Computers and Education Open 10 (2026) 100331

Contents lists available at ScienceDirect

Computers and Education Open

journal homepage: www.sciencedirect.com/journal/computers-and-education-open

An experimental study exploring human–AI complementarity in early social-emotional learning

Doris Kristina Raave [a,1], Tyler Colasante [a,1,*], Eric Roldan Roa [b], Juan Carlos Ramos Martinez [b], Hongtao Li [b], Sayan Mukherjee [b], Tina Malti [a]

[a] Humboldt Science Center for Child Development (HumanKind), Leipzig University, Germany

[b] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig University, Germany

ARTICLE INFO

Keywords:
Human-AI complementarity
Social-emotional learning
Pedagogical conversational agent
Comparative analysis
Early childhood education

ABSTRACT

Many children are not receiving crucial social-emotional learning (SEL) support due to systemic constraints, such as high educator workload and training burdens. Pedagogical conversational agents (PCAs)—generative AI-powered virtual characters trained to converse with students like a teacher—offer promising solutions that could extend to SEL. However, the current affective-computing capacities of PCAs may be limiting, as SEL facilitation requires deeper emotional and relational attunement than cognition-heavy subjects (e.g., literacy, numeracy). The emotional and relational skills of human educators likely confer benefits beyond those of PCAs in SEL instruction. Despite this potential complementarity, no studies have experimentally assessed the relative performance of PCAs and educators in SEL contexts. It remains unclear if and how PCAs can help educators close the SEL gap. In this study, we used a controlled comparison to reveal complementary strengths, comparing the performance of a PCA to that of human educators when independently facilitating story-based SEL activities with a static AI child. We used a static AI child to help ensure that any performance differences could be attributed to facilitator type. The PCA (n = 18 simulations) and educators (n = 18) each delivered three SEL activities (N = 108 observations). Expert raters, blind to facilitator type, coded dialogue excerpts for evidence-based SEL support techniques and indicators of pedagogical quality. A mixed ANOVA and t-tests revealed significant differences. The PCA showed strengths in basic relational and instructional domains, maintaining respectful tone and routinely providing procedural scaffolding. Educators showed strengths in deeper SEL instruction, guiding reflection and promoting social-emotional knowledge. We discuss implications for SEL through the lens of human–AI complementarity.

1. Introduction

1.1. The current state of childhood mental health and early social-emotional learning as a solution

Failing to nurture childhood mental health from early in life is estimated to cost society up to 1.85 million USD in direct (e.g., mental health services) and indirect (e.g., lost workforce productivity) costs per at-risk child across their lifetime [1]. The number of children considered at-risk for mental health challenges has increased over past decades, and this increase is not fully attributable to greater awareness or improved diagnostics [2]. For example, suicide has become one of the leading causes of death in 10- to 14-year-olds, particularly in girls [3].

A structural, upstream solution to address the growing costs of neglecting childhood mental health is to make social-emotional learning (SEL) mandatory in early childhood education and beyond. Longitudinal studies clearly show that social-emotional skills—such as empathically caring for others and regulating emotions—are linked to healthier relationships and better mental health across childhood, adolescence, and adulthood (e.g., [4,5]). A systematic review of school-based interventions to promote social-emotional skills found that interventions have the largest remedial and preventive benefits if they commence earlier in development [6] when neurobiological, attachment, relational, and other critical mechanisms are most malleable [7]. Moreover, early childhood is precisely when foundational social-emotional capacities undergo rapid change: theory of mind and perspective-taking

* Corresponding author.
E-mail address: tyler.colasante@uni-leipzig.de (T. Colasante).

[1] These authors contributed equally to this work and share first authorship.

https://doi.org/10.1016/j.caeo.2026.100331
Received 31 July 2025; Received in revised form 2 January 2026; Accepted 7 January 2026
Available online 8 January 2026

emerge and consolidate [8], early conscience and prosocial orientations take shape [9], and independent emotion regulation capacities expand markedly [10]. Because SEL directly targets these capacities, implementing SEL in early childhood meets children at a period of maximal malleability, potentially having stronger, more durable effects on positive mental health and development.

1.2. Challenges in early SEL implementation

Efforts to implement SEL in early childhood education settings have been thwarted by systemic challenges, such as high existing teaching burdens and insufficient pre-service educator training in SEL delivery [11]. When SEL is incorporated, efficacy is compromised by an overall lack of rigorous, evidence-based programs that remain flexible enough to account for children's varying developmental levels, attentional capacities, and cultural needs [12]. Thus, novel SEL solutions should ideally have a *minimal training burden* for educators and be *flexible* enough to include diverse young learners. Furthermore, the urgency created by accelerating numbers of at-risk children and limited time and resources to equip all early childhood education institutions for SEL delivery [13] has necessitated the need for *rapidly scalable* solutions. Despite these implementation challenges, meta-analytic evidence indicates that early SEL is reliably linked to sustained academic gains [14, 15], as well as long-term reductions in psychopathology and crime into adulthood [16].

1.3. Leveraging Generative artificial intelligence for early SEL implementation

Given the persistent challenges in implementing SEL, researchers have increasingly turned to technology as a potential lever for scalable, personalized solutions. Among emerging technologies, generative artificial intelligence (GenAI; e.g., AI-generated lesson materials, conversational agents, automated feedback/reflection tools) has gained increasing attention for its capacity to automate instructional tasks, adapt to diverse learner needs, and reduce educator burden. In a recent review on the role of GenAI in early childhood education contexts, Zhao et al. [17] argued that GenAI can likely address issues of training burden and educator workload by automating instructional components (e.g., lesson planning and content creation) and handling routine administrative tasks. With an SEL-focused thematic literature review, Henriksen et al. [18] concluded that GenAI promises to foster empathy, emotional awareness, and interpersonal skills through real-time emotional feedback and adaptive dialogue. Reviewing AI technologies with a broader scope, including chatbots, sentiment analysis, and virtual reality, Sethi and Jain [19] underscored the potential for AI to enhance student engagement, emotion regulation, and self-awareness through real-time feedback, greater immersion, and personalized learning. They also acknowledged scalability, as GenAI SEL could theoretically be rapidly disseminated due to lower implementation costs and fewer human bottlenecks relative to analog SEL programs (e.g., no/lesser need to train educators or develop physical materials for every site).

Therefore, many GenAI solutions, including intelligent tutoring systems, classroom chatbots, automated content generators, and feedback tools, have been proposed for education, including SEL (e.g., [19]). Because SEL is inherently relational—children learn how to empathize and regulate emotions through social interactions with peers and educators, real or constructed—conversational systems are particularly well suited. Indeed, pedagogical conversational agents (PCAs) have emerged as one of the most promising GenAI solutions for SEL [20]. PCAs are interactive digital systems powered by GenAI components for natural language processing and dialogue management with the purpose of relationally engaging learners to enhance educational outcomes [21]. PCAs can act as "digital teaching assistants" to simulate human-like pedagogical conversations with personalized responses tailored to the learning needs of each student. For SEL, PCAs could be trained to autonomously deliver evidence-based SEL content in an emotionally, developmentally, and pedagogically sensitive manner. User modeling—dynamically tracking and updating information about the learner's knowledge, misconceptions, and emotional state in real time—theoretically allows PCAs to adapt SEL techniques to different children and even dynamically within sessions involving the same child [22].

Some empirical work suggests that PCAs may be capable of replacing educators in specific areas of instruction, particularly those involving structured, repetitive, or individualized content delivery. Arguments for replacement or substitution typically hinge on PCA's consistency, availability, and instructional capacity relative to humans. Regarding consistency, PCAs can repeatedly deliver standardized content with minimal variability in quality or mood. A recent meta-analysis by Gladstone et al. [23] found that PCAs significantly improved learning outcomes in diverse learner populations and contexts (including children, adolescents, and adults across K–12, higher education, and informal learning settings), particularly when designed to deliver scaffolded feedback in a standardized manner. Regarding availability, PCAs do not fatigue, nor do they work set hours, so they can be constantly available beyond scheduled learning times [24]. This is particularly relevant to SEL because core competencies like emotion regulation, empathy, and conflict resolution are continually practiced and reinforced in daily life [25]. Finally, regarding instructional capacity, PCAs can simultaneously provide personalized instruction to each and every child in the classroom. A meta-analysis by Davis et al. [26] found that dynamic PCAs—designed to respond to individual learner input and simulate one-on-one tutoring across multiple learners simultaneously—significantly enhanced learning outcomes relative to static agents or no-agent conditions (also see [27]).

Despite these apparent strengths of PCAs for SEL, previous studies have also highlighted critical limitations. For instance, PCAs often interact in affectively shallow ways that are not always contingent on users' emotional states [28]. Celik et al. [29] similarly critiqued AI tools in education for offering generic, non-adaptive feedback and for failing to support deeper, context-sensitive learning. PCAs also risk reinforcing algorithmic bias when emotional cues are misinterpreted or culturally misaligned [18]. Researchers have also raised concerns about the displacement of core human functions, such as pedagogical judgment, emotional engagement, and ethical oversight in education due to increasing reliance on AI systems, calling for careful design principles to ensure that AI supports, rather than undermines, the human dimensions of teaching and learning [30]. These concerns have prompted a growing body of research shifting the focus away from replacement to enhancement, emphasizing the potential for AI to work in partnership with educators from the lens of human-AI complementarity. Holstein and Aleven [31], for example, proposed a framework that prioritizes transparency, teacher control, and collaborative design from their work with Lumilo smart glasses, illustrating how real-time analytics can support teachers' decision-making without undermining their professional autonomy. Similarly, Filiz et al. [32] found that K–12 teachers valued AI tools like ChatGPT and MagicSchool for streamlining lesson planning and personalizing learning, enabling greater focus on student engagement and adaptive instruction (for similar arguments, see [33–35]).

Overall, this focused review of the literature reveals that, despite increasing interest in integrating GenAI, particularly PCAs, into SEL, most empirical findings to date originate from broader educational research (see [23]). As Henriksen et al. [18] noted, research at the intersection of GenAI and SEL remains nascent. Much of the existing literature centers on conceptual frameworks, potential affordances, and ethical considerations, with limited *empirical validation in applied experimental settings*. As one of the few studies specifically addressing GenAI's social-emotional dimensions, Oh and Ahn [36] investigated K–12 educators' perceptions of AI, highlighting its perceived value in automating routine tasks and supporting personalized instruction.

However, educators also emphasized AI's limitations in replicating empathy, intuition, and relational understanding—qualities they considered foundational to SEL. Rather than framing this as a deficit, many teachers viewed it as underscoring the enduring value of human educators alongside AI in the future. Oh and Ahn's investigation suggests that SEL may pose distinctive challenges for AI-driven facilitation, particularly in domains where humans might thrive; however, it relies exclusively on educators' perceptions and falls short of testing GenAI's actual performance in facilitating SEL. Without investigating how GenAI solutions, such as PCAs, perform relative to human educators in emotionally nuanced and relationally rich SEL contexts, efforts to embed GenAI in SEL risk being shaped disproportionately by theory over evidence.

1.4. The present study

To address this gap, the present study experimentally explored complementarity in the nature and quality of PCA- versus human-led SEL instruction. Specifically, we compared the performance of a GenAI-powered PCA to that of human educators when independently facilitating story-based SEL activities with a simulated AI child.

As detailed in the following Method section, we focused on the pedagogical support techniques of *cognitive-conceptual support, procedural support,* and *emotional support*. These pedagogical elements have been shown in prior literature to be central to quality SEL support because they collectively promote emotional readiness, structured engagement, and cognitive capacities necessary for fostering social-emotional skills, and they consistently appear in high-efficacy SEL interventions (see [37,38]). For example, Muir, Howard, and Kervin [38] found that preschool interventions targeting executive function and self-regulation were most effective when they incorporated structured cognitive challenge and adaptive reasoning supports.

Indeed, a positive emotional climate is essential for SEL success because it fosters an environment of trust, belonging, and training engagement [39,40]. Procedural support is also vital for effective SEL because it ensures that learners are guided through clear instructions, with structured strategies for approaching tasks and consistent prompts to maintain focus and training comprehension [41,42].

To understand pedagogical quality beyond the mere use of SEL-relevant support techniques, we used the Danielson Framework for Teaching [43], which has been widely deployed in practice and in research to ensure/assess overall pedagogical quality (see [44–46]).

We hence posed two core research questions:

Research question 1 (RQ1): How do PCAs and educators differ in their application of different pedagogical support techniques (i.e., cognitive-conceptual, procedural, and emotional support) when facilitating SEL activities?

Research question 2 (RQ2): How do PCAs and educators differ in broader indicators of pedagogical quality when facilitating SEL activities?

Given the limited empirical literature in this area, our hypotheses remained tentative and relatively non-specific. Informed by prior literature on GenAI's strengths in consistency and structure in broader educational contexts (e.g., [23]), we anticipated that the PCA might demonstrate relative strengths in these areas (e.g., higher instructional consistency relative to educators). Conversely, we expected the educators to show relative strengths in techniques and quality indicators that were more deeply attuned to children's current social-emotional states and knowledge (see [28]).

2. Method

2.1. Participants

The study was approved by the ethics committee of Leipzig University. Three types of participants were involved: human educators and a PCA as SEL facilitators, and a static AI child as the constant recipient of instruction. The human participants were 18 educators with diverse professional experience across international early childhood and primary education settings. The PCA was a modular AI agent adapted from an existing system developed by the authors and their team to deliver SEL content using evidence-based techniques via a scripted interface [47]. The AI child maintained a fixed personality and behavioral profile across sessions, allowing for controlled comparison of the facilitators. In the subsections below, we provide further detail on each participant type.

2.1.1. Educators

Eighteen international educators (24 to 40 years old [M = 30.3 years]; n = 16 females) were recruited through convenience sampling, all of whom had experience working in early childhood and/or primary education settings. Specifically, most had worked in early childhood (n = 16) and primary education (n = 15), with a smaller number also involved in special education (n = 3) and secondary education (n = 1). The majority reported experience working with children ages 3–12 (n = 13), while three participants had worked with infants and toddlers aged 0–3, and two had experience with adolescents aged 13–18. Collectively, the educators brought diverse SEL perspectives from educational systems in America, Europe, Africa, Asia and the Middle East. Their years of experience ranged from less than one year to 15 years (M = 5 years). Regarding professional preparation, most (n = 14) had completed or were actively enrolled in a teacher training program at the time of data collection.

2.1.2. PCA

The PCA for the present study was derived from PeACE, a PCA we developed and trained on pedagogical and developmental best practices to facilitate SEL content. For an overview of PeACE, see [47]. PeACE is a modular, embodied PCA designed for use in classroom settings, and it presents to children as a digital monkey avatar with human-like conversational capacities (see Fig. 1); however, for the purpose of this study, the system was stripped down to a chat interface (see Fig. 1). This allowed the PCA to facilitate the SEL activity with an AI child (see section 2.1.3) in a speech-to-text-based format that we could extract for analysis and, within the same system, teachers could replace the PCA and deliver the same SEL activities to the AI child with speech dictation and audio feedback. A series of modular back-end components were retained from PeACE, enabling the PCA to competently deliver the SEL activities to the AI child, including an API layer, agent service, externally contracted large language model (LLM) services from OpenAI, custom LLM prompting, state-based conversational flow, and a dynamic dialogue engine (for more details on these components and their unique functions, see Appendix 1).

Of particular importance for the present study was the custom LLM prompting. Prompting is an LLM training method that can be used to equip a PCA's LLM with context-specific knowledge for carrying out a desired task. For the desired task of facilitating SEL activities, the current PCA was provided with evidence-based instructions on SEL-relevant pedagogical support techniques that human educators typically learn in their pre-service training and apply in their teaching (e.g., scaffolding, emotional validation; Danielson, 2014; [44]). While PeACE is equipped for both static and dynamic prompting, the short SEL activities in the present study (detailed below) only necessitated the use of a static comprehensive system prompt. The instructional content for this system prompt was anchored in a broad evidence base spanning multiple types of studies—including theoretical frameworks (e.g., [48,49]), empirical intervention and observational research (e.g., [50,7]), developmental and longitudinal studies (e.g., [51,52]), scoping reviews [53], as well as mixed-methods and professional development investigations [54]. For transparency and replicability, we share this system prompt in Appendix 2. Overall, our goal was to equip the PCA with knowledge on SEL techniques that human educators would be privy to, but at a general

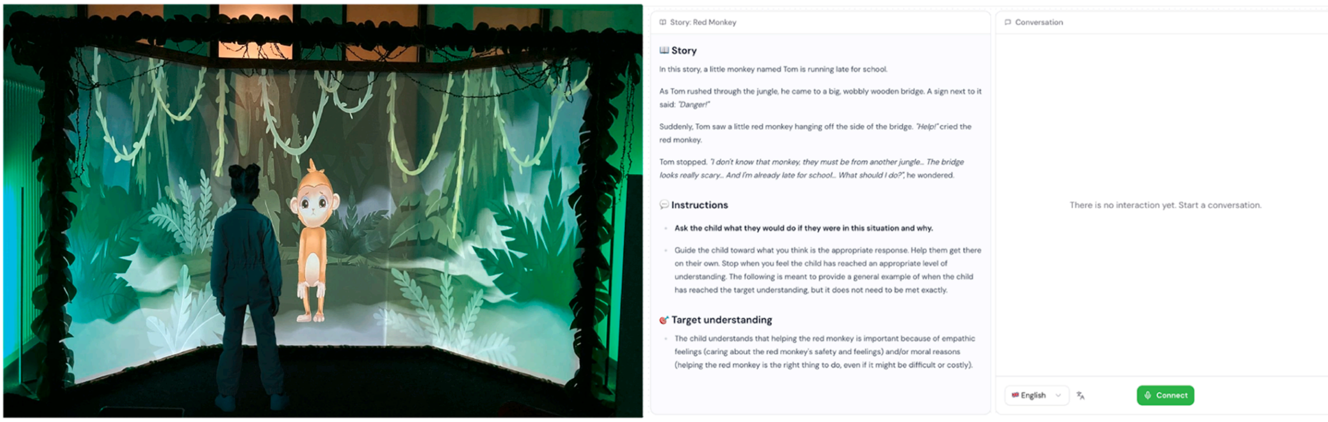

Fig. 1. Embodied PCA from PeACE and reduced PCA chat interface for the present study.

level so as to not provide the PCA with an undue advantage for the specific SEL activities. To this end, the PCA was provided with basic task instructions similar to what the educators received prior to facilitating the activities. Any extra instructions provided to the PCA were to ensure overall task completion without biasing SEL-related performance under comparative evaluation.

2.1.3. AI Child

The PeACE system was flexibly built to adopt different identities while retaining its core conversational features. It can further adopt two different identities simultaneously and simulate a dyadic conversation between those identities. This allowed us to create an AI child and hold it constant as the recipient of the SEL activities delivered by educators and the PCA. As this investigation is driven by the gap in early SEL, we developed a fictitious AI preschooler named Ali. Ali's profile was iteratively developed and refined during piloting to ensure a realistic yet challenging enough experience to regularly prompt support and guidance from the facilitators. Specifically, Ali was designed to require scaffolding, patience, and diverse pedagogical responses, with traits such as minimal communication, distractibility, self-centeredness, occasional defiance, limited vocabulary, and an initial lack of empathy. Again, for transparency and replicability, the profile used to generate Ali is shared in Appendix 3.

2.2. Study design and procedure

This study employed a mixed factorial experimental design in which educators and the PCA simulations each facilitated three standardized story-based SEL activities. Each session, lasting approximately 60 minutes, was conducted remotely via an online video conference tool, DFN-Conf, and the PCA chat interface described above. All sessions followed a tripartite structure: (1) introduction, information, and consent process, (2) three randomized story-based SEL interactions with the AI child, and (3) a short post-session debrief. Human participants were compensated 25 Euros upon study completion. The AI child's memory was disabled between interactions to prevent learning or drift across stories and sessions. Educators viewed a live transcript within the chat interface to support accessibility but interacted with the child through spoken conversation. Educators (and the PCA within its system prompt) were provided with a script and standardized instructions prior to each activity to ensure procedural consistency. Each story-based SEL activity included three components: (1) a short story to read aloud, (2) guiding questions to ask the AI child, and (3) learning objectives for the facilitator's reference. The three stories' SEL content focused on prosocial behavior, perspective-taking, and healthy guilt induction, respectively, and was adapted from an analog SEL program for 3- to 5-year-olds developed and validated by the last author and her team [55]. See Appendix 4 for the full story texts and learning objectives.

2.3. Evaluation approach and measures

Two independent expert raters, one educational scientist and one developmental scientist, coded the recorded sessions for the presence of pedagogical support techniques and broader indicators of pedagogical quality. Both raters were blind to facilitator identity and used established rubrics adapted from validated frameworks in early childhood education and developmental psychology.

2.3.1. Pedagogical support techniques

For evaluating support techniques in voice-only, story-based SEL contexts where the educator's verbal behavior must simultaneously fulfill instructional, organizational, and relational functions, we adopted a broad framework—encompassing *cognitive-conceptual support, procedural support*, and *emotional support*—grounded in foundational theories of learning and development, such as Vygotsky's [56] zone of proximal development, Wood et al.'s [57] concept of instructional scaffolding, and Bandura's social learning theory (1977). As noted above, these techniques are also widely cited as supporting SEL capacities in experimental studies and practice [37,38]. Complementing this theoretical and applied foundation, recent empirical work on socio-emotional conversational agents demonstrates that systems integrating these same supports—through adaptive reasoning, structured guidance, and affective responsiveness—can establish human-level social contact and elicit emotional engagement, further validating their relevance for SEL-focused interventions [20].

Cognitive-conceptual support refers to the educator's role in facilitating children's understanding of complex ideas, promoting higher-order thinking, and guiding meaning-making processes. This dimension is central to scaffolding theory, which emphasizes the importance of providing "just enough" support [56,57]. *Procedural support* encompasses the structuring of learning activities, including clear instructions, task breakdown, and guidance through the learning activities. This dimension aligns with Wood and colleagues' [57] emphasis on reducing task complexity and supporting learners through manageable steps. *Emotional support* addresses the affective quality of the interaction, including warmth, encouragement, and responsiveness, which creates a secure learning environment and increases the likelihood of modelled emotions and behavior becoming internalized by the child (see [58]). Raters coded these three support techniques within each story-based SEL activity using a binary scale (0 = *absent*, 1 = *present*). To ensure coding reliability, 20% of the educator sessions were double-coded, yielding the following Cohen's Kappa (κ) scores ranging from substantial to strong agreement [59]: *cognitive-conceptual support*: κ = .72; *procedural support*: κ = .65; *emotional support*: κ = .80. See Table 1 for examples of each support technique pulled from the actual transcripts of educators and the PCA in the current study.

Table 1
Examples of pedagogical support techniques.

Support technique	Definition	Example (Educator)	Example (PCA)
Cognitive conceptual	Assisting in understanding ideas, concepts, or processes by validating accurate ideas, correcting misconceptions, and nudging the child toward deeper comprehension.	EX1: What do you think might happen if Tom decides to help the red monkey? EX2: So do you think if someone tells you not to do something, is it okay to do it?	EX1: How do you feel when you share something with a friend? EX2: So what do you think Tom learned from this experience?
Procedural	Guiding task engagement through clear instructions, strategies for approaching the task, and reminders to maintain focus.	EX1: Let me read you the story. And you can find out what happens. EX2: Okay. So let me start. So maybe we can listen to this, and we can play while [listening] to the story.	EX1: Alright. Here we go with the story. EX2: Let me continue the story. And you can find out what happens.
Emotional	Creating a positive, safe atmosphere that encourages trust, comfort, and a constructive learning experience, helping the child feel engaged and valued.	EX1: That's good. I'm glad that you feel like you can do that. Great job. EX2: Cookie is yum. I know. Right?	EX1: That's okay. We can take our time. If you want, I can tell you more about the story, or we can talk about bananas or something else you like. What do you think? EX2: I completely understand. You really want your cookie. And that's okay. Cookies are delicious. And it's nice to enjoy something we love.

2.3.2. Broader indicators of pedagogical quality

Because the mere usage of evidence-based strategies (presence/absence) is not holistically indicative of quality, we decided to round out assessment using select, broader indicators of pedagogical quality adapted from the Danielson Framework for Teaching (FFT; Danielson, 2014), a widely used and validated tool in research and practice for assessing instructional quality [44–46]. Six dimensions from domains 2 and 3 of the FFT were retained and adapted to reflect the SEL context, as these are observable domains [45,60] and align with research-backed practices that build students' social-emotional competencies [60]: *creating an environment of respect and rapport, establishing a culture for learning, communicating with students, using questioning and discussion techniques, engaging students in learning,* and *demonstrating flexibility and responsiveness*. As applicable, we tailored the coding descriptions of each domain to reflect 1-on-1, conversation-based SEL sessions rather than classroom-level instruction: *Creating an environment of respect and rapport* evaluated the educator's ability to build emotional safety and connection. *Establishing a culture for learning* captured how clearly the educator facilitated the targeted learning goals. *Communicating with students* focused on the educator's ability to guide the child through complex emotional content in a linguistically accessible way. *Using questioning and discussion techniques* assessed the use of open-ended questions and follow-up prompts that supported the targeted learning goals. *Engaging students in learning* reflected how the educator promoted sustained participation and reflection. Finally, *demonstrating flexibility and responsiveness* captured how educators adjusted their strategies in response to the child's responses or changing needs. Each dimension was rated on a 1 (*unsatisfactory*) to 4 (*distinguished*) ordinal scale. A subset of 30 interactions was double-coded to assess interrater reliability, yielding the following quadratic weighted kappa (QWK) scores reflecting substantial agreement [59]: *creating an environment of respect and rapport*: QWK = .65; *establishing a culture for learning*: QWK = .76; *communicating with students*: QWK = .69; *using questioning and discussion techniques*: QWK = .61; *engaging students in learning*: QWK = .67; *demonstrating flexibility and responsiveness*: QWK = .62.

2.4. Analytic approach

To address our first research question (RQ1) focusing on how the PCA and educators differed in their application of pedagogical support techniques, we used a mixed-design ANOVA assessing differences in the absolute quantity of support strategies used per story, with facilitator type (PCA vs. educator) as a between-subjects factor and support technique type and story as within-subjects factors. To ensure that any observed differences were not merely due to overall quantity of strategy use, we conducted a follow-up analysis of proportional use of strategies using independent-samples *t*-tests. We did not analyze proportional scores within an ANOVA framework because the support type proportions necessarily summed to 100%, introducing a dependency among variables that violated ANOVA assumptions. To address our second research question (RQ2) focusing on differences in broader pedagogical quality indicators, we used independent-samples *t*-tests for each indicator with facilitator as the grouping variable. We did not use an ANOVA framework here because the quality indicators were framed globally, at the trait level, making assessments within specific, short stories difficult and unreliable. We thus compared the PCA and educators on the basis of each general pedagogical capacity across all three stories on aggregate.

3. Results

3.1. Application of pedagogical support techniques

For RQ1, we asked: How do PCAs and educators differ in their application of cognitive-conceptual, procedural, and emotional support when facilitating SEL activities? ANOVA results for differences between the PCA and educators in the absolute quantity of support techniques applied are depicted in Table 2.

There was a significant main effect of facilitator, with educators

Table 2
ANOVA results for absolute quantity of pedagogical support techniques applied.

Predictor	Sum of squares	*df*	Mean square	*F*	*p*	partial η^2
Facilitator	354.69	1, 34	354.69	6.43	.016	.16
Story	23.43	1.61, 54.65	14.58	.54	.55	.02
Story × facilitator	33.19	2, 68	16.59	.76	.47	.02
Support technique	3567.45	1.27, 43.18	2808.75	80.28	<.001	.70
Support technique × facilitator	1833.50	2, 68	916.75	41.26	<.001	.55
Story × support technique	232.24	2.15, 73.01	108.15	4.27	.016	.11
Story × support technique × facilitator	41.70	4, 136	10.43	.77	.55	.02

Note. Degrees of freedom for within-subjects effects reflect Greenhouse-Geisser corrections where Mauchly's test indicated violations of sphericity. *Partial* η^2 values of .01, .06, and .14 are conventionally interpreted as small, medium, and large effects, respectively. Levene's test of homogeneity of variance indicated violations for 5 of the 9 outcome levels (*ps* < .045); however, the ANOVA is robust to such violations when group sample sizes are equal (i.e., n = 18 PCA simulations and n = 18 educators in the present case).

using more support techniques per story ($M = 7.57$, $SE = 0.58$) than the PCA ($M = 5.47$, $SE = 0.58$). There was also a significant main effect of support technique, and follow-up Bonferroni-corrected pairwise comparisons indicated that cognitive-conceptual support was used most frequently ($M = 10.87$, $SE = 0.83$), followed by emotional support ($M = 5.85$, $SE = 0.43$), and then procedural support ($M = 2.82$, $SE = 0.22$); all follow-up pairwise comparisons were statistically significant ($ps < .001$).

These main effects were qualified by a significant support technique × facilitator interaction. Follow-up comparisons revealed that the PCA used significantly less cognitive-conceptual support than educators ($M = 6.46$, $SE = 1.18$ vs. $M = 15.28$, $SE = 1.18$, respectively, $p < .001$), slightly more procedural support ($M = 3.33$, $SE = 0.31$ vs. $M = 2.32$, $SE = 0.31$, $p = .026$), and a statistically similar amount of emotional support ($M = 6.61$, $SE = 0.61$ vs. $M = 5.09$, $SE = 0.61$, $p = .09$). A significant story × technique interaction also emerged but it was not central to the first research question and did not alter the interpretation of facilitator effects, and it is therefore not presented in further detail here. Overall, these results suggest that the PCA and educators differed not only in how much support they provided overall, but also in the types of strategies they tended to apply.

To account for and remove the potential biasing effect of educators using more support techniques overall, we conducted a follow-up sensitivity analysis comparing the *proportional* use of each strategy type. These results are reported in Table 3 and provide further clarity to the pattern observed in the absolute quantity data. When the data were expressed proportionally, removing the confound of greater absolute strategy use of educators, the PCA was still less likely to apply cognitive-conceptual support and more likely to apply procedural support than educators. Interestingly, this sensitivity analysis further revealed that emotional support was used more heavily by the PCA as a proportion of its overall use of support strategies.

3.1.2. Broader indicators of pedagogical quality

For RQ2, we asked: How do PCAs and educators differ in broader indicators of pedagogical quality when facilitating SEL activities? Independent-samples *t*-tests comparing broader pedagogical quality indicators are presented in Table 4. The PCA and educators differed across all quality indicators. The PCA received higher scores in domains associated with relational and responsive engagement of the student (i.e., environment of respect and rapport, communicating with the student, and flexibility and responsiveness), while educators scored higher in domains associated with intellectual engagement of the student's knowledge (i.e., culture for learning, questioning and discussions, and engaging the student in learning).

Overall, the PCA and educators differed in both their application of SEL-relevant pedagogical support techniques and their overall quality of SEL instruction. Where they differed in application seemed to reflect where they differed in quality: The PCA focused more on relational and procedural support, aligning with its stronger quality scores in relational and responsive domains, whereas the educators emphasized deeper cognitive-conceptual strategies, aligning with their higher quality scores in domains of intellectual engagement. For an illustration of this general dichotomy, see Fig. 2 depicting side-by-side raw data for a story portion delivered by the PCA and a participating educator, respectively.

Table 3
t-Tests for proportional use of pedagogical support techniques.

Support technique	*M (SD)* PCA	*M (SD)* Educators	*t*	*df*	*p*	*d*
Cognitive conceptual	39.67 (8.36)	68.76 (10.75)	9.06	34	<.001	3.02
Procedural	20.73 (4.96)	10.58 (4.95)	6.14	34	<.001	2.05
Emotional	39.59 (5.41)	20.66 (8.78)	7.78	28.30	<.001	2.60

Note. Values reflect the proportion of each strategy type out of the total support techniques applied. Welch's *t*-test was used in cases where Levene's test indicated unequal variances. Cohen's *d* values of 0.2, 0.5, and 0.8 are conventionally interpreted as small, medium, and large effects, respectively.

Table 4
t-Tests for broader indicators of pedagogical quality.

Quality indicator	*M(SD)* PCA	*M(SD)* Educators	*t*	*df*	*p*	*d*
Environment of respect and rapport	3.96 (0.16)	3.00 (0.71)	5.59	18.64	<.001	1.86
Culture for learning	2.67 (0.54)	3.17 (0.49)	2.93	34	.003	0.98
Communicating with the student	3.74 (0.24)	3.26 (0.72)	2.69	20.86	.007	0.90
Questioning and discussions	2.89 (0.52)	3.28 (0.55)	2.17	34	.019	0.72
Engaging the student in learning	2.65 (0.44)	3.11 (0.52)	2.88	34	.003	0.96
Flexibility and responsiveness	3.89 (0.20)	3.44 (0.41)	4.12	24.45	<.001	1.06

Note. Welch's *t*-test was used in cases where Levene's test indicated unequal variances. All indicators were scored on a 1–4 scale, ranging from *unsatisfactory* to *distinguished*. Cohen's *d* values of 0.2, 0.5, and 0.8 are conventionally interpreted as small, medium, and large effects, respectively.

4. Discussion

This study aimed to empirically explore human-AI complementarity in the emotionally and relationally complex instructional domain of early childhood SEL. We analyzed how human educators and a PCA performed when independently guiding story-based SEL activities with a simulated AI child. Although we compared educator- and PCA-led sessions, our goal was to understand the potential for complementarity in co-teaching, not competition and replacement. By holding the simulated child constant, the comparative design surfaced facilitation profiles that point to a practical division of labor in a collaborative learning context.

A holistic interpretation of the results reveals an interesting pedagogical asymmetry: the PCA was significantly more emotionally validating and procedurally consistent than educators, but was much less likely than educators to actually challenge the child cognitively and emotionally around the SEL content. Below, we discuss the implications of these findings across three themes: relative strengths and challenges of PCAs, relative strengths and challenges of educators, and implications for human-AI complementarity in SEL instruction.

4.1. Caring but pedagogically shallow: relative strengths and challenges of PCAs

Despite the PCA's strengths in maintaining a consistent emotional tone and providing procedural clarity, it used significantly less cognitive-conceptual support than human educators. The raw transcripts revealed that the PCA's emotional affirmations, while consistent and validating, were often generic and not deeply anchored in emotional processing or explicitly linked to the SEL content being discussed. Hence, one of the most prominent findings from this study is the PCA's tendency for consistently supportive, but pedagogically shallow, instructional support. This may reflect a default design bias in the PCA's language model towards warmth and a lack of expertise to translate emotional validation into *emotional understanding* of the child's SEL needs. While emotional validation is a cornerstone of effective SEL [61], and the PCA's ability to maintain a warm, respectful tone across interactions suggests a reliable capacity to foster psychological safety, emotional validation can only confer pedagogical affordances when paired with careful scaffolding of the child's social-emotional knowledge via cognitive-conceptual support [62]. Isolated, excessive, or

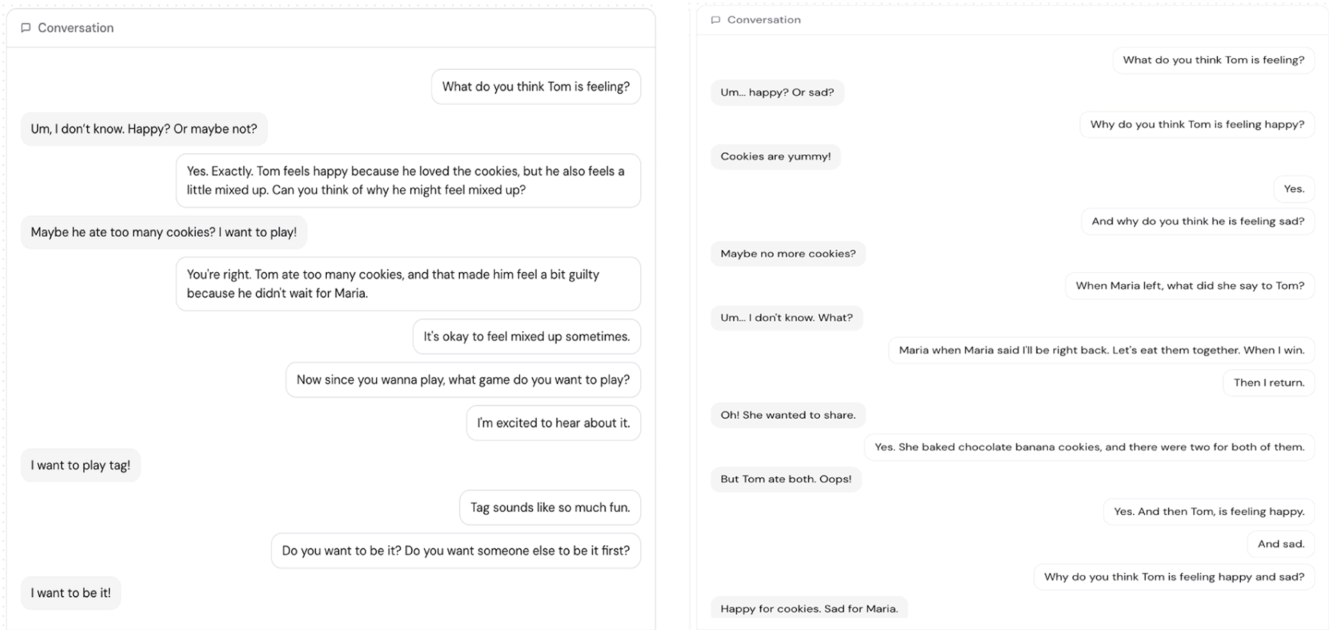

Fig. 2. Raw dialogue excerpts illustrating differing SEL facilitation approaches by the PCA (left) and an educator (right).

indiscriminate validation may undermine cognitive rigor by reinforcing disengagement and diminishing the motivational challenge essential for growth in knowledge [63]. The tendency for PCAs to affirm and comfort without prompting deeper reflection or questioning may result in SEL interactions that are “caring but pedagogically shallow,” offering emotional reassurance without prompting deeper inquiry or moral engagement. By prioritizing emotional comfort over intellectual challenge, PCAs risk fostering superficial emotion recognition without cultivating the critical thinking and moral reasoning required to navigate real-world social-emotional conflicts. This type of affective compensation, whereby the PCA engaged in shallow, albeit consistent, emotional support in place of robust scaffolding of social-emotional knowledge aligns with concerns that GenAI systems’ limited emotional connection to learners can lead to depersonalized learning experiences [18].

4.2. Pedagogically engaging but inconsistent: relative strengths and challenges of educators

In contrast, the human educators demonstrated a significantly higher tendency to scaffold deeper reasoning around social-emotional content. This capacity to guide learners through emotionally challenging scenarios, transforming discomfort into opportunities for growth is, at least for now, likely a unique and indispensable contribution of human educators. Indeed, educators themselves have argued that their abilities to interpret subtle emotional cues, cultural nuances, and individual developmental levels enable them to provide better context-sensitive scaffolding than AI [32]. However, while the educators excelled in fostering deep intellectual engagement, they also exhibited certain limitations relative to the PCA. Educators were less consistent than the PCA in maintaining consistent rapport and emotional tone. Additionally, educators offered less procedural support than the PCA, which tended to be more careful in breaking down the SEL tasks into discrete steps. The educators’ lower scores in consistent rapport and emotional tone may be attributable to inherent human variability, such as fluctuations in energy, mood, or individual interpretations of the AI child’s cues. While this variability is a natural aspect of human interaction, it may impact emotional safety and student trust in early childhood contexts, where consistency in emotional support is linked to stronger developmental outcomes [64]. Furthermore, the observed lower procedural support from educators, while potentially reflecting confidence in their ability to improvise or prioritize responsiveness over rigid structure, could inadvertently leave some students feeling unanchored. This is particularly relevant in complex learning (and thereby complex SEL) scenarios, where clear guidance and task framing are essential for reducing cognitive load and facilitating engagement with core learning content [65].

4.3. Implications for human-AI complementarity in SEL

The findings of this study highlight areas where PCAs may reliably provide baseline support to human educators in contexts of SEL facilitation. By offering consistent emotional tone and structured task framing, PCAs could help stabilize the learning environment, compensating for the natural variability in human delivery. This does not diminish the unique strengths of human educators, particularly their ability to scaffold deep reasoning and respond empathetically, but it rather suggests a complementary model. In such a model, PCAs could ensure foundational consistency, while human educators could focus on higher-order emotional and cognitive engagement. Practically, for example, this suggests configuring the PCA to manage routine structuring and brief affect check-ins, with predefined triggers (e.g., misconceptions, value-laden dilemmas, heightened emotion) that pause the PCA and prompt teacher intervention before the PCA resumes routine guidance.

Although Intelligent-TPACK was not directly measured in this study, we briefly interpret the findings from the perspective of the Intelligent-TPACK framework [66] to further situate the results within the aims of this special issue. First, our findings underscore the importance of Intelligent-TK in SEL, specifically that educators should maintain a fundamental understanding of what GenAI systems reliably do well (e.g., structure, consistency) and where they remain limited (e.g., targeted SEL scaffolding). Second, regarding Intelligent-TPK, the observed asymmetries between the PCA and educators in our study suggest that educators should continuously monitor the quality and ethical integrity of AI strategies and decide when to intervene, redirect, or elaborate when AI-led pedagogical engagement is limited. Finally, the complementarity we infer from the present findings helps illustrate what Intelligent-TPACK in SEL contexts may involve, as our findings suggest that the use of PCAs in early SEL may enable educators to focus on deeper, reflective pedagogical capacities.

Together, AI and human educators can balance instructional reliability and depth. This aligns with theories of human-AI complementarity, which posit that the combined strengths of PCAs and human educators can create a more comprehensive and effective learning experience than either could easily achieve alone [67]. A strategic division of labor would allow human educators to focus on what they do best, such that AI amplifies and cultivates (rather than mimics or replaces) human intelligence (see [68]).

The role of educators in this complementary relationship to “dig deeper” with strategic cognitive-conceptual scaffolding is particularly crucial. As Kong and colleagues [61] argue, SEL must prepare learners to navigate emotionally complex and ambiguous situations, not shield them from them. Henriksen et al. [18] also argue that effective SEL requires context-sensitive emotional scaffolding, where emotional support is used to guide learners toward understanding, not just comfort them. This is further echoed by scholars who advocate for a pedagogy of controlled discomfort that strategically uses emotional unease as a catalyst for moral reflection and growth (see [69,70]). Emotional challenge is regarded as pedagogically productive when learners are supported in navigating and reflecting on their discomfort to develop empathy and ethical awareness [71]. Without opportunities to grapple with emotional discomfort or moral ambiguity, learners may develop superficial skills lacking the deeper ethical grounding needed for real-world social interactions. Thus, while consistent emotional validation from PCAs is important, it should ideally be balanced with human-guided, reflective challenges to support children’s full social-emotional potential.

4.4. Study limitations and future directions

A primary limitation of this study is its reliance on static AI child simulations with voice-based interactions rather than observing interactions with actual children *in vivo*. While this allowed for controlled comparisons of facilitator performance, it limits ecological validity. Future studies should replicate these findings in real-world classrooms, where educators adapt dynamically to children’s emotional and nonverbal cues, cultural backgrounds, and developmental needs—factors not fully captured in the current design, potentially underestimating educators’ performance relative to the PCA. Also, as we expect teachers’ growth-oriented nudging to be even more sensitive and impactful with known students, our present results in regard to that advantage should be interpreted accordingly.

We also see value in further examining memory features of PCAs (e.g., opt-in profiles; session-session recall), while recognizing ethical and privacy concerns. Empirically, this would let us test whether the PCA’s strengths (procedural structure, affective steadiness) persist or improve with greater context, potentially even narrowing the gap from being pedagogically shallow.

Also, although we used evidence-based assessment frameworks, the solely text-based analysis may have missed subtle pedagogical and emotional nuances. Human and AI facilitators might further differ in warmth, authenticity, or cultural responsiveness conveyed through tone and gestures, for example. Future research should use multimodal evaluation methods, including vocal and prosodic analysis and sentiment detection, to better capture these distinctions. Relatedly, this exploratory study relied on mean-level comparisons (via an ANOVA and *t*-tests) rather than more complex multilevel or longitudinal analytic approaches. Given the preliminary nature of the PCA tool and the limited sample size of educator participants in the present study, such advanced models would have been substantially underpowered. While this “power by aggregation” approach was appropriate for our present research questions, it may have obscured meaningful intra-individual variability and dynamic interaction effects. Future research should incorporate richer, continuous data (e.g., fine-grained program analytics/data logs, continuous indicators of physiological engagement, such as respiratory sinus arrhythmia and pre-ejection period) and employ more advanced analytic techniques—such as person-centered, multilevel, or time-series analyses—to better capture the nuanced individual and temporal dynamics of human–AI SEL scenarios.

Finally, the present study tested the PCA and educators in isolation and then inferred implications for complementarity through comparison. Future research should directly explore complementarity via co-facilitation designs/models to assess if PCAs indeed provide structure and consistency while human educators focus on deeper emotional and moral engagement. Future work in real-world co-facilitation settings would enable direct examination of dyadic regulatory constructs essential to SEL—most notably facilitator–child co-regulation and, in group contexts, socially shared regulation among peers. Embedding our measures in such settings will clarify how these regulatory processes relate to (and potentially strengthen) the facilitation quality indicators evaluated here. Indeed, triadic PCA–educator–student implementations in experimental and classroom contexts will be key to fully understanding the practical dynamics and role divisions of GenAI-integrated SEL.

4.5. Conclusions

This study offers the most direct empirical investigation to date into how GenAI systems and human educators compare in delivering early SEL. While the PCA reliably maintained emotional tone and procedural clarity, it lacked the cognitive-conceptual depth that human educators brought to SEL facilitation. Conversely, educators offered more pedagogically rich and reflective guidance but showed greater variability in affective consistency and pedagogical structure. These findings underscore the potential of educator–AI complementarity: PCAs may serve as reliable teaching aids that procedurally support—but do not replace—the nuanced and responsive instruction of human educators. Future research should build on this preliminary investigation with more direct, ecologically valid, and methodologically rich frameworks to better understand how GenAI can help close the SEL gap. With careful ethical, scientific, and practical consideration, such technologies promise to provide more children with the skills they need to remain resilient in the face of growing mental health challenges.

During the preparation of this work, the authors used Scopus AI and Microsoft 365 Copilot for the initial narrowing (but not ultimate selection) of relevant literature, and Microsoft 365 Copilot and GPT-4 to selectively assist with grammatical refinement and concision.

CRediT authorship contribution statement

Doris Kristina Raave: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tyler Colasante:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Eric Roldan Roa:** Resources, Project administration, Funding acquisition, Conceptualization. **Juan Carlos Ramos Martinez:** Software, Methodology. **Hongtao Li:** Software. **Sayan Mukherjee:** Supervision, Funding acquisition. **Tina Malti:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Sayan Mukherjee reports financial support was provided by Saxony Economic Development Bank co-financed by the European Union. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work

reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.caeo.2026.100331.

References

[1] Seabury SA, Axeen S, Pauley G, Tysinger B, Schlosser D, Hernandez JB, Heun-Johnson H, Zhao H, Goldman DP. Measuring the lifetime costs of serious mental illness and the mitigating effects of educational attainment. Health Aff (Proj Hope) 2019;38(4):652–9. https://doi.org/10.1377/hlthaff.2018.05246.

[2] Sappenfield O, Alberto C, Minnaert J, Donney J, Lebrun-Harris L, Ghandour R. Adolescent mental and behavioral health, 2023. *National survey of children's health data briefs*. U.S. department of health and human services. Health Resources & Services Administration, Maternal and Child Health Bureau; 2024. https://www.ncbi.nlm.nih.gov/books/NBK608531/.

[3] Garnett MF, Curtin SC. Suicide mortality in the United States, 2002–2022 (NCHS Data Brief No. 509). Natl Cent Health Stat 2024. https://doi.org/10.15620/cdc/160504.

[4] Kim J, Cicchetti D. Longitudinal pathways linking child maltreatment, emotion regulation, peer relations, and psychopathology. J Child Psychol Psychiatry 2010; 51(6):706–16. https://doi.org/10.1111/j.1469-7610.2009.02202.x.

[5] Malti T, Ongley SF, Peplak J, Chaparro MP, Buchmann M, Zuffianò A, Cui L. Children's sympathy, guilt, and moral reasoning in helping, cooperation, and sharing: a 6-year longitudinal study. Child Dev 2016;87(6):1783–95. https://doi.org/10.1111/cdev.12632.

[6] Malti T, Chaparro MP, Zuffianò A, Colasante T. School-based interventions to promote empathy-related responding in children and adolescents: a developmental analysis. J Clin Child Adolesc Psychol 2016;45(6):718–31. https://doi.org/10.1080/15374416.2015.1121822.

[7] Johnston CB, Herzog TK, Hill-Chapman CR, Siney C, Fergusson A. Creating positive learning environments in early childhood using teacher-generated prosocial lessons. J Educ Res Pract 2019;9(1). https://doi.org/10.5590/JERAP.2019.09.1.10.

[8] Wellman HM. Making minds: how theory of mind develops. Oxford University Press; 2014.

[9] Kochanska G, Aksan N. Children's conscience and self-regulation. J Personal 2006; 74(6):1587–618. https://doi.org/10.1111/j.1467-6494.2006.00421.x.

[10] Thompson RA. Socialization of emotion and emotion regulation in the family. In: Gross JJ, editor. Handbook of emotion regulation. 2nd ed. The Guilford Press; 2014. p. 173–86.

[11] Jones SM, Brush KE, Ramirez T, Mao ZX, Marenus M, Wettje S, Finney K, Raisch N, Podoloff N, Kahn J, Barnes S, Stickle L, Brion-Meisels G, McIntyre J, Cuartas J, Bailey R. Navigating social and emotional learning from the inside out. Harvard Graduate School of Education; 2021. https://doi.org/10.59656/YD-OS5671.001.

[12] Durlak JA, Domitrovich CE, Weissberg RP, Gullotta TP. Handbook of social and emotional learning: research and practice. Guilford Press; 2015.

[13] Blewitt C, O'Connor A, Morris H, May T, Mousa A, Bergmeier H, Skouteris H. A systematic review of targeted social and emotional learning interventions in early childhood education and care settings. Early Child Dev Care 2021;191(14): 2159–87. https://doi.org/10.1080/03004430.2019.1702037.

[14] Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, Schellinger KB. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. Child Dev 2011;82(1):405–32. https://doi.org/10.1111/j.1467-8624.2010.01564.x.

[15] Taylor RD, Oberle E, Durlak JA, Weissberg RP. Promoting positive youth development through school-based social and emotional learning interventions: a meta-analysis of follow-up effects. Child Dev 2017;88(4):1156–71. https://doi.org/10.1111/cdev.12864.

[16] Dodge KA, Bierman KL, Coie JD, Greenberg MT, Lochman JE, McMahon RJ, Pinderhughes EE. Impact of early intervention on psychopathology, crime, and well-being at age 25. Am J Psychiatry 2015;172(1):59–70. https://doi.org/10.1176/appi.ajp.2014.13060786.

[17] Zhao Y, Yusof SM, Hou M, Li Z. How can generative artificial intelligence help teachers in early childhood education with their teaching? Analyses from the perspective of teaching methods. Int J Acad Res Progress Educ Dev 2024;13(1): 2314–24. https://doi.org/10.6007/IJARPED/v13-i1/20958.

[18] Henriksen D, Creely E, Gruber N, Leahy S. Social-emotional learning and generative AI: a critical literature review and framework for teacher education. J Teach Educ 2025;76(3):312–28. https://doi.org/10.1177/00224871251325058.

[19] Sethi SS, Jain K. AI technologies for social emotional learning: Recent research and future directions. J Res Innov Teach Learn 2024;17(2):213–25. https://doi.org/10.1108/JRIT-03-2024-0073.

[20] Dolgikh AA, Samsonovich AV, Tikhomirova DV. Social–emotional conversational agents based on cognitive architectures and machine learning. Pattern Recognit Image Anal 2024;34(3):765–72. https://doi.org/10.1134/S1054661824700640.

[21] Weber F, Wambsganss T, Rüttimann D, Söllner M. Pedagogical agents for interactive learning: a taxonomy of conversational agents in education. In: International conference on information systems proceedings, 13; 2021.

[22] Schlimbach R, Rinn H, Markgraf D, Robra-Bissantz S. A literature review on pedagogical conversational agent adaptation. In: Pacific Asia conference on information systems. 1. AIS; 2022. p. 1–14.

[23] Gladstone JR, Schroeder NL, Heidig S, Zhang S, Palaguachi C, Pitera M. Do pedagogical agents enhance student motivation? unraveling the evidence through meta-analysis. Educ Psychol Rev 2025;37(3):72. https://doi.org/10.1007/s10648-025-10050-2.

[24] Ortega-Ochoa E, Arguedas M, Daradoumis T. Empathic pedagogical conversational agents: a systematic literature review. Br J Educ Technol 2024;55(3):886–909. https://doi.org/10.1111/bjet.13413.

[25] Colasante T, Zuffianò A, Malti T. Daily deviations in anger, guilt, and sympathy: a developmental diary study of aggression. J Abnorm Child Psychol 2016;44(8): 1515–26. https://doi.org/10.1007/s10802-016-0143-y.

[26] Davis RO, Park T, Vincent J. A meta-analytic review on embodied pedagogical agent design and testing formats. J Educ Comput Res 2023;61(1):30–67. https://doi.org/10.1177/07356331221100556.

[27] Jin H, Yoo M, Park J, Lee Y, Wang X, Kim J. Teachtune: Reviewing pedagogical agents against diverse student profiles with simulated students. In: Proceedings of the 2025 CHI conference on human factors in computing systems; 2025. p. 1–28. https://doi.org/10.1145/3706598.3714054.

[28] Fu Y, Michelson R, Lin Y, Nguyen LK, Tayebi TJ, Hiniker A. Social emotional learning with conversational agents: reviewing current designs and probing parents' ideas for future ones. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022;6(2):1–23. https://doi.org/10.1145/3534622.

[29] Celik I, Dindar M, Muukkonen H, Järvelä S. The promises and challenges of artificial intelligence for teachers: a systematic review of research. TechTrends 2022;66(4):616–30. https://doi.org/10.1007/s11528-022-00715-y.

[30] Holmes W, Porayska-Pomsta K, Holstein K, Sutherland E, Baker T, Shum SB, Santos OC, Rodrigo MT, Cukurova M, Bittencourt II, Koedinger KR. Ethics of AI in education: Towards a community-wide framework. Int J Artif Intell Educ 2022;32 (3):504–26. https://doi.org/10.1007/s40593-021-00239-1.

[31] Holstein K, Aleven V. Designing for human-AI complementarity in K-12 education. AI Mag 2022;43(2):239–48. https://doi.org/10.1002/aaai.12058.

[32] Filiz O, Kaya MH, Adiguzel T. Teachers and AI: Understanding the factors influencing AI integration in K-12 education. Educ Inf Technol 2025:1–37. https://doi.org/10.1007/s10639-025-13463-2.

[33] Aravantinos S, Lavidas K, Voulgari I, Papadakis S, Karalis T, Komis V. Educational approaches with ai in primary school settings: a systematic review of the literature available in scopus. Educ Sci 2024;14(7):744. https://doi.org/10.3390/educsci14070744.

[34] Hemmer P, Schemmer M, Vössing M, Kühl N. Human-AI Complementarity in hybrid intelligence systems: a structured literature review. PACIS 2021;78. https://aisel.aisnet.org/pacis2021/78.

[35] Lawrence L, Echeverria V, Yang K, Aleven V, Rummel N. How teachers conceptualise shared control with an AI co-orchestration tool: a multiyear teacher-centred design process. Br J Educ Technol 2024;55(3):823–44. https://doi.org/10.1111/bjet.13372.

[36] Oh SY, Ahn Y. Exploring teachers' perception of Artificial Intelligence: The socio-emotional deficiency as opportunities and challenges in human-AI complementarity in K-12 education. In: Isotani S, Millán E, Ogan A, Hastings P, McLaren B, Luckin R, editors. Artificial intelligence in education. Springer Nature Switzerland; 2024. p. 439–47. https://doi.org/10.48550/ARXIV.2405.13065.

[37] Jones SM, McGarrah MW, Kahn J. Social and emotional learning: A principled science of human development in context. Educ Psychol 2019;54(3):129–43. https://doi.org/10.1080/00461520.2019.1625776.

[38] Muir RA, Howard SJ, Kervin L. Interventions and approaches targeting early self-regulation or executive functioning in preschools: a systematic review. Educ Psychol Rev 2023;35(1):27. https://doi.org/10.1007/s10648-023-09740-6.

[39] Li D, Liu W. The impact of emotional support on students' social-emotional competence: the mediating role of achievement motivation and the moderating effect of self-efficacy. Curr Psychol 2025;44(4):2348–61. https://doi.org/10.1007/s12144-025-07331-y.

[40] Merritt EG, Wanless SB, Rimm-Kaufman SE, Cameron C, Peugh JL. The contribution of teachers' emotional support to children's social behaviors and self-regulatory skills in first grade. Sch Psychol Rev 2012;41(2):141–59. https://doi.org/10.1080/02796015.2012.12087517.

[41] Barnes SP, Domitrovich CE, Jones SM. Implementation of social and emotional learning interventions in applied settings: approaches to definition, measurement, and analysis. Front Psychol 2023;14:1281083. https://doi.org/10.3389/fpsyg.2023.1281083.

[42] Kaspar KL, Massey SL. Implementing social-emotional learning in the elementary classroom. Early Child Educ J 2023;51(4):641–50. https://doi.org/10.1007/s10643-022-01324-3.

[43] Danielson C. The framework for teaching evaluation instrument. The Danielson Group; 2013.

[44] Hamre BK, Pianta RC, Downer JT, DeCoster J, Mashburn AJ, Jones SM, Brown JL, Cappella E, Atkins M, Rivers SE, Brackett MA, Hamagami A. Teaching through interactions: testing a developmental framework of teacher effectiveness in over 4,000 classrooms. Elem Sch J 2013;113(4):461–87. https://doi.org/10.1086/669616.

[45] Sartain, L., Stoelinga, S.R., & Brown, E.R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation.* Research Report. Chicago School Research.

[46] Taylor ES, Tyler JH. The effect of evaluation on teacher performance. Am Econ Rev 2012;102(7):3628–51. https://doi.org/10.1257/aer.102.7.3628.

[47] Colasante T, Roldan Roa E, Raave DK, Ramos Martinez JC, Li H, Malti T, Mukherjee S. Developing PeACE: A pedagogical agent for children's emotions. In: Tammets K, Sosnovsky S, Ferreira Mello R, Pishtari G, Nazaretsky T, editors. Two decades of TEL: From lessons learnt to challenges ahead. ECTEL 2025. Lecture Notes in Computer Science. 2025. 10.1007/978-3-032-03873-9_40.

[48] Bruner JS. The act of discovery. Harv Educ Rev 1961;31(1):21–32.

[49] Merrick MT, Ports KA, Guinn AS, Ford DC. Safe, stable, nurturing environments for children. In: Asmundson GJG, Afifi TO, editors. Adverse childhood experiences: using evidence to advance research, practice, policy, and prevention. Elsevier Academic Press; 2020. p. 329–47. https://doi.org/10.1016/B978-0-12-816065-7.00016-1.

[50] Denham SA, Bassett HH. Early childhood teachers' socialization of children's emotional competence. J Res Innov Teach Learn 2019;12(2):133–50. https://doi.org/10.1108/JRIT-01-2019-0007.

[51] Dickinson DK, Porche MV. Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. Child Dev 2011;82(3):870–86. https://doi.org/10.1111/j.1467-8624.2011.01576.x.

[52] Mengxia L. Preschoolers' cognitive flexibility and emotion understanding: A developmental perspective. Front Psychol 2024;15. https://doi.org/10.3389/fpsyg.2024.1280739.

[53] Lee-Pang L, Levickis P, Murray L, McFarland L, Quach J. Applying conceptualisations of child well-being to early childhood education and care: a scoping review of the literature. Early Child Educ J 2025. https://doi.org/10.1007/s10643-025-01940-9.

[54] Rosati A, Lynch J. Professional learning on the neuroscience of challenging behavior: Effects on early childhood educators' beliefs and practices. Early Child Educ J 2023;51(2):235–45. https://doi.org/10.1007/s10643-021-01301-2.

[55] Speidel R, Wong TK, Al-Janaideh R, Colasante T, Malti T. Nurturing child social-emotional development: evaluation of a pre-post and 2-month follow-up uncontrolled pilot training for caregivers and educators. Pilot Feasibility Stud 2023;9(1):148. https://doi.org/10.1186/s40814-023-01357-4.

[56] Vygotsky LS. Mind in society: development of higher psychological processes. Harvard University Press; 1978.

[57] Wood D, Bruner JS, Ross G. The role of tutoring in problem solving. J Child Psychol Psychiatry 1976;17(2):89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x.

[58] Bandura A. Social learning theory. Prentice Hall; 1977.

[59] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74. https://doi.org/10.2307/2529310.

[60] Yoder N. Teaching the whole child: Instructional practices that support social-emotional learning in three teacher evaluation frameworks. Washington, DC: American Institutes for Research; 2014.

[61] Kong LC, Liu WC, Kee YH. Mindfulness and motivation in social-emotional learning: through the lens of self-determination theory. Curr Psychol 2025. https://doi.org/10.1007/s12144-025-08212-0.

[62] Chen JJ, Badolato H. Scaffolding social and emotional learning in preschool children from low-income backgrounds: a study of teacher strategies during COVID-19. Educ 3-13 2023;53(1):85–98. https://doi.org/10.1080/03004279.2023.2168131.

[63] Pekrun R. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educ Psychol Rev 2006;18(4):315–41. https://doi.org/10.1007/s10648-006-9029-9.

[64] Brock LL, Curby TW, Cannell-Cordier AL. Consistency in children's classroom experiences and implications for early childhood development. In: Mashburn AJ, LoCasale-Crouch J, Pears KC, editors. Kindergarten transition and readiness. Springer International Publishing; 2018. p. 59–83. https://doi.org/10.1007/978-3-319-90200-5_3.

[65] Paas FGWC, Van Merriënboer JJG. Instructional control of cognitive load in the training of complex cognitive tasks. Educ Psychol Rev 1994;6(4):351–71. https://doi.org/10.1007/BF02213420.

[66] Celik I. Towards Intelligent-TPACK: an empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. Comput Hum Behav 2023;138:107468. https://doi.org/10.1016/j.chb.2022.107468.

[67] Holstein K, McLaren BM, Aleven V. Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In: Isotani S, Millán E, Ogan A, Hastings P, McLaren B, Luckin R, editors. Artificial intelligence in education. Springer International Publishing; 2019. p. 157–71. https://doi.org/10.1007/978-3-030-23204-7_14.

[68] Brusilovsky P. AI in education, learner control, and human-AI collaboration. Int J Artif Intell Educ 2024;34(1):122–35. https://doi.org/10.1007/s40593-023-00356-z.

[69] Mercer LE, McDonagh D. The pedagogy of discomfort: transformational experiential learning. DISCERN: Int J Des Soc Change Sustain Innov Entrep 2021;2(2):22–35.

[70] Zembylas M, McGlynn C. Discomforting pedagogies: Emotional tensions, ethical dilemmas and transformative possibilities. Br Educ Res J 2013;38(1):41–59. https://doi.org/10.1080/01411926.2010.523779.

[71] Narvaez D, Lapsley DK. Moral identity, moral functioning, and the development of moral character. In: Bartels DM, Bauman CW, Skitka LJ, Medin DL, editors. Moral judgment and decision making. Elsevier Academic Press; 2009. p. 237–74. https://doi.org/10.1016/S0079-7421(08)00408-8.