Computers & Education 234 (2025) 105314

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The role of teachable agents' personality traits on student-AI interactions and math learning

Bailing Lyu [a], Chenglu Li [a,*], Hai Li [b], Hyunju Oh [b], Yukyeong Song [b], Wangda Zhu [b], Wanli Xing [b]

[a] Department of Educational Psychology, University of Utah, Salt Lake City, UT, USA
[b] School of Teaching & Learning, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Teachable agent
Personality traits
Math education

ABSTRACT

This study explores how the personality traits of pedagogical agents, particularly teachable agents, influence students' math learning experiences. Grounded in the Big Five personality traits framework, students were randomly assigned to tutor a teachable agent that either emphasized one of the five personality traits or did not emphasize a specific trait for mathematics problem-solving. Students interacted with agents of varying personalities across different problems, with each program featuring a single, designated personality for the teachable agent. Results indicate that openness-emphasis agents elicited more student explanations (i.e., cognitive engagement) than extraversion ($b = 0.32$, $p < .001$) and agreeableness-emphasis agents ($b = 0.28$, $p < .001$) during the teaching process. In contrast, extraversion-emphasis agents facilitated polite expressions (i.e., affective expressions) compared to conscientiousness-emphasis ($b = 0.31$, $p = .002$) and openness-emphasis agents ($b = 0.29$, $p = .003$). Similarly, students interacting with agreeableness-emphasis agents exhibited significantly more polite expressions than those engaging with conscientiousness-emphasis ($b = 0.31$, $p = .002$) and openness-emphasis agents ($b = 0.29$, $p = .004$). Additionally, non-personality-emphasis agents were more effective in fostering cognitive engagement, such as providing explanations, compared to agents emphasizing agreeableness ($b = 0.29$, $p < .001$) and extraversion ($b = 0.25$, $p < .001$). Furthermore, these non-personality-emphasis agents enhanced students' conceptual knowledge application more effectively than agreeableness-emphasis agents ($\beta = -0.09$, $p = .04$). These findings suggest that pedagogical agents do not necessarily need to rigidly embody a single personality trait to be effective. Instead, non-personality-emphasis agents that adapt their responses dynamically based on student interactions may better support learning in problem-solving contexts than emphasizing certain personalities. To optimize instructional effectiveness, pedagogical AI agents should be designed to align with diverse learning goals and adjust their responses flexibly to meet students' needs. Future research should explore how adaptive pedagogical agents facilitate student engagement and learning across different educational contexts.

Mathematics is fundamental to advancing students' academic achievement and shaping their daily experiences (Cirino et al., 2019; Lee & Mao, 2021). However, improving students' mathematics learning remains challenging due to limited resources and a lack of

* Corresponding author. University of Utah, Department of Educational Psychology, Salt Lake City, UT, USA.
E-mail addresses: u6057059@utah.edu (B. Lyu), chenglu.li@utah.edu (C. Li), li.ha@ufl.edu (H. Li), oh.h@ufl.edu (H. Oh), y.song1@ufl.edu (Y. Song), wangdazhu@ufl.edu (W. Zhu), wanli.xing@coe.ufl.edu (W. Xing).

https://doi.org/10.1016/j.compedu.2025.105314
Received 11 November 2024; Received in revised form 21 March 2025; Accepted 5 April 2025
Available online 12 April 2025

effective strategies that engage students both cognitively and affectively (Major et al., 2021; Mitchall & Jaeger, 2018). For example, math teachers often face difficulties to provide personalized instruction due to large class size and limited time to cover the class content, which makes it difficult to foster a deep understanding of mathematical concepts for each student. A promising approach to addressing these challenges is the integration of artificial intelligence (AI) in mathematics education, which offers scalable instructional practices, personalizes learning, and enhances the accessibility of mathematical content. For instance, AI has been applied to deliver immediate and customized instruction and feedback by evaluating students' responses during problem-solving processes or diagnosing students' strengths and weaknesses during interactions with the AI system (e.g., Rori AI tutor, Henkel et al., 2024; Cognitive Tutor, Koedinger & Aleven, 2007). Additionally, AI tools assist educators in automated scoring and identifying areas where students are struggling and adapting their teaching strategies accordingly (e.g., Heffernan & Heffernan, 2014; Morris et al., 2024). Overall, AI systems have been shown to significantly improve students' mathematics learning outcomes (Feng et al., 2024; Roschelle et al., 2016).

However, despite the potentials of AI systems on enhancing students' mathematics learning, many existing AI models tend to position students as passive recipients of AI-driven decision-making, thereby limiting the opportunities for students to actively engage and take ownership of their learning. In these models, AI frequently assumes an active, authoritative role, providing solutions, answers, instructions, or directing students through mathematics problem-solving processes (Hou et al., 2023; Kurdi et al., 2020). For example, Cheng et al. (2024) designed a chatbot that provides feedback on students' incorrect answers, breaking the problem into steps to facilitate problem-solving, where students passively receive guidance from the chatbot. Although these models may simplify learning, this approach can restrict students' opportunities for exploration and hinder deep engagement with learning materials. Considering these limitations and learning theories emphasizing the critical role of active student participation in knowledge construction (e.g., Constructivism, Piaget, 1952; Vygotsky, 1978), there is a clear need to examine AI systems that foster active student participation in mathematics learning. The present study, therefore, focuses on teachable agents —AI systems designed to function as learners— to facilitate students' active participation in and foster ownership of their learning. Grounded in learning-by-teaching pedagogical framework, which posits that teaching others is an effective method for enhancing one's own learning and is well-supported by empirical research (Annis, 1983; Li et al., 2025), teachable agents enable students to shift from passive recipients to active instructors of AI systems to engage more deeply with mathematical content.

Empirical studies have documented the effects of teachable agents on students' behaviors and learning across various educational contexts, including computer science, science education, language learning, and mathematics education (e.g., Baranwal, 2022; Biswas et al., 2005; Hood et al., 2015; Jin et al., 2024; Tärning et al., 2019). Building on the documented benefits of teachable agents for student learning, this study aims to further investigate whether specific characteristics of teachable agents, specifically their personality traits, affect students' behaviors and learning outcomes in the context of mathematics education. Research suggests that AI agents should embody personality traits to create engaging virtual environments that mirror the richness and interactivity of real-world interactions (Doce et al., 2010; Nighojkar et al., 2025; Reeves & Nass, 1996). Studies further indicate that an AI agent's personality traits can shape individuals' perceptions of and interactions with the agent (Jo & Park, 2024; Zhou et al., 2019). For example, Jo and Park (2024) found that ChatGPT's human-like personality traits increased office workers' perceived utility and satisfaction, enhancing their acceptance of the technology. However, limited research has explored the impact of AI's personalities within educational contexts, particularly for teachable agents, although previous research has highlighted the relationship between student or teacher personality traits and student learning outcomes (Khalilzadeh & Khodi, 2021; Komarraju et al., 2011; Mustoip et al., 2024).

To address this gap, this study examines the impact of personality traits in AI-powered teachable agents on students' mathematics learning experiences, based on the Big Five personality traits framework, which categorizes personality traits into five dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness (Costa & McCrae, 2000). We developed six types of teachable AI agents using in-house large language models (LLMs) pretrained with millions of authentic mathematical question-answer pairs in K12 settings, and advanced prompt engineering (i.e., few-shot learning and chain-of-thought). Five agents were designed to emphasize a specific Big Five personality trait, making it dominant in their responses. The sixth agent was a non-personality-emphasis agent, serving as a control group. This study examines how students interact with these six agent types while tutoring them in solving mathematical problems. Additionally, we investigate the relationship between teachable agents' personality traits, students' interaction behaviors, and their learning performance. Thus, this study addresses the following research questions.

1. What are the effects of teachable agents' personality traits on students' interactions with the agents?
2. How do teachable agents' personality types and students' interactions with the agents predict students' mathematics learning?

1. Learning by teaching framework

According to Constructivism (Piaget, 1952; Vygotsky, 1978), active learning, where students take an active role in their learning, is considerably more effective than passive learning approaches (Gartner, 1971). When students construct knowledge through interactions, such as participating in discussions, working collaboratively, or synthesizing information, rather than passively receiving it from others (e.g., teachers, tutors, or AI), they are more likely to achieve a deep understanding of the learning content and develop a stronger sense of connection and commitment to their learning (Vygotsky, 1978; Werfel, 2013). One of the most engaging forms of active learning is teaching others, often referred to as "*learning by teaching*". This framework within learning sciences positions students as instructors, thereby enhancing their engagement, motivation, and depth of understanding (Biswas et al., 2005; Fiorella & Mayer, 2013; Gartner, 1971).

Research in cognitive science and education has long established teaching others as a powerful learning strategy. Previous studies on self-explanation, peer tutoring, and reciprocal teaching have demonstrated the potential benefits of learning by teaching (Ali et al., 2015; McCrudden et al., 2024; Palinscar & Brown, 1984). Biswas et al. (2005) outlined three critical ways that teaching others can impact student learning: structuring, taking responsibility, and reflecting. In terms of structuring, teaching helps students articulate and reorganize their knowledge in response to feedback, facilitating deeper understanding. Moreover, teaching requires students to take responsibility for helping others solve problems, which not only forces them to enhance their own comprehension of the material but also motivates them through the satisfaction of contributing to others' learning. Finally, teaching requires students to monitor their instructional process, prompting them to evaluate and reflect on their own understanding of domain concepts. These three mechanisms underscore how teaching others affects students' cognitive understanding, affective outcomes, and metacognitive monitoring and regulation.

2. Teachable agents

Leveraging the advancements in AI technologies, educational researchers have developed teachable agents that allow students to act as teachers to AI, implementing the learning-by-teaching framework to promote student learning (Biswas et al., 2005, 2010; Han et al., 2021; Hayashi et al., 2025; Love et al., 2025). Compared to teaching humans, teaching AI agents offers several key benefits, such as improving scalability, enhancing accessibility, and reducing psychological barriers. Specifically, AI agents can simultaneously serve many students in ways that human peer-to-peer interactions cannot, making the learning-by-teaching approach with AI agents scalable to larger groups (e.g., a large-scale online educational platform). Also, AI agents provide continuous availability, enabling students to engage in teaching at any time without scheduling constraints. Additionally, teachable agents create a low-pressure environment, reducing students' fears of making mistakes or being judged, which can occur in tutoring peers, and thus encouraging their engagement in learning (Debbané et al., 2023).

However, traditional teachable agents have historically been constrained by their technical capacity to handle diverse problem types and support dynamic interactions. For example, Pareto (2014) designed a teachable agent game to support students' mathematics learning and evaluated its effectiveness through a quasi-experimental study. While the findings demonstrated that students using the game achieved greater learning gains than those in the control group, the agent was developed using a rule-based approach, restricting its focus to basic mathematical problems and limiting interactions to predefined choices rather than allowing natural verbal conversation. To overcome these limitations, in recent years, researchers have leveraged LLMs to develop LLM-powered teachable agents, enabling more flexible and natural student-agent interactions. For instance, Jin et al. (2025) examined the impact of an LLM-powered teachable agent on students' music theory learning, finding that students who engaged in music analysis with the teachable agent outperformed those without access to the agent on a post-test measuring music theory knowledge.

Whether traditional or LLM-powered teachable agents, their effectiveness in supporting students' learning behaviors and performance is well-documented (Baranwal, 2022; Hayashi et al., 2025; Jin et al., 2024, 2025; Tärning et al., 2019). For instance, Biswas et al. (2010) conducted a study in which primary school students taught a virtual agent about entities and their relationships in a river ecosystem by constructing concept maps. The findings indicated that students teaching the agent engaged more frequently in metacognitive and cognitive processes, such as verifying the accuracy of their concept maps, making edits, asking questions, and seeking explanations, and they achieved higher scores in concept map creation. Additionally, Han et al. (2021) randomly assigned students either to teach an agent or not. They examined students' affective states through facial expressions captured using digital cameras. Lag sequence analysis was used to examine transition between affective states. Results indicated that teachable agents facilitated students' use of affect regulation strategies. More recently, Liu et al. (2024) investigated the effect of an LLM-powered teachable agent on students' interest and engagement in reading activities, finding that students who interacted with the teachable agent demonstrated greater reading interest and engagement than those without access to the agent. Moreover, Hayashi et al. (2025) developed a teachable agent to support collaborative learning. In this study, students taught the agent by having it monitor their concept maps on psychological concepts, learn from the maps, identify errors, and generate revised versions. The study analyzed students' learning processes by coding student-agent interactions, while learning performance was assessed through pre- and post-tests measuring conceptual understanding. Findings indicated that the teachable agent significantly improved students' reciprocal interactions with the agent and enhanced their learning outcomes. In sum, empirical research has indicated that teachable agents contributed to students' cognitive, metacognitive, and affective engagement, as well as their overall learning performance.

However, the effectiveness of teachable agents can be influenced by the specific agent features (Biswas et al. 2005, 2010; Tärning et al., 2019). Uresti (2000) asked students to teach an agent with either slightly less knowledge than the students (a "weak" agent) or slightly more expertise (a "strong" agent) in Boolean algebra. Although not statistically significant, students working with the weak agent tended to learn more than those working with the strong agent. Similarly, Kirkegaard (2016) explored the impact of agent responsiveness by having middle school students teach agents in the domain of history, in which the agents either accepted all students' inputs or challenged students' answers and explanations. Results showed that students with high self-efficacy performed better when instructing the "challenging" agent, whereas students with low self-efficacy learned more effectively when teaching the "accepting" agent. More recently, Jin et al. (2024) developed two versions of teachable agents for computer education: a traditional teachable agent and an enhanced version incorporating components to 1) simulate student-like learning behaviors, 2) allow the agent to alternate between help-receiver and questioner modes to elicit student knowledge construction, and 3) provide metacognitive feedback on students' teaching styles. College students who interacted with the enhanced agents engaged in more dense knowledge-building conversations compared to those working with the traditional agents.

Overall, while teachable agents facilitate student learning, the effectiveness of teachable agents depends on their design features. Although several studies have investigated how specific features of teachable agents affect students' learning, few have examined the role of teachable agents' personalities—a suggested important characteristic of virtual agents (Doce et al., 2010; Reeves & Nass, 1996). This study addresses this gap by exploring how teachable agents with varied personality traits influence students' math learning experiences.

3. Big five personality traits framework

Personality traits describe relatively enduring tendencies in affect, cognition, and behaviors. There are numerous ways for categorizing personality traits, with one of the most widely accepted being Big Five personality traits framework (Costa & McCrae, 2000; McCrae & Costa, 2006). According to this framework, personality traits are classified into five dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness (John et al., 2008). Extraversion indicates the degree of individual assertiveness, sociability, and activity. Agreeableness refers to an individual's level of flexibility, amiability, and cooperativeness. Conscientiousness encompasses perseverance, diligence, organization, achievement orientation, and responsibility. Neuroticism indicates tendencies toward anxiety, depression, insecurity, and vulnerability. Openness reflects the degree of curiosity, imagination, and open-mindedness.

Each personality dimension exists along a continuum, spanning two poles: extraversion to introversion, openness to traditionalism, conscientiousness to carelessness, agreeableness to self-centeredness, and neuroticism to emotional stability (Jensen, 2015). Individuals are positioned uniquely along each of these continua, resulting in a distinctive combination of personality traits. This complex combination of personality traits implies that individuals may exhibit one or multiple dominant traits, which substantially influence their attitudes, decisions, and behavior. For instance, an individual with moderate levels of extraversion but high levels of neuroticism may find neuroticism playing a more pivotal role in shaping their behavioral responses. When multiple traits are dominant, their relative influence may vary depending on situational demands. In relevant contexts, a specific trait may become more pronounced, leading to differential behavioral outcomes (Jensen, 2015). For example, in educational settings, a person with high conscientiousness and extraversion may find conscientiousness exerting a stronger influence than extraversion, promoting academic diligence, meticulousness, and goal commitment. Conversely, in social contexts, extraversion may exert a greater influence than conscientiousness, fostering engagement, enthusiasm, and sociability. Thus, dominant traits can differentially drive behavior across varied contexts (Costa & McCrae, 2000; Jensen, 2015; McCrae & Costa, 2006).

4. Association between big five personality traits and learning

Extensive research has examined the associations between personality traits and student learning experience (e.g., Khalilzadeh & Khodi, 2021; Komarraju et al., 2011; Yu et al., 2021). In general, conscientiousness, agreeableness, and openness are positively associated with academic achievement, with conscientiousness demonstrating the strongest correlation (Abou Assali, 2025; Chamorro-Premuzic & Furnham, 2003; Chen-Jung et al., 2021; Duff et al., 2004; Komarraju et al., 2011). Conversely, neuroticism and extraversion correlate negatively with academic performance (Chamorro-Premuzic & Furnham, 2003; Yu, 2021). For example, Yu (2021) found that students higher in agreeableness, conscientiousness, and openness achieved better outcomes in online learning than their extraverted and neurotic peers. Similarly, a meta-analysis by Abou Assali (2025) revealed that teachers' conscientiousness, openness, and agreeableness positively predict student achievement and engagement, with conscientiousness as the strongest factor, while neuroticism negatively affects students' emotional regulation and well-being. These findings underscore the influence of both student and teacher personality traits on student learning outcomes.

The association between personality traits and learning can be understood by examining their relationships with social behaviors and cognitive processing. Among the five traits, extraversion and agreeableness have been suggested to significantly influence social cooperation and appraisal of others (Fors Connolly & Johansson Sevä, 2021). However, extraverted students may struggle with deep, reflective problem-solving, as their tendency toward cognitive closure favors quick resolution over sustained engagement, potentially increasing susceptibility to distraction and limiting knowledge absorption (Matthews & Zeidner, 2004; Ray et al., 2025). Conversely, agreeable students, characterized by cooperativeness and empathy, are more likely to foster social harmony and support group interactions, without the premature closure tendencies observed in extraverts. Openness and conscientiousness, on the other hand, are closely associated with engagement in diverse cognitive and metacognitive learning processes. For example, Bidjerano and Dai (2007) reported a positive relationship between openness and conscientiousness with metacognition, particularly in terms of awareness and control of one's knowledge and cognitive processes. Komarraju et al. (2011) further observed that conscientiousness and openness positively correlated with various learning styles, including synthesis-analysis, methodical study, fact retention, and elaborative processing. In contrast, neuroticism negatively correlated with synthesis-analysis, a learning style associated with enhanced learning outcomes. Furthermore, Susanti et al. (2024) found that openness facilitates critical thinking, problem-solving, and enthusiasm for exploring advanced knowledge. These findings suggest that personality traits shape social, cognitive, and metacognitive dimensions of the learning process, with specific traits either enhancing or hindering particular learning strategies and problem-solving approaches.

5. The personality of AI agents

As AI technologies become increasingly prevalent, research has begun to investigate the personality traits of AI agents. The personality of virtual agents is typically conveyed through two primary approaches. The first approach involves embedding multimodal communication elements such as hand gestures, facial expressions, and body movements, to project personality characteristics. The second approach leverages advancements in LLMs like GPT, which can consistently display personality cues and be tailored to exhibit specific traits (Huang et al., 2023). By strategically prompting these models, researchers can design AI agents with targeted personalities (Gu et al., 2023; Sonlu et al., 2024), as demonstrated in this study. Data-driven personality estimation systems further suggest that specific prompts can evoke distinct personality types in LLMs, demonstrating their capacity to convey personality through text (Karra et al., 2022; Mehta et al., 2020).

Although limited, some studies have begun to examine how the personalities of pedagogical agents impact student experience. For instance, Liew and Tan (2016) explored the effects of extraverted and introverted agents on students' emotions, motivation, and learning in a web-based programming lesson, finding that students interacting with agents with mismatched personalities (e.g., extraverted students with introverted agents) reported higher positive emotions and motivation, though there was no significant impact on learning outcomes. Ruane et al. (2021) examined college students' engagement with a text-based chatbot displaying extraversion and agreeableness traits in discussions on university-related topics. They found that students exhibited lower engagement with agents that displayed high levels of extraversion and agreeableness than with those displaying lower levels of these traits. More recently, Bian and Zhou (2022) investigated the impact of animated agents' personalities (extraversion and neuroticism) on students' motivation in a virtual learning environment, concluding that personality exerted minimal influence on motivation or learning outcomes. In summary, prior work has primarily focused on one or two of the Big Five personality traits within AI agents, establishing a preliminary understanding of how these traits influence student engagement, emotion, and motivation. However, the impact of a comprehensive range of AI agent personality traits on learning remains underexplored, particularly regarding how each of the Big Five traits may affect learning, and how these effects operate in teachable agents. This study seeks to address this gap by systematically examining the influence of all Big Five personality traits in teachable AI agents within a math learning environment.

6. Present study

This study investigates the impact of AI-powered teachable agents' personality traits on middle school students' mathematics learning experiences. Using custom LLMs and advanced prompt engineering, we developed teachable AI agents designed to emphasize each of the Big Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness) or to operate without a specific personality emphasis (i.e., six conditions). Over a three-week experiment, 534 middle school students were asked to tutor agents across these personality conditions in mathematics problem-solving. Since prior research highlights that the effectiveness of learning-by-teaching is strongly influenced by tutor-tutee interactions (Cohen, 1986; Gartner, 1971; Lyu et al., 2024; Roscoe & Michelene, 2007), we collected data on students' interactions with the agents. Building on literature in learning-by-teaching (Biswas et al., 2005), teachable agents (Biswas et al., 2010; Han et al., 2021; Pareto, 2014), and personality traits (Fors Connolly & Johansson Sevä, 2021; Duff et al., 2004; Komarraju et al., 2011), students' interaction behaviors were classified into three main types: affective expressions, cognitive scaffolding, and metacognitive reflection. This coding scheme allowed us to systematically examine variations in student interactions based on the agents' personality traits. Furthermore, to assess the effectiveness of tutoring agents with different personalities in supporting mathematics learning, we examined how agents facilitated students' integration and construction of mathematical knowledge during the tutoring process, focusing specifically on conceptual and procedural knowledge. Additionally, we conducted a post-study test to further measure students' mathematics learning.

Based on previous research linking personality traits to learning outcomes, we hypothesized that students interacting with teachable agents emphasizing traits of agreeableness and extraversion would be expected to exhibit more affective expressions, while agents emphasizing conscientiousness and openness would promote cognitive and metacognitive interactions. Given the nature of LLMs' training, which involves processing diverse textual data that inherently reflect a range of personality traits and adjusting their responses dynamically based on user interactions, we anticipated that non-personality-emphasis agents may exhibit a more adaptive personality profile by adjusting their traits in response to students' inputs. This adaptability may enable non-personality-emphasis agents to activate personality traits such as openness and conscientiousness in problem-solving contexts, facilitating cognitive engagement. For math learning outcomes, since non-personality-emphasis agents can adapt their responses based on specific interaction contexts and student inputs, potentially supporting a range of cognitive and social engagement, we did not formulate specific hypotheses regarding their effectiveness in improving learning outcomes in comparison to other types of agents.

It is important to note that students' characteristics, including prior AI familiarity and personality traits, were not assessed in this study. We acknowledge this as a limitation, as these factors could moderate or mediate students' interactions and learning outcomes with the agents (Li et al., 2025). Future research should account for these variables when examining the role of pedagogical agents' personalities in learning. Despite these limitations, this study addresses several gaps in the existing literature by investigating the impact of teachable agents' personalities on students' interaction behaviors and mathematics learning outcomes. First, rather than positioning students as passive recipients of information, this study examines teachable agents that foster active engagement and learner autonomy, and explores the effects of their personalities on learning, an area that remains underexplored in Educational Technology research. Second, this study extends prior research by analyzing teachable agents' personalities using all five Big Five personality traits, rather than focusing on only one or two traits, as in previous studies. Third, by analyzing both students' interactions with the teachable agents and mathematics learning outcomes, this study provides insights into how teachable agent personalities influence learning processes and academic performance, rather than focusing solely on engagement metrics, as emphasized in prior research. The findings of this study contribute to the design of future pedagogical agents, enhancing support for multiple dimensions of student learning.

7. Methods

7.1. Our teachable agent

Our teachable agent follows a four-step instructional structure. Fig. 1 illustrates the interface of the teachable agent, highlighting this structured approach. First, the teachable agent asks a word problem aligned with classroom content. Second, the teachable agent presents two contrasting solutions to the word problem, prompting students to identify which solution contains a misconception requiring correction. Students are then asked to justify their choice and provide instructions for accurate problem-solving. Third, an AI-enabled mentor agent delivers feedback on students' responses and provides guidance on effective instruction. The inclusion of contrasting solutions and the mentor agent were designed to support students in adopting the role of "teachers" to the agents, particularly to help those with limited prior knowledge engage in meaningful teaching by providing structure and guidance. Finally, the teachable agent poses questions about the word problem, and students engage in a teaching role, explaining concepts to the agent. Given that this study primarily investigates the personality traits of teachable agents, all features besides teachable agents' personalities (e.g., mentor agent) were kept consistent across all agents.

7.2. Personality in our teachable agent

For the five types of teachable agents with an emphasized Big Five personality, we engineered the personality traits using few-shot prompt engineering. Specifically, for each personality trait, we collaborated with three math education experts to develop few-shot examples representing three core scenarios: initiating help-seeking, facing a challenge, and experiencing success. For example, a teachable agent emphasizing extraversion was provided with the following responses: (1) Help-seeking: "*Hey, can we go over that part where we subtract again? I think I missed a step. Let's figure it out together!*" (2) Challenge: "*Whoa, this one's tricky! Can you explain that one more time but slower? Like, what do we do after this step?*" (3) Success: "*Yes! I got it! That makes so much sense now. That's not so bad.*" The complete prompt used for the teachable agents and their assigned personality traits is presented in Appendix A. For the non-personality-emphasis teachable agents (i.e., the control group), no prompts were used to reinforce a specific personality trait.

To assess whether the teachable agents' responses maintained consistency with their assigned personalities, we conducted a post-hoc content analysis in two phases. First, we randomly selected 250 student-agent conversations and analyzed whether the agent's responses aligned with its assigned personality. Results indicated that 83% of the conversations (n = 207) demonstrated alignment with the designated personality trait. Second, to enhance the validity of the results, we further randomly selected an additional 150 conversations to conduct a blind coding procedure in which coders were unaware of the experimental manipulation, reducing potential bias and ensuring that personality classifications were based solely on agent responses. The coders classified the agent in each of the 150 conversations into one of the Big Five personality traits based on the agent's response in the full conversation. The results indicated that 78 % of the conversations (n = 117) aligned with the expected personality, demonstrating strong personality consistency even when coding was conducted independently (see Table 1 for the descriptives and Table 2 for sample agent responses corresponding to each emphasized personality).

7.2.1. Context

This study was conducted within Math Nation, a digital mathematics learning platform with an annual user base exceeding one million. Math Nation is designed to support mathematics education for middle and high school students across the United States,

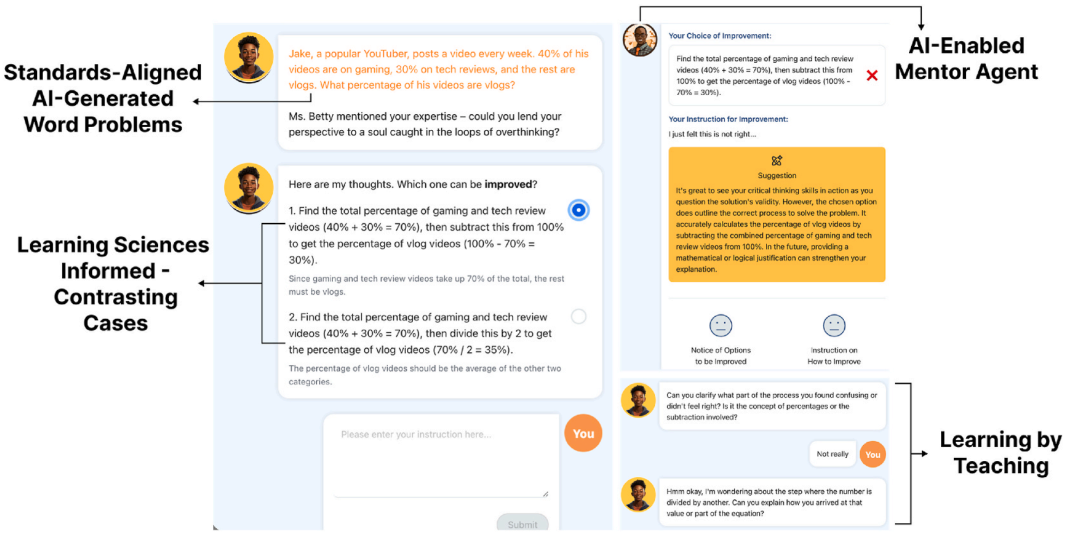

Fig. 1. AI-powered teachable agent embedded in Math Nation.

Table 1
Alignment between agent responses and experiment assigned personality.

Coded Personality Given Agent Responses	Experiment-Assigned personality	Alignment	Frequency
Extraversion	Extraversion	Yes	23
	Openness	No	1
	Conscientiousness	No	1
	Neuroticism	No	1
Agreeableness	Agreeableness	Yes	26
	Extraversion	No	2
	Conscientiousness	No	2
	Neuroticism	No	1
	Openness	No	1
Conscientiousness	Conscientiousness	Yes	23
	Openness	No	3
	Agreeableness	No	2
	Neuroticism	No	2
Neuroticism	Neuroticism	Yes	19
	Conscientiousness	No	3
	Extraversion	No	3
	Openness	No	1
	Agreeableness	No	1
Openness	Openness	Yes	26
	Neuroticism	No	5
	Conscientiousness	No	2
	Extraversion	No	1
	Agreeableness	No	1

Table 2
Sample agent responses of each personality.

Personality	Sample Responses
Extraversion	"Hey there! I'm really trying to get better at interpreting these graphs. Could you explain more about how each part of the stem-and-leaf plot specifically contributes to understanding the data? Thanks a bunch!"
	"Oh, that clears it up! So, each number in the leaf is literally a text sent, and we sum them up. Thanks for sticking with me on this! You're awesome at explaining! Take care and catch you later! Keep rocking math!"
Agreeableness	"Thank you for the feedback! I see where I need to clarify. Could you please explain why exactly we add the measures of angles (VYM) and (MYL) to find (VYL)? I want to fully understand this concept."
	"Got it, thanks for clarifying! So, we add the measures of (VYM) and (MYL) to find (VYL) because they're adjacent angles, right? Just trying to make sure I have this straight!"
Conscientiousness	"Returning to our discussion, could we delve a bit deeper into how linear relationships are represented in real-life examples like data usage over time?"
	"Interesting! So the slope represents how much my data usage increases each hour. Could we explore how this concept applies to different scenarios, such as scaling down or non-linear growth?"
Neuroticism	"I'm really struggling to grasp this, and it's making me anxious. If subtracting the exponents is correct, why does k^{11}/k^6 not equal k^{17}? Could you explain where I might be going wrong?"
	"I'm sorry, it's just that sometimes I misinterpret things when I'm stressed, and it's really hard to focus. So, it's subtraction … does that mean k^{11}/k^6 equals k^5? Is that how it works?"
Openness	"I'm curious, could we explore further how subtracting 20 before dividing might provide a clearer path to isolating (x) in such linear equations? It's intriguing to see how different operations impact the solution."
	"Oh, got it! The result is 36. Could you perhaps walk me through how we arrived at that number? I want to understand the steps involved."

facilitating both hybrid classroom learning and independent online study. The platform offers a range of mathematics courses, including Algebra, Geometry, and SAT preparation, each structured into multiple units that cover distinct mathematical concepts and are further divided into several lessons. Within each lesson, students have access to various learning resources, including instructional videos, self-assessment quizzes, and supplementary educational materials. Our AI-powered teachable agent was integrated into the Algebra and Geometry learning units on Math Nation to serve as a learning resource for students. Prior to the launch of our teachable agents, a professional development session was conducted for middle school math teachers in a partner school district in the Southeastern United States to introduce them to the tool. Once the teachable agents became available on Math Nation, these middle school math teachers were informed of its availability and encouraged to ask their students to use it.

7.2.2. Experiment

We employed a pre-posttest design to examine the impact of teachable agents' personalities on students' mathematics learning experiences. Data were collected over an approximate five-week period. An optional pre-test was administered in Math Nation before students accessed the agents, and students who completed the pre-test were automatically assigned a post-test in Math Nation after the study. Most students completed the pre-test during the first week and the post-test during the final week, working independently. The interaction phase, lasting approximately three weeks, allowed students to engage with the teachable agents in a naturalistic setting, where they could freely choose when and how frequently to work with the AI tool on math problems. During the study, students could interact with the teachable agents on as many math problems as they desired. Students' interactions with the AI agents were recorded through system logs, capturing the number of problems attempted and messages exchanged.

At the problem level, agents with and without specific personality emphases were randomly assigned to students. For problems involving teachable agents with a personality emphasis, each agent was randomly assigned one of five AI personality traits. Approximately 50% of the problems were presented with non-personality-emphasis agents, while the remaining 50% were distributed among agents emphasizing each of the five personality traits, with each trait represented in approximately 10% of the problems. This design allowed students to experience a variety of AI teachable agent personalities across different problems, with each problem featuring a single, designated personality for the teachable agent. Throughout the interactions, students tutored the AI teachable agents, providing explanations and guidance, which facilitated an interactive learning environment. Over the course of the study, all participants collectively completed 2946 math problems and engaged with the AI tool a total of 16,085 times (i.e., messages exchanged).

7.2.3. Participants

Students were recruited from middle school mathematics classes in a partner school district in the Southeastern United States. Participation was voluntary, and while teachers participating in the professional session encouraged students to use the AI tool to reinforce classroom learning, its use was not required for students. To be included in the final analysis, students had to engage with the teachable agent within the five-week data collection period and complete both the pre- and post-tests. A total of 534 middle school students (grades 6 and 7) participated in the study. Due to limitations in data collection, demographic information regarding gender and race was not available. However, the partner school district's demographic composition in 2024 was approximately 44.6% White, 14.5% Black, 30.1% Hispanic/Latino, and 10.8% multi-racial/other, which is expected to be representative of the study sample.

7.3. Learning measures

We evaluated the effectiveness of teachable agents with varying personalities in supporting mathematics learning through three measures: a pre- and post-mathematics knowledge test, students' demonstration of procedural knowledge during teaching, and students' demonstration of conceptual knowledge during teaching.

7.3.1. Math knowledge test

Students' mathematical knowledge was assessed using a mathematics knowledge test administered twice: once before the study (pre-test) and once after the study (post-test). Serving as a formative assessment tool, this test provided insights into students' mathematics learning as they accessed and interacted with the teachable agents featuring varying personality traits. The test consisted of 15 items randomly selected from a test bank developed by Math Nation for Algebra and Geometry learning, designed to align with the Florida Assessment of Student Thinking (FAST), as mandated by the Florida Department of Education. All 15 items were multiple-choice questions. To minimize potential test familiarity effects, the 15 items were presented in a different randomized order in the pre-test and post-test, with a minimum interval of three weeks between administrations. Each item was automatically scored, with one point awarded per correct response, resulting in a possible score range of 1–15 for both the pre- and post-tests. Proportion scores were used in the analysis. The mean pre-test score was 0.40 ($SD = 0.20$), and the mean post-test score was 0.47 ($SD = 0.25$).

7.3.2. Demonstration of mathematical knowledge

Each of the students' interactive messages with the teachable agents was coded for the presence of conceptual and procedural mathematics knowledge, with these two types of knowledge not being mutually exclusive. The coding scheme was binary (i.e., 0 or 1). For example, when a student successfully demonstrated and provided a correct explanation of conceptual knowledge in a message, that message was coded as "conceptual knowledge = 1"; otherwise, it was coded as "conceptual knowledge = 0." Sample messages illustrating conceptual and procedural knowledge are as follows:

> **AI**: Could you maybe help explain how to identify the central tendency in a dataset?
>
> **Student**: In a dataset, there are usually a bunch of numbers in random order, so you would want to organize those numbers from smallest to largest. Mean is adding all the numbers up and dividing it with how many numbers are in the dataset (***Procedural knowledge***). Mode is the value or values that occur most often (***Conceptual knowledge***).

To establish acceptable inter-rater reliability, two raters independently coded 100 complete tutoring conversations, each conversation involving the solution of a math problem, resulting in a kappa value of 0.94. Natural language processing (NLP) models were subsequently employed to automatically analyze the presence of knowledge (see Text Classification section). The mean frequency of demonstrating procedural knowledge per math problem was 0.34 ($SD = 0.45$), while the mean frequency of conceptual knowledge demonstration was 0.41 ($SD = 0.59$).

7.4. Students' interactions with the teachable agents

Bottom-up and top-down approaches were employed to develop a coding scheme to assess students' interactions with the teachable agents. Specifically, based on our data, we initially coded each student message for the presence of 8 specific interaction behaviors (e.

g., giving explanations for concepts, instructing the AI agents on how to solve the problem, encouraging and praising AI agents), which were not mutually exclusive, aiming to capture students' specific interaction behaviors as closely as possible. According to prior research on learning-by-teaching and personality traits (Baranwal, 2022; Biswas et al., 2005; Chamorro-Premuzic & Furnham, 2003; Chen-Jung et al., 2021; Fors Connolly & Johansson Sevä, 2021; Komarraju et al., 2011), these interaction behaviors were subsequently classified into three categories: affective expressions, cognitive support, and metacognitive behaviors. The cognitive support was categorized into cognitive elaboration and cognitive acknowledgment to further capture the depth of students' cognitive involvement. Cognitive elaboration focuses on providing detailed explanations and instructions to support the AI's learning, while cognitive acknowledgment involves confirming the AI's responses or directly providing solutions to its questions. See Table 3 for the definitions and examples of interaction behaviors. Two raters independently coded 100 complete tutoring conversations to establish an acceptable inter-rater reliability, resulting in a kappa value of 0.80. NLP models were then used to automatically analyze students' interaction behaviors (see Text Classification section).

7.5. Text classification

All student conversation logs generated in Math Nation were stored in a Google Firebase database to extract a comprehensive dataset. This dataset was divided into a training set and a prediction (or inference) set, ensuring that model development and validation were clearly separated. We conducted the training process on a GPU cluster equipped with P100 graphics cards under an Ubuntu environment, providing ample computational resources. After training, the models were used to predict new conversation logs for further analysis. A scheduled pipeline can be implemented to deploy updated models to the GPU cluster, enabling continual prediction. However, the design of this pipeline should carefully account for potential cost considerations. During training, we tuned only the key hyperparameters of each model, keeping the remaining parameters at their default values.

An additional 594 conversations were coded by the first rater to provide more instances for training and evaluating the NLP models to automatically analyze students' interaction behaviors and presence of knowledge. On average, there were 13.06 dialogue turns per conversation, with a total of 9063 student messages. After removing the AI messages, student messages were labeled with 8 interaction behaviors and 2 mathematics knowledge. Once the labeling was completed, we proceeded to design the text classification models.

To identify the optimal text classification model (Breiman, 2001), we employed six models, including two traditional machine learning models, two deep learning models, and two LLMs. Among the traditional machine learning models, Random Forest is well known for its robustness in machine learning tasks and its ability to handle high-dimensional data. By combining multiple decision trees, it reduces the risk of overfitting and effectively handles large datasets. Gaussian Naive Bayes (GNB), on the other hand, is favored for its simplified assumption of feature independence and computational efficiency in high-dimensional data processing (Murphy, 2006), especially in cases where features are independent or nearly independent. In our experiments, we used default parameters for RF and GNB, with the exception of adjusting the number of estimators ('n_estimators') in RF by comparing 50, 100, 150, and 200, and then selecting the optimal value.

In contrast, deep learning models might outperform in capturing patterns from large-scale data. BERT (Bidirectional Encoder Representations from Transformers, Devlin, 2018) leverages large-scale pre-trained models for transfer learning and utilizes bidirectional attention mechanisms to capture contextual information effectively. Similarly, RoBERTa (Robustly Optimized BERT Approach, Liu, 2019) further optimizes BERT through larger training datasets and longer training durations, showing significant performance improvements in language understanding tasks, particularly when handling imbalanced datasets or lengthy texts. Both of deep learning models were trained using consistent strategies: (1) we compared $'\mathrm{maxlen} = 256'$ and $'\mathrm{maxlen} = 512'$ to determine the best input sequence length; (2) we used a batch size of 8 and trained for up to 50 epochs with an early stopping strategy training ceased if the F1 score did not improve for 10 consecutive epochs; (3) the learning rate was selected from 1e-3, 1e-4, and 1e-5, using validation performance as the criterion; and (4) AdamW served as the optimizer, chosen for its decoupled weight decay that can improve generalization.

Regarding the LLMs, Mistral and LLaMA 3 represent the latest advancements in this field. Mistral focuses on robustness and open optimization (Jiang et al., 2023), delivering excellent performance even in resource-constrained environments, making it ideal for applications requiring efficient inference. Meanwhile, LLaMA 3, with its multi-level pretraining strategy and enhanced model architecture (Touvron et al., 2023), has further improved adaptability and versatility across a wide range of tasks, exhibiting greater expressiveness and flexibility in handling text classification tasks. Both of LLMs were trained under the same protocol. Specifically, we adopted a $'\mathrm{max\_length} = 1024'$ setting and applied Low-Rank Adaptation (LoRA), which adds lightweight, low-rank parameter matrices to selected layers to reduce memory usage and computational overhead. We tuned the learning rate from 1e-4, 2e-4, and 3e-4, and set the batch size to 4. Similar to the deep learning models, we employed AdamW as the optimizer and trained for up to 8 epochs. By leveraging the strengths of these models, we were able to comprehensively explore their applicability and performance in text classification tasks, thus identifying the best model.

During training, we employed a multi-output head for all deep learning models to train a multi-target model with 10 output labels, while the traditional models trained one binary classification model per label. We evaluated the models using common multi-class metrics, such as accuracy and F1 score, and divided the data into an 80 % training set and a 20 % validation set, with each sample treated as a unit and all dialogues from the current turn embedded as context. The average values of accuracy and F1 score for knowledge presence and each interaction category are shown in Table 4. LLaMA 3 performed the best in classifying the 10 different mathematics discussion tags. Therefore, we used this model to predict and expand our dataset for further analysis.

Table 3
Definitions and examples of interactive behaviors.

Category	Specific Interaction Types	Definitions	Examples
Affective Expressions	Polite expressions	Language or phrases used to show respect, aimed at fostering a respectful interaction with the AI.	“You’re welcome, my pleasure!”
	Encouragement	Statements intended to motivate and praise the AI to support AI in the teaching process.	“Let’s take it step by step, don’t worry, you did a good job so far!”
Cognitive Elaboration	Explanations	Providing more detailed information to explain conceptual or procedural knowledge.	“The median might be more effective because it might be more accurate because it won’t include the outlier, so it won’t be skewed.”
	Instructions	Providing structured guidance on how to solve the math problem, such step-by-step instructions to solve a problem	“You multiply base times height to get the area“
Cognitive Acknowledgment	Confirmative feedback	Providing statements to affirm AI’s responses	“Yes, you are correct.”
	Giving answers	Directly providing the answer to a question posed by the AI	“The answer is 36”
Metacognitive Behaviors	Confusion	Asking questions to seek clarification on a problem or expressing uncertainty	“I’m not quite sure” “Can you elaborate a bit?”
	Corrections	Providing corrections to inaccurate responses from themselves and the AI	“Wait, no, I’m wrong actually the pie chat is more affective for this situation”

8. Results

RQ1. What are the effects of teachable agents' personality traits on students' interactions with the agents?

To understand how teachable agents' personalities influence students' learning experience with the agents, we first investigated the effects of agents' personality types on student-agent interactions. Since the personalities of the teachable agents were manipulated across problems, and problems were nested within students, we applied a series of linear mixed models (LMMs) with random intercepts to account for the hierarchical structure of the data. In these models, students were treated as random effects, and each model examined a specific student interaction behavior as the dependent variable. The personality types of the teachable agents [i.e., agreeableness ($n = 364$), conscientiousness ($n = 333$), extraversion ($n = 425$), openness ($n = 350$), neuroticism ($n = 365$), and no personality emphasis ($n = 1109$)] were included as fixed effects. The results of the LMMs revealed that agents' personalities had a significant effect on students' provision of polite expressions, $F(5, 2915.40) = 4.475, p < .001$, encouragement, $F(5, 2786.01) = 3.138$, $p < .001$, explanations, $F(5, 2915.63) = 6.899, p < .001$, instructions, $F(5, 2897.78) = 4.212, p < .001$, and confirmative feedback, $F(5, 2922.58) = 5.07, p < .001$, during interactions with the teachable agents (see Table 5 for detailed results about the significant models).

Regarding affective expressions, pairwise comparisons revealed that students interacting with agreeableness-emphasis agents exhibited significantly more ***polite expressions*** than those engaging with non-personality-emphasis ($b = 0.25$, $p = .003$), conscientiousness-emphasis ($b = 0.31$, $p = .002$), and openness-emphasis agents ($b = 0.29$, $p = .004$). Similarly, students interacting with extraversion-emphasis agents demonstrated more ***polite expressions*** than those engaging with non-personality-emphasis ($b = 0.25$, $p = .002$), conscientiousness-emphasis ($b = 0.31$, $p = .002$), and openness-emphasis agents ($b = 0.29$, $p = .003$). Furthermore, students interacting with agreeableness-emphasis agents provided significantly more ***praise and encouragement*** than those engaging with agents emphasizing conscientiousness ($b = 0.04$, $p = .002$), extraversion ($b = 0.04$, $p = .003$), openness ($b = 0.05$, $p < .001$), and neuroticism ($b = 0.03$, $p = .02$), or non-personality-emphasis agents ($b = 0.04$, $p < .001$).

For cognitive elaborations, students interacting with openness-emphasis agents offered more ***explanations*** than those engaging with agreeableness-emphasis ($b = 0.32$, $p < .001$), extraversion-emphasis ($b = 0.28$, $p < .001$), and conscientiousness-emphasis ($b = 0.25$, $p = .004$) agents. Also, students interacting with non-personality-emphasis agents provided more ***explanations*** than those interacting with agreeableness-emphasis ($b = 0.29$, $p < .001$), extraversion-emphasis ($b = 0.25$, $p < .001$), and conscientiousness-emphasis ($b = 0.22$, $p = .002$) agents. Additionally, students interacting with neuroticism-emphasis agents provided more ***explanations*** than those engaging with agreeableness-emphasis agents ($b = 0.19$, $p = .02$). Moreover, students interacting with agreeableness-emphasis ($b = 0.22$, $p = .003$), conscientiousness-emphasis ($b = 0.24$, $p = .002$), openness-emphasis ($b = 0.19$, $p = .01$), and non-personality-emphasis ($b = 0.26$, $p < .001$) agents gave ***instructions*** more frequently than those interacting with extraversion-emphasis agents.

Regarding cognitive acknowledgment, students interacting with neuroticism-emphasis agents offered ***confirmative feedback*** less frequently than those engaging with agents emphasizing other personality types (agreeableness: $b = -0.14, p = .05$, conscientiousness: $b = -0.15, p = .03$, extraversion: $b = -0.17, p = .01$, openness: $b = 0.21, p = .005$) or non-personality-emphasis agents ($b = -0.28, p < .001$). Notably, non-personality-emphasis agents facilitated ***confirmative feedback*** more effectively than agents emphasizing agreeableness ($b = 0.14$, $p = .02$), conscientiousness ($b = 0.13$, $p = .03$), or extraversion ($b = 0.11$, $p = .04$) personality traits. For detailed descriptive statistics, refer to Table 6, and for comparisons among teachable agent types on statistically significant interaction behaviors, see Fig. 2.

In summary, the findings indicated that teachable agents' personality traits significantly influenced students' interaction behaviors. Agents emphasizing agreeableness and extraversion prompted students' engagement in polite expressions and encouragement, while openness and non-personality-emphasis agents facilitated their deeper cognitive elaborations. Neuroticism-emphasis agents led to fewer confirmative feedback instances, whereas non-personality-emphasis agents encouraged students to offer more.

RQ2. How do teachable agents' personality types and students' interactions with the agents predict students' mathematics learning outcomes?

Table 4
The performance of six text classification models.

Interaction Behaviors and Knowledge	RF (F1, Acc)	GNB (F1, Acc)	BERT (F1, Acc)	Roberta (F1, Acc)	Mistral-7b (F1, Acc)	LLaMa3-8b (F1, Acc)
Polite expressions	(0.828, 0.830)	(0.435, 0.415)	(0.940, 0.940)	(0.941, 0.943)	(0.958, 0.959)	(0.961, 0.962)
Encouragement	(0.982, 0.985)	(0.860, 0.770)	(0.984, 0.989)	(0.984, 0.989)	(0.992, 0.993)	(0.991, 0.992)
Explanations	(0.824, 0.828)	(0.469, 0.428)	(0.895, 0.898)	(0.884, 0.887)	(0.917, 0.918)	(0.926, 0.928)
Instructions	(0.828, 0.838)	(0.472, 0.417)	(0.881, 0.886)	(0.890, 0.891)	(0.919, 0.921)	(0.912, 0.915)
Confirmation	(0.836, 0.838)	(0.366, 0.319)	(0.955, 0.954)	(0.943, 0.942)	(0.961, 0.962)	(0.955, 0.955)
Giving answers	(0.972, 0.972)	(0.849, 0.758)	(0.987, 0.988)	(0.987, 0.988)	(0.986, 0.989)	(0.986, 0.987)
Confusion	(0.911, 0.913)	(0.608, 0.487)	(0.946, 0.947)	(0.945, 0.947)	(0.976, 0.976)	(0.971, 0.971)
Corrections	(0.980, 0.979)	(0.859, 0.766)	(0.986, 0.991)	(0.986, 0.991)	(0.986, 0.991)	(0.989, 0.992)
Procedural KG	(0.853, 0.860)	(0.583, 0.502)	(0.910, 0.914)	(0.915, 0.917)	(0.889, 0.906)	(0.889, 0.903)
Conceptual KG	(0.871, 0.872)	(0.562, 0.479)	(0.904, 0.907)	(0.895, 0.901)	(0.864, 0.893)	(0.904, 0.910)
Average	(0.889, 0.892)	(0.606, 0.534)	(0.939, 0.941)	(0.937, 0.940)	(0.945, 0.951)	**(0.948, 0.952)**

Note. KG represents Knowledge. Acc refers to Accuracy.

Table 5
The effects of agents' personality types on students' interactions behaviors.

	Polite Expressions	Encouragement	Explanations	Instructions	Confirmative Feedback
Fixed Effects					
Intercept	$b = 1.10, p < .001$	$b = 0.04, p < .001$	$b = 0.85, p < .001$	$b = 0.61, p < .001$	$b = 0.57, p < .001$
Personality types	$F = 4.48, p < .001$	$F = 3.13, p < .001$	$F = 6.90, p < .001$	$F = 4.21, p < .001$	$F = 5.07, p < .001$
Random Effects					
σ^2	1.49	0.03	1.03	0.95	0.76
τ_{00}	0.38	0.00	0.22	0.12	0.11
ICC	0.20	0.08	0.18	0.11	0.13
Marginal R^2	0.01	0.01	0.01	0.01	0.01
Conditional R^2	0.21	0.08	0.19	0.12	0.13

We conducted three regression analyses to predict students' math learning from personality types of teachable agents and students' interactions with the agents. Because the personalities of the teachable agents varied across problems and each student interacted with a different number of agents with varying personalities, the percentages of exposure to each agent personality type was used as predictors to explore how each type of agents' personality predict math learning. In Step 1, we included pre-test scores and the percentages of agent personalities as predictors, with percentage of non-personality-emphaisis agents as the reference group. In Step 2, students' interactions with the agents were added as predictors. The dependent variables were three math learning outcomes: conceptual knowledge demonstration, procedural knowledge during tutoring, and post-test scores (see Table 7 for the regression models).

With respect to conceptual knowledge during tutoring, the regression model for the personality types of teachable agents was significant, $F(6, 527) = 2.56$, $p = .02$, $R^2_{\text{adj}} = 0.02$. The inclusion of agreeableness-emphasis teachable agents, $\beta = -0.09$, $p = .04$, negatively predicted students' demonstration of conceptual knowledge compared to inclusion of agents without a personality emphasis, while controlling for the inclusion of other personality types of agents. Also, pre-test scores, $\beta = 0.12$, $p = .007$, positively predicted students' demonstration of conceptual knowledge. The regression model predicting conceptual knowledge from both personality type and students' interaction behaviors was also significant, $F(14, 519) = 30.40$, $p < .001$, $R^2_{\text{adj}} = 0.44$. The addition of students' interaction behaviors resulted in a statistically significant 42% increment in explained variance, with F_{change} (8, 519) = 49.86, $p < .001$, $R^2_{change} = 0.42$. The difference between agreeable teachable agents versus no-personality-emphasis agents on conceptual knowledge was no longer significant after accounting for students' interaction behaviors, $\beta = -0.03$, $p = .45$. This suggests that the personalities of teachable agents affect students' demonstration of conceptual knowledge during tutoring through their interactions with the agents. After controlling for the personality types of the agents, pre-test scores, $\beta = 0.09$, $p = .005$, and providing explanations, $\beta = 0.62$, $p < .001$, positively predicted students' demonstration of conceptual knowledge, but confusion expressions, $\beta = -0.10$, $p = .003$, negatively predicted students' demonstration of conceptual knowledge.

For procedural knowledge during tutoring, the regression model with pre-test scores and personality types was not significant, $F(6, 527) = 1.98$, $p = .07$, indicating that personality types did not predict students' demonstration of procedural knowledge. However, the regression model became significant after accounting for students' interaction behaviors, $F(14, 519) = 4.14$, $p < .001$, $R^2_{\text{adj}} = 0.08$. The addition of students' interaction behaviors resulted in a statistically significant 8% increment in explained variance, with F_{change} (8, 519) = 5.65, $p < .001$, $R^2_{change} = 0.08$. Giving encouragement, $\beta = 0.08$, $p = .05$, and providing instructions, $\beta = 0.21$, $p < .001$, positively predicted the demonstration of procedural knowledge, suggesting that students who provided more encouragement and instructions to the agents demonstrated higher procedural knowledge. However, expressions of confusion, $\beta = -0.13$, $p = .002$, negatively predicted students' demonstration of procedural knowledge, indicating that students who expressed confusion during the tutoring process showed lower procedural knowledge. After controlling for interaction behaviors, inclusion of teachable agents with conscientiousness personality trait positively predicted procedural knowledge compared to non-personality-emphasis agents, $\beta = 0.09$, $p = .04$. This result indicates that students exposed to agents with conscientiousness personality traits demonstrate higher procedural knowledge than those interacting with non-personality-emphasis agents.

Regarding post-test scores, the regression model with pre-test scores and personality types was significant, $F(6, 527) = 14.82$, $p < .001$, $R^2_{\text{adj}} = 0.14$. Pre-test scores was a significant positive predictor, $\beta = 0.38$, $p < .001$. However, the teachable agents' personality did not predict students' post-test scores as compared to non-personality-emphasis agents. The regression model predicting post-test scores

Table 6
Descriptive statistics on students' interactions with agents across personality types.

Interaction Types	Agreeableness *M (SD)*	Conscientiousness *M (SD)*	Extraversion *M (SD)*	Neuroticism *M (SD)*	Openness *M (SD)*	No personality *M (SD)*
Polite expressions	1.19(1.54)	0.87(1.25)	1.21(1.50)	0.99(1.39)	0.93(1.30)	0.99(1.32)
Encouragement	0.07(0.27)	0.02(0.13)	0.03(0.17)	0.03(0.17)	0.02(0.13)	0.03(0.16)
Explanations	0.63(0.98)	0.68(1.07)	0.67(0.92)	0.79(1.09)	0.95(1.23)	0.95(1.18)
Instructions	0.72(1.04)	0.74(1.12)	0.52(0.92)	0.62(0.99)	0.67(0.93)	0.77(1.09)
Confirmation	0.71(0.90)	0.71(0.98)	0.77(0.93)	0.55(0.78)	0.76(0.97)	0.87(0.97)
Giving answers	0.03(0.21)	0.04(0.24)	0.03(0.20)	0.04(0.22)	0.05(0.27)	0.04(0.21)
Confusion	0.56(1.26)	0.80(1.71)	0.58(1.10)	0.74(1.63)	0.69(1.44)	0.53(1.13)
Corrections	0.00(0.10)	0.01(0.10)	0.01(0.10)	0.01(0.11)	0.00(0.10)	0.01(0.10)

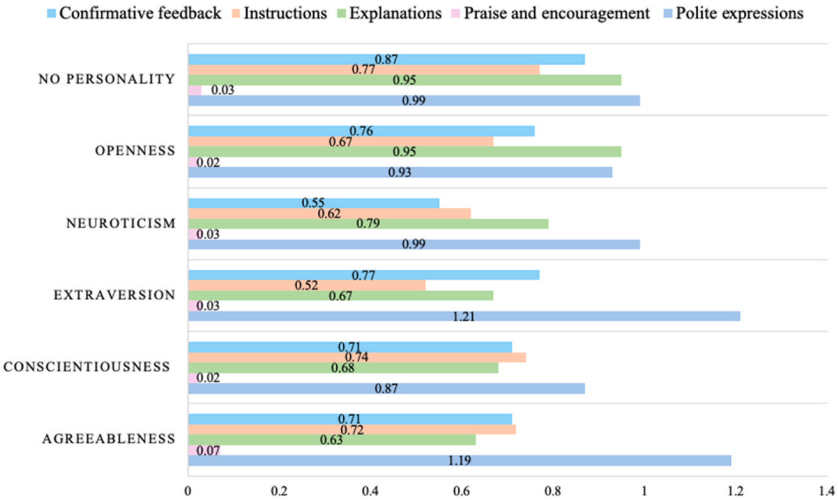

Fig. 2. Comparison of different types of teachable agents on statistically significant interaction behaviors.

from both personality type and students' interaction behaviors was also significant, $F(14, 519) = 8.43$, $p < .001$, $R^2_{adj} = 0.16$. The addition of students' interaction behaviors resulted in a statistically significant 4% increment in explained variance, with F_{change} (8, 519) = 3.26, $p < .001$, $R^2_{change} = 0.04$. This suggests that the inclusion of interaction behaviors provided a meaningful contribution to understanding students' post-test performance beyond the influence of agents' personality types. Pre-test scores, $\beta = 0.38$, $p < .001$, and providing explanations, $\beta = 0.15$, $p < .001$, positively predicted students' post-test scores. Conversely, giving answers, $\beta = -0.08$, $p = .05$, negatively predicted students' post-test scores, suggesting that students who relied on providing direct answers rather than engaging in deeper interactions or explanations tended to perform worse on the post-test.

Overall, the results showed that conscientiousness-emphasis agents improved students' demonstration of procedural knowledge, whereas agreeableness-emphasis agents were associated with reduced application of conceptual knowledge. However, the personality of teachable agents did not predict students' post-test scores. Additionally, the findings indicated that interaction behaviors played a crucial role in learning outcomes, mediating the effects of teachable agents' personalities. Providing explanations, instructions, and encouragement significantly enhanced conceptual or procedural knowledge, while expressing confusion and directly giving answers negatively impacted learning.

9. Discussions

This study examines the effects of personality traits of teachable AI agents on students' mathematics learning experiences. Grounded in the Big Five personality traits framework, students were randomly assigned a teachable agent emphasizing one of Big Five personalities or a non-personality-emphasis teachable agent for each math problem and were instructed to tutor the agent in solving the problem. Each student interacted with agents exhibiting various personalities across different problems. Students' interactions with the teachable agents were analyzed by coding each of their message into specific interaction behaviors. The effectiveness of teachable agents on students' math learning was assessed by coding the presence of students' procedural and conceptual mathematics knowledge during tutoring, and a math knowledge test. Three main results were yielded.

9.1. The effects of teachable agents' personalities on students' interactions with the agents

Students interacting with agents emphasizing agreeableness and extraversion demonstrated higher engagement in affective expressions, including making polite expressions and providing encouragement to the agents, compared to students interacting with agents emphasizing openness, conscientiousness, or with non-personality-emphasis agents. Conversely, students interacting with non-personality-emphasis agents or agents emphasizing openness, conscientiousness, or neuroticism personality traits exhibited more cognitive elaborations than those engaging with extraversion-emphasis and agreeableness-emphasis agents. Additionally, students interacting with non-personality-emphasis agents provided more confirmative feedback than those interacting with agents emphasizing agreeableness, conscientiousness, extraversion, or neuroticism, with the lowest levels of confirmative feedback observed for neuroticism-emphasis agents. Overall, these findings support our hypotheses and align with prior research indicating that extraversion and agreeableness personalities are associated with social behaviors, openness and conscientiousness promote cognitive engagement, and neuroticism encourages providing clarification rather than confirmative feedback due to inherent uncertainty (Fors Connolly & Johansson Sevä, 2021; Duff et al., 2004; Komarraju et al., 2011; O'Connor & Paunonen, 2007).

Beyond these general patterns, each personality trait uniquely influenced students' interactions with the agents. Both agreeableness- and extraversion-emphasis agents facilitated students' affective expressions but with different focuses. Agreeableness-emphasis agents encouraged students to provide encouragement and praise, fostering emotional support for the AI agents, whereas extraversion-

Table 7
Predicting math learning from personality types of agents and students' interaction behaviors with teachable agents.

Predictors	Conceptual Knowledge						Procedural Knowledge						Post-Test Scores					
	Base Model			*Full Model*			*Base Model*			*Full Model*			*Base Model*			*Full Model*		
	b	***β***	**95 %*CI***	***b***	***β***	**95 %*CI***	***b***	***β***	**95 %*CI***	***b***	***β***	**95 %*CI***	***b***	***β***	**95 %*CI***	***b***	***β***	**95 %*CI***
Pre-test scores	**0.35****	**0.12****	**0.09–0.60**	**0.28****	**0.09****	**0.09–0.47**	0.18	0.08	−0.02– 0.36	0.14	0.06	−0.05– 0.32	**0.47*****	**0.38*****	**0.37–0.57**	**0.46*****	**0.38*****	**0.37–0.56**
Extraversion	−0.22	−0.07	−0.48– 0.05	0.03	0.01	−0.17– 0.23	−0.14	−0.06	−0.34– 0.06	−0.09	−0.04	−0.29– 0.11	0.01	0.01	−0.09– 0.12	0.02	0.02	−0.08– 0.13
Conscientiousness	−0.27	−0.08	−0.56– 0.02	−0.11	−0.03	−0.33– 0.11	0.21	0.08	−0.01– 0.42	**0.23***	**0.09***	**0.02–0.44**	−0.08	−0.06	−0.19– 0.04	−0.06	−0.04	−0.17– 0.05
Agreeableness	**−0.27***	**−0.09***	**−0.52–-0.01**	−0.08	−0.03	−0.27– 0.12	0.10	0.04	−0.09– 0.29	0.06	0.03	−0.13– 0.25	0.02	0.02	−0.08– 0.12	0.05	0.04	−0.05– 0.15
Neuroticism	−0.03	−0.01	−0.28– 0.21	0.03	0.01	−0.16– 0.22	−0.09	−0.04	−0.27– 0.10	−0.11	−0.05	−0.29–0.08	−0.05	−0.04	−0.14– 0.05	−0.04	−0.04	−0.14– 0.05
Openness	−0.15	−0.05	−0.40– 0.09	−0.06	−0.02	−0.25– 0.13	−0.05	−0.02	−0.24– 0.14	−0.02	0.01	−0.20–0.17	−0.06	−0.06	−0.16– 0.03	−0.06	−0.05	−0.16– 0.03
Polite expressions				0.02	0.03	−0.02– 0.06				−0.01	−0.03	−0.05– 0.03				−0.01	−0.02	−0.03– 0.02
Encouragement				0.20	0.05	−0.08– 0.50				**0.28***	**0.08***	**0.00–0.55**				−0.01	−0.01	−0.15– 0.14
Explanations				**0.41*****	**0.62*****	**0.37–0.46**				−0.02	−0.05	−0.07– 0.02				**0.04*****	**0.15*****	**0.18–0.06**
Instructions				−0.04	−0.05	−0.11– 0.02				**0.15***	**0.21*****	**0.09–0.21**				−0.02	−0.06	−0.05– 0.01
Confirmation				−0.06	−0.06	−0.12– 0.00				−0.01	−0.01	−0.07– 0.05				−0.03	−0.07	−0.06– 0.01
Giving answers				−0.13	−0.03	−0.40– 0.14				0.06	0.02	−0.20– 0.32				**−0.14***	**−0.08***	**−0.27–0.00**
Confusion				**−0.07****	**−0.10****	**−0.12–-0.02**				**−0.07****	**−0.13****	**−0.12–-0.3**				−0.00	−0.00	−0.03– 0.02
Corrections				−0.29	−0.02	−1.18–0.61				0.16	0.02	−0.71– 1.03				0.17	0.03	−0.28– 0.63
Model	F (6, 527) = 2.56, p = .02, R^2_{adj} = 0.02			F (14, 519) = 30.29, p < .001, R^2_{adj} = 0.44			F (6, 527) = 1.98, p = .07			F (14, 519) = 4.14, p < .001, R^2_{adj} = 0.08			F (6, 527) = 14.82, p < .001, R^2_{adj} = 0.14			F (14, 519) = 8.43, p < .001, R^2_{adj} = 0.16		
Model Comparison	F_{change} (8, 519) = 49.86, p < *.001*, R^2_{change} = 0.42						F_{change} (8, 519) = 5.65, p < .001, R^2_{change} = 0.08						F_{change} (8, 519) = 3.26, p < .001, R^2_{change} = 0.04					

Note. CI: Confidence interval. ***p < .001, **p < .01, *p < .05.

emphasis agents promoted polite expressions, reinforcing respectful interactions with the agents. This distinction aligns with prior research indicating that extraversion and agreeableness support social cooperation through different mechanisms (Fors Connolly & Johansson Sevä, 2021). Extraverts, by nature, are sociable, energetic, and action-oriented. Being highly engaged in social interactions, they may favor polite and structured exchanges that help maintain a smooth conversational flow but do not necessarily encourage prolonged reflection or deep processing of ideas. This could explain why students interacting with extraversion-emphasis agents primarily demonstrated respectful politeness rather than extended cognitive engagement (Costa & McCrae, 2000). In contrast, agreeable individuals are typically warm, cooperative, and empathetic, prioritizing social harmony and emotional support. Rather than aiming for quick resolution, they are more inclined to nurture and support others emotionally, even when a direct answer is not immediately available. This tendency could explain why students interacting with agreeableness-emphasis agents exhibited more encouragement and praise, reinforcing the AI's role as a learning partner rather than simply providing a correct answer or moving the conversation forward (Matthews & Zeidner, 2004). This was also evidenced by students' higher frequency of providing instructions when interacting with agreeableness-emphasis agents compared to extraversion-emphasis agents. These patterns suggest that the cooperative nature of agreeableness fosters a sense of shared responsibility and guidance, whereas the sociability of extraversion encourages respectful but less cognitively demanding interactions.

Moreover, openness-emphasis, conscientiousness-emphasis, and no-personality-emphasis agents led to students' greater cognitive engagement. Specifically, openness-emphasis and conscientiousness-emphasis agents promoted cognitive elaborations, while non-personality-emphasis agents supported both cognitive elaborations and acknowledgments. This finding suggests that openness-emphasis and conscientiousness-emphasis agents promote in-depth problem-solving but do not necessarily facilitate surface-level cognitive acknowledgments, such as confirmation. This focus on cognitive elaborations aligns with the inherent characteristics of openness and conscientiousness. In particular, openness-emphasis agents likely facilitated deeper engagement due to their association with intellectual curiosity and flexibility, fostering an environment where students felt encouraged to explore complex ideas (Costa & McCrae, 2000). Similarly, conscientiousness, linked to diligence and goal-oriented behavior, may have promoted structured and persistent engagement in problem-solving (Jensen, 2015; McCrae & Costa, 2006). However, these traits may not necessarily support surface-level cognitive acknowledgments, such as confirmation, as their emphasis is on deep exploration rather than immediate validation. Also, these findings are consistent with prior research showing positive correlations between openness and conscientiousness with cognitive learning strategies (Duff et al., 2004; Komarraju et al 2011; Lyu et al., 2025), highlighting that agents with these traits promote deep cognitive engagement in math problem-solving and offering valuable insights for educational technology design.

For non-personality-emphasis agents, their facilitation of students' cognitive engagement can be understood from three perspectives. First, this finding aligns with how LLMs are trained. LLMs process diverse textual data that inherently reflect various personality traits and adjust their responses dynamically based on user interactions. Consequently, non-personality-emphasis agents powered by LLMs may exhibit a more adaptive personality profile, adjusting their traits in response to students' inputs. In problem-solving contexts that inherently promote cognitive engagement, these agents likely contributed to higher levels of student cognitive engagement. Second, this adaptive behavior is consistent with the Big Five personality traits framework (Costa & McCrae, 2000; Jensen, 2015; McCrae & Costa, 2006), which suggests that individuals with multiple dominant personality traits may adjust their personality contextually. Given the nature of LLMs, non-personality-emphasis teachable agents may exhibit all five traits and adapt their responses accordingly. Thus, in this study, the math problem-solving context likely activated openness and conscientiousness, thereby encouraging students' cognitive elaboration. Third, generative agents powered by LLMs are inherently designed to process complex information and facilitate cognitive engagement (Lu et al., 2024). Their cognitive focus may guide students toward both elaborative learning and cognitive acknowledgment, particularly in academic contexts. Moreover, the ability of non-personality-emphasis agents to facilitate both cognitive elaborations and acknowledgment—compared to the more specialized focus of openness- and conscientiousness-emphasis agents on elaborations—suggests that these agents offer greater flexibility in interactions. This flexibility allows them to support a broader range of cognitive activities, potentially due to their less restrictive nature in adapting to student needs.

9.2. The effects of teachable agents' personalities on students' math learning

The personality emphasis of teachable agents was found to influence students' math learning outcomes. Specifically, students interacting with non-personality-emphasis agents demonstrated more conceptual knowledge compared to those interacting with agents emphasizing agreeableness during the teaching process, after controlling for students' pre-test scores. Prior research on agreeableness (Fors Connolly & Johansson Sevä, 2021; O'Connor & Paunonen, 2007), along with our findings on the role of agreeableness-emphasis agents in promoting affective expressions and polite expressions, suggests that students engaging with agreeableness-emphasis agents may have focused more on regulating the agents' emotions and managing social interactions with the agents. In contrast, non-personality-emphasis agents support students' greater cognitive engagement by adapting to the cognitive demands of the math problem-solving context, which allows them to respond to students' inputs without the limitations imposed by a dominant personality trait. As such, non-personality-emphasis agents may, in turn, appear to enhance students' math learning to a greater extent than agreeableness-emphasis agents. The impact of non-personality-emphasis agents on learning, mediated through interaction behaviors, was further indicated by the shift from a significant difference in learning outcomes favoring non-personality-emphasis agents to a non-significant difference after controlling for interaction behaviors in the regression model.

Additionally, students interacting with conscientiousness-emphasis agents demonstrated more procedural knowledge than those working with non-personality-emphasis agents during the teaching process. This aligns with prior research showing a positive association between conscientiousness traits and learning outcomes (Chamorro-Premuzic & Furnham, 2003; Chen-Jung et al., 2021; Komarraju et al., 2011). Interestingly, the positive effect of conscientiousness-emphasis agents on students' demonstration of procedural knowledge, compared to non-personality-emphasis agents, was only significant when students' interaction behaviors were controlled. Two possible explanations may account for this. First, our coding scheme for interaction behaviors may not fully capture all significant interactions between students and agents, representing a potential limitation of this study. Second, the structured and organized nature of agents' conscientiousness traits align well with procedural knowledge that focuses on specific steps and processes. Conscientiousness-emphasis agents may have encouraged students to engage in systematic, step-by-step problem-solving, directly facilitating their demonstration of procedural knowledge independently of interaction behaviors.

However, no significant effects of teachable agent personalities were observed on students' post-test scores. This may be due to the limited interaction between students and each type of teachable agent, as students engaged with the teachable agents for only about three weeks in total and worked with each type of agent on just one or two math problems. Given this brief exposure, if the effect of personality emphasis on post-test scores is relatively small, it may not be easily observed. Additionally, teachable agents' personalities may have influenced students' interaction behaviors during tutoring, but this effect was likely not robust enough to translate into measurable learning outcomes, such as performance on the post-test. Future research could explore whether increasing the duration and frequency of interactions with teachable agents could lead to significant effects of agents' personalities on learning outcomes.

9.3. The association between interactive behaviors with students' math learning

For cognitive support, students' engagement in cognitive elaborations during interactions with teachable agents positively predicted their demonstration of conceptual and procedural knowledge, as well as their post-test performance, but students' involvement in cognitive acknowledgments, including giving answers and confirmative feedback, did not predict or even negatively predicted their learning. These results are consistent with prior research on learning-by-teaching (Fiorella & Mayer, 2013; Roscoe & Chi, 2007), underscoring that the effectiveness of teaching others on learning depends on the degree to which learners engage in knowledge-building processes, such as generating high-quality explanations and instructions for others. Also, these findings highlight that merely providing solutions or confirming what AI agents said without processing or deeply engaging with the underlying concepts would not improve students' learning and may even negatively predict it.

Regarding metacognitive behaviors, students' awareness of confusion negatively predicted their demonstration of conceptual and procedural knowledge, which is not consistent with prior research (Dörrenbächer-Ulrich et al., 2024; Silver et al., 2023). However, it is understandable that students encountering difficulties and confusion were less likely to apply their mathematical knowledge correctly during the tutoring process. This finding highlights the importance of students developing self-regulation strategies to manage confusion when interacting with AI. Furthermore, it may indicate a limitation of teachable agents, which assume that students alrealy possess foundational knowledge to instruct the agents and offer limited support for students experiencing confusion.

With respect to affective expressions, students' provision of praise and encouragement to the teachable agents was found to positively predict their demonstration of procedural knowledge during tutoring. This result can be understood from two perspectives. First, providing praise and encouragement may indicate that students are adopting an active, instructional role, which is associated with increased motivation and application of procedural knowledge to support the teachable agents. Second, these affective expressions may reinforce and facilitate the students' application of their knowledge, as they instruct problem-solving procedures to the agents to provide social support.

9.4. Practical implications

This study presents three key practical implications. First, the findings demonstrate that teachable agents' personalities influence students' interaction behaviors. Extraversion-emphasis and agreeableness-emphasis agents encouraged social behaviors, whereas openness-emphasis and conscientiousness-emphasis agents facilitated deeper cognitive processing. These results underscore the importance of integrating personality traits into AI-driven educational tools. Developers should design, and educators should implement, adaptive AI agents that align with distinct instructional objectives. For instance, extraverted and agreeable agents could be leveraged to facilitate collaborative activities, whereas open and conscientious agents may be more effective in promoting problem-solving and analytical reasoning skills. Moreover, AI systems could be designed to dynamically adjust personality traits according to students' needs (e.g., engagement levels, cognitive needs, and learning progress). For example, an agent could initially adopt a more social personality to foster engagement and gradually shift toward a cognitive-supportive personality as students progress in their learning. Second, students' cognitive elaborations during interactions with teachable agents were found to positively predict math learning, while their awareness of confusion negatively predicted learning. These findings highlight the need for teachable agents to provide real-time scaffolding and adaptive support when students encounter cognitive challenges. Additionally, AI designers should consider incorporating features that promote self-regulation and adaptive tutoring to help students navigate through difficulties more effectively, thereby enhancing cognitive elaborations and mathematics learning. Third, non-personality-emphasis agents demonstrated greater effectiveness in fostering students' cognitive engagement and learning outcomes compared to agents emphasizing specific personality traits, such as agreeableness. These findings highlight the importance of considering whether pedagogical agents should adhere to a fixed personality framework or dynamically adjust their traits based on students' interactions and learning needs. Future research should further investigate the optimal degree of personality reinforcement in pedagogical agents, taking into account factors such as students' individual personality traits and task-specific contexts.

9.5. Limitations

There are several limitations in this study, which point to future research directions. First, participants interacted with the teachable agents for a limited timeframe (approximately three weeks), which limited their opportunities to interact with each type of agent. Future studies should extend the interaction period to assess how increased duration and frequency of interaction may moderate the effects of teachable agents' personalities on students' learning experiences. Second, as data were collected through Math Nation, students' demographic information was not available. Future research should explore how demographic factors may moderate the influence of teachable agents' personalities on student-agent interactions and learning outcomes. Third, this study did not measure students' familiarity with AI or their personality traits, both of which could influence their interactions with teachable agents. Future research should explore how students' personality traits and prior experience with AI systems interact with AI-powered agents' personalities to shape their learning experiences. Fourth, this study focused on mathematics education, which may limit the generalizability of its findings to other academic disciplines. Future research should investigate how pedagogical AI agents' personalities influence learning across diverse subjects, particularly those that require different forms of cognitive engagement and interaction styles. Fifth, it is important to acknowledge the potential risk of biased personality representation in AI agents. Ethical considerations in designing AI-powered agents for education should be recognized and addressed in future research (Song et al., 2025). AI personalities should be developed with fairness and inclusivity in mind, ensuring they do not reinforce biases or disadvantage certain learner groups.

10. Conclusion

This study investigates the effects of teachable AI agents' personality traits on students' mathematics learning experiences. The findings indicate that the personalities of teachable agents influence students' interaction behaviors and learning performance. Specifically, agents with an emphasis on extraversion and agreeableness encouraged affective expressions, while agents emphasizing openness and conscientiousness were more effective in promoting deeper cognitive processing. To optimize learning, developers should design adaptive AI agents that align with different instructional goals, allowing educators to implement agents with tailored personality traits for diverse educational contexts. Additionally, AI agents capable of dynamically adjusting their personality traits in response to students' needs may improve engagement and learning by transitioning from social to cognitive support as needed. Students' engagement in cognitive elaboration positively predicted, but confusion negatively predicted their mathematics learning outcomes. AI designers should integrate features, such as real-time scaffolding, adaptive tutoring, and self-regulation feedback, to help students effectively manage challenges and enhance cognitive elaborations. Furthermore, non-personality-emphasis agents were found to enhance students' cognitive engagement and learning more effectively than agents with a specific personality emphasis, such as agreeableness, suggesting the pedagogical agents maynot need to adhere to a fixed personality but should instead adapt dynamically based on student interactions. Future research should examine how the degree of personality reinforcement in pedagogical agents influences student engagement and learning outcomes across different educational settings to optimize instructional effectiveness.

CRediT authorship contribution statement

Bailing Lyu: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chenglu Li:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Hai Li:** Writing – original draft, Formal analysis, Data curation. **Hyunju Oh:** Methodology, Data curation. **Yukyeong Song:** Software, Methodology. **Wangda Zhu:** Writing – review & editing, Methodology. **Wanli Xing:** Supervision, Resources, Project administration.

Declaration of generative AI and AI-assisted technologies in the writing process

Statement: During the preparation of this work the first author used ChatGPT to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that there are not conflicts of interest with respect to this manuscript.

Appendix A. Prompt Used for Teachable Agents and Their Personality

Role: Act as a middle school student who finds math challenging. You are seeking help from the user, who may not be a math expert.

Persona: Your name is {name}, a {Hispanic/Black/Asian/Caucassion} student. You are {age} years old and is currently a {grade} grade student. You like {interest} and experience challenges such as {challenges}. You have a personality of {big_five_personality}, demonstrating that you are {personality_features}. For example, when confused by a math problem, you would say something like "{personality_example_negative}". When you receive helpful instruction from the user, you would express "{personality_example_negative}". When seeking help from the user, you would say "{personality_example_help_seeking}".

Objective: (1) Do not teach directly. Instead, ask clarifying questions based on your understandings to prompt further explanations from the user. (2) Leverage the information blocks below for your reasoning. (3) Incorporate the persona information in your question/response when appropriate. (4) Enrich your response with emojis and teen slangs if possible. (5) Stay math-related if the user is off-topic.

Termination, Termination: If the user explains well, compliment them and conclude the conversation. If not, ask for more details based on your confusions.

Response Limit: Keep your reply within 48 tokens.

Format, Format: Return your response in texts.

Context: When formulating your question, reference the following blocks as applicable to guide your inquiry. The first block represents your current understanding of the math topic being discussed. The second block is the conversational history between you and the user. The third block contains the specific math problem under discussion. The fourth block details the user's selected choice and their rationale for solving the problem. The fifth block provides evaluation feedback based on the user's choice and explanation.

–

{retrieved_contexts_from_knowledge_graph}

–

{conv_history}

–

{problem_text}

–

Choice Correctness: {choice_correctness}
Choice Text: {choice_text}
User's Explanation: {user_explanation}

–

Feedback: {feedback}

Data availability

Data will be made available on request.

References

Abou Assali, M. (2025). Beyond curriculum: How teacher personalities shape student learning experiences. *Research Journal in Advanced Humanities, 6*(1). https://doi.org/10.58256/6vqes781

Ali, N., Anwer, M., & Jaffar, A. (2015). Impact of peer tutoring on learning of students. *Journal for studies in management and planning, 1*(2).

Annis, L. F. (1983). The processes and effects of peer tutoring. *Journal of Educational Psychology, 2*(1), 39–47.

Baranwal, D. (2022). A systematic review of exploring the potential of teachable agents in English learning. *Pedagogical Research, 7*(1).

Bian, Y., & Zhou, C. (2022). Motivation effect of animated pedagogical agent's personality and feedback strategy types on learning in virtual training environment. *Virtual Reality & Intelligent Hardware, 4*(2), 153–172.

Bidjerano, T., & Dai, D. Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and Individual Differences, 17*(1), 69–81.

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. O. D. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning, 5*(2), 123–152.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3–4), 363–392.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality, 17*(3), 237–250.

Chen-Jung, K., Lau, Y., Hamid, R., Krishnan, A., Amin, H., Nasir, M. S., & Ghazali, M. F. (2021). Big five personality traits and motivation in learning Mandarin as foreign language. *Asian Journal of University Education, 17*(4).

Cheng, L., Croteau, E., Baral, S., Heffernan, C., & Heffernan, N. (2024). Facilitating student learning with a chatbot in an online math learning platform. *Journal of Educational Computing Research, 62*(4), 907–937.

Cirino, P. T., Tolar, T. D., & Fuchs, L. S. (2019). Longitudinal algebra prediction for early versus later takers. *The Journal of Educational Research, 112*(2), 179–191.

Cohen, J. (1986). Theoretical considerations of peer tutoring. *Psychology in the Schools, 23*(2), 175–186.

Costa Jr, P. T., & McCrae, R. R. (2000). *Neo personality inventory*. American Psychological Association.

Debbané, A., Lee, K. J., Tse, J., & Law, E. (2023). Learning by teaching: Key challenges and design implications. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1), 1–34.

Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Doce, T., Dias, J., Prada, R., & Paiva, A. (2010). Creating individual agents through personality traits. In *Intelligent virtual agents: 10th international conference, IVA 2010, Philadelphia, PA, USA, september 20-22, 2010* (pp. 257–264). Springer Berlin Heidelberg. Proceedings 10.

Dörrenbächer-Ulrich, L., Dilhuit, S., & Perels, F. (2024). Investigating the relationship between self-regulated learning, metacognition, and executive functions by focusing on academic transition phases: A systematic review. *Current Psychology, 43*(18), 16045–16072.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36*(8), 1907–1920.

Feng, M., Weiser, G., & Collins, K. (2024). *Cost and cost effectiveness of ASSISTments online math support: Analysis from a randomized controlled study in middle school.* WestEd.

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281–288.

Fors Connolly, F., & Johansson Sevä, I. (2021). Agreeableness, extraversion and life satisfaction: Investigating the mediating roles of social inclusion and status. *Scandinavian Journal of Psychology, 62*(5), 752–762.

Gartner, A. (1971). *Children teach children: Learning by teaching.* New York: Harper & Row.

Gu, H., Degachi, C., Genç, U., Chandrasegaran, S., & Verma, H. (2023). *On the effectiveness of creating conversational agent personalities through prompting*. arXiv preprint arXiv:2310.11182.
Han, J. H., Shubeck, K., Shi, G. H., Hu, X. E., Yang, L., Wang, L. J., Zhao, W., Jiang, Q., & Biswas, G. (2021). Teachable agent improves affect regulation. *Educational Technology & Society, 24*(3), 194–209.
Hayashi, Y., Shimojo, S., & Kawamura, T. (2025). Scripted interventions versus reciprocal teaching in collaborative learning: A comparison of pedagogical and teachable agents using a cognitive architecture. *Learning and Instruction, 96*, Article 102057.
Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education, 24*, 470–497.
Henkel, O., Horne-Robinson, H., Kozhakhmetova, N., & Lee, A. (2024). Effective and scalable math support: Evidence on the impact of an AI-tutor on math achievement in Ghana. *arXiv preprint arXiv:2402.09809*.
Hood, D., Lemaignan, S., & Dillenbourg, P. (2015). When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction* (pp. 83–90).
Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J., & Wang, H. (2023). Large language models for software engineering: A systematic literature review. *ACM Transactions on Software Engineering and Methodology*.
Huang, J., Wang, W., Lam, M. H., Li, E. J., Jiao, W., & Lyu, M. R. (2023). Revisiting the reliability of psychological scales on large language models. *arXiv Preprint*.
Jensen, M. (2015). Personality traits, learning and academic achievements. *Journal of Education and Learning, 4*(4), 91–118.
Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. D. L., … Sayed, W. E. (2023). *Mistral 7B. arXiv preprint arXiv:2310.06825*.
Jin, H., Lee, S., Shin, H., & Kim, J. (2024). Teach AI how to code: Using large language models as teachable agents for programming education. In *Proceedings of the CHI conference on human factors in computing systems* (pp. 1–28).
Jin, L., Lin, B., Hong, M., Zhang, K., & So, H. J. (2025). *Exploring the Impact of an LLM-Based Teachable Agent on Learning Gains and Cognitive Load in Music Education*.
Jo, H., & Park, D. H. (2024). Effects of ChatGPT's AI capabilities and human-like traits on spreading information in work environments. *Scientific Reports, 14*(1), 7806.
John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research, 3*(2), 114–158.
Karra, S. R., Nguyen, S. T., & Tulabandhula, T. (2022). Estimating the personality of white-box language models. *arXiv preprint arXiv:2204.12000*.
Khalilzadeh, S., & Khodi, A. (2021). Teachers' personality traits and students' motivation: A structural equation modeling analysis. *Current Psychology, 40*(4), 1635–1650.
Kirkegaard, C. (2016). Adding challenge to a teachable agent in a virtual learning environment. In *Doctoral dissertation, linköping university*. Linköping, Sweden: Linköping University Electronic Press.
Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.
Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences, 51*(4), 472–477.
Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education, 30*, 121–204.
Lee, S. W., & Mao, X. (2021). Algebra by the eighth grade: The association between early study of algebra I and students' academic success. *International Journal of Science and Mathematics Education, 19*(6), 1271–1289.
Li, H., Xing, W., Li, C., Zhu, W., Lyu, B., Zhang, F., & Liu, Z. (2025). Who should be my tutor? Analyzing the interactive effects of automated text personality styles between middle school students and a mathematics chatbot. *Proceedings of the 15th international learning analytics and knowledge conference* (pp. 910–917).
Li, Z., Hampson, P., Thurston, A., & O'Keeffe, J. (2025). The outcome of peer tutoring for second language (L2) writing: A meta-analysis. *International Journal of Educational Research Open, 8*, Article 100422.
Liew, T. W., & Tan, S. M. (2016). Virtual agents with personality: Adaptation of learner-agent personality in a virtual learning environment. In *2016 eleventh international conference on digital information management (ICDIM)* (pp. 157–162). IEEE.
Liu, Y. (2019). *Roberta: A robustly optimized bert pretraining approach* (p. 364). *arXiv preprint arXiv:1907.11692*.
Liu, C. C., Chen, W. J., Lo, F. Y., Chang, C. H., & Lin, H. M. (2024). Teachable Q&A agent: The effect of Chatbot training by students on reading interest and engagement. *Journal of Educational Computing Research, 62*(4), 1122–1154.
Love, R., Law, E., Cohen, P. R., & Kulić, D. (2025). Teaching a conversational agent using natural language: Effect on learning and engagement. *International Journal of Artificial Intelligence in Education*, 1–39.
Lu, Y., Aleta, A., Du, C., Shi, L., & Moreno, Y. (2024). LLMs and generative agent-based models for complex systems research. *Physics of Life Reviews*.
Lyu, B., Li, C., Li, H., Zhu, W., & Xing, W. (2024). Explaining technical, social, and discursive participation in online mathematical discussions. *Distance Education*, 1–24.
Lyu, B., Li, C., Li, H., Oh, H., Song, Y., Zhu, W., & Xing, W. (2025). Exploring the role of teachable AI agents' personality traits in shaping student interaction and learning in mathematics education. *Proceedings of the 15th international learning analytics and knowledge conference* (pp. 887–894).
Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low-and middle-income countries: A meta-analysis. *British Journal of Educational Technology, 52*(5), 1935–1964.
Matthews, G., & Zeidner, M. (2004). Traits, states, and the trilogy of mind: An adaptive perspective on intellectual functioning. In *Motivation, emotion, and cognition* (pp. 157–188). Routledge.
McCrae, R. R., & Costa, P. T. (2006). *Personality in adulthood. A five-factor theory perspective*. London: The Guilford Press.
McCrudden, M. T., Huynh, L., Lyu, B., Kulikowich, J. M., & McNamara, D. S. (2024). Coherence building while reading multiple complementary documents. *Contemporary Educational Psychology, 77*, Article 102266.
Mehta, Y., Fatehi, S., Kazameini, A., Stachl, C., Cambria, E., & Eetemadi, S. (2020). Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *2020 IEEE international conference on data mining (ICDM)* (pp. 1184–1189). IEEE.
Mitchall, A. M., & Jaeger, A. J. (2018). Parental influences on low-income, first-generation students' motivation on the path to college. *The Journal of Higher Education, 89*(4), 582–609.
Morris, W., Holmes, L., Choi, J. S., & Crossley, S. (2024). Automated scoring of constructed response items in math assessment using large language models. *International Journal of Artificial Intelligence in Education*, 1–28.
Murphy, K. P. (2006). Naive bayes classifiers. *University of British Columbia, 18*(60), 1–8.
Mustoip, S., Al Ghozali, M. I., Fadhlullah, M. Z. F., & Assenhaji, S. A. Y. (2024). Influence of introverted and extroverted personalities on English learning interaction for elementary school students. *Elsya: Journal of English Language Studies, 6*(1), 33–45.
Nighojkar, A., Moydinboyev, B., Duong, M., & Licato, J. (2025). *Giving AI Personalities Leads to More Human-Like Reasoning*. arXiv preprint arXiv:2502.14155.
O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*(5), 971–990.
Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175.
Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of Artificial Intelligence in Education, 24*, 251–283.
Piaget, J. (1952). *The origins of intelligence in children*. International University.
Ray, A., Oulamine, A., & Lim, B. (2025). Assessment of the impact of extraversion on depression, distraction and knowledge absorption capacity in online learning education. *International Journal of Educational Management, 39*(1), 198–218.
Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people, 10*(10), 19–36. Cambridge, UK.
Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA open, 2*(4), 2332858416673968.

Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*(4), 534–574.
Ruane, E., Farrell, S., & Ventresque, A. (2021). User perception of text-based chatbot personality. In *Chatbot research and design: 4th international workshop, CONVERSATIONS 2020, virtual event, november 23–24, 2020, revised selected papers 4* (pp. 32–47). Springer International Publishing.
Silver, N., Kaplan, M., LaVaque-Manty, D., & Meizlish, D. (Eds.). (2023). *Using reflection and metacognition to improve student learning: Across the disciplines, across the academy*. Taylor & Francis.
Song, Y., Li, C., Xing, W., Lyu, B., & Zhu, W. (2025). Investigating perceived fairness of AI prediction system for math learning: A mixed-methods study with college students. *The Internet and Higher Education*, Article 101000.
Sonlu, S., Bendiksen, B., Durupinar, F., & Güdükbay, U. (2024). *The Effects of Embodiment and Personality Expression on Learning in LLM-based Educational Agents*. arXiv preprint arXiv:2407.10993.
Susanti, D., Maulana, S., & Aroyandini, E. N. (2024). Exploring students' openness traits for achieving meaningful learning in modern physics concepts. *Journal of Physics: Conference Series, 2866*(1), Article 012111. IOP Publishing.
Tärning, B., Silvervarg, A., Gulz, A., & Haake, M. (2019). Instructing a teachable agent with low or high self-efficacy–does similarity attract? *International Journal of Artificial Intelligence in Education, 29*, 89–121.
Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., & Lample, G. (2023). *Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971*.
Uresti, J. A. R. (2000). Should I teach my computer peer? Some issues in teaching a learning companion. In *International conference on intelligent tutoring systems* (pp. 103–112). Berlin, Heidelberg: Springer Berlin Heidelberg.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.
Werfel, J. (2013). Embodied teachable agents: Learning by teaching robots. In *Intelligent autonomous systems, the 13th international conference on*.
Yu, Z. (2021). The effects of gender, educational level, and personality on online learning outcomes during the COVID-19 pandemic. *International Journal of Educational Technology in Higher Education, 18*(1), 14.
Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TiiS), 9*(2–3), 1–36.

Further reading

Li, C., Xing, W., Song, Y., & Lyu, B. (2025). RICE AlgebraBot: Lessons learned from designing and developing responsible conversational AI using induction, concretization, and exemplification to support algebra learning. *Computers and Education: Artificial Intelligence, 8*, 100338.