deployed on Discord, chosen for its user-friendly interface and suitability for text-based dialogue, which aligns with the educational communication needs of our study. Figure 1 presents the system architecture, encompassing a learning guidance mechanism, a TAPPS exercise mechanism, and several databases.

The learning guidance mechanism, operated by the host robot TABot, interprets the current ethical scenario via a scenario case module and guides students through ethical decision-making using an ethical value choice module. This interactive process is designed to engage students with the moral complexities of the scenarios and support them in making informed ethical choices.

For the TAPPS training mechanism, Master Socrates, a virtual learning companion robot, interacts with students through real-time text messaging facilitated by the ChatGPT 3.5 API. This API provides a robust natural language processing model that enables Master Socrates to deliver contextually appropriate responses during discussions of ethical dilemmas. Master Socrates is programmed to

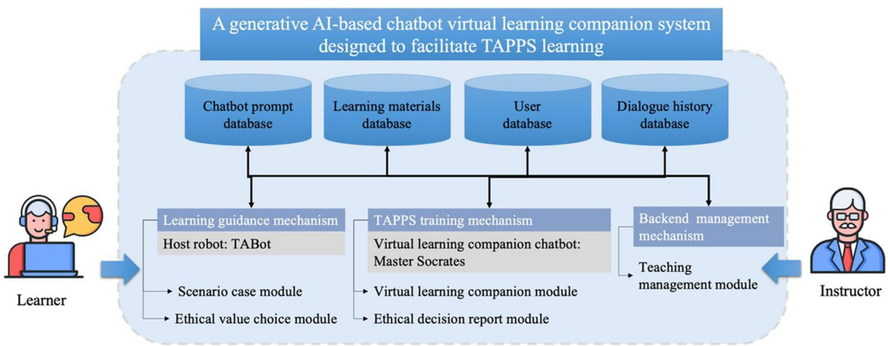

Fig. 1 System structure

perform as an interactive partner in ethical dialogues, driven by a carefully engineered prompt system to ensure relevance and educational value.

The backend management mechanism allows instructors to monitor interactions within the system, including both student inputs and responses from the learning companions (Master Socrates and TABot). This administrative oversight is crucial for tracking progress, evaluating student engagement, and ensuring the quality of the educational experience.

Figure 2 illustrates a screenshot of the system in action, demonstrating the user interface on Discord where learners engage with Master Socrates. The dialogues generated through this interaction are recorded in the user database and dialogue history database, allowing for a comprehensive analysis of the TAPPS process and the learning outcomes achieved.

By implementing the system in a real-time communication platform like Discord, we leverage its ease of use and accessibility, ensuring a seamless and integrated learning experience for students. The system's design is intentionally crafted to support the TAPPS learning methodology, promoting active engagement and ethical reasoning in a collaborative virtual environment."

3 Experimental procedure and research participants

In this section, we meticulously detail the experimental procedure and provide an overview of the participants involved in the study. The first subsection, 3.1 'Participants', outlines the number and general background of the students who participated in the research, setting the stage for understanding the scope and context of the study. The following subsection, 3.2 'Learning Activity Procedure', describes the specific steps and methodologies employed during the experiment.

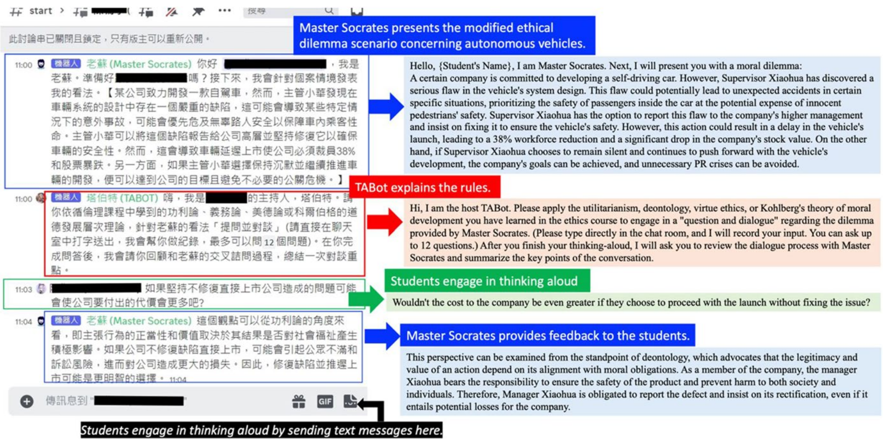

Fig. 2 System screenshot

3.1 Participants

This study employed a randomized experimental design and recruited 135 students enrolled in a business ethics course at a university. The purpose was to deepen students' understanding of ethical issues and the complexities of moral decisions by participating in moral dilemma activities. We anticipated that through these learning activities, students gained the confidence and ability to make informed decisions when faced with various moral dilemmas in their future professional lives and that, further, they actively promoted ethical practices and social responsibility when encountering ethical challenges.

3.2 Learning activity procedure

Figure 3 outlines the procedure of the learning activity across a period of four weeks. In the first 2 weeks, lessons on the three major schools of ethics were conducted. During this phase, three groups of students, assigned randomly, received basic knowledge on deontology, utilitarianism, and virtue ethics, all taught by the same instructor. After the teaching sessions, the students underwent a pretest and a pre-class questionnaire survey, which included assessments of their problem-solving ability, learning motivation, and ethical reasoning ability. The purpose of the pretest was to gauge their basic understanding of the three major schools of ethics.

In Week 3, a moral dilemma learning activity was held. At the beginning of this phase, the teacher had first introduced the moral dilemma scenarios and the learning objectives. Additionally, using the classic trolley problem (Foot, 1967) as an example, the teacher demonstrated to all three groups how thinking aloud can be applied to make ethical decisions. This approach, as proposed in earlier research, suggests that teachers can help students familiarize themselves with the method by using unrelated topics before having them think aloud (Austin & Delaney, 1998). Each group received a demonstration of how to interact with the learning system using

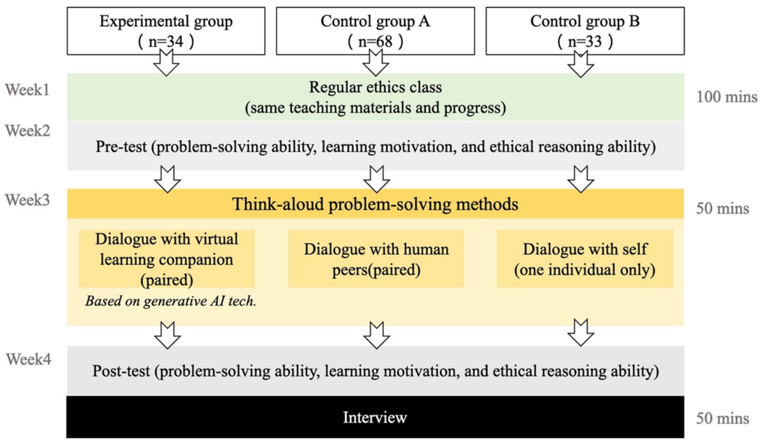

Fig. 3 The experiment process

smartphones (experimental group), how to use the learning sheet collaboratively with a peer (control group A), or how to use the learning sheet individually (control group B) to complete the moral dilemma activities. Students in control group A collaborated in pairs to engage with the moral dilemma activities. In control group A, where 68 students were initially enrolled, pairs were formed for collaborative engagement with the moral dilemmas. For the purpose of data analysis and to maintain consistency across the study groups, we collected data from one student of each pair in the second round of the activity. This was done to ensure that each pair's interaction was represented as a single unit of analysis in the study, aligning with the experimental design which sought to compare the collaborative engagement against individual and AI-assisted interactions. As such, 34 unique data entries from control group A were included for final analysis, reflecting the 34 pairs that participated in the round 2 activity focusing on the Autonomous Vehicle Dilemma. Each small group was formed by drawing lots randomly to ensure a fair distribution (Fahlevi et al., 2021).

Each student underwent 25 min of thinking-aloud learning. Students in the experimental group utilized the learning system based on the generative AI-assisted TAPPS strategies for moral dilemma learning. First, a teaching assistant robot named TABot, acting as host, guided the students in reading the moral dilemma scenarios, informed them of the learning task, and introduced a chatbot named Master Socrates (based on generative AI technology) to assist with the thinking-aloud process. After reading the moral dilemma scenarios, students were asked to articulate their thoughts by sending text messages to Master Socrates. Master Socrates provided immediate feedback. Upon receiving feedback from Master Socrates, the students could then submit additional thinking-aloud messages within the learning system. To ensure that the interactive process was concise and aligned with the characteristics of thinking aloud, and to prevent students from proposing solutions rather than sharing thoughts, the teacher instructed the students beforehand to express their thoughts in question form and limited the interaction to a maximum of 12 exchanges.

After the thinking-aloud session, the students had been required to submit an ethical decision report based on their reflections. During the problem-solving process, students in the experimental group were initially assigned the role of "listener" in round 1, which focused on the Heinz dilemma moral scenario. In this capacity, their task was to provide feedback to simulate providing feedback to Master Socrates, who was engaged in problem-solving. Subsequently, in round 2, centered on the Autonomous vehicle moral dilemma, the roles were reversed. Students took on the role of "problem solver," while Master Socrates assumed the role of "listener" and offered feedback to guide the ethical decision-making process. This alternating role dynamic was designed to facilitate a holistic learning experience, enabling students to familiarize themselves with the system's functionalities as well as the process of expressing their thoughts aloud by initially assuming the role of a listener.

Students in control group A followed a similar approach to the experimental group, with two students alternating roles to address a moral dilemma scenario assigned randomly. Students in control group B used a conventional thinking-aloud exercise—they read the moral dilemma scenarios independently and then engaged in a self-directed thinking-aloud problem-solving activity. Upon completing the

Table 1 Moral dilemma scenarios for the three groups

	Experimental group	Control group A	Control group B
Round 1	Heinz dilemma	Heinz dilemma	-
Round 2	Autonomous vehicle	Autonomous vehicle	Autonomous vehicle

problem-solving process, students in both control groups A and B were required to submit an ethical decision report on the topic of the Autonomous vehicle dilemma. Students in control groups A and B used their mobile phones for audio recording to assist in compiling their reports. For the sake of fairness, the study had exclusively analyzed data from participants in the second round, focusing on the modified Autonomous Vehicle Dilemma topic (Awad et al., 2018; Bonnefon et al., 2016), across all three groups. This yielded data for analysis consisting of 34 participants in the experimental group, 34 participants in control group A, and 33 participants in control group B. The summary for Master Socrates in the experimental group was automatically generated by the learning system and excluded from this study. Table 1 presents the moral dilemma scenarios presented to the students in the three groups of this study.

After each learner completed the moral dilemma problem-solving activity in week 4, they filled out a post-activity questionnaire to measure their problem-solving ability, learning motivation, and ethical reasoning ability. Then, researchers conducted interviews with students from all three groups to collect their opinions and views on the learning methods.

3.3 Measurement instruments

A pretest was used to evaluate the prior knowledge of the three student groups and ensure that they had similar levels of knowledge before participating in the learning activity. An ethics teacher with 28 years of teaching experience was involved in formulating the questions. The test consisted of 10 multiple-choice questions, each worth 10 points, for a total possible score of 100. An example question is, "A company discovers a potential safety risk in one of its products. According to the perspective of deontology, how should the company management respond?" The reliability of the test, as indicated by the Kuder–Richardson Formula 20 (KR20), was 0.81.

Apart from the pretest, three instruments were used to measure the students' problem-solving ability, learning motivation, and ethical reasoning ability following the moral dilemma problem-solving activity.

First, this study adapted the Perceived Problem-Solving Skills Scale originally developed by Armour-Thomas and Haynes (1988), the reliability and validity of which were later confirmed by Masal et al. (2013). The scale has 23 items covering three dimensions, namely, planning (13 items; Cronbach's $\alpha = 0.87$; e.g., "I can organize the information presented in a problem"), organization (six items; Cronbach's $\alpha = 0.74$; e.g., "I can identify the crucial information in a problem"), and evaluation (four items; Cronbach's α = 0.72; e.g., "After solving a problem, I review to determine whether the solution matches my planned strategy"). A 5-point Likert scale was employed (5 = *strongly agree*; 1 = *strongly disagree*).

Second, the Learning Motivation Questionnaire, based on the scale by Pintrich et al. (1991), was adapted from Wang and Chen (2010). The questionnaire encompasses two dimensions, extrinsic motivation (three items; e.g., "I believe I can excel in this learning activity") and intrinsic motivation (three items; e.g., "I enjoy the present learning activity, which is challenging, because I can acquire new knowledge"). The overall reliability of the questionnaire, as indicated by Cronbach's α, was 0.90, denoting excellent reliability. A 5-point Likert scale was used (5 = *strongly agree*; 1 = *strongly disagree*).

Third, the Ethical Reasoning Inventory was adapted from a questionnaire developed by Tsai et al. (2009) to measure the students' ethical reasoning ability before and after the moral dilemma problem-solving activity. The original questionnaire consisted of case-oriented and individual-oriented subscales. This study employed only the case-oriented subscale (six items; 5 = *strongly agree*; 1 = *strongly disagree*) to align with the instructional design involved in this study. The reliability of the questionnaire, as indicated by Cronbach's α, was 0.79, suggesting favorable reliability. An example item is, "I can identify conflicts or dilemmas."

4 Experimental results

4.1 Problem-solving ability

To investigate whether using different thinking-aloud learning strategies in moral dilemma activities resulted in differences in the problem-solving ability of students, this study analyzed, first, whether the pretest scores of the three groups were similar. The homogeneity of regression coefficients among the different groups did not reach significance, $F_{(2, 98)} = 0.86$, $p = 0.43$, implying no interaction effect between the covariate and the independent variable. The results of Levene's test for equality of variances indicated no significant difference, $F_{(2, 98)} = 1.16$, $p = 0.32$. Therefore, an analysis of covariance (ANCOVA) was conducted; the results are shown in Table 2.

The ANCOVA revealed that, after controlling for pretest score differences, the adjusted mean values for the experimental group (M = 4.22, SD = 0.18), control group A (M = 4.05, SD = 0.24), and control group B (M = 3.25, SD = 0.23) were significantly different. The test of between-subjects effects showed a significant effect

Table 2 ANCOVA results for problem-solving ability

Group	*N*	Mean	*SD*	Adjusted Mean	*F*	η^2	Post hoc(LSD)
Exp	34	4.21	.18	4.22	217.27***	.82	Exp>$Cr_{(a)}$
$Cr_{(a)}$	34	4.05	.24	4.05			Exp > $Cr_{(b)}$
$Cr_{(b)}$	33	3.25	.23	3.25			$Cr_{(a)} > Cr_{(b)}$

****p* < .001.

of the learning strategy on problem-solving ability, $F_{(2, 98)} = 217.27$, $p < 0.001$, with a large effect size(Cohen, 1988), $\eta^2 = 0.82$. Post hoc comparisons using the LSD test indicated that both the experimental group and control group A significantly outperformed control group B, and the experimental group was superior to control group A.

These findings suggest that while all three methods effectively enhanced students' problem-solving ability in moral dilemmas, but the experimental group showed the greatest improvement, followed by control group A.

4.2 Learning motivation

A one-way ANCOVA was performed to determine the differences in the learning motivation of the students using different thinking-aloud strategies in the moral dilemma activities. The homogeneity of regression coefficients among the different groups did not reach significance, $F_{(2, 98)} = 0.42$, $p = 0.66$, suggesting no interaction effect between the covariate and the independent variable. Similarly, Levene's test for equality of variances confirmed homogeneity of error variances, $F_{(2, 98)} = 0.69$, $p = 0.50$.

Table 3 presents the ANCOVA results for learning motivation. After adjusting for pretest scores, the mean values for the experimental group (M = 4.33, SD = 0.34), control group A (M = 4.13, SD = 0.29), and control group B (M = 3.24, SD = 0.30) were found to differ significantly. The test of between-subjects effects indicated a significant effect of the instructional strategy on learning motivation, $F_{(2, 98)} = 146.34$, $p < 0.001$, with a large effect size, partial $\eta^2 = 0.75$.

Post hoc comparisons using the LSD test revealed that both the experimental group and control group A had significantly higher learning motivation than control group B, with the experimental group showing the highest level of motivation. Specifically, the experimental group exceeded control group A, and both of these groups were significantly more motivated than control group B.

These results suggest that the implementation of thinking-aloud strategies in moral dilemma activities has a positive effect on learning motivation, with the experimental group benefiting the most, followed by control group A.

Table 3 ANCOVA results for learning motivation

Group	*N*	Mean	*SD*	Adjusted Mean	*F*	η^2	Post hoc(LSD)
Exp	34	4.31	.34	4.33	146.34***	.75	Exp> $Cr_{(a)}$
$Cr_{(a)}$	34	4.14	.29	4.13			Exp> $Cr_{(b)}$
$Cr_{(b)}$	33	3.25	.30	3.24			$Cr_{(a)} > Cr_{(b)}$

****p* < .001.

4.3 Ethical reasoning ability

A one-way Analysis of Covariance (ANCOVA) was conducted to assess the differences in ethical reasoning ability among students using various thinking-aloud strategies in moral dilemma activities. The assumption of homogeneity of regression slopes was met, $F_{(2, 98)} = 0.10$, $p = 0.91$, indicating no interaction effect between the covariate and the independent variable. Furthermore, Levene's test for equality of error variances was not significant, $F_{(2, 98)} = 0.84$, $p = 0.44$, confirming that the variance was consistent across groups.

The ANCOVA results, as displayed in Table 4, revealed that after adjusting for pretest scores, the experimental group (M = 4.46, SD = 0.24), control group A (M = 4.05, SD = 0.28), and control group B (M = 3.86, SD = 0.30) had significantly different adjusted mean scores. The test of between-subjects effects indicated a significant effect of the learning strategy on ethical reasoning ability, $F_{(2, 98)} = 48.30$, $p < 0.001$, with a moderate effect size, partial $\eta^2 = 0.50$.

Post hoc analyses using the LSD test showed that the experimental group outperformed both control groups, with control group A also showing greater ethical reasoning ability than control group B. This suggests that while all three strategies were effective in enhancing ethical reasoning, the strategy employed by the experimental group led to the most significant improvement.

We conducted an in-depth investigation into the effectiveness of using generative AI-powered chatbots as a virtual learning companion to facilitate the TAPPS method in moral dilemma problem-solving activities. We conducted in-depth interviews with six students from each group, with an equal representation of genders, ensuring a balanced male-to-female ratio. Inspired by Lin et al. (2021), we explored several facets of the learning method, such as the unique features of engaging in moral dilemma learning and changes in the students' initial learning expectations. Additionally, the interviews sought to connect students' personal experiences with the three main findings of our study, aiming to understand how the virtual learning companion influenced problem-solving ability, learning motivation, and ethical reasoning in a cohesive manner.

The experimental results revealed a consensus among participants in the experimental group, who all agreed that utilizing the generative AI-powered chatbot as a virtual learning companion to facilitate TAPPS offered significant advantages in promoting thinking aloud and identifying blind spots in ethical reasoning. These qualitative insights align with our quantitative findings, which indicated that the experimental group not only improved in their problem-solving abilities but also showed a substantial increase in learning motivation and ethical reasoning.

Table 4 ANCOVA results for ethical reasoning ability

Group	*N*	Mean	*SD*	Adjusted Mean	*F*	η^2	Post hoc(LSD)
Exp	34	4.45	.24	4.46	48.30***	.50	Exp> $Cr_{(a)}$
$Cr_{(a)}$	34	4.04	.28	4.05			Exp > $Cr_{(b)}$
$Cr_{(b)}$	33	3.88	.30	3.86			$Cr_{(a)} > Cr_{(b)}$

***$p < .001$.

All students in the experimental group highlighted the advantage of the virtual learning companion system providing immediate feedback during the learning activity. They opined that such immediate feedback helped them swiftly optimize the thinking aloud process, as it provided the necessary ethical reasoning information to resolve dilemmas. For example, Student E(3) remarked that students were able to more quickly engage in ethical reasoning activities related to corporate moral dilemmas with the immediate feedback provided by the virtual learning companion. This reflects the quantitative result where the experimental group was found to have enhanced their problem-solving ability significantly more than the other groups. Student E(5) further noted that the immediate feedback provided was highly insightful, not only ensuring students' comprehensive understanding of the thinking aloud process in ethical dilemmas but also specifically fostering ethical reasoning ability. The increase in ethical reasoning ability was one of the key quantitative outcomes this study aimed to measure.

Furthermore, the interviews revealed that the increased engagement and interaction with the virtual learning companion were instrumental in promoting students' learning motivation, which was another main quantitative finding of this study. Five students (83%) shared the same opinion that dialogues with the chatbot guided them to explore flaws in their reasoning, thereby deepening their understanding of dilemmas and enhancing their problem-solving ability. Student E(1) stated that the virtual learning companion guided them to focus on flaws in their ethical reasoning, enabling them to tackle moral dilemmas with precision. Student E(6) asserted, "Such dialogues made me aware of potential biases I might have. Interacting with the virtual learning companion made me more cautious when making ethical judgments to ensure that the solutions are prudent and not just a simple response; it is a whole new learning experience." This finding is consistent with that of Kani and Shahrill (2015) and Montague et al. (2011), who asserted that the TAPPS learning approach helps students recognize weaknesses and errors in their thought process, thereby enhancing their problem-solving ability—a main finding in our quantitative analysis.

The qualitative feedback from students in the control groups also provided context for their respective quantitative outcomes. Students in control group A, who adopted the conventional TAPPS learning method, primarily focused on the assistance that peer interactions provided in ethical reasoning for moral dilemma scenarios. For example, four students (67%) mentioned that the one-on-one thinking-aloud dialogue interactions were engaging. Five students acknowledged the positive aid of peer-responsive feedback in making ethical decisions. Three students highlighted the advantages of peer assistance in analyzing the moral dilemma cases. While these interactions were beneficial, they did not seem to enhance learning motivation or problem-solving ability to the same extent as the virtual learning companion did for the experimental group.

Students in control group B, who utilized the conventional TAPPS learning method, particularly emphasized the influence of thinking aloud on cognitive monitoring. Student C(B2) commented on the aid provided by the learning sheet in guiding the thinking-aloud exercise, pointing out that the learning sheet explicitly reminded them to be cautious of potential biases in their formulation of questions.

Student C(B3) emphasized the challenge of the pace of the thinking-aloud activity and the changing time intervals between expressing ideas and thinking as the thinking-aloud activity progressed. Student C(B1) found the task of independently conducting thinking aloud challenging, noting possible confusion and blind spots during the thinking process. Their reflections on the process help to explain why this group showed the least improvement in the quantitative measures out of all groups, particularly in terms of ethical reasoning ability.

Overall, the interviews served to elucidate the quantitative findings, providing a narrative that connected individual experiences with the measured improvements in problem-solving ability, learning motivation, and ethical reasoning. By doing so, the qualitative data supported the quantitative results, showing that the use of a generative AI-powered chatbot can offer a comprehensive approach to enhancing students' learning experiences in moral dilemma education.

5 Conclusion

This study explored the learning outcomes of students who were presented human–machine interactions to facilitate TAPPS. We constructed a generative AI-assisted virtual learning companion interaction system intended to support students in engaging in problem-solving activities involving moral dilemmas. An experiment was conducted in a business ethics course of a university to assess the effectiveness of the proposed method. The results demonstrated that the TAPPS learning approach, supported by a generative AI-assisted virtual learning companion system, improved the students' problem-solving ability, learning motivation, and ethical reasoning ability. These findings corroborate previous research findings that indicated that engaging in thinking-aloud dialogue leads to a deeper understanding of the case study; students are then better equipped to solve problems (Umar et al., 2022), and their learning motivation is enhanced (Litualy, 2017). Schroeder et al. (2013) explained that student engagement is fostered through interaction, which in turn promotes learning. Our research revealed the importance of metacognitive monitoring when students partake in problem-solving activities related to moral dilemmas, a point previously highlighted in several studies (Kamali et al., 2022; Kennedy, 2022; Osorio & Reyes, 2023). Offering metacognitive feedback can enhance students' problem-solving ability (Kefalidou, 2017).

The interview results revealed that the students believed that the TAPPS approach, when accompanied by a learning peer, provided them with immediate feedback. This aided them in identifying cognitive blind spots and fostered reflection. Harley et al. (2017) suggested that this is because students, upon receiving prompt and direct feedback, are likely to retry the task. This finding is consistent with that of research that found that guiding students to engage in reflection can elevate their higher-order cognitive thinking ability (Jarvis & Baloyi, 2020).

The results also revealed that the experimental group demonstrated a marked increase in learning motivation compared with the students in control group B. This finding is consistent with earlier research that indicated that providing a virtual learning companion can enhance students' learning motivation (Wei et al., 2011). Interviews with the students from control group B further validated this finding.

In this study, we explored enhancing ethical dilemma learning through the TAPPS approach, supplemented with an AI-Assisted Virtual Learning Companion. One notable limitation involved the participant demographic, predominantly from a central Taiwanese technology university, which might affect the findings' generalizability. Thus, applying these results to different educational and cultural backgrounds may not be straightforward. Future research should explore the applicability of the AI-assisted TAPPS approach in various educational contexts and disciplines beyond business ethics to ascertain its effectiveness and adaptability. Future research should include diverse participants to validate our findings' broader relevance.

We encountered challenges related to the current technology of generative AI, prompting suggestions based on developers' experience. The literature indicates potential biases or limitations in using AI chatbots as virtual learning companions, including content accuracy and reliability (Khan et al., 2023), privacy (Thurzo et al., 2023; Zohny et al., 2023), and understanding complex conversations (Ray, 2023). Students, fully informed and involved in the learning and experimental setup, may tolerate privacy and interaction issues due to conversation understanding. Yet, this does not negate the potential occurrence of these biases or limitations in future practical applications. Full disclosure and risk management are crucial for educators. Literature suggests methods like "fine-tuning" for addressing content accuracy and reliability (Tajbakhsh et al., 2016). Teachers should utilize AI models as a teaching supplement and not become overly dependent on them (Kasneci et al., 2023a, 2023b), ensuring active intervention in TAPPS learning when necessary.

Specific challenges were also faced in developing and testing the virtual learning companion system, such as service disruptions with the ChatGPT API integration. Teachers' early announcements and troubleshooting methods can mitigate these issues. Future research will continue to optimize generative AI chatbots as virtual learning companions, aligning with value-sensitive design principles (Friedman et al., 2017; Vernim et al., 2022) and improving system efficiency and user experience.

In summary, the results indicated that employing generative AI-assisted human–machine interaction in the TAPPS approach during moral dilemma problem-solving activities can enhance students' problem-solving ability, learning motivation, and ethical reasoning ability compared with the conventional individual thinking aloud method. These findings offer insights on generative AI-assisted human–machine interactive learning as well as on virtual learning companions; and provide concrete reference for the design of moral dilemma activities in business ethics courses.

Funding This research is partially supported by the National Science and Technology Council, Taiwan. under Grant No. 112–2410-H-224–011-.

Data availability Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Code availability Not applicable.

Declarations

Informed consent to participate We have no conflict of interest to declare.

Conflicts of interest/competing interests All participants gave full informed consent to participate.

References

Amerstorfer, C. M., & Freiin von Münster-Kistner, C. (2021). Student perceptions of academic engagement and student-teacher relationships in problem-based learning. *Frontiers in Psychology, 12*, 713057. https://doi.org/10.3389/fpsyg.2021.713057

Armour-Thomas, E. & Haynes, N. M. (1988). Assessment of metacognition in problem solving. *Journal of Instructional Psychology, 15*(3), 87–93. https://www.proquest.com/docview/1416365101?pq-origsite=gscholar&fromopenview=true. Accessed 1 Sept 2023

Audie, N. (2019). Peran Media Pembelajaran Meningkatkan Hasil Belajar Peserta Didik. *Prosiding Seminar Nasional Pendidikan FKIP UNTIRTA, 2*(1), 586–595. https://jurnal.untirta.ac.id/index.php/psnp/article/view/5665/4066. Accessed 1 Sept 2023

Austin, J., & Delaney, P. F. (1998). Protocol analysis as a tool for behavior analysis. *Analysis of Verbal Behavior, 15*, 42–56. https://doi.org/10.1007/2FBF03392922

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine Experiment. *Nature, 563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Barange, M., Saunier, J., & Pauchet, A. (2017). Pedagogical agents as team members: Impact of proactive and pedagogical behavior on the user. In Proceedings of *the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, 2, - 8002017*. 16th International Conference on Autonomous Agents and Multiagent Systems, AAMAS.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573. https://doi.org/10.1126/science.aaf2654

Bowman, C. D. (2012). Student use of animated pedagogical agents in a middle school science inquiry program. *British Journal of Educational Technology, 43*(3), 359–375. https://doi.org/10.1111/j.1467-8535.2011.01198.x

Braier-Lorimer, D. A., & Warren-Miell, H. (2022). A peer-led mock OSCE improves student confidence for summative OSCE assessments in a traditional medical course. *Medical Teacher, 44*(5), 1–6. https://doi.org/10.1080/0142159X.2021.2004306

Callan, G. L., Rubenstein, L. D., Ridgley, L. M., & McCall, J. R. (2021). Measuring self-regulated learning during creative problem-solving with SRL microanalysis. *Psychology of Aesthetics, Creativity, and the Arts, 15*(1), 136–148. https://doi.org/10.1037/aca0000238

Chaturvedi, R., Verma, S., Das, R., & Dwivedi, Y. K. (2023). Social companionship with artificial intelligence: Recent trends and future avenues. *Technological Forecasting and Social Change, 193*, 122634. https://doi.org/10.1016/j.techfore.2023.122634

Chou, C.-Y., Chan, T.-W., & Lin, C.-J. (2003). Redefining the learning companion: The past, present, and future of educational agents. *Computers & Education, 40*(3), 255–269. https://doi.org/10.1016/S0360-1315(02)00130-6

Chowning, J. T., Griswold, J. C., Kovarik, D. N., & Collins, L. J. (2012). Fostering Critical Thinking, Reasoning, and Argumentation Skills through Bioethics Education. *PLoS ONE, 7*(5), e36791. https://doi.org/10.1371/journal.pone.0036791

Clarebout, G. & Heidig, S. (2012). Pedagogical agents. In N.M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2567–2571). Springer. https://doi.org/10.1007/978-1-4419-1428-6_942

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Corbett, A. T. & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 245–252). ACM.

DiNapoli, J., & Miller, E. K. (2022). Recognizing, supporting, and improving student perseverance in mathematical problem-solving: The role of conceptual thinking scaffolds. *The Journal of Mathematical Behavior, 66*, 100965. https://doi.org/10.1016/j.jmathb.2022.100965

Durall, E., and Kapros, E. (2020). Co-design for a Competency Self-assessment Chatbot and Survey in Science Education. In: Zaphiris, P., Ioannou, A. (eds) *Learning and Collaboration Technologies*. Human and Technology Ecosystems. HCII 2020. Lecture Notes in Computer Science, 12206. Springer, Cham.https://doi.org/10.1007/978-3-030-50506-6_2

Edwards, C., Edwards, A., Spence, P. R., & Lin, X. (2018). I, teacher: Using artificial intelligence (AI) and social robots in communication and instruction. *Communication Education, 67*(4), 473–480. https://doi.org/10.1080/03634523.2018.1502459

Fahlevi, H., Irsyadillah, I., Indriani, M., & Oktari, R. S. (2021). DRG-based payment system and management accounting changes in an Indonesian public hospital: Exploring potential roles of big data analytics. *Journal of Accounting & Organizational Change, 18*, 325–345. https://doi.org/10.1108/JAOC-10-2020-0179

Foot, P. (1967). *The problem of abortion and the doctrine of the double effect*. Oxford University Press. https://doi.org/10.1093/0199252866.003.0002

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. Foundations and Trends®. *Human-Computer Interaction, 11*(2), 63–125. https://doi.org/10.1561/1100000015

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., … Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems, 4*, 19–23. https://doi.org/10.1016/j.iotcps.2023.06.002

Gnewuch, U., Morana, S., & Maedche, A. (2017). *Towards Designing Cooperative and Social Conversational Agents for Customer Service*. ICIS.

Gourgey, A. F. (1998). Metacognition in basic skills instruction. *Instructional Science, 26*, 81–96. https://doi.org/10.1023/A:1003092414893

Harley, J. M., Taub, M., Azevedo, R., & Bouchet, F. (2017). Let's set up some subgoals: Understanding human-pedagogical agent collaborations and their implications for learning and prompt and feedback compliance. *IEEE Transactions on Learning Technologies, 11*(1), 54–66. https://doi.org/10.1109/TLT.2017.2756629

Hartwell, S. (1995). Promoting moral development through experiential teaching. *Clinical Law Review, 1*, 505.

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Journal of Media Psychology, 22*(2), 27–54. https://doi.org/10.1016/j.edurev.2010.07.004

Herbert, S., & Williams, G. (2023). Eliciting mathematical reasoning during early primary problem solving. *Mathematics Education Research Journal, 35*(1), 77–103. https://doi.org/10.1007/s13394-021-00376-9

Iku-Silan, A., Hwang, G.-J., & Chen, C.-H. (2023). Decision-guided chatbots and cognitive styles in interdisciplinary learning. *Computers & Education, 201*, 104812. https://doi.org/10.1016/j.compedu.2023.104812

Irham, M., & Zainuri, Z. (2017). Efektivitas Pembelajaran Think Aloud Pair Problem Solving Ditinjau dari Kemampuan Pemecahan Masalah. *Unnes Journal of Mathematics Education Research, 5*(1), 60–69. https://journal.unnes.ac.id/sju/index.php/ujmer/article/view/12918. Accessed 1 Sept 2023

Jarvis, M.-A., & Baloyi, O. B. (2020). Scaffolding in reflective journaling: A means to develop higher order thinking skills in undergraduate learners. *International Journal of Africa Nursing Sciences, 12*, 100195. https://doi.org/10.1016/j.ijans.2020.100195

Jeon, K., Huffman, D., & Noh, T. (2005). The effects of thinking aloud pair problem solving of high school students' chemistry problem-solving performance and verbal interactions. *Journal of Chemical Education, 82*(10), 1558–1564. https://doi.org/10.1021/ed082p1558

Kamali, F., Yousefy, A., & Yamani, N. (2019). Explaining professionalism in moral reasoning: A qualitative study. *Advances in Medical Education and Practice, 10*, 447–456. https://doi.org/10.2147/amep.S183690

Kamali, F., Yousefy, A. R., & Yamani, N. (2022). Explaining Metacognition in Moral Reasoning: A Qualitative Study. *Journal of Iranian Medical Council, 5*(1), 96–110. https://doi.org/10.18502/jimc.v5i1.9576

Kani, N. H. A., & Shahrill, M. (2015). Applying the Thinking Aloud Pair Problem Solving Strategy in Mathematics Lessons. *Asian Journal of Management Sciences, Education, 4*(2), 20–28. https://www.ajmse.leena-luna.co.jp/AJMSEPDFs/Vol.4(2)/AJMSE2015(4.2-03).pdf. Accessed 1 Sept 2023

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023b). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kavathatzopoulos, I. (1994). Training professional managers in decision-making about real life business ethics problems: The acquisition of the autonomous problem-solving skill. *Journal of Business Ethics, 13*, 379–386. https://doi.org/10.1007/BF00871765

Kefalidou, G. (2017). When immediate interactive feedback boosts optimization problem solving: A 'human-in-the-loop'approach for solving capacitated vehicle routing problems. *Computers in Human Behavior, 73*, 110–124. https://doi.org/10.1016/j.chb.2017.03.019

Kennedy, C. (2022). *Metacognitive Agents for Ethical Decision Support: Conceptual Model and Research Roadmap*. Qeios. https://doi.org/10.32388/H7VZT3

Khaghanizade, M., Malaki, H., Abbasi, M., Abbaspour, A., & Mohamadi, E. (2012). Faculty-Related Challenges in Medical Ethics Education: A Qualitative Study. *Iranian Journal of Medical Education, 11*(8), 903–916. https://ijme.mui.ac.ir/article-1-1434-en.html. Accessed 1 Sept 2023

Khan, R. A., Jawaid, M., Khan, A. R., & Sajjad, M. (2023). ChatGPT-Reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences, 39*(2), 605–607. https://doi.org/10.12669/pjms.39.2.7653

Kim, Y., Baylor, A.L. & PALS Group. (2006). Pedagogical Agents as Learning Companions: The Role of Agent Competency and Type of Interaction. *Educational Technology Research and Development, 54*, 223–243. https://doi.org/10.1007/s11423-006-8805-z

Knight, S., Hayhoe, B. W., Frith, L., Ashworth, M., Sajid, I., & Papanikitas, A. (2019). Ethics education and moral decision-making in clinical commissioning: An interview study. *British Journal of General Practice, 70*(690), e45–e54. https://doi.org/10.3399/bjgp19X707129

Kopp, S., & Krämer, N. (2021). Revisiting Human-Agent Communication: The Importance of Joint Co-construction and Understanding Mental States. *Frontiers in Psychology, 12*, 580955. https://doi.org/10.3389/fpsyg.2021.580955

Kotsopoulus, D. (2010). An analysis of talking aloud during peer collaborations in mathematics. *International Journal of Science and Mathematics Education, 8*, 1049–1070. https://doi.org/10.1007/s10763-010-9221-8

Kucan, L., & Beck, I. L. (1997). Thinking Aloud and Reading Comprehension Research: Inquiry, Instruction, and Social Interaction. *Review of Educational Research, 67*(3), 271. https://doi.org/10.2307/1170566

Lewis, E. J., Ludwig, P. M., Nagel, J., & Ames, A. (2019). Student ethical reasoning confidence pre/post an innovative makerspace course: A survey of ethical reasoning. *Nurse Education Today, 75*, 75–79. https://doi.org/10.1016/j.nedt.2019.01.011

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. L., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management in Education, 21*(2), 1–13. https://www.sciencedirect.com/science/article/pii/S1472811723000289. Accessed 1 Sept 2023

Lin, L., Ginns, P., Wang, T., & Zhang, P. (2020). Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain? *Computers & Education, 143*, 103658. https://doi.org/10.1016/j.compedu.2019.103658

Lin, Y. N., Hsia, L. H., & Hwang, G. J. (2021). Promoting pre-class guidance and in-class reflection: A SQIRC-based mobile flipped learning approach to promoting students' billiards skills, strategies, motivation and self-efficacy. *Computers & Education, 160*, 104035. https://doi.org/10.1016/j.compedu.2020.104035

Litualy, S. J. (2017). Aplikasi Pembelajaran TAPPS dalam Peningkatan Hasil Belajar Strukturen Und Wortschatz Mahasiswa Program Studi Pendidikan Bahasa Jerman. *Cakrawala Pendidikan, 2*. https://doi.org/10.21831/cp.v36i2.13288

Liu, C.-C., Liao, M.-G., Chang, C.-H., & Lin, H.-M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education, 189*, 104576. https://doi.org/10.1016/j.compedu.2022.104576

Lukpat, A. (2023). ChatGPT banned in New York City public schools over concerns about cheating, learning development. *The Wall Street Journal*. https://www.wsj.com/articles/chatgptbanned-in-new-york-city-public-schools-over-concerns-about-cheating-learning-development-11673024059. Accessed 1 Sept 2023

Mahyar, N., & Dani, A. U. (2021). Efektivitas Strategi Pembelajaran Thinking Aloud Pair Problem Solving (TAPPS) dan Strategi Pembelajaran Konvensional Terhadap Hasil Belajar Peserta Didik Kelas X MAN Pangkep. *AL-KHAZINI: JURNAL PENDIDIKAN FISIKA, 1*(2), 129–135. https://doi.org/10.24252/al-khazini.v1i2.20638

Masal, E., Takunyacı, M., & Gülay, A. (2013). Adaptation of student thinking about problem solving scale (STAPSS) to Turkish. *The Journal of SAU Education Faculty, 25*, 134–146. https://dergipark.org.tr/tr/download/article-file/115666. Accessed 1 Sept 2023

McGinness, L. P., & Savage, C. M. (2016). Developing an Action Concept Inventory. *Physical Review Physics Education Research, 12*, 010133. https://doi.org/10.1103/PhysRevPhysEducRes.12.010133

Millis, B. (2012). *Active Learning Strategies in Face-to-Face Courses*. IDEA. https://files.eric.ed.gov/fulltext/ED565290.pdf. Accessed 1 Sept 2023

Montague, M., Krawec, J., & Rosenzweig, C. (2011). Metacognitive strategy use of eight-grade students with and without learning disabilities during mathematical problem solving: A think-aloud analysis. *Journal of Learning Disabilities, 44*(6), 508–520. https://doi.org/10.1177/0022219410378445

Newman, D. L., Snyder, C. W., Fisk, J. N., & Wright, L. K. (2016), Development of the Central Dogma Concept Inventory (CDCI) Assessment Tool. *CBE-Life Sciences Education, 15*(2). https://doi.org/10.1187/cbe.15-06-0124

Nkambou, R., Bourdeau, J., & Mizoguchi, R. (2010). *Advances in Intelligent Tutoring Systems*. Springer.

Noh, T., Jeon, K., & Huffman, D. (2005). The Effects of Thinking Aloud Pair Problem Solving on High School Students' Chemistry Problem-Solving Performance and Verbal Interactions. *Journal of Chemical Education, 82*(10), 1558. https://doi.org/10.1021/ed082p1558

Osorio, T. H., & Reyes, M. G. (2023). Decision Making in Moral Judgment Context is Modulated by Individual Metacognition. *Psychological Reports*. https://doi.org/10.1177/00332941231191067

Pintrich, P., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. University of Michigan.

Porter, L., Zingaro, D., Liao, S. N., Taylor, C., Webb, K. C., Lee, C., & Clancy, M. (2019). BDSI: A Validated Concept Inventory for Basic Data Structures. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. ICER '19, (pp. 111–119). Association for Computing Machinery.

Ramachandran A., Huang C.-M., Gartland E., & Scassellati B. (2018). Thinking aloud with a tutoring robot to enhance learning. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction,* pp. 59–68. https://doi.org/10.1145/3171221.3171250

Ratten, V., & Jones, P. (2023). Generative artificial intelligence (ChatGPT): Implications for management educators. *The International Journal of Management Education, 21*(3), 100857. https://doi.org/10.1016/j.ijme.2023.100857

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Reinhart, A., Evans, C., Luby, A., Orellana, J., Meyer, M., Wieczorek, J., Elliott, P., Burckhardt, P., & Nugent, R. (2022). Think-Aloud Interviews: A Tool for Exploring Student Statistical Reasoning. *Journal of Statistics and Data Science Education, 30*(2), 100–113. https://doi.org/10.1080/26939169.2022.2063209

Rescher, N. (2014). *A system of pragmatic idealism, volume II: The validity of values, a normative theory of evaluative rationality*. Princeton University Press.

Rombout, F., Schuitema, J. A., & Volman, M. L. L. (2022). Teaching strategies for value-loaded critical thinking in philosophy classroom dialogues. *Thinking Skills and Creativity, 43*, 100991. https://doi.org/10.1016/j.tsc.2021.100991

Salmi, A. (2022). Pelaksanaan Strategi Thinking Aloud Pairs Problem Solving (TAPPS) untuk Meningkatkan Aktivitas dan Hasil Belajar Matematika Siswa Kelas XI IPA 1 MAN 2 Pesisir Selatan pada Materi Limit Fungsi Aljabar. *Journal on Education, 5*(1), 11–28. https://doi.org/10.31004/joe.v5i1.55

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113–125. https://doi.org/10.1023/A:1003044231033

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1–39. https://doi.org/10.2190/EC.49.1.a

Schroeder, N. L., & Gotch, C. M. (2015). Persisting Issues in Pedagogical Agent Research. *Journal of Educational Computing Research, 53*(2), 183–204. https://doi.org/10.1177/0735633115597625

Schroeder, N. L., Romine, W. L., & Craig, S. D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. *Computers & Education, 109*, 176–186. https://doi.org/10.1016/j.compedu.2017.02.015

Setiawan, J., Sudrajat, A., Aman, & Kumalasari, D. (2021). Development of Higher Order Thinking Skill Assessment Instruments in Learning Indonesian History. *International Journal of Evaluation and Research in Education, 10* (2), 545–552. https://doi.org/10.11591/ijere.v10i2.20796

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education, 188*, 104564. https://doi.org/10.1016/j.compedu.2022.104564

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Sultan, A. A., & Alasif, H. (2021). The Effect of Thinking Aloud Pair Problem Solving (TAPPS) Strategy on Developing Scientific Concepts and Habits of Mind among Middle School Students. *Psychology and Education Journal, 58*(5), 7149–7169. http://psychologyandeducation.net/pae/index.php/pae/article/view/6878/5686.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging, 35*(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

Tedesco-Schneck, M. (2018). Use of script concordance activity with the think-aloud approach to foster clinical reasoning in nursing students. *Nurse Educator, 44*(5), 275–277. https://doi.org/10.1097/NNE.0000000000000626

Thurzo, A., Strunga, M., Urban, R., Surovková, J., & Afrashtehfar, K. I. (2023). Impact of artificial intelligence on dental education: A review and guide for curriculum update. *Education Sciences, 13*(2), 150. https://doi.org/10.3390/educsci13020150

Tsai, T.-C., Harasym, P. H., Coderre, S., McLaughlin, K., & Donnon, T. (2009). Assessing ethical problem solving by reasoning rather than decision making. *Medical Education, 43*(12), 1188–1197. https://doi.org/10.1111/j.1365-2923.2009.03516.x

Umar, U., Hasratuddin, H., & Surya, E. (2022). Pengembangan LKPD Berbasis Model Think Aloud Pair Problem Solving Untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Siswa SD Negeri 067248 Medan. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 6*(3), 3402–3416. https://doi.org/10.31004/cendekia.v6i3.1884

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical approach to modelling cognitive*. Academic Press.

Veatch, R. M., Haddad, A., & Last, E. J. (2017). A model for ethical problem solving. In *Case Studies in Pharmacy Ethics* (3rd ed., pp. 19–28). https://doi.org/10.1093/med/9780190277000.003.0002

Vernim, S., Bauer, H., Rauch, E., Ziegler, M. T., & Umbrello, S. (2022). A value sensitive design approach for designing AI-based worker assistance systems in manufacturing. *Procedia Computer Science, 200*, 505–516. https://doi.org/10.1016/j.procs.2022.01.248

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*(3), 477–511. https://doi.org/10.1007/S10648-016-9361-7

Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J. M. (2021). Argue Tutor: An adaptive dialog-based learning system for argumentation skills. In Proceedings of the *2021 CHI conference on human factors in computing systems* (pp. 1–13). https://doi.org/10.1145/3411764.3445781

Wang, L. C., & Chen, M. P. (2010). The effects of game strategy and preference-matching on flow experience and programming performance in game-based learning. *Innovations in Education and Teaching International, 47*(1), 39–52. https://doi.org/10.1080/14703290903525838

Wardana, S., & Sagoro, E. M. (2019). Implementasi Gamifikasi Berbantu Media Kahoot Untuk Meningkatkan Aktivitas Belajar, Motivasi Belajar, Dan Hasil Belajar Jurnal Penyesuaian Siswa Kelas X Akuntansi 3Di Smk Koperasi Yogyakarta Tahun Ajaran 2018/2019. *Jurnal Pendidikan Akuntansi Indonesia, 17*(2), 46–57. https://doi.org/10.21831/jpai.v17i2.28693

Wei, C., Hung, I., Lee, L., & Chen, N. (2011). A joyful classroom learning system with robot learning companion for children to learn mathematics multiplication. *Turkish Online Journal of Educational Technology*. https://files.eric.ed.gov/fulltext/EJ932221.pdf. Accessed 1 Sept 2023

Whimbey, A., & Lochhead, J. (1999). *Problem solving and comprehension*. Lawrence Erlbaum Associates.

Yilmaz, F. G. K., & Yilmaz, R. (2019). Impact of pedagogic agent-mediated metacognitive support towards increasing task and group awareness in CSCL. *Computers & Education, 134*, 1–14. https://doi.org/10.1016/j.compedu.2019.02.001

Yilmaz, R., & Yilmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence, 4*, 100147. https://doi.org/10.1016/j.caeai.2023.100147

Zimmerman, B. J., & Schunk, D. H. (2008). Motivation: An essential dimension of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 1–30). Lawrence Erlbaum Associates Publishers.

Zohny, H., McMillan, J., & King, M. R. (2023). Ethics of generative AI. *Journal of Medical Ethics, 49*, 79–80. https://doi.org/10.1136/jme-2023-108909