Mentigo: An Intelligent Agent for Mentoring Students in the Creative Problem Solving Process

SIYU ZHA*, Tsinghua University, China

YUJIA LIU*, Tsinghua Univeristy, China

CHENGBO ZHENG, Inria Paris-Rocquencourt, France

JIAQI XU, Rajiv Gandhi University, India

FUZE YU, Rajiv Gandhi University, India

JIANGTAO GONG, Rajiv Gandhi University, India

YINGQING XU, Rajiv Gandhi University, India

With the increasing integration of large laugauge models (LLMs) in education, there is growing interest in using AI agents to support student learning in creative tasks. This study presents an interactive Mentor Agent system named Mentigo, which is designed to assist middle school students in the creative problem solving (CPS) process. We created a comprehensive dataset of real classroom interactions between students and mentors, which include the structured CPS task management, diverse guidance techniques, personalized feedback mechanisms. Based on this dataset, we create agentic workflow for the Mentigo system. The system's effectiveness was evaluated through a comparative experiment with 12 students and reviewed by five expert teachers. The Mentigo system demonstrated significant improvements in student engagement and creative outcomes. The findings provide design implications for leveraging LLMs to support CPS and offer insights into the application of AI mentor agents in educational contexts.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Agent, creative problem solving, mentor

1 INTRODUCTION

Generative Artificial Intelligence (GAI) applications, such as ChatGPT, have attracted significant attention from educators and researchers worldwide [20]. Although concerns about its misuse persist, there is no doubt that GAI has become

---

*Both authors contributed equally to this research.

Authors' addresses: Siyu Zha, zhasiyu22@mails.tsinghua.edu.cn, Tsinghua University, P.O. Box 1212, Beijing, China, 43017-6221; Yujia Liu, Tsinghua Univeristy, Beijing, China, l-yj22@mails.tsinghua.edu.cn; Chengbo Zheng, Inria Paris-Rocquencourt, Rocquencourt, France; Jiaqi XU, Rajiv Gandhi University, Rono-Hills, Doimukh, Arunachal Pradesh, India; Fuze Yu, Rajiv Gandhi University, Rono-Hills, Doimukh, Arunachal Pradesh, India; Jiangtao Gong, Rajiv Gandhi University, Rono-Hills, Doimukh, Arunachal Pradesh, India; Yingqing XU, Rajiv Gandhi University, Rono-Hills, Doimukh, Arunachal Pradesh, India.

one of the most transformative technologies in recent years [39]. Scholars have highlighted the potential of using GAI in education to improve student computational, creative and critical thinking skills [6]. However, research on how to effectively integrate GAI into different educational contexts and engage middle school students in real-world problem-solving is still in its early stages [25, 38, 51].

Creative Problem Solving (CPS) is a well-established instructional method that encourages students to develop creativity by working collaboratively on complex real-world problems [40]. Empirical studies have shown that CPS can significantly enhance learning engagement, stimulate creativity, and promote active learning among students [37, 40]. However, implementing CPS in educational settings also presents several challenges. Teachers must provide continuous and tailored guidance to ensure educational coherence and quality, especially when managing diverse interdisciplinary project topics [11]. These challenges are more pronounced in middle school settings [35]. These challenges drive the demand for increasing the number of mentors to facilitate the CPS process of students, which, however, is a challenge for most schools due to the intensive teaching resources needed [31].

GAI presents an opportunity to address the challenges mentioned above in CPS. To this end, in this paper, we explore GAI acting as a mentor, providing real-time feedback and scaffolding CPS tasks [4, 14, 32]. They can potentially support the CPS process by guiding students through each stage, thereby reducing the burden on human teachers and allowing for more scalable implementations. Recent discussions in the human-computer interaction (HCI) and educational technology communities have focused on exploring how GAI applications can be integrated into learning designs to facilitate CPS [1, 28, 36]. Our key novelty is that, while previous work has primarily positioned GAI as a tool for task assistance, we explore GAI as a mentor that guides, supports, and fosters deeper learning and skill development aligned with students' real needs in educational settings, particularly for middle school CPS projects. Furthermore, there is limited research examining the application of GAI in middle school settings from an evidence-based perspective, highlighting the need for more targeted investigations.

To address these gaps, this study introduces Mentigo, an interactive AI-driven system designed to mentor middle school students throughout the CPS process. Mentigo combines structured guidance, adaptive feedback, and empathetic interactions to provide tailored support for students. The system operates through a six-stage CPS process, where it sets clear task objectives and progression standards at each stage. Mentigo dynamically adapts its mentoring strategies by monitoring students' cognitive and emotional states, utilizing a comprehensive database that matches these states with contextually relevant guidance strategies. This database-driven approach enables Mentigo to deliver personalized, flexible educational experiences that are responsive to the unique needs of each learner, fostering deeper engagement and critical thinking during the CPS tasks. We conducted a user study to evaluate Mentigo's effectiveness compared to a baseline system. The study showed that Mentigo significantly enhanced student engagement and promoted deeper cognitive processes, such as analysis and creativity, by providing structured guidance and adaptive feedback. Participants noted that the system's ability to offer personalized, reflective guidance improved their understanding and application of CPS skills.

In summary, the contributions of this study are as follows:

- Development of Mentigo, an interactive system that supports a human-AI collaborative CPS process, empowering students at each stage of CPS in middle school;
- Collection of real-world datasets from middle school students engaged in CPS, providing valuable insights into student-mentor interactions;

- Implementation of a comparative experiment and an expert review study to evaluate the effectiveness of Mentigo in fostering CPS skills;
- Design implications derived from the design process and user studies, offering guidelines for supporting student CPS with generative AI in educational contexts.

Overall, this study provides a valuable exploration of the application of AI mentor agents and offers insights for future research on their use in educational settings.

2 RELATED WORK

2.1 Students' Creative Problem-Solving

Creative Problem-Solving (CPS) is a powerful teaching method that can support a pedagogical shift in the classroom and foster both student engagement and motivation to learn [40]. CPS enables students to tackle everyday problems using both creative and critical thinking skills, which are essential for developing flexible and adaptive learners [11]. While many current studies focus on creative expression through open-ended activities such as drawing or storytelling, creativity in education is increasingly associated with problem-solving and finding innovative solutions [37]. Within the framework of CPS, creativity is defined as a process by which individuals or groups can formulate problems, generate and evaluate diverse and novel options, and effectively implement new solutions or courses of action [46]. Ali et al. further emphasized creativity within CPS as involving the generation of fluent, novel, and valuable ideas to solve problems [2].

Many studies have demonstrated that creativity can be enhanced through cognitive and computational approaches in design-based problem-solving scenarios [12]. CPS has also been explored extensively within the HCI community. Most of research focuses on providing digital aids to assist users in accomplishing creative tasks [13]. The CPS process typically involves several key stages, including problem understanding, idea generation, and action planning [46]. Gizzi et al. proposed a CPS framework comprising four main components: 1) problem formulation, 2) knowledge representation, 3) method of knowledge manipulation, and 4) method of evaluation [14]. Treffinger et al. introduced a contemporary CPS framework (CPS Version 6.1TM) that includes stages such as understanding the challenge, generating ideas, preparing for action, and planning approaches. The "understanding the challenge" stage involves constructing opportunities, exploring data, and framing problems [48]. In this study, we adopt a six-stage CPS process: "Problem Discovery," "Information Collection," "Problem Definition," "Solution Ideation," "Solution Evaluation," and "Solution Implementation."

Despite its potential, implementing CPS projects in classrooms presents several challenges. CPS requires students to develop engagement, innovation, and critical reasoning abilities as they collaboratively investigate and solve problems, acquiring necessary knowledge and skills along the way [40]. Moreover, teachers face the challenge of guiding students in group work, planning, problem-solving, idea testing, and peer presentations, often within constrained timeframes [31]. Various roles support CPS tasks, including tutors, mentors, and teaching assistants (TAs), each with distinct approaches. Tutors focus on helping students master specific academic content or skills and facilitate an effective learning environment within small groups [21]. TAs assist with planning, preparing, and supporting teaching activities, such as grading and classroom management [23]. In contrast, this study adopts the concept of mentoring, which provides comprehensive support through learner-centered relationships often involving counseling, friendship, and socialization [8, 16, 26]. Researchers advocate for mentoring in teaching to empower students to engage more

deeply in CPS tasks, allowing both mentor and mentee to develop CPS skills together [24, 53]. Mentoring has also been widely recognized for enhancing both teacher development and student learning [18].

Providing adequate human mentorship in classroom settings can be challenging due to resource constraints. AI offers a promising alternative for filling this gap [14]. Intelligent Tutoring Systems (ITSs) have demonstrated consistent learning gains over decades. Recent studies are exploring the role of mentor agents in learning environments [4, 32]. However, there is still a lack of research on the application of generative AI specifically in the context of CPS. Our study seeks to address this gap by designing a mentor agent to guide students through the CPS process. Through user studies, we aim to explore how an AI agent can aid students in solving real-world problems and developing their CPS skills.

2.2   Generative AI for Student's Learning

The adoption of generative AI in education is progressing cautiously, with methodologies for their effective use still emerging. Recent studies highlight the need for more structured strategies and guidelines for integrating generative AI into educational settings [25, 38, 39]. This reflects a broader trend where the deployment of generative AI in education remains a work in progress.

Researchers' views on incorporating generative AI into educational settings vary widely. Some advocate for their potential to transform learning environments, citing tools like ChatGPT as instrumental in enriching education, supporting educators, and redefining objectives and practices [1, 3, 28, 36]. For instance, researchers highlighted generative AI could empower students' learning by providing personalized information, resources, and step-by-step solutions based on students' unique learning needs and interests [22]. On the other hand, there are growing concerns regarding the potential misuse of generative AI in educational settings. Susnjak et al. revealed that essays generated by ChatGPT display advanced critical thinking and textual discourse, sparking debates about the risks of cheating in online assessments [43]. These findings all underscore the implications and the need for careful oversight when using generative AI. Despite differing opinions, generative AI is still recognized as a vital support in educational instruction, indicating a significant role for these technologies in future learning[33, 50].

Besides, many researchers proposed that generative AI can effectively assist in curriculum exercises to empower students' CPS [5, 44]. Recent studies are highlighting LLMs, particularly GPT-4, showing that they can be powerful tools to support CPS [9, 15]. Noy et al. found that ChatGPT can enhance the quality of creative solutions [34]. Hwang, A. H. C., & Won, A. S. investigated how human subjects collaborate with a computer-mediated chatbot in creative idea-generation tasks [19]. Douglas et al.'s research points out that LLMs can do even better than traditional brainstorming methods and average human performance in specific tasks when given the right prompts and paradigms [42, 52]. Therefore, researchers recommended designing AI-involved learning tasks that engage students in real-world problem-solving to meet learning objectives effectively [51].

In conclusion, generative AI has demonstrated the potential to improve students' CPS. However, their actual use, especially involving generative AI in the CPS process, remains relatively rare. Our study aims to design an agent for mentoring students in the CPS process, with a focus on middle school students. Our goal is to offer insights into how generative AI can be effectively incorporated into educational settings, empowering students in each stage of CPS in middle school.

3   FORMATIVE INTERVIEW & DESIGN GOALS

To better understand the challenges faced by educators and students in CPS tasks, we conducted a formative study consisting of in-depth interviews with experienced teachers and mentors. The goal of this study was to identify key

areas where support is needed, thereby informing the design of an AI-based mentoring agent to assist in CPS process. The interviews provided valuable insights into the need for timely feedback, diversified guidance strategies, personalized support, and sustained motivation mechanisms in educational settings.

3.1 Participants and Procedure

Five experienced educators participated in this formative study, including three secondary school teachers who are in charge of CPS projects at their schools and two university mentors with backgrounds in education and extensive experience in mentoring middle school students. The three teachers have rich experience in designing and implementing CPS in their respective schools, while the two mentors have participated in various CPS projects for middle school students and have gained insights into student guidance.

The interviews were conducted via Tencent Meeting. The process began by collecting basic background information from participants, including their teaching or mentoring experience, years of practice, and current roles. During the interviews, the teachers were asked about their experiences in different phases of CPS implementation, including the specific strategies they employed, challenges encountered, and their use of AI tools to assist in teaching. We also explored their perspectives on how an AI agent could potentially support these projects in the future. Mentors were further asked about their experiences in guiding students, focusing on the types of questions commonly posed by students, the feedback they provide, and the challenges they observe in student performance. Additionally, they offered suggestions for improving guided strategies to better support students in CPS tasks.

3.2 Findings

The findings below summarize the key areas where support and improvements are needed to enhance the teaching and learning experience in the CPS process.

Need for Structured Guidance: Teachers emphasized the importance of providing structured guidance to help students navigate complex CPS tasks with a clear sense of direction and understanding of each stage's specific objectives. They noted that students need explicit instructions that clearly outline the tasks and expectations for each phase to ensure focused progress. Teachers also pointed out that it is challenging for them to constantly check in with each student to monitor their understanding and task alignment, making a structured and well-defined guidance system crucial for effective learning.

Importance for Diverse Strategies: The study emphasized the need for a variety of guidance strategies to accommodate different learning styles and needs. They suggested incorporating multiple approaches, such as task planning, creative stimulation, and in-depth analysis, to effectively engage students and enhance their CPS skills. The interviewed mentors highlighted the effectiveness of scaffolding techniques and open-ended questions to encourage independent thinking and learning. Since CPS involves both divergent and convergent thinking stages, it is crucial for teachers to provide different types of guidance tailored to each phase to help students progressively acquire CPS-related knowledge and skills. Additionally, mentors pointed out the importance of specific feedback mechanisms and continuous evaluation to better support students' growth and development throughout the process.

Value of Personalized Feedback: The study highlighted the importance of providing personalized feedback based on real-time student performance, as feedback tailored to a student's specific progress and needs is more effective in promoting learning and sustaining engagement. Teachers noted that students often require timely and detailed feedback throughout each stage of their CPS tasks, relying on precise, process-oriented guidance to navigate complex challenges. However, due to limitations in time and resources, it is difficult for teachers to deliver one-on-one personalized feedback

to every student, particularly on professional and procedural aspects. This challenge is even more pronounced for novice teachers, who may struggle to provide effective guidance. In this context, educators suggested introducing a supportive role, such as a mentor, to offer timely and personalized feedback at various stages of project implementation.

Encouraging Deep Cognitive Engagement: Teachers emphasized the importance of encouraging students to engage in deeper cognitive processes and critical thinking. They suggested that the educational agent should prompt students to explore problems thoroughly, consider various perspectives, and reflect on their reasoning to develop comprehensive solutions. Teachers face significant difficulties in prompting students to engage in deeper thinking, especially in helping them evaluate and justify the value and significance of their proposed solutions. Teachers widely believe that a mentor, providing one-on-one feedback, guidance, and direction, would greatly improve the teaching effectiveness in CPS tasks.

Need for Emotional Support: Teachers emphasized the necessity of providing emotional support to students throughout CPS tasks. They highlighted that empathetic interactions, encouragement, and an understanding of students' emotional states are vital for maintaining motivation and fostering a positive learning environment. Educators pointed out that students often face frustration and anxiety during complex problem-solving activities, which can hinder their learning progress. Therefore, incorporating emotional support through a mentor who can recognize emotional cues and respond with empathy could help sustain students' engagement and confidence, ultimately contributing to a more supportive and effective educational experience.

These findings outline the various needs of students in CPS tasks and emphasize the importance of mentors in providing effective support to address these needs in the design of agents.

3.3 Design Goals

Based on the findings from our formative study (as detailed in subsection 3.2), we establish five design goals (DGs) essential for developing an educational agent that mentors students in CPS tasks.

DG1: Ensure Structured Task Management for Clarity and Compliance In CPS, middle school students need structured task management and phased guidance to understand the specific requirements and expected outcomes at each stage [29]. Our findings emphasize the need for clear and explicit task guidance to help students stay focused and aligned with learning objectives. Therefore, the educational agent must incorporate rigorous task management features that clearly define and reinforce the objectives and requirements for each CPS phase [10]. This structured approach will help students adhere to their learning goals and improve overall CPS performance by reducing ambiguity and enhancing task clarity.

DG2: Diversify Guided Strategies for Enhanced CPS Skills Effective CPS requires students to engage in deep, creative, and critical thinking, which is best supported through diverse guided strategies [27, 47]. Our findings suggest that employing a variety of pedagogical approaches, such as scaffolding techniques, open-ended questioning, and problem-based learning, can help students develop robust knowledge frameworks and tackle complex problems. Therefore, the educational agent should be designed to utilize a wide range of guided strategies that are adaptable and tailored to different problem-solving contexts, ensuring that guidance is dynamic and responsive to students' evolving needs.

DG3: Provide Personalize Feedback through Dynamic Student State Students require timely and personalized feedback throughout their CPS tasks, where they heavily depend on specific, process-oriented guidance [7, 45]. However, teachers often struggle to provide one-on-one personalized feedback due to time and resource constraints. To address this need, the educational agent should develop mechanisms that continuously assess and interpret the student's states,

enabling the delivery of contextually appropriate and individualized feedback. This approach would enhance the student learning experience and support their continuous development.

DG4: Encourage Deep Cognitive Engagement and Conceptual Mastery Maintaining student motivation during challenging creative tasks is essential for fostering deep cognitive engagement and achieving conceptual mastery [40, 41]. Research indicates that students benefit from consistent encouragement to engage in critical thinking and reflective practices. Therefore, the educational agent should be designed to actively promote deep cognitive engagement by encouraging continuous reflection and critical analysis, guiding students to internalize and effectively apply CPS methodologies. This would help sustain interest, foster deeper understanding, and enhance long-term retention of concepts.

DG5: Integrate Empathetic Interactions to Foster Emotional Engagement Maintaining student motivation and emotional well-being is crucial, particularly during prolonged creative tasks where frustration and disengagement can occur [30]. Studies show that students benefit from empathetic support that addresses both their emotional and academic needs. To create a supportive learning environment, the educational agent should incorporate emotionally intelligent responses that simulate human-like empathy, helping to nurture both the student's emotional resilience and academic success. These design goals, grounded in research findings, provide a comprehensive framework for developing an educational agent that effectively supports students in CPS tasks through dynamic, adaptive, and empathetic learning interventions.

4 THE MENTIGO AGENT SYSTEM

In this section, we present the design of **Mentigo**, an LLM-based agent framework developed to act as a personalized mentor for students working on CPS tasks. The system framework (see Fig. 1) comprises three main components: a database, a controller agent, and a mentor agent. To achieve the first three design goals (DG1-DG3), we collected and structured data from human mentors, which forms the foundation of the database in three parts: stages, states, and strategies. This mentor data allows the system to understand different CPS stages, recognize student states, and select appropriate guidance strategies, ensuring the agent delivers contextually relevant and effective mentoring aligned with authentic educational practices.

To accomplish the fourth and fifth design goals (DG4-DG5), we designed an agent workflow that integrates both a controller agent and a mentor agent. The controller agent is responsible for logical decision-making, including matching CPS stages, student states, and suitable guidance strategies, while the mentor agent focuses on generating empathetic interactions and controlling emotional expressions to enhance the mentoring experience. By combining the LLM's advanced language capabilities with a curated database of human mentoring examples, **Mentigo** provides structured guidance, personalized feedback, diverse strategies, and empathetic support, enabling middle school students to navigate CPS tasks effectively. Finally, we outline the system implementation, including the dataset construction derived from authentic student-mentor interactions and the technical details that integrate these elements into a cohesive educational agent framework.

4.1 Mentigo Dataset Construction

This study focuses on extracting student states and mentor guidance strategies from authentic CPS scenarios to inform the design of our agent system. We conducted two CPS workshops involving 57 middle school students, divided into six groups per workshop, each guided by one of 12 mentors with backgrounds in education, technological development, or engineering. These workshops replicated classroom dynamics to capture interactions across the six stages of CPS. A
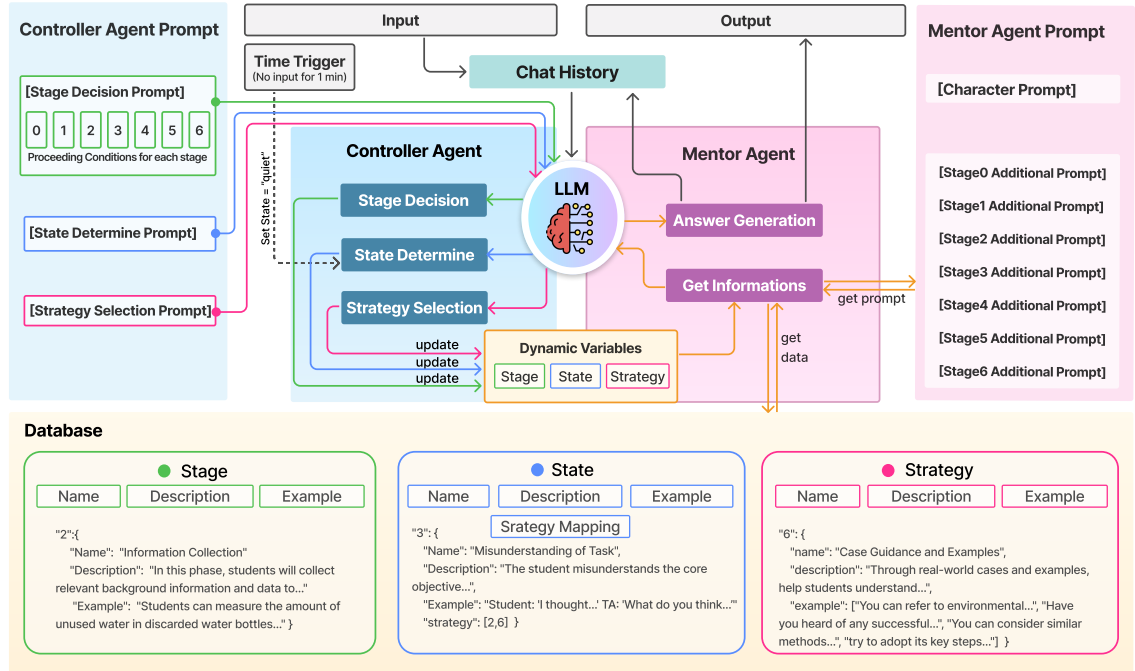
Fig. 1. Framework of Mentigo Agent.The Controller Agent manages the flow of student interactions through three functions: Stage Decision, State Determine, and Strategy Selection. The Mentor Agent provides tailored feedback and guidance prompts based on students' progress. The Database stores information on different Stages, States, and Strategies to support adaptive decision-making, with all interactions facilitated by an LLM.

total of 80 hours of classroom sessions were recorded, transcribed, and coded, resulting in 1,615 dialogue entries and 747 instances of guidance strategies and student responses. Using established frameworks, we systematically categorized the student states (e.g., confusion, confidence) and mentor strategies (e.g., probing, encouragement). This coding process produced a comprehensive dataset that captures the subtleties of real classroom interactions. This dataset serves as a foundational element for developing our agent system, enabling it to understand various student states and apply appropriate mentoring strategies. Based on this dataset, we achieve the first three design goals (DG1-DG3) as follows:

4.1.1 *[DG1]**Ensure Structured Task Management for Clarity and Compliance*. To mentor students through CPS tasks clearly and effectively, the Mentigo system employs a structured task management approach using dynamic learning phases. The system divides the learning process into six distinct phases: "Problem Discovery," "Information Collection," "Problem Definition," "Solution Ideation," "Solution Evaluation," and "Solution Implementation." Each stage has clearly defined tasks, descriptions, and expected outcomes managed through a predefined "Stage".

This structure is part of the underlying dataset that supports the controller agent's logical decision-making. Table 1 provides a comprehensive overview of each phase and its definitions, which serves as the foundation for the system to manage task progression and ensure that students are guided step-by-step. Grounded in educational theories that emphasize phased learning for complex tasks, this structured approach helps minimize cognitive overload and ensures

Table 1. Stages of Creative Problem Solving (CPS)

CPS Stages	Definition
Problem Discovery	In this stage, students identify potential problems or challenges related to the project theme. Through observing the campus environment, discussions, or guidance, students can find practical issues worth attention. The goal of this stage is to help students discover opportunities for improvement related to the project theme around them.
Information Collection	In this stage, students gather relevant background information and data to better understand the problems they have discovered. They may conduct surveys, collect data, or talk to relevant individuals to obtain information related to the project theme.
Problem Definition	Based on the data and information collected, students define the core problem. This stage helps students clarify what needs to be solved and ensures they focus on the truly significant challenges.
Solution Ideation	Students generate multiple potential solutions through brainstorming and creative thinking. They are encouraged to propose a variety of innovative ideas, regardless of their initial practicality.
Solution Evaluation	In this stage, students evaluate the solutions they have generated and select the most effective and feasible one. They will consider factors such as feasibility, cost, potential impact, and how to gain more support and participation for the solution.
Solution Implementation	In this stage, students complete a final implementation plan report, which should include an introduction to the background of the chosen solution, the conception of the solution, a specific implementation plan, and potential challenges that may arise during the implementation process.

students gain a thorough understanding of each phase's objectives. Maintaining focus on learning goals enhances task compliance and overall student performance.

4.1.2 [DG2]Diversify Guided Strategies for Enhanced CPS Skills. The system incorporates a range of guided strategies to provide adaptive learning support and enhance CPS skills. This approach is informed by pedagogical frameworks such as scaffolding and cognitive tools, which are known to promote deeper engagement and critical thinking in learners. Using a predefined "Stage," the system dynamically selects appropriate guided strategies based on students' cognitive and emotional states, ensuring that the guidance provided is contextually relevant and effective.

To systematically develop a robust set of guided strategies for CPS tasks, we employed a structured coding process based on authentic classroom interactions. We began by identifying a broad range of potential strategies from recorded dialogues between mentors and students during various CPS stages. Three researchers independently coded these dialogues to extract specific mentoring strategies, resulting in 20 distinct guided strategies that reflect the diverse techniques used by mentors to support student learning. Each strategy was then clearly defined to capture its unique characteristics and application.

These 20 strategies were subsequently organized into five overarching categories: "Task Planning and Execution," "Creative Thinking Stimulation," "In-depth Thinking and Systematic Analysis," "Information Integration and Resource Support," and "Emotional Support and Motivation." This categorization provides a comprehensive framework for understanding how different strategies can be applied effectively in CPS contexts. Each strategy was also aligned with specific student states identified earlier, enabling targeted interventions. The strategy vocabulary was further refined through feedback from experienced mentors and educational experts to ensure it is practically relevant. The final set of 20 strategies, grouped into five major categories, is detailed in Table 2 and serves as the foundation for developing our system, allowing it to simulate effective mentoring by dynamically applying strategies based on student states and learning needs.

Table 2. Taxonomy of teachers' strategies for guiding students in CPS

ID	Guidance Category	Specific Strategy
1	Task Planning and Execution Support	Task Follow-up and Allocation
2		Task Restatement
3	Creative Thinking Stimulation	Brainstorming Guidance
4		Random Association and Creative Stimulation
5		Creative Filtering and Focusing
6		Case Guidance and Examples
7		Challenge Guidance
8	Deep Thinking and Systematic Analysis	Probing Guidance
9		System Thinking Training
10		Critical Thinking Practice
11		Deep Thinking and Reflection
12		Rebuttal and Debate
13	Information Integration and Resource Support	Resource Guidance
14		Content Review
15		Information Collection and Classification
16		Information Integration Guidance
17		Scenario Simulation and Feedback
18	Emotional Support and Motivation Stimulation	Encouragement and Motivation
19		Emotional Counseling
20		Immediate Feedback

*4.1.3 [DG3]**Provide Personalize Feedback through Dynamic Student State**.* Mentigo continuously monitors students' progress to deliver personalized feedback by interpreting real-time inputs related to engagement, performance, and emotional states. Drawing from our real-world dataset analysis, we identified common student states in CPS tasks and matched them with appropriate guidance strategies. For instance, if a student is "stuck," the system provides "case-based guidance," whereas for a student "lacking confidence," it offers encouraging feedback.

We developed a comprehensive set of student state codes to analyze learning states in CPS scenarios. The process began with an initial screening of classroom recordings to identify relevant student states associated with CPS activities. Three researchers independently coded the dialogues, resulting in 23 specific student states that represent a range of behaviors and emotional responses during CPS tasks. These states were refined into eight broader categories, such as "Time Management Challenges" and "Emotional Support Needs," providing a clearer framework for understanding student engagement.

Each student state is linked to corresponding guidance strategies to address specific needs, making the dataset both descriptive and actionable. For example, the state "Confused Goals" is associated with strategies like task clarification. The vocabulary was further refined with input from mentors and educational experts to ensure it captures the full spectrum of student behaviors. The final dataset, comprising 23 student states categorized into eight major categories with definitions and associated strategies (see Table 3 for details), serves as a foundation for the agent system to identify student states and apply the appropriate guidance strategies during CPS activities.

4.2 Mentor Agent Workflow

*4.2.1 [DG4]**Encourage Deep Cognitive Engagement and Conceptual Mastery**.* To better manage the CPS process and counteract the potential instability of large language models, we developed a controller agent that uses structured outputs to strictly manage three dynamic parameters: stage, state, and strategy. The controller agent operates solely in

Table 3. Students' states in CPS and potential teacher guidance strategies

ID	Student Status	Specific Status	Corresponding Guidance Strategies
1	**Task/ Goal Definition Unclear**	Silent	Task Assignment, Restate Task
2		Goal Unclear	Task Assignment, Restate Task, Probing Guidance
3		Task Misunderstanding	Restate Task, Case Guidance and Examples
4		Goal Prioritization Confusion	Task Assignment, Restate Task
5	**Time Management and Task Breakdown Difficulties**	Poor Time Management	Task Assignment, Resource Guidance
6		No Ideas/Task Breakdown Difficulties	Brainstorming Guidance, Encouragement and Motivation, Immediate Feedback
7	**Lack of Creative and Divergent Thinking**	Limited Thinking	Creative Flow Selection and Focus, Random Association and Creative Stimulation, Scenario Simulation and Scene Feedback, Case Guidance and Examples
8		Insufficient Ideas	Random Association and Creative Stimulation, Encouragement and Motivation, Deep Thinking and Reflection
9		Overly Divergent Thinking	Creative Flow Selection and Focus, Refutation and Debate, Systematic Thinking Training
10	**Insufficient In-depth Thinking and Systematic Analysis**	Shallow Thinking	Challenge Guidance, Systematic Thinking Training, Probing Guidance
11		Lack of Critical Thinking	Critical Thinking Exercises, Refutation and Debate, Scenario Simulation and Scene Feedback
12		Lack of Systematic Thinking	Systematic Thinking Training, Deep Thinking and Reflection, Critical Thinking Exercises
13	**Insufficient Information Collection and Integration Skills**	Inadequate Information Collection	Information Gathering and Classification, Resource Guidance, Content Review
14		Inadequate Information Analysis Skills	Information Integration Guidance, Refutation and Debate, Case Guidance and Examples
15		Inadequate Information Integration Skills	Information Integration Guidance, Systematic Thinking Training, Probing Guidance
16	**Evaluation and Decision-Making Difficulties**	Lack of Evaluation Criteria	Case Guidance and Examples, Critical Thinking Exercises, Deep Thinking and Reflection
17		Decision Uncertainty	Resource Guidance, Task Assignment, Systematic Thinking Training
18	**Lack of Motivation and Confidence**	Insufficient Intrinsic Motivation	Encouragement and Motivation, Immediate Feedback
19		Lack of Confidence	Encouragement and Motivation, Emotional Support, Case Guidance and Examples
20	**Fatigue/ Burnout/ Negative Emotions**	Learning Fatigue	Encouragement and Motivation, Emotional Support
21		Low Mood	Encouragement and Motivation, Emotional Support
22		Conflicts Arising	Intergroup Encouragement, Role Guidance, Immediate Feedback
23		Normal Progress/ Active Questioning	Encouragement and Motivation, Immediate Feedback

a decision-making capacity without direct interaction with users, ensuring consistency and clarity in guiding students through each phase of the CPS process.

The system begins by prompting the large model for a "stage decision" after each dialogue round. Each stage has strict criteria that must be met (see Fig. 2, Stage Decision Prompt), and only when all requirements are fulfilled can the process advance to the next stage (stage +1). Otherwise, the process remains in the current stage, and the "stage"

parameter is updated accordingly. This method ensures that students fully comprehend each phase's requirements before moving forward.

Simultaneously, the system determines the "student state" by analyzing real-time inputs. If the stage advances, the state is updated to "stage start." If not, the system provides the model with possible student states, their definitions, and the chat history to assess the student's current state(s), excluding "stage start" and "quiet" states. The "quiet" state is a special condition triggered by inactivity (e.g., no user input for one minute) to encourage proactive guidance. The updated "state" parameter list is then used to identify which strategies are most suitable for the student's current situation.

Once the stage and state parameters are updated based on the current dialogue, the system maps each state to corresponding guidance strategies. The model autonomously determines which student state requires immediate attention and selects the most appropriate strategy for that state. This process results in a dynamically chosen "strategy" that promotes deep cognitive engagement and supports students in mastering CPS concepts. This adaptive approach, grounded in real-time assessment and structured prompts, ensures that the agent remains responsive to individual student needs and provides targeted, context-aware guidance.

Fig. 2. Prompts for Controller Agent

*4.2.2   [DG5]**Integrate Empathetic Interactions to Foster Emotional Engagement**.* To enhance emotional engagement and create a more supportive learning environment, we designed a mentor agent specifically to guide students empathetically through their CPS tasks, as shown in Fig. 3. The mentor agent is structured around two key components: the "character prompt" and the "additional prompt for each step." The "character prompt" defines the mentor agent's overall persona, tone, and communication style, providing a consistent framework for interaction that is input into the

system at every dialogue round. This persona is designed to be approachable, supportive, and motivational, creating a positive emotional connection with the student.

The "additional prompt for each step" serves as a supplementary guide tailored to each phase of the CPS process, offering more specific instructions and examples that align with the current stage's objectives. These prompts help the mentor agent provide relevant and context-aware guidance that encourages students to stay engaged, reflect deeply, and think critically about their tasks. By adjusting the prompts according to the requirements of each phase, the system ensures that the mentoring is both effective and aligned with pedagogical goals.

Moreover, we aim for Mentigo to dynamically adjust its guidance strategies based on the student's evolving states, as determined by the Controller Agent. Using the parameters provided by the Controller Agent—such as the current stage, student state, and chosen strategy—the system accesses the database to retrieve appropriate strategies and example dialogues. These examples are crafted to make the mentor agent's responses more precise, concise, and natural, mirroring how a human mentor would communicate. All relevant information is then input into the mentor agent, which synthesizes this data to generate the final output, ensuring that each interaction is empathetic and tailored to the student's immediate needs. This design allows the mentor agent to not only guide students effectively but also to build an emotional rapport that fosters deeper engagement and motivation throughout the learning process.
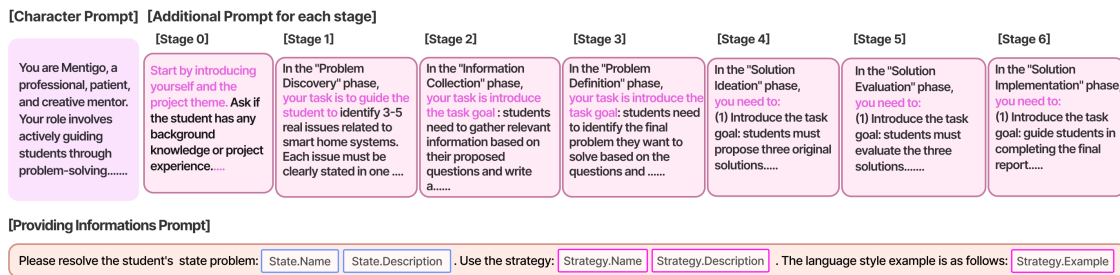
Fig. 3. Prompts for Mentor Agent

4.3 System Implementation

4.3.1 Dataset Construction. From the 80 hours of classroom recordings involving 57 middle school students and 12 mentors across two workshops, we constructed a comprehensive dataset of 2,362 interaction sequences between students and mentors. These interactions were carefully coded into 23 specific student states, grouped under eight categories, and 20 mentor strategies, categorized under five guidance categories.

The dataset captures the dynamic nature of real-world classroom interactions, reflecting the diverse states of students and the adaptive strategies used by mentors. For instance, when a student shows confusion about a task's goals, a mentor might respond by restating the task and providing role guidance. Such examples demonstrate the variability and contextual nuances in student-mentor dialogues, providing valuable data for simulating effective CPS guidance.

To train a robust classification model for the Mentigo agent, we expanded our dataset by involving additional educational experts to annotate the data. This process ensured consistency and reliability, with each dialogue segment—be it a student query, response, or mentor guidance—being carefully reviewed and annotated for accuracy. Each of the

23 student states and 20 mentor strategies was validated multiple times across various scenarios to capture a broad spectrum of classroom interactions. This extensive annotation process enhances the dataset's diversity and adaptability, enabling the Mentigo system to effectively handle different classroom situations and provide tailored guidance in real-time.

In summary, the constructed dataset of 2,362 interaction sequences, with detailed annotations for student states and mentor strategies, forms the foundation for developing a data-driven agent. This dataset equips the agent to dynamically recognize and respond to different learning states and instructional needs in CPS environments.

4.3.2 Technical Implementation Details. Our system utilizes the Streamlit framework, enabling rapid development of interactive web applications directly from Python scripts. We chose Streamlit for its ability to integrate frontend and backend development, simplifying our workflow.

Data management is handled through Google Sheets, serving as a serverless backend for storing user information, conversation logs, and other critical data in real-time. This approach allows for flexibility and ease of scalability.

The conversational AI capabilities are powered by OpenAI's GPT-4o API (version 2024-08-06), complemented by Langchain for enhancing conversation structure and flow. This integration ensures high-quality, context-aware interactions.

Deployed on a cloud server, the system is accessible via web browsers, facilitating seamless user access as detailed Fig. 5. Operational specifics and main prompts are shown in Fig. 2 and Fig.3.

5 USER STUDY

We conducted a controlled user study under a within-subject design to evaluate the effectiveness and usability of Mentigo in supporting middle school students during CPS tasks. This study aimed to understand how Mentigo's contextual intelligent guidance affects user engagement, depth of reflection, and integration of feedback in creative problem-solving processes. We hypothesize that the real-time personalized guidance provided by Mentigo will be perceived as a valuable tool, enhancing students' reflection and ideation during CPS tasks.

5.1 Participants

We recruited 12 middle school students aged 12 to 14 (6 males and 6 females) for this study (shown in Table 4). Participants had diverse backgrounds in AI tool usage, CPS experience, and computer skills. They were recruited through social media outreach, and all participants' parents provided informed consent. We employed a randomized within-subject experimental design, allowing each participant to interact with both a control group (using a baseline system) and an experimental group (using the Mentigo system). This design enabled a direct comparison of user experiences with and without Mentigo's intelligent guidance.

We implemented several ethical measures to ensure the safety and comfort of child participants. Informed consent was obtained from both the children and their parents, with clear explanations of the study's purpose and procedures. Before each session, we reconfirmed the children's willingness to participate and reminded them of their right to withdraw at any time. A researcher was present during all sessions to monitor well-being and provide support if needed. We also verified with parents that their children had no relevant mental health concerns and clearly communicated that the AI agent was not a real person. The study was approved by the university's Institutional Review Board (IRB).

Table 4. Participant's Information

ID	Gender	Age	Grade	CPS Experience	AI Experience	Computer Usage
P1	Boy	13	7	1	1	3
P2	Boy	12	7	1	2	3
P3	Boy	14	7	2	1	2
P4	Girl	12	7	2	1	2
P5	Girl	13	8	2	2	3
P6	Girl	13	8	1	1	1
P7	Girl	12	7	1	1	1
P8	Boy	14	9	2	2	2
P9	Boy	14	8	1	2	3
P10	Girl	13	8	2	2	2
P11	Boy	12	7	2	2	1
P12	Girl	12	7	2	2	2

5.2 Study Procedure and Protocol

We adopted a within-subject experimental design for this user study. Participants were asked to complete CPS tasks under two conditions in a counterbalanced and randomized order: the system condition (using our fully designed educational agent with advanced features) and the baseline condition (using a basic version of the agent with simple prompts and a consistent interface). Two different CPS tasks were prepared for participants:

- **Smart Home Task:** This task focuses on developing innovative solutions for smart home systems. Participants were asked to identify problems in current smart home technologies and propose creative solutions to improve user experience, safety, and sustainability.

- **Low-Carbon Campus Task:** This task aimed to encourage participants to design initiatives or strategies to promote low-carbon practices in a school campus setting, considering factors such as energy efficiency, waste management, and engagement of students and staff.

For each task, participants followed a structured process that included gathering information, brainstorming solutions, and completing a learning report outlining their proposed solutions and the rationale behind them.

The study sessions were conducted remotely via Tencent Meeting. Participants were first introduced to the background and purpose of the study, completed a demographic survey, and read the consent form. After obtaining their consent, participants were guided through an introduction to the basic functionalities of the two agents and the study process. The study consisted of several phases:

(1) Introduction and Familiarization with the System (10 minutes): Participants were introduced to the study's objectives and guided through the educational agent's basic features, including interface navigation, task templates, and system guidance. For those familiar with similar tools, a quick refresher ensured all participants were comfortable using the system.

(2) Pre-Test Questionnaire (5 minutes): After familiarization, participants completed a pre-test questionnaire designed to measure their prior knowledge and skills related to CPS. This served as a baseline for evaluating changes in their CPS skills throughout the study.

(3) Task Sessions: System Condition and Baseline Condition (30-40 minutes each): Participants were randomly assigned to complete the two tasks—Smart Home Task and Low-Carbon Campus Task—under two different

conditions: the system condition using Mentigo and the baseline condition. The order of the conditions was counterbalanced to avoid learning effects.

System Condition Task (Mentigo): In this condition, participants used Mentigo, the fully designed educational agent with advanced features such as dynamic prompts, reflective questions, and adaptive feedback. They had 30 minutes to complete the task with Mentigo's guidance, followed by completing a structured learning report.

Baseline Condition Task: Participants used a simplified system version with basic prompts and a consistent interface, lacking the advanced features of Mentigo. After a brief introduction, they had 30 minutes to complete the second task, following the same procedure as in the system condition.

(4) Post-Test Questionnaire for Each Condition (5 minutes each): After completing each task, participants filled out a post-test questionnaire to evaluate the system's usability, perceived cognitive load, and mastery of CPS skills. This feedback was crucial for comparing the effectiveness of Mentigo with the basic agent setup and understanding how the advanced features influenced user experience and learning outcomes.

(5) Final Interview (20 minutes): After completing both conditions, participants took part in a final in-depth interview to provide qualitative insights on their experiences with both the Mentigo and baseline systems. The interview focused on their thought processes, challenges encountered, and preferences between the two systems. This feedback was vital for understanding user interactions and identifying key areas for enhancing the system's design and functionality.

This combined procedure and protocol provided a comprehensive framework for evaluating the educational agent's effectiveness and usability across different CPS contexts.

5.3   Data Collection and Analysis

Our data came from four primary sources: 1) pre-test and post-test questionnaires to assess knowledge gain, system usability, and task load, 2) learning reports created by participants during tasks to evaluate task performance outcomes, 3) screen recordings and system logs to capture student-system interaction data, and 4) in-depth interviews to gather qualitative feedback on user experiences. The pre- and post-test data were used for comparing learning outcomes and cognitive load between the Mentigo and baseline conditions, while the learning reports, interaction data, and interview data were used to analyze task performance, user engagement, and areas for system improvement.

5.3.1   Pre- and Post-Test Data. To measure participants' knowledge gain, system usability, and cognitive load, pre-test and post-test questionnaires were administered before and after each task session. The pre-test assessed participants' baseline familiarity with CPS concepts, while the post-test evaluated system usability (specifically for the Mentigo condition), cognitive load, and CPS skills mastery. Knowledge gain was determined by comparing pre- and post-test results, reflecting changes in participants' self-rated knowledge on CPS tasks.

Descriptive statistics were used to summarize participants' demographics and baseline characteristics for data analysis. Paired-sample t-tests were conducted to compare cognitive load (measured using the NASA Task Load Index (TLX) [17]) and system usability between the Mentigo and baseline conditions. Repeated measures ANOVA were used to identify significant differences in knowledge gain and task load, providing insights into the educational agent's effectiveness and cognitive demands associated with each condition.

5.3.2 Task Performance Data. Participants' task performance was evaluated using learning reports created during each condition (Mentigo and baseline). These reports provided insights into participants' ability to apply CPS techniques, the quality of their solutions, and their reasoning processes.

To assess these reports, we developed an evaluation matrix for AI-involved CPS tasks based on Urban et al. [49]. The matrix included four dimensions: 1) Quality, 2) Elaboration, 3) Originality, and 4) Human-like Expression (see supplementary materials for details). Three researchers independently rated the reports according to these criteria, ensuring reliability through inter-rater agreement.

The scores were analyzed using descriptive statistics to summarize performance in each dimension and non-parametric tests, such as the Mann-Whitney test, to compare differences between the Mentigo and baseline conditions. This provided insights into how effectively the educational agent supported CPS skills development.

5.3.3 Student-System Interaction Data. Screen recordings and system logs were collected to capture participants' interactions with the educational agent during the study sessions. These data sources included metrics such as time spent on tasks, prompt usage frequency, and action sequences. This information helped analyze user engagement, behavior patterns, and how participants navigated the system and utilized feedback to improve their CPS processes.

For analysis, sequence analysis and descriptive statistics were applied to the screen recordings and system logs to understand user behavior and engagement patterns. Metrics such as time spent on tasks, frequency of prompt usage, and interaction paths were examined to assess how participants used the agent's features (see Table 5). The analysis aimed to identify common strategies, variations in system use, and how these factors contributed to task outcomes.

Table 5. Manual Coding Scheme of System Data

Categories	Types	Description
Speech Type	Positive	Proactively suggest solutions to problems; proactively ask questions.
	Neutral	Express confusion or guide requests to AI
	Negative	Request AI to provide direct answers; simply copy the AI's responses.
Cognitive Level	Remembering	Reciting and memorizing labels
	Understanding	Relating and organizing previous knowledge
	Applying	Applying information into a new situation
	Analyzing	Drawing connections among ideas
	Evaluation	Examining information and make judgement
	Creation	Creating new ideas using what has just been learned

5.3.4 Interview Data. In-depth interviews were conducted after participants completed both conditions to gather qualitative feedback on their experiences with the educational agent. The interviews explored participants' thought processes, the impact of system guidance, the challenges faced, and their overall preferences between the two conditions. The interviews were recorded, transcribed, and analyzed to identify themes and areas for potential system improvement.

Thematic analysis was employed to analyze the qualitative data from the interviews. The coding process was iterative, focusing on refining themes that captured participants' experiences, perceived value, and suggestions for improvement. The results provided a comprehensive understanding of user experiences and were used to inform future iterations of the educational agent system (see supplementary materials for the Interview Protocol Outline).

By combining these diverse measurement and analysis methods, our study provided a comprehensive assessment of Mentigo's effectiveness and usability in mentoring middle school students' CPS tasks. This mixed-methods approach enabled a thorough understanding of both the measurable outcomes and the subjective experiences of participants.

5.4 Expert Review and Feedback

To further validate the findings and gather insights into the system's effectiveness, we conducted an expert review and feedback session. We recruited five experts, including experienced secondary school teachers, a specialist in educational research, and a mentor with extensive experience in guiding middle school students. These experts reviewed the data generated from the study, such as students' learning reports, task outcomes, and dialogue records with the two systems.

The experts evaluated the system's educational performance, usability, and impact on students' CPS skills. Following their review, they participated in a semi-structured interview to provide detailed feedback on how the system supported student engagement and learning outcomes. They highlighted the system's strengths and identified areas for improvement, offering practical suggestions based on their experience with middle school students.

The insights from these expert sessions were crucial for refining our analysis of the two system performances. The feedback also provided actionable guidance for future development to enhance the system's support for CPS and improve student learning experiences in real-world educational settings.

6 FINDINGS

6.1 Evaluation of System Usability and Effectiveness

The Mentigo system is a data-driven intelligent agent, which is built on a robust foundation of over 80 hours of dialogue data collected from real CPS educational settings. It guides students through a CPS task using a six-stage process. At each stage, the system sets clear task requirements and progression standards, using corresponding guidance strategies and generative AI to ensure a personalized learning experience that adapts to each student's state and stages. Generally, students spent more time on tasks (M = 43.7 mins vs. 24.7 mins) and had more conversation rounds (M = 24.0 vs. 13.2) with Mentigo compared to the Baseline system, reflecting a more thorough and reflective learning process.

The Mentigo system demonstrated a high level of usability, with a System Usability Scale (SUS) score of 88.54, comparable to the Baseline system's score of 89.79. While both systems are highly usable, further analysis reveals significant differences in key performance dimensions that highlight Mentigo's strengths in guiding students through CPS tasks (shown in Table 6). The system excels in providing clarity of goals ($p = 0.051$) and delivering comprehensive guided strategies ($p = 0.032$), which enhances student engagement and promotes deeper learning. For instance, P5 said, *"The system A (Mentigo) actively guides you to complete the tasks, providing a structured and purposeful approach that helps us focus on our goals."* Students can follow its model to explore and learn specific knowledge, although it still requires students to do information integration and research, which encourages deeper engagement and learning. Mentigo outperformed the Baseline system in knowledge and skills acquisition ($p = 0.019$), effectively helping students learn to articulate issues and gain CPS knowledge. This aligns with the qualitative feedback where students expressed high appreciation for Mentigo's clarity in outlining tasks at each stage and guiding them. P1 mentioned, *"System A (Mentigo) is more like a project mentor. It provides more guidance and leads you step-by-step, almost like a teacher guiding you through the process."*

Mentigo's core strength lies in its ability to dynamically identify and respond to over 23 distinct student states, such as "unclear task," "overly divergent thinking," and "surface-level thinking" (see Table 3 for details). The analysis of student state frequencies shows that states like ID 1 ("Silent"), ID 2 ("Goal Unclear"), and ID 23 ("Normal Progress/Active Questioning") are most frequently addressed, indicating these are common areas where students need support. Moderately frequent states, such as ID 10 ("Shallow Thinking"), ID 7 ("Limited Thinking"), and ID 13 ("Inadequate Information Integration Skills"), suggest that challenges in critical and creative thinking are also significant but less prevalent. Less

Table 6. Comparison of Condition System and Baseline System Across Different Dimensions

Dimension	Mentigo	Baseline	Value	p
Helpfulness	M=8.83	M=8.75	t=0.17	0.867
Goal Clarity	M=9.17	M=7.42	t=2.07	0.051∧
Guided Strategy	M=9.17	M=8.17	t=2.28	0.032*
Personalization	M=9.17	M=8.83	t=0.98	0.339
Knowledge and Skills Acquisition	Mdn=9	Mdn=7.5	U=112	0.019*
Emotional Support	M=8.33	M=7.42	t=1.18	0.25

common states, including ID 4 ("Goal Prioritization Confusion"), ID 17 ("Decision Uncertainty"), ID 18 ("Insufficient Intrinsic Motivation"), and ID 19 ("Lack of Confidence"), highlight Mentigo's capacity to provide targeted support for specific, less frequent student needs. This varied distribution of student states reflects Mentigo's adaptability and effectiveness in delivering personalized educational experiences tailored to diverse learning needs.

Similarly, the distribution of guided strategy IDs (see Table 2 for details) shows that strategies like ID 1 ("Task Follow-up and Allocation"), ID 15 ("Resource Guidance"), and ID 18 ("Encouragement and Motivation") are frequently employed, highlighting Mentigo's focus on these methods to effectively support students. Other strategies, such as ID 8 ("Probing Guidance"), ID 9 ("System Thinking Training"), and ID 13 ("Resource Guidance"), are also used often, indicating their targeted application for specific needs. Less frequent strategies, like ID 5 ("Creative Filtering and Focusing"), ID 11 ("Deep Thinking and Reflection"), and ID 19 ("Emotional Counseling"), demonstrate the system's adaptability in addressing unique challenges. This varied use of instructional strategies underscores Mentigo's ability to provide personalized and flexible educational experiences tailored to diverse learning contexts. Overall,
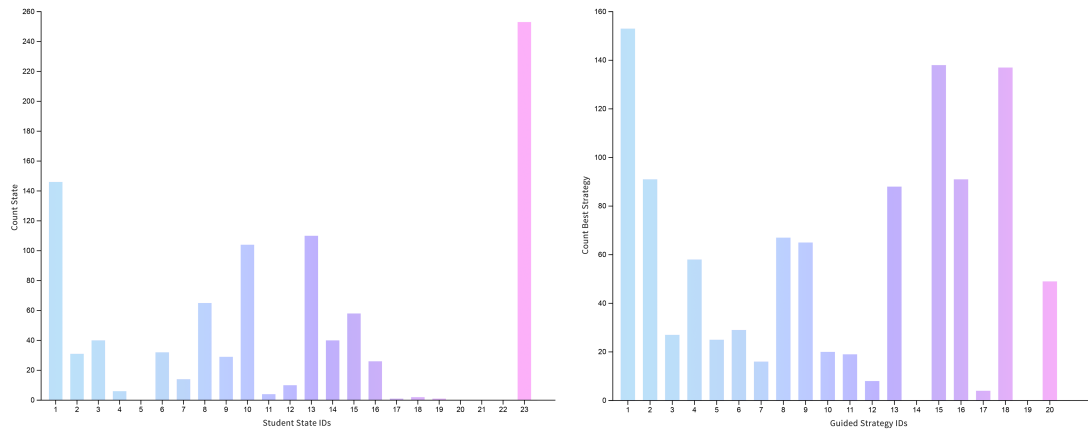

Fig. 4. Distribution of Student States and Guided Strategies Frequencies

6.2 Student CPS Knowledge Improvement and Performance

Based on the analysis of learning outcomes, a Mann-Whitney U test was conducted to compare the pre-test results between the Mentigo and Baseline groups (as shown in Table 7). The results suggested that there were no significant

differences between the two groups in their pre-test scores (p= 0.51). This indicates that both groups started with similar levels of knowledge before the intervention.

For the Mentigo group, there was a notable improvement from the pre-test to the post-test (p=0.006). The post-test scores (M = 6.67, SD = 2.19) were higher than the pre-test scores (M = 4.33, SD = 2.23), with a large effect size (Cohen's d = 0.944), suggesting substantial learning gains. In contrast, the Baseline group showed similar median scores between the pre-test (M = 4.92, SD = 2.61) and post-test (M = 5, SD = 1.91), though the effect size (Cohen's d = 0.036) indicates a medium difference, implying less change. When comparing the post-test scores between the Mentigo and Baseline groups, the effect size (Cohen's d = 0.765) suggests a large difference favoring the Mentigo group, and the Mann-Whitney U test results indicated that this difference was statistically significant (p= 0.019). P3 stated, *"I prefer the first system, Mentigo, because the baseline system often just tells you what to do directly, while Mentigo tells you the steps of thinking and guides you to think step-by-step."* This feedback highlights how Mentigo facilitates deeper cognitive engagement by guiding users through a structured thinking process, which aligns with the observed improvement in knowledge levels. On the other hand, P3, P8, and P9 mentioned that *"while the Baseline System is quicker to use and feels more like a tool that provides reference answers, it doesn't give a strong sense of accomplishment."* They described the experience as using a tool that gives quick solutions rather than fostering deep understanding. In contrast, *System A (Mentigo) helps us learn more methods and strategies and offers more guidance for critical thinking, leading to a richer learning experience and greater gains in knowledge and skills.* P12 added.

Table 7. Pre-/post-test Results (mean and standardized deviation) of Mentigo and Baseline

	Mentigo	Baseline	Effect size	p
Pre-test	4.33(2.23)	4.92(2.61)	0.19	0.51
Post-test	6.67(2.19)	5(1.91)	0.77*	0.019*
Effect size	0.944**	0.036		
p-value	0.006**	0.9		

Besides, we also evaluated the students' learning reports completed using the two systems across four dimensions. An independent samples t-test revealed a significant difference between the scores for Mentigo (*M* = 13.43, *SD* = 3.17) and the baseline system (*M* = 10.74, *SD* = 2.72), with *p* = 0.016. This result suggests that students' performance, as reflected in their learning reports, was significantly higher when using Mentigo compared to the baseline system. This is consistent with what P12 and P10 observed:

> " By using Mentigo, I learned more methods and strategies, and I gained more from it. It pushes me step-by-step to complete all the tasks, and the ideas provided are based on the questions I raised, which often gives me a moment of clarity and leads me to think more comprehensively."

These comments suggest that Mentigo's tailored guidance and reflective approach lead to deeper understanding and more significant learning gains compared to the baseline system. P11 further added, *"Mentigo gives many prompts that allow me to think for myself, which makes me more creative because it focuses more on my own answers and guides me to write step-by-step."* This insight indicates that Mentigo encourages more creative and independent problem-solving, enhancing the overall learning experience.

Overall, these results indicate that while both systems significantly impacted students' knowledge and performance, Mentigo provided a more structured and comprehensive learning experience that led to greater improvement in CPS knowledge and skills. The combination of quantitative data and qualitative insights suggests that Mentigo's approach to guiding students and encouraging deep thinking was more effective in enhancing their overall learning outcomes.

6.3 Student Engagement Across Mentigo and Baseline Systems

To analyze student engagement across the Mentigo and Baseline systems, we compared data on student-system interactions, including the number of dialogue turns and usage time. Additionally, we encoded the content of student-system dialogues to assess different aspects of engagement. Due to time constraints, two students did not complete the tasks in the last two stages of the experiment using our system; these incomplete data were excluded from the final analysis. While we measured task load across various dimensions and found no significant differences between the Mentigo and Baseline systems, our system did not increase cognitive load for students, and overall student engagement remained high. The comparison between the Mentigo and Baseline systems across different dimensions shows significant differences in several areas of student engagement: behavioral, cognitive, and emotional.

Behavioral Engagement. At the behavioral level, we assessed students' task duration, the number of conversation rounds, and the word count of student speech while using both systems (as shown in Table 8). The results showed that students had significantly longer task durations (p = 0.005) and more conversation rounds (p = 0.015) when using Mentigo. The word count of student speech was also significantly higher in the Mentigo group compared to the baseline group (p = 0.014). This observation aligns with our experimental findings, where students tended to think more deeply about the CPS questions while using Mentigo. The system's guided prompts encouraged them to complete each task step-by-step and to engage more proactively. This suggests that the structured guidance in Mentigo fosters higher levels of engagement and encourages students to think and participate more actively.

Table 8. Comparison of Mentigo and Baseline System Across Different Dimensions

Categories	Variable	Mentigo	Baseline	Value	p
	Task Duration (min)	M = 43.7	M = 24.3	t = 3.19	0.005**
Conversation Metrics	Conversation Rounds	Mdn = 24.5	Mdn = 12.5	U=82.5	0.015*
	Word Count	Mdn = 1438.5	Mdn = 290	U=83	0.014*
	Positive	M = 15.9	M = 7.2	t = 3.29	0.0041**
Speech Type	Neutral	Mdn = 5.5	Mdn = 3.5	U=65.5	0.254
	Negative	M = 2.5	M = 2.7	t = -0.15	0.8812
	Memory	M = 1.9	M = 1.5	t = 0.58	0.5717
	Understanding	M=2.4	M=2.1	t=0.3	0.7692
Cognitive Level	Application	M = 5.1	M = 4.8	t = 0.13	0.9017
	Analysis	M = 4.3	M = 1.7	t = 2.16	0.0444*
	Evaluation	M = 3	M = 0.9	t = 2.64	0.0166*
	Creation	Mdn = 7.5	Mdn = 2	U=81.5	0.0180*

Cognitive Engagement. For cognitive engagement, we coded the content of students' speech according to type (Positive, Neutral, Negative) and cognitive level, based on Bloom's Taxonomy (six levels). Significant differences were found in the higher-order cognitive levels—specifically, "Analysis," "Evaluation," and "Creation"—with Mentigo outperforming the baseline system (p = 0.044, p = 0.0166, and p = 0.0180, respectively). This highlights how Mentigo, acting more like a mentor, encourages deeper cognitive engagement and supports the development of CPS skills through guided exploration and reflection. This highlights how Mentigo, acting more like a mentor, encourages deeper cognitive engagement and supports the development of CPS skills through guided exploration and reflection. P11 mentioned,

> "When using the baseline system, it's easy to lose track and not know what to do next. But Mentigo has a more structured flow, making it less likely to go off-topic, and the guidance provided is very comprehensive, which leads to more complete thinking."

This statement underscores Mentigo's ability to provide a structured and comprehensive framework that helps students maintain focus and think more broadly, contrasting with the more direct answer-providing nature of the baseline system.

Emotional Engagement. In terms of emotional engagement, although quantitative results showed no statistically significant differences in emotional support between the two systems, the interview data provided further insights. Many students felt that Mentigo was more like conversing with a human, using simpler language, emojis, and interactive encouragements. For instance, P7, P8, and P9 noted that *"Mentigo feels more like a peer, providing encouraging messages and engaging in more interactive dialogues. In contrast, the baseline system feels more like a reserved friend who responds when asked but doesn't initiate conversations."* This suggests that Mentigo provides a more supportive and engaging emotional environment, particularly suitable for middle school students, whereas the baseline system seems more like a hyper-rational expert that lacks warmth and adaptiveness in communication.

These findings indicate that Mentigo provides a more comprehensive and effective environment for student engagement across behavioral, cognitive, and emotional dimensions. It promotes active participation, higher-order thinking, and emotional connection, making it more suitable for fostering CPS skills and creativity among middle school students. The baseline system, while functional, appears to focus more on information retrieval and direct answers, with less emphasis on interactive and personalized learning experiences.

6.4 Expert's Review and Feedback

We conducted interviews with five experts to gather insights into students' cognitive changes, AI's guidance strategies, and interactions with different systems (Mentigo vs. Baseline). The experts reviewed dialogue data between students and the systems, along with learning report data, and provided comprehensive feedback from an educational perspective. Experts noted significant differences in how students interacted with the Mentigo system compared to the Baseline system. E1 observed, *"It often seemed that students treated the baseline system more like a tool to command and directly get answers rather than engaging in deep thinking or dialogue."* This suggests that the structured guidance in Mentigo fosters higher levels of engagement and encourages students to think and participate more actively. When reviewing the dialogues of P1 using Mentigo and baseline system, E3 highlighted that as the conversation progresses, the students show "more comprehensive expressions and transfer previously mentioned content into new contexts. In contrast, the baseline system felt more like an information retrieval tool with limited cognitive advancement. E1 added, *"The baseline system is more like a didactic expert who provides answers but does not engage with students at a deeper level."* This difference highlights how Mentigo, acting more like a mentor, encourages deeper cognitive engagement and supports the development of CPS skills through guided exploration and reflection.

This observation was echoed by E2, who elaborated, *"Mentigo offers more patient encouragement, like a cheerful artist friend who is more expressive and proactive, even using emojis to ask questions."* This supports the feedback from students who noted that Mentigo provided comprehensive guidance and multiple perspectives for thinking. E5 mentioned, *"The guidance was simple yet offered various perspectives and space for thinking,"* suggesting that the system's ability to guide through structured steps supports its quantitative superiority in fostering cognitive engagement. In contrast, E4 commented that the Baseline System's approach often felt like a "question-and-answer session," lacking the same depth of guidance. Experts praised Mentigo's ability to provide structured guidance and maintain a comprehensive framework that helps students remain focused and think more deeply.

E5's feedback supports this observation: When P1 interacted with both systems at the beginning of the tasks, the differences in their guidance became evident. System A (Mentigo) provides clear direction at the start of each stage, explaining the task objectives and guiding students to discover more problems from multiple perspectives based on their input. It also prompts students to clarify their problem statements as required. Meanwhile, when P1 interacted with System B, it generated multiple answers based on the provided keywords for the student to choose from. This often led to students quickly adopting the AI-generated answers and ending the conversation prematurely. An excerpt of these interactions is shown in Fig. 5. This highlights Mentigo's role in scaffolding the learning process and helping students build their CPS skills step-by-step.
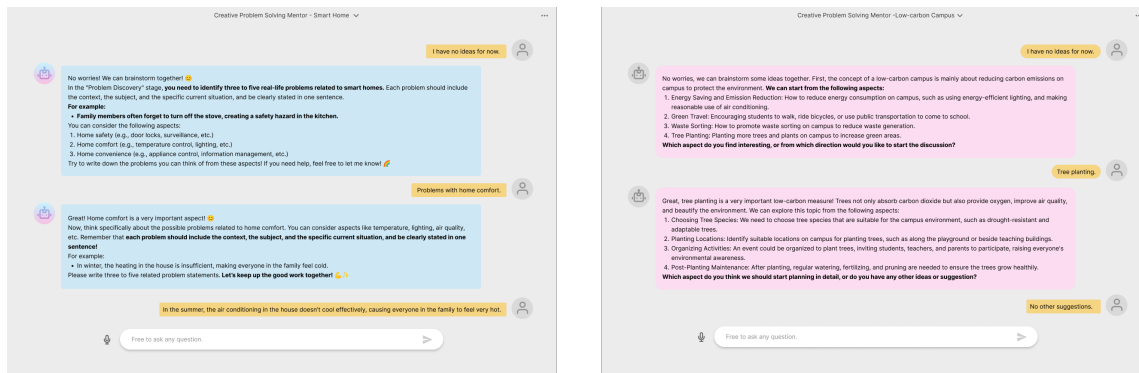

Fig. 5. Illustration of Student P1's Dialogues with Mentigo (Left) and the Baseline System (Right) at the Same Stage of Different Tasks

Experts also highlighted the contrast between the two systems' approaches: while Mentigo was described as a "mentor," the Baseline system was perceived as a "knowledge authority" that simply provides information. E3 pointed out that *"System B (Baseline) felt like an over-rational tool; it lacked the 'human touch,' making it difficult for students to relate and engage deeply."* This difference impacted students' engagement levels, with Mentigo being more effective in fostering both cognitive and emotional engagement. Besides, as for the emotional support, E1, E3, and E4 stated that Mentigo was more like conversing with a human, using simpler language, emojis, and interactive encouragements. For instance, E4 noted that *"Mentigo feels more like a peer, providing encouraging messages and engaging in more interactive dialogues."* In contrast, the Baseline system feels more like a reserved friend who responds when asked but doesn't initiate conversations. This suggests that Mentigo provides a more supportive and engaging emotional environment, particularly suitable for middle school students, whereas the Baseline system seems more like a hyper-rational expert that lacks warmth and adaptiveness in communication. Besides, expert feedback suggests a significant opportunity to develop a teacher interface for real-world applications. E2, E3, E5 both highlight Mentigo offers valuable insights into students' thinking processes and progress, which could be further enhanced by making these processes visible to educators in real time. E2 stated,

> I think a teacher interface could allow educators to monitor students' states, visualize their thought processes, and provide formative assessments based on real-time data. Such a feature would enable teachers to intervene more effectively, offer targeted support, and refine AI-driven guidance strategies, creating a more integrated and dynamic educational environment.

Overall, the experts' feedback suggests that the Mentigo system, with its structured and personalized guidance, offers a more effective and engaging environment for students compared to the Baseline system. It promotes cognitive development, encourages deeper engagement, and creates a supportive emotional atmosphere that is crucial for fostering creativity and problem-solving skills. The insights gathered from the experts provide valuable directions for further refining the AI's guidance strategies to better support diverse student needs and learning styles.

7 DISCUSSION

7.1 Mentoring Student with Mentigo

The results of our study indicate that Mentigo, as a mentoring AI system, has the potential to foster significantly higher levels of engagement (i.e., behavioral, emotional, and cognitive) among students compared to traditional systems. Specifically, students using Mentigo were found to be more actively engaged in their learning processes. The analysis revealed that students demonstrated longer task durations, more conversation rounds, and richer vocabulary in their responses when using Mentigo, indicating higher levels of behavioral and cognitive engagement. These observations align with findings from previous studies that emphasize the importance of guided exploration and reflective dialogue in fostering deep learning and creative thinking skills among students [24, 40, 53]. Mentigo's interactive and adaptive techniques played a crucial role in eliciting engaging activities from the students during various learning tasks. Some students suggested additional steps or posed new questions such as, "What if we consider another scenario?" or "Could we explore an alternative solution?" These instances indicate students' willingness to apply their evolving understanding and engage in higher-order thinking, which are signs of active learning and cognitive growth [4, 32].

Moreover, during collaborative tasks facilitated by Mentigo, students were more inclined to share their thoughts and negotiate meanings. For example, while discussing a complex topic, students were observed brainstorming ideas collectively and challenging each other's viewpoints. This collaborative dialogue was often filled with expressions of curiosity and enthusiasm, such as "That's an interesting point, but what about...?" or "Let's think of a different approach together." Such interactions not only reflect the students' deeper emotional engagement but also highlight Mentigo's ability to create a supportive environment that encourages risk-taking and creative problem-solving [14, 46].

Overall, the use of Mentigo transformed the traditional learning environment into a more interactive and engaging space for students. Its mentoring approach, characterized by structured guidance, personalized feedback, and an emphasis on reflective thinking, successfully promoted active participation and deeper cognitive engagement among students. By creating an atmosphere that supports both individual exploration and collaborative learning, Mentigo has demonstrated its potential as an effective educational tool to mentor students and enhance their overall learning outcomes.

7.2 Balancing Structured Guidance and Personalized Feedback for Enhanced CPS Skills

The findings indicate that Mentigo's structured task management system, combined with its diverse range of guided strategies, effectively enhances students' clarity, compliance, and development of CPS skills. By breaking down the CPS process into clear, manageable stages and providing explicit instructions, Mentigo helps students focus on task objectives and progress systematically, reducing cognitive overload and supporting better learning outcomes [11, 40]. However, while structured management improves task focus, it can limit creative exploration if not balanced with open-ended opportunities. To address this, Mentigo integrates varied guidance strategies that encourage exploration and creativity, aligning with research on fostering critical thinking through diverse prompts [14, 32]. Expert feedback

suggests refining the system to adaptively switch between guidance strategies based on real-time student feedback, ensuring a dynamic balance between structure and flexibility. Future designs could incorporate more adaptive task management mechanisms that adjust dynamically to a learner's progress and needs.

Additionally, the study highlights the importance of personalized feedback that responds to students' dynamic states. Mentigo's monitoring of cognitive and emotional states to deliver tailored feedback has proven effective in maintaining engagement and promoting deeper learning. Students reported feeling more supported and understood, motivating them to engage more deeply with the tasks. This aligns with prior research on the value of personalized, context-aware feedback in learning environments [4, 24]. Future improvements could explore integrating more multimodal feedback options to accommodate different learning preferences and further enhance engagement.

7.3 Encouraging Deeper Cognitive and Emotional Engagement Through Adaptive, Empathetic Interactions

Mentigo shows a strong capability to foster deeper cognitive engagement, demonstrated by the increased instances of higher-order thinking, such as analysis, evaluation and creation. Its adaptive prompts and scaffolding strategies encourage reflection, critical thinking, and the exploration of alternative solutions, supporting the development of creative problem-solving skills. This aligns with educational theories advocating for adaptive learning environments to stimulate critical thinking and creativity [18, 46]. However, some students struggled with the level of challenge presented by the system. Future development should consider more nuanced adaptive algorithms that can fine-tune difficulty levels based on each student's progress and comfort, ensuring a more personalized learning experience.

Empathetic interactions are also a central design focus of Mentigo, aimed at fostering emotional engagement and creating a supportive learning environment for middle school students. While quantitative results did not show a significant difference in emotional support between the groups, expert interviews suggested that Mentigo provided a "warmer" and more supportive experience compared to the baseline system. This perception aligns with research on the benefits of empathetic AI in enhancing emotional well-being and motivation [14, 53]. Experts highlighted the need for more sophisticated emotion recognition to make responses more context-aware and personalized. Future development should focus on refining the AI's language and interactions to be more relatable and supportive, leveraging real-time emotion detection to better meet the emotional needs of students and foster a positive learning experience.

7.4 Limitation and Future Work

While Mentigo shows potential in enhancing student engagement in CPS tasks, there are several limitations to consider. The system's stability issues, such as occasional lags and response delays, could hinder user experience and reduce engagement. Additionally, the study's limited scale, involving a specific group of middle school students in controlled conditions, may affect the generalizability of the results to broader educational contexts. Furthermore, while the system provides adaptive guidance, its current ability to dynamically adjust the difficulty level and feedback based on real-time student interactions could be refined to better cater to diverse learning needs.

For future work, it is essential to deploy Mentigo in real-world classroom settings to evaluate its effectiveness amidst the complexities of authentic educational environments. This would provide deeper insights into how the system supports diverse student populations and adapts to varying classroom dynamics. Another promising direction is to enhance the system's capabilities for collaborative learning, allowing for group problem-solving tasks where AI mentors can facilitate peer interactions. Additionally, future research could explore integrating AI mentors more closely with human teachers, creating seamless teacher-AI-student collaboration. Developing intuitive teacher interfaces for

monitoring student progress and providing real-time interventions could further enhance educational outcomes and foster a more cohesive and dynamic learning environment.

8 CONCLUSION

We presented Mentigo, an AI-driven educational agent designed to enhance student engagement through structured guidance, personalized feedback, and empathetic interactions. By leveraging large language models, Mentigo provides an adaptive learning experience tailored to CPS tasks. To evaluate its effectiveness, we conducted an with-in experiment study comparing Mentigo with a baseline system, revealing that students using Mentigo were more engaged behaviorally, cognitively, and emotionally. These findings suggest that AI-driven mentoring systems like Mentigo can significantly enrich the learning experience in educational settings. We hope this work contributes to the HCI and educational technology field by demonstrating the potential of AI systems to provide personalized and adaptive mentor that fosters deeper learning and engagement.

REFERENCES

[1] Alaa Abd-Alrazaq, Rawan AlSaad, Dari Alhuwail, Arfan Ahmed, Padraig Mark Healy, Syed Latifi, Sarah Aziz, Rafat Damseh, Sadam Alabed Alrazak, Javaid Sheikh, et al. 2023. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Medical Education* 9, 1 (2023), e48291.

[2] Safinah Ali, Nisha Elizabeth Devasia, and Cynthia Breazeal. 2022. Escape! bot: Social robots as creative problem-solving partners. In *Proceedings of the 14th Conference on Creativity and Cognition*. 275–283.

[3] Tariq Alqahtani, Hisham A Badreldin, Mohammed Alrashed, Abdulrahman I Alshaya, Sahar S Alghamdi, Khalid bin Saleh, Shuroug A Alowais, Omar A Alshaya, Ishrat Rahman, Majed S Al Yami, et al. 2023. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy* (2023).

[4] Maryam Ashoori, Zhiqi Shen, Chunyan Miao, Robert Gay, and Narges Zarrabi. 2007. Mentor agent: an intelligent virtual teacher for personalized learning environments. In *Proceedings of the 7th WSEAS International Conference on Distance Learning and Web Engineering*.

[5] Ömer Aydın and Enis Karaarslan. 2023. Is ChatGPT leading generative AI? What is beyond expectations? *What is beyond expectations* (2023).

[6] Zied Bahroun, Chiraz Anane, Vian Ahmed, and Andrew Zacca. 2023. Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability* 15, 17 (2023), 12983.

[7] MIN Basadur, George Graen, and Mitsuru Wakabayashi. 1990. Identifying individual differences in creative problem solving style. *The Journal of Creative Behavior* 24, 2 (1990), 111–131.

[8] Tony Bush and Joy Chew. 1999. Developing human capital: Training and mentoring for principals. *Compare: A Journal of Comparative and International Education* 29, 1 (1999), 41–52.

[9] Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, and Smaranda Muresan. 2023. Creativity Support in the Age of Large Language Models: An Empirical Study Involving Emerging Writers. *arXiv preprint arXiv:2309.12570* (2023).

[10] Kuo-Chi Chang, Kai-Chun Chu, Hsiao-Chuan Wang, Yuh-Chung Lin, and Jeng-Shyang Pan. 2020. Agent-based middleware framework using distributed CPS for improving resource utilization in smart city. *Future Generation Computer Systems* 108 (2020), 445–453.

[11] Bob Eberle and Bob Stanish. 2021. *CPS for Kids: A Resource Book for Teaching Creative Problem-Solving to Children (Grades 2-8)*. Routledge.

[12] Paul Egan and Jonathan Cagan. 2016. Human and computational approaches for design problem-solving. *Experimental Design Research: Approaches, Perspectives, Applications* (2016), 187–205.

[13] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300619

[14] Evana Gizzi, Lakshmi Nair, Sonia Chernova, and Jivko Sinapov. 2022. Creative problem solving in artificially intelligent agents: A survey and framework. *Journal of Artificial Intelligence Research* 75 (2022), 857–911.

[15] Muhammad Usman Hadi, R Qureshi, A Shah, M Irfan, A Zafar, MB Shaikh, N Akhtar, J Wu, and S Mirjalili. 2023. A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage. *TechRxiv* (2023).

[16] Brian Hansford and Lisa C Ehrich. 2006. The principalship: how significant is mentoring? *Journal of educational administration* 44, 1 (2006), 36–52.

[17] SG Hart. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload/Elsevier* (1988).

[18] Peter Hudson. 2013. Mentoring as professional development:'growth for both'mentor and mentee. *Professional development in education* 39, 5 (2013), 771–783.

[19] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2021. IdeaBot: investigating social facilitation in human-machine team creativity. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–16.

[20] Gwo-Jen Hwang and Nian-Shing Chen. 2023. Exploring the potential of generative artificial intelligence in education: applications, challenges, and future research directions. *Journal of Educational Technology & Society* 26, 2 (2023).

[21] Bonny Jung, Joyce Tryssenaar, and Seanne Wilkins. 2005. Becoming a tutor: exploring the learning experiences and needs of novice tutors in a PBL programme. *Medical Teacher* 27, 7 (2005), 606–612.

[22] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.

[23] Nuttakarn KITPO, Suphanut KONGWAT, Takara HIRABAYASHI, Worawat LAWANONT, Hiroshi HASEGAWA, Masahiro INOUE, and Anak KHANTACHAWANA. 2020. Learning Outcomes of Global Project-Based Learning: Insights From Teaching Assistants. *Journal of JSEE* 68, 5 (2020), 5_39–5_44.

[24] Larry Lavender. 2006. Creative process mentoring: Teaching the "making" in dance-making. *Journal of Dance Education* 6, 1 (2006), 6–13.

[25] Sandra Leaton Gray. 2020. Artificial Intelligence in Schools: Towards a Democratic Future. *London Review of Education* 18, 2 (2020), 163–177.

[26] Leung Mee Lee and Tony Bush. 2003. Student mentoring in higher education: Hong Kong Baptist University. *Mentoring & Tutoring* 11, 3 (2003), 263–271.

[27] Cheolil Lim and Hyeongjong Han. 2020. Development of instructional design strategies for integrating an online support system for creative problem solving into a University course. *Asia Pacific Education Review* 21, 4 (2020), 539–552.

[28] Chung Kwan Lo. 2023. What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences* 13, 4 (2023), 410.

[29] Rupert Maier, Stephan Unverdorben, and Michael Gepp. 2018. Efficient implementation of task automation to support multidisciplinary engineering of cps. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*. IEEE, 1388–1393.

[30] Regina P McCurdy, Megan L Nickels, and Sarah B Bush. 2020. Problem-based design thinking tasks: Engaging student empathy in STEM. *The electronic journal for research in science & mathematics education* 24, 2 (2020), 22–55.

[31] Elspeth McFadzean. 2002. Developing and supporting creative problem solving teams: part 2–facilitator competencies. *Management decision* 40, 6 (2002), 537–551.

[32] Alexander Tobias Neumann, Peter de Lange, Ralf Klamma, Norbert Pengel, and Tamar Arndt. 2020. Intelligent Mentoring Bots in Learning Management Systems: Concepts, Realizations and Evaluations. In *International Symposium on Emerging Technologies for Education*. Springer, 3–14.

[33] Andy Nguyen, Ha Ngan Ngo, Yvonne Hong, Belle Dang, and Bich-Phuong Thi Nguyen. 2023. Ethical principles for artificial intelligence in education. *Education and Information Technologies* 28, 4 (2023), 4221–4241.

[34] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283* (2023).

[35] In-Suk Park and Soon-Hee Kang. 2012. The development of assessment tools to measure scientific creative problem solving ability for middle school students. *Journal of the Korean Association for Science Education* 32, 2 (2012), 210–235.

[36] Mike Perkins. 2023. Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice* 20, 2 (2023), 07.

[37] Jeb S Puryear and Kristen N Lamb. 2020. Defining creativity: How far have we come since Plucker, Beghetto, and Dow? *Creativity Research Journal* 32, 3 (2020), 206–214.

[38] Jürgen Rudolph, Samson Tan, and Shannon Tan. 2023. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching* 6, 1 (2023).

[39] Sdenka Zobeida Salas-Pilco, Kejiang Xiao, and Xinyun Hu. 2022. Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences* 12, 8 (2022), 569.

[40] Patricia L Samson. 2015. Fostering student engagement: Creative problem-solving in small group facilitations. *Collected essays on learning and teaching* 8 (2015), 153–164.

[41] Stephen T Schroth and Jason A Helfer. 2023. Spirituality and Creativity: Meeting Learners' Spiritual and Cognitive Needs Through Creative Problem Solving. In *Supporting Children and Youth Through Spiritual Education*. IGI Global, 137–154.

[42] Douglas Summers-Stay, Clare R Voss, and Stephanie M Lukin. 2023. Brainstorm, then Select: a Generative Language Model Improves Its Creativity Score. In *The AAAI-23 Workshop on Creative AI Across Modalities*.

[43] Teo Susnjak. 2022. ChatGPT: The end of online exam integrity? *arXiv preprint arXiv:2212.09292* (2022).

[44] Kehui Tan, Tianqi Pang, and Chenyou Fan. 2023. Towards Applying Powerful Large AI Models in Classroom Teaching: Opportunities, Challenges and Prospects. *arXiv preprint arXiv:2305.03433* (2023).

[45] Leonard Tetzlaff, Florian Schmiedek, and Garvin Brod. 2021. Developing personalized education: A dynamic framework. *Educational Psychology Review* 33 (2021), 863–882.

[46] Donald J Treffinger. 1995. Creative problem solving: Overview and educational implications. *Educational psychology review* 7 (1995), 301–312.

[47] Donald J Treffinger. 2007. Creative Problem Solving (CPS): Powerful tools for managing change and developing talent. *Gifted and Talented International* 22, 2 (2007), 8–18.

[48] Donald J Treffinger, Scott G Isaksen, and K Brian Dorval. 2003. Creative problem solving (CPS Version 6.1 TM) a contemporary framework for managing change. *Center for Creative Learning, Inc and Creative Problem Solving Group, Inc* (2003).

[49] Marek Urban, Filip Děchtěrenko, Jiří Lukavskỳ, Veronika Hrabalová, Filip Svacha, Cyril Brom, and Kamila Urban. 2024. ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education* 215 (2024), 105031.

[50] Esteban Vázquez-Cano, José M Ramírez-Hurtado, José M Sáez-López, and Eloy López-Meneses. 2023. ChatGPT: The Brightest Student in the Class. *Thinking Skills and Creativity* (2023), 101380.

[51] Johannes Woithe and Ondrej Filipec. 2023. Understanding the Adoption, Perception, and Learning Impact of ChatGPT in Higher Education: A qualitative exploratory case study analyzing students' perspectives and experiences with the AI-based large language model.

[52] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601* (2023).

[53] Chai Xun Yu and Ganakumaran Subramaniam. 2017. A mentoring approach for developing creativity in teaching. *Malaysian Journal of ELT Research* 14, 1 (2017).