AI

MDPI

Article

AI as a Practice Partner: A Feasibility Study of MentaClassAI, a Conversational LLM Tool for Training Educators' Mentalizing Responses to Child Dysregulation

Gali Chelouche-Dwek * and Peter Fonagy

Research Department of Clinical, Educational and Health Psychology, University College London, London WC1H 0AP, UK; p.fonagy@ucl.ac.uk
* Correspondence: gali.dwek.20@ucl.ac.uk

Abstract

Background: Teachers routinely encounter children whose behaviour reflects emotional distress and dysregulation, yet they have limited opportunities to practise the relational skills required to respond effectively. These challenges are particularly pronounced in Alternative Provision (AP), which serves children who frequently present with histories of trauma, neurodevelopmental differences, and complex emotional and behavioural needs. Mentalization, the capacity to understand behaviour in terms of underlying mental states, is central to effective relational practice in such contexts. Conversational Artificial Intelligence (AI) may offer a scalable means of supporting this form of skills development, but its feasibility as a teacher-training modality remains largely unexplored. **Methods:** This mixed-methods proof-of-concept feasibility study evaluated MentaClassAI, a novel AI-based training tool in which educators engaged in simulated voice conversations with AI child characters portraying classroom dysregulation and subsequently received individualised, mentalization-informed feedback. Eleven staff members from a single AP school (four teachers and seven teaching assistants) completed a single training session and were allocated to either a psychoeducation video condition ($n = 6$) or a no-video condition ($n = 5$). The video condition received a brief introduction to mentalization and epistemic trust prior to engaging with the simulation. Pre- and post-engagement measures included the Reflective Functioning Questionnaire (RFQ-8) and a Teacher Self-Efficacy Scale. Post-engagement measures included an 18-item acceptability questionnaire, a Technology Acceptance Model scale, and open-ended questions analysed using thematic analysis. **Results:** Acceptability was high, with 84.8% of questionnaire responses falling within the positive range (overall $M = 5.60/7$). Feedback accuracy ($M = 6.55$) and clarity ($M = 6.36$) received the highest ratings. Participants reported higher teacher self-efficacy after the session than before ($d = 1.20$, $p = 0.003$), with 10 of 11 participants demonstrating improvement. Self-reported hypomentalizing was lower after the session ($d = -0.86$, $p = 0.017$). Between-condition differences (video versus no-video) were not statistically significant. The video condition scored numerically higher on the directional indicators. Qualitative analysis identified five themes: the value of consequence-free rehearsal; the specificity and usefulness of feedback; appreciation of the focus on the child's emotional experience; limitations in the ecological diversity of AI child characters; and a desire for more naturalistic interaction. **Conclusions:** These findings provide preliminary support for the feasibility and acceptability of AI-based mentalization practice for AP staff. The principal value of the tool appears to lie not only in the simulation itself but in the quality of the reflective feedback generated. Although based on a small sample, the observed pre–post changes provide an encouraging signal that may justify a controlled trial. The contribution of pre-session psychoeducation to training outcomes remains an important question for future research.

Check for updates

Academic Editor: Olufemi A. Omitaomu

Received: 1 July 2026
Revised: 31 July 2026
Accepted: 5 August 2026
Published: 8 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: mentalization; reflective functioning; teacher training; artificial intelligence; large language models; simulation-based learning; conversational agent; professional development; alternative provision; MentaClassAI

1. Introduction

1.1. Behavioural and Emotional Dysregulation in the Classroom

Consider a child who slides under her desk, wraps her arms around her knees, and says flatly that she is not coming out for PE. The teacher has only a few seconds to decide how to respond. The child is not simply being difficult. She may be frightened, ashamed, or overwhelmed. Whether the teacher responds to the feeling beneath the behaviour or to the behaviour alone may shape how the next hour unfolds and, over time, what the child learns to expect from adults when distressed.

Moments such as these are a routine feature of classroom life. Conduct and emotional difficulties are among the most common reasons children come to professional attention. Population studies place the prevalence of conduct and oppositional difficulties at approximately 5–10% of school-aged children, with many more displaying subclinical behavioural difficulties that disrupt learning [1]. Disruptive behaviour is among the most frequently cited predictors of teacher stress, burnout, and attrition [2,3], while children whose behaviour is met punitively rather than understood may be at increased risk of exclusion and longer-term harm [4,5]. These difficulties are particularly concentrated in Alternative Provision (AP), which serves children who have not been successfully accommodated within mainstream educational settings [6]. Yet the underlying challenge of responding effectively to a dysregulated child is one faced by almost all teachers.

Research suggests that children return most readily to a regulated state when adults respond with warmth, curiosity, and an understanding that behaviour communicates an underlying emotional experience rather than simple defiance [7,8]. Teachers who respond in this way appear to support more emotionally secure classrooms and manage dysregulation more effectively [9,10]. However, many teachers report feeling underprepared for such encounters, and conventional behaviour-management training has often emphasised control and consequence over the relational capacities upon which effective de-escalation depends [2,3]. The gap between what classrooms demand and what teacher training provides is the problem this paper addresses, beginning with the question of what such responses actually require of a teacher.

1.2. The Mentalizing Teacher

A useful way of conceptualising the capacity these moments require is mentalization: the imaginative activity through which behaviour, both one's own and that of others, is understood in terms of underlying mental states such as feelings, intentions, desires, and beliefs [11–13]. Originally developed within attachment theory and psychoanalytic developmental psychology, the construct has subsequently been operationalised across clinical, educational, and social care contexts [11,12]. A teacher who mentalizes holds in mind that a child's behaviour, however challenging, expresses an internal experience that is real and worthy of understanding. Rather than focusing solely on stopping the behaviour, the teacher seeks to understand what the child may be feeling, communicates that effort to understand, and helps the child experience themselves as understood [14].

In classroom settings, this stance takes several recognisable forms. It includes naming and validating emotional experience (*"I can see this is really hard for you"*), using a calm voice and reduced demands to support down-regulation, maintaining curiosity about

what lies beneath behaviour rather than reacting only to its surface features, and using playfulness, humour, or gentle reframing to interrupt escalating cycles [9]. It also includes modelling reflective thinking aloud and repairing relationships when an initial response proves ineffective. Observational studies have begun to document these behaviours in schools and to associate them with improved in-the-moment outcomes, while a review of mentalization-based work in education reported medium-to-large gains across social, emotional, and behavioural domains [10,15]. An independent evaluation of a teacher-focused mentalization programme reported preliminary gains in pupil outcomes [16]. Relational teacher-training programmes that cultivate attuned adult responses have similarly been associated with reductions in aggression and disruption [17], consistent with broader findings from the social–emotional learning and classroom-management literatures [18,19].

Three related terms are used deliberately throughout this paper. Mentalization denotes the general reflective capacity [11–13]. A mentalizing response denotes the observable teacher behaviours through which that capacity is expressed in the moment; in this study, these are operationalised as validating the child's emotional experience, remaining curious about its meaning, co-regulating and reducing demand, holding the child's mind tentatively rather than with certainty, collaborating, and repairing after a misstep. Operationalising mentalizing as observable, mental-state-focused adult behaviour follows established independent work on mind-mindedness and on reflective functioning in applied settings [20,21]. The six categories used here are an adaptation of that approach to the classroom; they were developed for the present study and have not been independently validated. Mentalizing feedback denotes the rubric against which those responses were described and returned to participants. The in-task and feedback components of the present study target mentalizing responses in this specific behavioural sense, rather than the broader capacity or the rubric alone.

1.3. A Capacity That Is Fragile Under Stress: Why It Must Be Practised, Not Only Explained

Importantly, mentalization is increasingly understood not as a fixed trait but as a dynamic capacity that can diminish under conditions of heightened arousal and interpersonal threat, precisely the circumstances that characterise classroom crises [13,22]. A teacher may therefore possess a sound conceptual understanding of mentalizing yet struggle to mentalize when a chair is thrown across the room. This distinction has important implications for training. Declarative knowledge about mentalization is not equivalent to the procedural fluency required to deploy it in real time under stress. Although there is growing evidence that mentalizing capacities can be strengthened through targeted practice [13,23], professional development in this area has remained largely theoretical rather than practice-based [10].

1.4. Practising Relational Skills at Scale: Simulation, Deliberate Practice, and AI

The question this raises is how such practice might actually be delivered. Simulation-based learning, in which trainees rehearse skills within structured environments designed to replicate critical features of real-world situations, has a strong evidence base in medicine and psychotherapy training [24,25]. Its active ingredients are well established: sufficient realism to activate the target skill, structured repetition, and immediate, specific feedback on performance [26]. Simulation is particularly valuable where real-world practice carries unacceptable risk. Responding to a genuinely distressed child is precisely such a situation, as teachers cannot ethically experiment with unfamiliar responses during an actual crisis.

What has been largely absent is a mechanism for delivering this kind of practice at scale. Reviews of AI in education describe rapid growth in applications alongside persistent gaps in rigorous evaluation of classroom-relevant outcomes [27,28]. Expert-led role-play

and supervision remain effective but costly and scarce, which helps explain why relational skills are often discussed rather than actively rehearsed. Recent LLM systems may alter this landscape. They can sustain coherent, emotionally plausible dialogue that adapts dynamically to user responses and can generate immediate, individualised, transcript-level feedback. Together, these capabilities make it possible to offer repeated, low-stakes practice conversations with simulated children, each followed by specific feedback, at a cost and level of accessibility that human-led training cannot readily match [29,30].

A growing body of work suggests that AI-based feedback systems are generally acceptable to trainees, can identify relevant behavioural indicators with reasonable accuracy, and may be experienced as less threatening than feedback delivered by human evaluators [31–33]. However, these possibilities also carry significant risks. AI feedback may be inaccurate, superficially plausible, or based on culturally narrow assumptions about child distress and behaviour [34]. Simulated children who de-escalate too easily may inadvertently train unrealistic expectations about how distressed children respond in practice. Whether teachers find such tools credible, usable, and educationally valuable cannot therefore be assumed.

To our knowledge, no published research has examined AI-based mentalization training for teachers responding to child dysregulation. A recent systematic review identified AI tools for training caregivers, educators, and therapists in mentalization-related skills as a promising but largely untested category of intervention [35]. The present study addresses this gap directly.

1.5. Preparing Teachers to Learn from AI Feedback: Epistemic Trust

Scalability, however, settles only part of the question. Whether feedback changes practice depends not only on its content but also on whether the learner is open to receiving it. Epistemic trust theory proposes that openness to new knowledge is contingent on whether the source of that knowledge is experienced as relevant, trustworthy, and responsive to the learner's perspective [36,37], an account that builds on independent work on epistemic vigilance and on natural pedagogy [38,39]. Applied to professional learning, this suggests that teachers may derive greater benefit from AI feedback if they are first primed to understand its relevance to their own practice.

A brief psychoeducational introduction to mentalization, dysregulation, and epistemic trust may therefore foster a more receptive stance towards the feedback that follows. There is also an interpersonal dimension to consider. Human supervisors inevitably bring social evaluation, authority, and relational history into feedback interactions, factors that can evoke defensiveness and limit openness to learning. AI feedback carries no such interpersonal stakes. Although often viewed as a limitation, this absence of evaluative pressure may, in some contexts, constitute a pedagogical advantage.

Epistemic trust may also influence uptake and sustained engagement. A tool that users experience as understanding their perspective is more likely to be accepted and used. For this reason, the present study examined not only change following practice but also participants' perceptions of the acceptability and usability of MentaClassAI. The study additionally explored whether viewing a brief psychoeducational video on mentalization and epistemic trust before engaging with the tool enhanced subsequent engagement and learning.

1.6. The Present Study

This study reports the development and proof-of-concept feasibility evaluation of MentaClassAI, a novel AI-based mentalization practice tool in which educators engage in voice-based simulated conversations with AI child characters portraying classroom dysreg-

ulation and subsequently receive structured, individualised feedback on the mentalizing quality of their responses. The system is intended as a training platform rather than a therapeutic intervention and is designed to complement, rather than replace, human-led professional development.

The primary aim was to establish whether AI-based mentalization practice is feasible and acceptable for educational staff and to examine how educators experience practising relational skills with an AI system. Particular attention was paid to how participants engaged with simulated child characters, how they responded to AI-generated feedback, and what learning value they perceived in the experience. The study was conducted within an AP setting, where dysregulation is especially common and opportunities for deliberate practice may be particularly valuable, although the questions addressed have broader relevance to teacher development.

Six research questions guided the study. RQ1 examined whether an AI-based mentalization practice tool is acceptable and usable for educators. RQ2 investigated whether there were preliminary indications of change in reflective functioning and teacher self-efficacy following a single training session. RQ3 explored whether a brief psychoeducational video on mentalization and epistemic trust enhanced engagement with and learning from the tool. RQ4 examined what participants perceived as most valuable about the system. RQ5 explored concerns, limitations, and barriers identified by participants. RQ6 sought to identify design recommendations arising from user feedback.

For the quantitative components of the study, three directional hypotheses were specified. H1 predicted that acceptability ratings would exceed the scale midpoint. H2 predicted improvements in reflective functioning and teacher self-efficacy from pre- to post-session. H3 predicted that participants who viewed the psychoeducation video would engage more actively with the tool and demonstrate more mentalizing responses during practice than those who did not. The remaining questions (RQ4–RQ6) were exploratory.

One conceptual argument runs through what follows and is worth stating at the outset. The tool is not offered as a substitute for a relationship with a child, and its value may not depend on the AI child being experienced as a convincing social other. The idea is that a simulated encounter can generate a record of the teacher's own responses when AI can act as a low-stakes practice partner, which can then be returned to them in a theoretically organised form. On this view the tool supports reflective self-recognition rather than relational immersion, a distinction that separates it from a generic AI role-play simulator and that we return to in the Discussion.

2. Methods

2.1. Design

An embedded mixed-methods feasibility design was employed [40]. The quantitative component combined a within-participant pre–post design (RFQ-8, Teacher Self-Efficacy Scale) with a between-participant manipulation. Participants were allocated to one of two conditions on entry to the tool. In the video condition, participants viewed a brief psychoeducation video introducing mentalization, dysregulation, and epistemic trust before beginning the practice conversation. In the no-video condition, participants proceeded directly to the simulation. This manipulation was designed to test whether brief priming, consistent with epistemic trust theory [36], enhanced engagement with and learning from the tool.

Both groups completed identical pre- and post-engagement measures, a post-session acceptability and usability questionnaire, a Technology Acceptance Model scale, and open-ended qualitative questions. Qualitative data were analysed using reflexive thematic analysis [41,42]. Given the small sample size, quantitative analyses (described in Section 2.6)

are interpreted cautiously. In particular, between-condition comparisons are treated as preliminary, hypothesis-generating signals rather than confirmatory tests.

The study received ethical approval from the UCL Research Ethics Committee, Division of Psychology and Language Sciences (Project ID 3284).

Data handling, security, and consent: Each session generated a real-time audio stream, a text transcript of the conversation, and the participant's questionnaire responses. All AI components were supplied by a single external provider, OpenAI. The voice conversation ran through the OpenAI Realtime API (gpt-4o-realtime-preview), which received the microphone audio, produced the transcription (gpt-4o-transcribe), and returned the child character's replies as synthesised speech; the feedback reports were generated with gpt-4o. No other external language, transcription, or speech provider was used. OpenAI therefore processed the session audio and the transcripts derived from it. Under the provider's API terms in force during the study, content submitted through the API is not used to train or improve its models. The study itself retained no audio; only the transcripts and questionnaire data were kept. All records were pseudonymised at the point of collection using participant codes, no names or other direct identifiers were held in the system or transmitted to the provider, and participants were asked not to refer to real pupils by name. Transcripts and questionnaire data were captured by the platform's Supabase backend and held for analysis on UCL-managed, encrypted, password-protected institutional storage in the UK, accessible to the researcher and the principal investigator only. In line with UCL policy, the data will be retained for five years after publication and then securely deleted. At consent, participants were told what would be recorded, that the child character and the feedback were generated by an external AI system that does not use responses to train its models, and that they could request deletion of their data within two weeks of participation, after which records were anonymised and incorporated into the analysis. These arrangements formed part of the ethical approval noted above.

2.2. Participants and Setting

Participants were teachers and teaching assistants employed at a single AP school in England serving children aged 6–11 years. Pupils attending the school were predominantly neurodiverse and frequently presented with co-occurring emotional and behavioural difficulties. Inclusion criteria required a current classroom-based role and sufficient spoken English to engage with the simulation. Because seven of the eleven participants were teaching assistants, the terms 'educators', 'educational staff', and 'participants' are used when referring to the sample as a whole; 'teachers' and 'teaching assistants' are used only where the distinction between the two roles is analytically relevant.

As staff members working within an AP setting characterised by a well-established reflective practice culture, participants were likely to have had some prior exposure to trauma-informed, attachment-aware, or mentalization-informed approaches through school-based CPD. This characteristic of the sample is considered further in the Limitations section.

Eleven participants completed the study, comprising four teachers and seven teaching assistants. Eight identified as male and three as female. All participants were aged between 18 and 35 years (18–25 years, $n = 8$; 26–35 years, $n = 3$). Length of experience in their current role ranged from 1–3 years ($n = 6$) to 4–7 years ($n = 5$). Six participants were allocated to the video condition and five to the no-video condition. Overall, this was a relatively young and early-career sample, a characteristic discussed further in the Limitations section.

2.3. The AI System

MentaClassAI is a web-based AI training platform through which participants progress through a fixed sequence comprising entry and condition allocation, optional

psychoeducation, scenario selection, a safe-space briefing, scenario setup, a voice-based simulated conversation with an AI child character, and a personalised feedback report. The system was developed iteratively by the research team in collaboration with a senior AI solutions architect. It was built primarily using Claude Code, with GitHub used for version control and Supabase for backend infrastructure and data storage. The design of both scenarios and feedback was informed by the mentalization literature and educational applications of mentalization-based approaches [9,15,43].

System interface and participant journey

Participants entered the system in their pre-assigned condition. Those allocated to the video condition first viewed a brief psychoeducation video of approximately two minutes' duration introducing mentalization, dysregulation, and epistemic trust. Participants in the no-video condition proceeded directly to scenario selection.

Participants then selected one of nine classroom scenarios. Each scenario was presented as a card displaying the child's name, age, classroom context, and emotional-state tags summarising the hypothesised internal state driving the presentation (Figure 1). A safe-space briefing subsequently framed the exercise as non-evaluative practice and outlined the structure of the session (Figure 2). Participants then viewed a scenario setup screen introducing the child's background, describing the unfolding classroom situation, and presenting the child's opening statement (Figure 3).

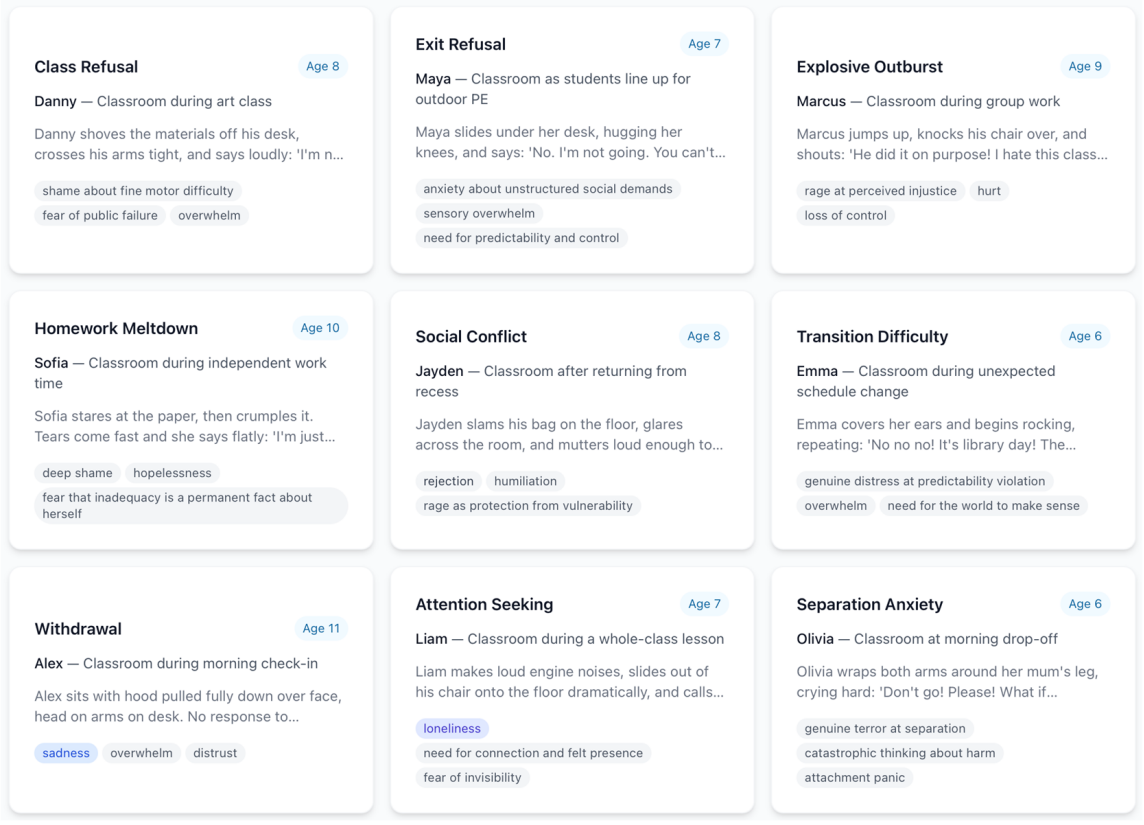

Figure 1. Scenario library. Each card displays the child's name and age, the classroom context, and emotional-state tags summarising the hypothesised internal state driving the presentation.

The live simulation was conducted through a voice-based interface displaying the active child character throughout the interaction (Figure 4). Following completion of the conversation, participants received an automatically generated personalised feedback report.

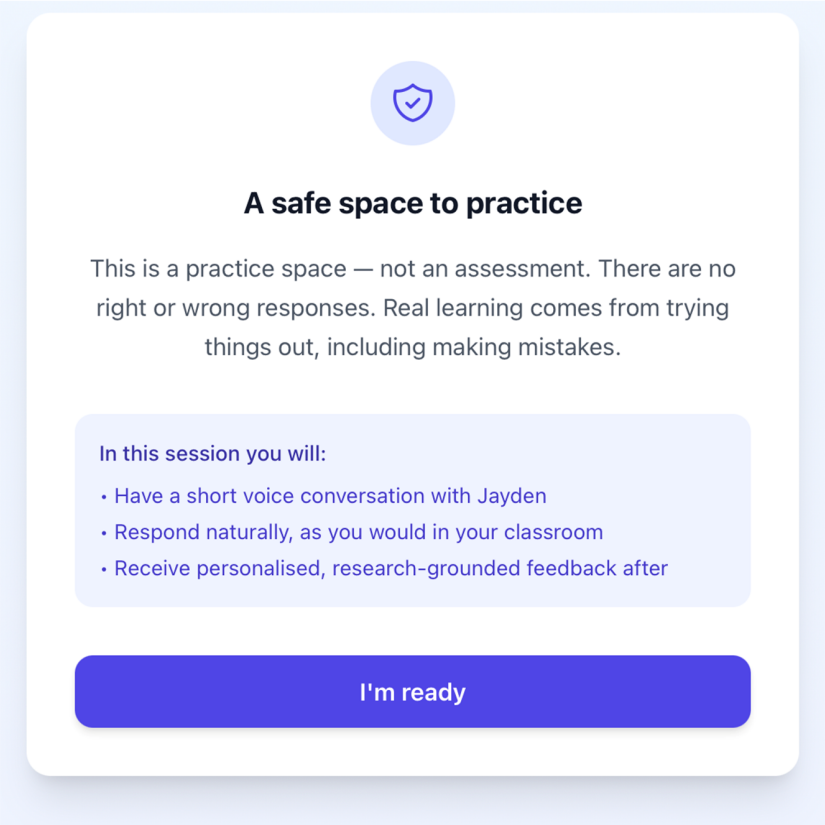

Figure 2. Safe-space briefing. The session is framed explicitly as non-evaluative practice, with a preview of what the participant will do.

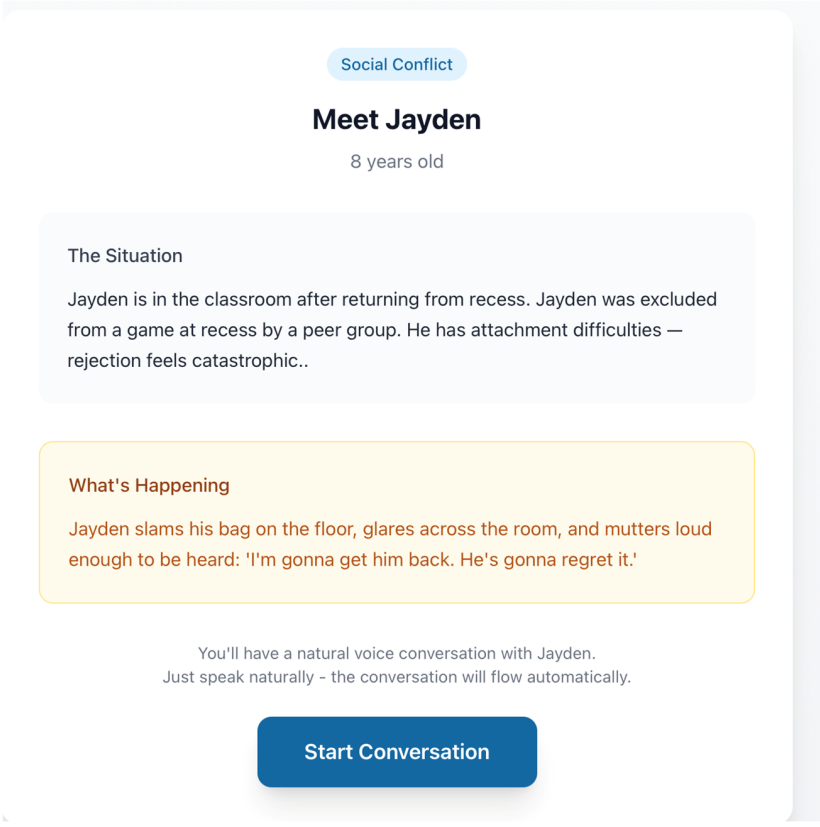

Figure 3. Scenario setup. The child is introduced by name and age, with a brief psychological background, a description of the situation, and the child's opening words.

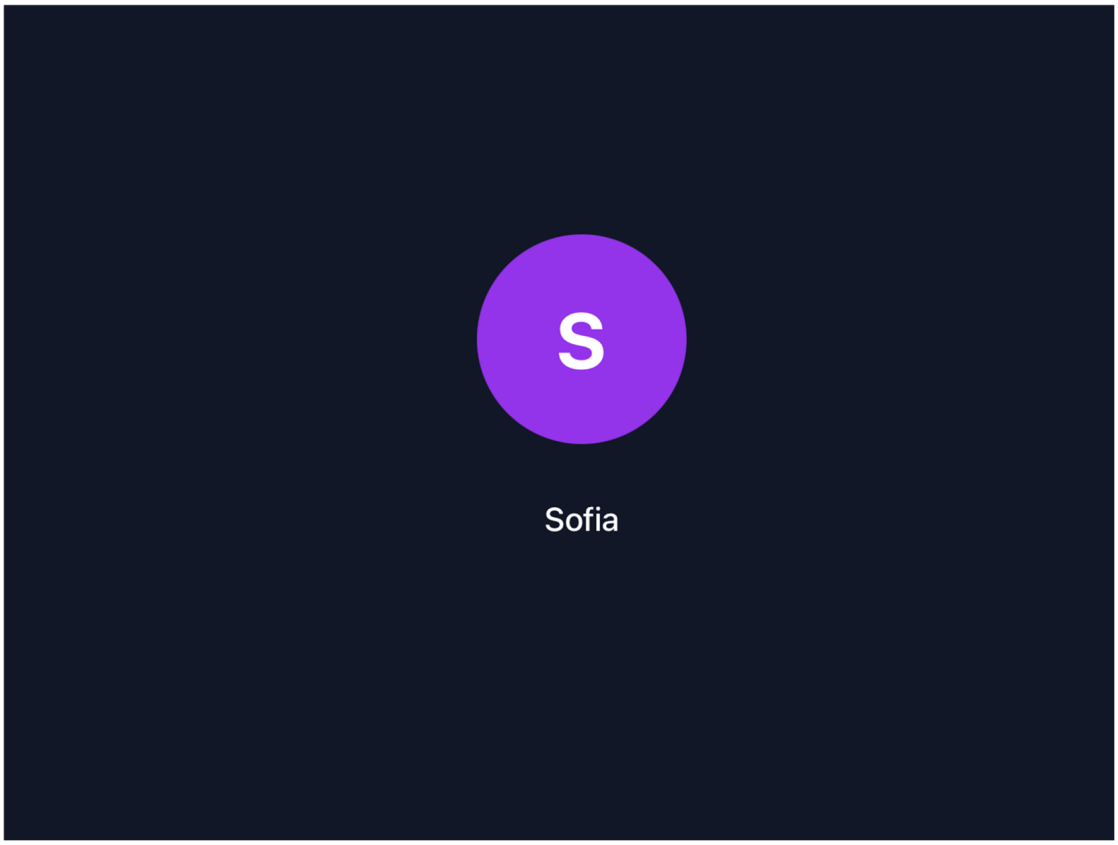

Figure 4. Voice conversation. The active AI child character is displayed during the live, voice-based exchange.

Child agent configuration and validation: Each AI child character was implemented as a large-language-model agent governed by a structured system prompt that specified the child's name and age, a brief developmental and emotional background, the presenting classroom situation, the hypothesised internal state driving the behaviour, and a set of in-role constraints: to remain in character as a distressed child rather than an assistant, to use short, age-appropriate language, to adjust the child's emotional state in response to the teacher's approach, and not to settle unrealistically quickly. The agent held no persistent memory across sessions; each conversation was independent and conditioned only on the selected scenario and the unfolding dialogue. Spoken input was transcribed to text, passed to the language model, and the reply was returned as speech through a text-to-speech interface. A representative system-prompt template is provided in Appendix A to support reproducibility. The characters were not formally validated against recordings of real children; instead, both the characters and the associated feedback were calibrated through the expert-in-the-loop process described below, in which experienced mentalization-based practitioners reviewed sample interactions, while participants' own ratings of conversational naturalness (Section 3) provide an initial, if limited, index of face validity. Systematic validation against real child-interaction data remains an important next step and is noted among the limitations.

Psychoeducation video

The psychoeducation video was designed to prime a mentalizing and receptive stance prior to practice. Consistent with evidence that prior mental-state framing can influence attributional style [13], and with epistemic trust theory's emphasis on perceived relevance [37], the video introduced mentalization, dysregulation, and epistemic trust in accessible language and illustrated these concepts through concrete examples of mental-

izing responses. The video did not rehearse the specific scenarios or feedback categories subsequently used within the practice task.

Scenario design

Participants selected from nine pre-constructed classroom scenarios portraying common presentations encountered within AP settings. These included class refusal (Danny, aged 8), exit refusal (Maya, aged 7), explosive outburst (Marcus, aged 9), homework-related distress (Sofia, aged 10), social conflict (Jayden, aged 8), transition difficulties following an unexpected schedule change (Emma, aged 6), withdrawal (Alex, aged 11), attention-seeking behaviour (Liam, aged 7), and separation anxiety (Olivia, aged 6).

Each scenario card displayed emotional-state tags summarising the hypothesised internal experience driving the child's behaviour. These formulations were derived from mentalization-based conceptualisations of classroom dysregulation [15].

The safe-space briefing explicitly framed the exercise as practice rather than assessment:

> *"This is a practice space, not an assessment. There are no right or wrong responses. Real learning comes from trying things out, including making mistakes."*

Presenting information about the child's emotional state prior to the interaction was intended to orient participants towards a mentalizing stance before the conversation began.

Voice conversation

Participants engaged in a live voice-based conversation with the AI child character using the platform's integrated audio interface. Conversations consisted of five participant turns and typically lasted between three and five minutes. The five-turn format was set during piloting to balance eliciting a codeable sample of participant responses against participant burden and the constraints of a single-session, proof-of-concept design; it was not intended to represent a complete classroom interaction.

The AI child responded dynamically in real time while maintaining the emotional logic of the selected scenario. No scripted correct responses were built into the system. Instead, the child character adapted plausibly to a range of teacher responses, including validating, directive, avoidant, or coercive approaches, with subsequent emotional reactions reflecting the interactional trajectory established by the participant.

An illustrative segment from a practice conversation is presented in Table 1, with a second example shown in Table 2.

AI-generated feedback

Following completion of the conversation, participants received a structured feedback report comprising four components:

1. An overall narrative summary of the participant's interactional approach.
2. A *Mentalizing strengths* section highlighting effective responses, supported by verbatim quotations from the participant's transcript and linked to specific mentalization concepts (e.g., empathic validation, co-regulation).
3. A *What to develop* section identifying less effective responses, accompanied by alternative formulations and explanatory rationale.
4. A *Try this tomorrow* section providing a transferable practice principle grounded in the mentalization literature.

The feedback taxonomy was derived from established mentalizing competency frameworks [9] adapted for educational settings [10,15]. The competency framework itself is externally established; the classroom adaptation, the feedback categories, and the rubric applied here were developed by the present authors and have not been independently validated. The feedback model underwent iterative refinement through an expert-in-the-loop process in which experienced mentalization-based practitioners reviewed sample interactions alongside generated feedback and calibrated the system to align with expert judgement.

Consequently, each report highlighted both strengths and areas for development, grounding every observation in the participant's own language while linking feedback explicitly to the theoretical framework underpinning the tool. An illustrative example of the feedback format is presented in Table 3.

Table 1. Illustrative segment from a practice conversation (Emma, age 6, transition difficulty), showing the teacher's spoken turns and the AI child character's responses.

Exchange	Teacher (Participant)	AI Child Character
3	"Someone else is using the library so we won't have space, we can do it later maybe."	*"But. . . but it's library time now! It has to be now! It's wrong if it's later! It's not fair, it's not fair. . ."*
4	"What would you like to do in a library?"	*"I like the books and the quiet. And it feels safe 'cause I know what happens there. It has to be now!"*
5	"What about if we find a book and a quiet space somewhere else to read where there's no noise?"	*"Find a book and quiet place now? Like the library but not the library? Maybe. . . I don't know. . . It's still wrong. . ."*

Table 2. A segment from a practice conversation (Maya, age 7, reluctance to leave the classroom for an outdoor activity), in which the teacher leads with validation and down-regulation. The first two exchanges are shown.

Exchange	Teacher (Participant)	AI Child Character
1	"Okay, Maya, I completely understand that. What I'm going to do is get everyone else to go out first, and then you and I will sit back here, talk a little bit, and see how we get on. There's no pressure to go outside immediately, but I'd like us to talk about it a little, so I can understand why you don't want to go out, because I can see that you don't. Is that something we can do?"	*(still a bit hesitant)* "It's just. . . really loud and. . . and too much. I don't like it when it's like that."
2	"Do you know what, Maya, I can totally understand that. Sometimes when I'm in really loud situations I find it very disruptive as well, and I totally get why you might not be comfortable with that. How about this: why don't we sit here for a few more minutes, just chat, and then I take you downstairs and we'll test it, and if you're not sure, then we can reassess."	*(softly)* "Okay, we can try that. I just. . . don't want it to be too much. But we can try."

Table 3. Illustrative excerpt from a feedback report (format example, not a real participant transcript).

Overall. You stayed calm and made several genuine attempts to understand what was going on for Jayden before moving to solve the problem. There is a clear opportunity to slow down and name his feeling before redirecting.
What you did well. You said: "It sounds like that felt really unfair to you." This is ***empathic validation****: you recognised Jayden's experience as legitimate, which helps a dysregulated child feel understood and lowers their arousal.*
What to develop. You said: "Okay, let's just get back to the table now." Here the redirection came before the feeling had been acknowledged, which can leave a child feeling *managed rather than understood. You might try: "You're really angry he's out, and that makes sense. When you're ready, we'll sort out what happens next together," pairing validation with collaboration.*
Try this tomorrow. Name the feeling before you name the task. A child can rarely think about what to do next until they feel that someone has understood how they got here.

2.4. Measures

Reflective Functioning Questionnaire (RFQ-8) [44]

The RFQ-8 is an eight-item self-report measure of reflective functioning rated on a seven-point scale (1 = strongly disagree, 7 = strongly agree). Consistent with recent evidence suggesting that the RFQ-8 primarily indexes hypomentalizing along a single dimension [45,46], lower uncertainty and higher certainty scores were both interpreted as reflecting lower hypomentalizing. The RFQ-8 was administered immediately before and after engagement with the training tool. As a brief self-report index of hypomentalizing, the RFQ-8 is relatively distal from the in-the-moment relational skill the tool seeks to train; the transcript-based in-task measure described below is more proximal, and is the more appropriate candidate primary outcome for future trials.

Teacher Self-Efficacy Scale [47]

Eight items were selected from Bandura's Teacher Self-Efficacy Scale [47] and administered pre- and post-engagement using a nine-point scale (1 = nothing, 9 = a great deal), with higher scores indicating greater perceived efficacy. Items were drawn from three domains relevant to responding to child dysregulation: behaviour management (three items), student engagement (three items), and relationship building (two items). A shortened version was used to minimise participant burden within a single-session design while retaining the domains most relevant to the present study. Internal consistency of the eight-item scale in the present sample was acceptable (Cronbach's $\alpha = 0.71$ at pre-test).

In-task mentalizing (transcript-based)

As an exploratory behavioural index, each of the participant turns in the voice conversation was coded from the transcript for the presence of a mentalizing response, defined as above, yielding a count from 0 to 5. A turn was coded as mentalizing if it contained at least one of the behaviours operationalised in Section 1.2 (validating the child's emotional experience, remaining curious about its meaning, co-regulating or reducing demand, holding the child's mind tentatively, collaborating, or repairing after a misstep), and as non-mentalizing if it was purely directive, corrective, or procedural. For example, "It sounds like that felt really unfair to you" was coded as mentalizing, whereas "Okay, let's just get back to the table now" was not. Coding followed these pre-specified criteria and was carried out by the first author; as a single-rater index at the feasibility stage, it was not blinded to condition. This index lies closer to the skill the tool seeks to train than the self-report measures and is reported here only for the between-condition comparison. Because it was developed for the present feasibility study, applied to a small sample, and not yet subjected to formal inter-rater reliability testing, it should be treated as preliminary. Developing it into a reliable primary or co-primary outcome is a priority for future trials.

Acceptability and usability questionnaire

Acceptability was assessed using an 18-item post-session questionnaire developed for this study and informed in part by Technology Acceptance Model constructs [48]. The measure comprised five subscales: perceived usefulness (five items), ease of use (four items), feedback trust (four items), emotional safety (three items), and intention to re-engage (two items). All items were rated on a seven-point agreement scale, with higher scores indicating more favourable evaluations. Internal consistency of the full 18-item scale in the present sample was acceptable (Cronbach's $\alpha = 0.79$); given the small sample, this estimate should be read as indicative.

Technology Acceptance Model (TAM) [48]

A further eight post-session items assessed perceived usefulness (four items) and perceived ease of use (four items) using a seven-point agreement scale. Higher scores reflected greater perceived usefulness and ease of use.

Open-ended questions

Four post-session questions invited qualitative reflection: *"What did you find most useful about the system?", "What did you find least useful or most limiting?", "Is there anything you would change about the AI child character?", and "Is there anything you would change about the feedback format or content?".*

2.5. Procedure

Participants first provided written informed consent and completed the pre-intervention questionnaire battery, which required approximately 10 min. Participants were then allocated to either the video or no-video condition. Allocation was determined by the researcher prior to participation and alternated according to order of recruitment, resulting in six participants in the video condition and five in the no-video condition.

Participants in the video condition viewed the psychoeducation video before proceeding to the simulation. Participants in the no-video condition proceeded directly to the training task. All participants then engaged independently with the AI system by selecting a scenario, completing the voice-based conversation, and reviewing their personalised feedback report. This stage typically required between 20 and 25 min.

Following system engagement, participants completed the post-intervention questionnaire battery, including the RFQ-8, Teacher Self-Efficacy Scale, acceptability questionnaire, TAM items, and open-ended questions. Completion required approximately 15–20 min. Total participation time was approximately 50–60 min.

2.6. Data Analysis

Quantitative outcomes were analysed using paired-samples *t*-tests comparing pre- and post-engagement scores on the RFQ-8 and Teacher Self-Efficacy measures. Cohen's *d*, calculated as the mean pre–post difference divided by the standard deviation of the difference scores, was used to estimate effect size magnitude. Because these effect sizes are based on the standard deviation of the paired difference scores, they correspond to the within-participant form of Cohen's d (d_z); d_z is not directly comparable to the conventional between-groups benchmarks (e.g., d = 0.80 for a large effect) and tends to be larger for an equivalent underlying effect, so the values reported here should be read accordingly. Given the small sample size, non-parametric Wilcoxon signed-rank tests were conducted as a robustness check.

Between-condition comparisons (video versus no-video) were analysed using the Mann–Whitney *U* test, with rank-biserial correlation reported as the corresponding effect size. Two-tailed alpha was set at 0.05. Consistent with feasibility-study conventions [49], no correction for multiple comparisons was applied. Statistical findings are therefore interpreted as exploratory and hypothesis-generating rather than confirmatory. Descriptive statistics are reported as means and standard deviations. Analyses were conducted in IBM SPSS Statistics version [29.0] (IBM Corp., Armonk, NY, USA).

Qualitative responses were analysed using reflexive thematic analysis [41,42]. Analysis proceeded through iterative phases of familiarisation with the data, generation of initial codes, development of candidate themes, review and refinement of themes, and final theme definition and naming.

3. Results

3.1. Sample Characteristics

Eleven participants completed the study, comprising four teachers and seven teaching assistants. Eight participants identified as male and three as female. All participants were aged between 18 and 35 years. Six reported between one and three years of experience in

their current role, while five reported between four and seven years. Six participants were allocated to the video condition and five to the no-video condition.

3.2. Quantitative Findings

3.2.1. Reflective Functioning (RFQ-8)

Reflective functioning was indexed using the mean of the seven uncertainty-keyed RFQ-8 items, with higher scores indicating greater uncertainty regarding mental states and therefore higher levels of hypomentalizing. This single-score approach follows recent evidence suggesting that the RFQ-8 primarily captures a hypomentalizing dimension rather than two independent certainty and uncertainty constructs [45,46].

Hypomentalizing decreased from pre-engagement (M = 2.82, SD = 0.36) to post-engagement (M = 2.49, SD = 0.33), a statistically significant pre–post change, $t(10) = -2.85$, $p = 0.017$, $d = -0.86$. The corresponding Wilcoxon signed-rank test confirmed the result ($p = 0.027$). Nine of the eleven participants demonstrated reductions in hypomentalizing, one increased, and one remained unchanged. Overall, the pattern is consistent with a reduction in self-reported hypomentalizing following a single session of AI-supported practice and feedback.

3.2.2. Teacher Self-Efficacy

Total teacher self-efficacy increased from pre-engagement (M = 6.67, SD = 0.59) to post-engagement (M = 7.03, SD = 0.55) on the nine-point scale, a statistically significant pre–post change, $t(10) = 3.98$, $p = 0.003$, $d = 1.20$.

All three subscales showed pre–post improvement. Behaviour management increased significantly, $t(10) = 3.32$, $p = 0.008$, $d = 1.00$; student engagement increased significantly, $t(10) = 2.96$, $p = 0.014$, $d = 0.89$; and relationship building increased significantly, $t(10) = 2.89$, $p = 0.016$, $d = 0.87$. The relationship-building effect should be interpreted cautiously, however, as only five participants showed change on this subscale, while six remained unchanged.

Wilcoxon signed-rank tests produced the same pattern of findings. Subscale comparisons are presented in Figure 5.

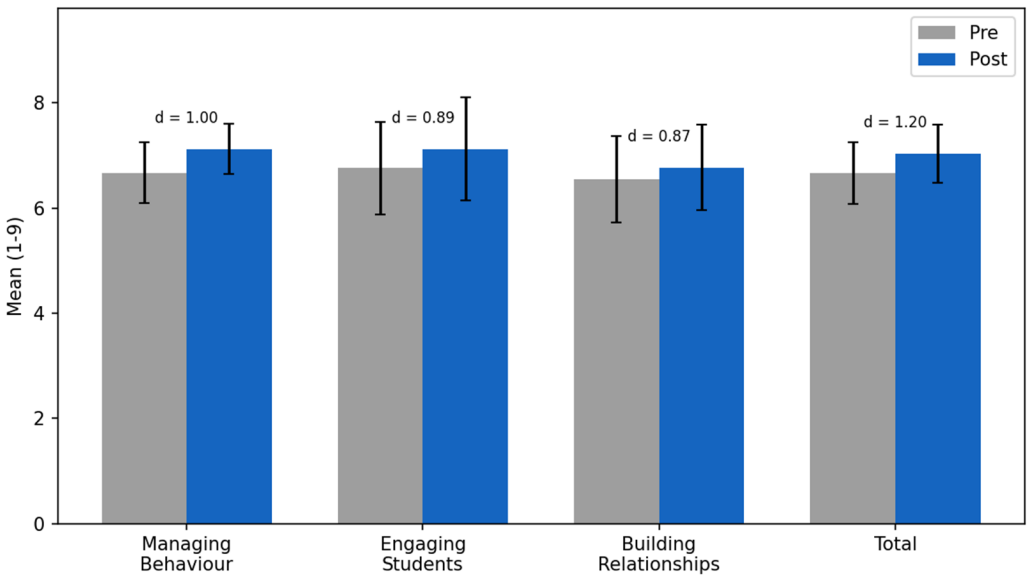

Figure 5. Teacher self-efficacy subscale scores pre- and post-engagement (N = 11). Grey bars represent pre-engagement means and blue bars represent post-engagement means, with error bars showing standard deviations. Cohen's d (effect size) is displayed above each pair.

3.2.3. Acceptability and Usability

Overall acceptability was high (M = 5.60, SD = 0.49). Across all questionnaire items, 84.8% of responses (168/198) fell within the positive range of the scale (Table 4). Item-level ratings are presented in Figure 6.

Table 4. Acceptability subscale means (N = 11).

Subscale	Items	M (SD)
Perceived Usefulness	5	5.78 (0.62)
Ease of Use	4	5.61 (0.70)
Feedback Trust	4	6.14 (0.63)
Emotional Safety	3	4.79 (0.79)
Intention to Re-engage	2	5.27 (0.93)
Overall	**18**	**5.60 (0.49)**

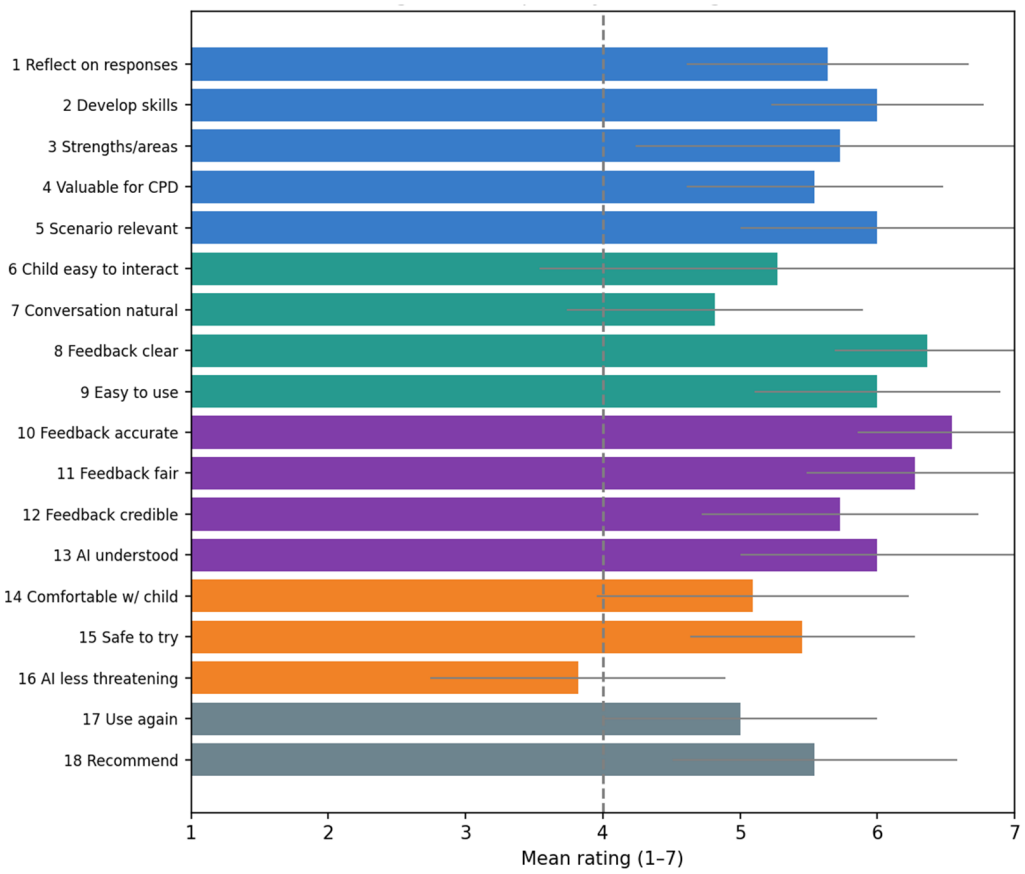

Figure 6. Acceptability item ratings by subscale (N = 11). Horizontal bars represent mean ratings on a 7-point scale, and error bars represent standard deviations. The dashed line indicates the scale midpoint (4.0). Items are colour-coded by subscale: blue, perceived usefulness; teal, ease of use; purple, feedback trust; orange, emotional safety; grey, intention to re-engage.

The highest-rated items concerned the quality of the feedback. Participants strongly endorsed *"The feedback accurately reflected what I actually said in the conversation"* (M = 6.55, SD = 0.69), *"The feedback was clearly written and easy to understand"* (M = 6.36, SD = 0.67), and *"The feedback felt fair"* (M = 6.27, SD = 0.79). Participants also rated the scenarios as relevant to their professional experience (M = 6.00, SD = 1.00) and found the system easy to use (M = 6.00, SD = 0.89).

The lowest-rated item was *"Receiving feedback from an AI felt less threatening than receiving feedback from a human supervisor or trainer"* (M = 3.82, SD = 1.08), the only item to fall below the scale midpoint. Ratings for conversational naturalness (M = 4.82, SD = 1.08) and comfort interacting with the AI child character (M = 5.09, SD = 1.14) were the next lowest and showed greater variability across participants. These findings suggest that participants were more confident in the quality of the feedback than in the realism of the interaction itself.

3.2.4. Technology Acceptance

Perceived ease of use on the TAM was relatively high (M = 5.82, SD = 0.87), whereas perceived usefulness was somewhat lower (M = 4.70, SD = 0.70). The overall TAM score was 5.26 (SD = 0.74). Consistent with the acceptability findings, participants appeared more confident in their ability to operate the system than in its potential to improve day-to-day classroom practice.

3.2.5. Prior Psychoeducation (Video vs. No-Video)

Six participants were allocated to the video condition and five to the no-video condition. The two groups were broadly comparable at baseline on reflective functioning and teacher self-efficacy measures.

No between-condition comparison reached statistical significance. However, participants who viewed the psychoeducation video demonstrated slightly higher levels of in-task mentalizing, somewhat larger gains in self-efficacy, and marginally higher acceptability ratings. Change in hypomentalizing was effectively identical across conditions.

Although these differences cannot be interpreted as evidence of an effect, all three engagement and outcome indicators favoured the video condition, and the self-efficacy difference corresponded to a medium-to-large effect size. Given the extremely small subgroup sizes (n = 6 and n = 5), even meaningful effects would have been unlikely to achieve statistical significance. In this context, the consistency of direction across theoretically related outcomes may be more informative than significance testing alone. These findings therefore provide a preliminary, theory-consistent signal that brief psychoeducational priming may enhance engagement with AI-based mentalization practice and warrants investigation in a larger, adequately powered study.

Because the open-ended questions did not ask specifically about the psychoeducation video, the qualitative data do not directly inform interpretation of this manipulation. Future studies would benefit from including qualitative questions exploring participants' experiences of the pre-session input. Summary statistics for these comparisons are presented in Table 5.

Table 5. Video vs. no-video condition comparisons (exact Mann–Whitney U).

Measure	Video (n = 6) M (SD)	No-Video (n = 5) M (SD)	U	*p*	Rank-Biserial
Mentalizing turns (in-task,/5)	4.17 (1.17)	3.60 (1.14)	10.0	0.472	0.33
Self-efficacy change	0.48 (0.35)	0.23 (0.19)	7.0	0.167	0.53
Hypomentalizing change	−0.31 (0.48)	−0.34 (0.26)	13.0	0.755	0.13
Acceptability (overall)	5.74 (0.47)	5.43 (0.50)	9.5	0.320	0.37

3.3. Qualitative Findings

These data come from brief written answers to four open-ended questions rather than from interviews, so the patterns below are best read as qualitative response pat-

terns rather than as a deep thematic account of experience. Five themes were generated from the open-ended responses. The first three capture aspects of the system that participants valued, while the final two reflect perceived limitations and recommendations for future development.

Theme 1: Practice as rehearsal

The most consistently valued feature was the opportunity to practise responding to dysregulated behaviour without real-world consequences. One participant wrote: *"It gives you a space to get things wrong without real consequences, you can try an approach, see how it lands, and adjust before you're with an actual child"* (P9). Two others expressed a similar idea: *"Talking through a situation without a child there means you can practice for when one is. Found useful breaking down what to do differently, particularly giving positive feedback as well as areas you could improve"* (P8), and *"Talking through a situation without a child present means you can practise for when one is there. I found it useful to break down what to do differently, getting positive feedback on what worked, as well as the areas I could improve"* (P10). One participant suggested that they *"should practice more scenarios to better improvement"* (P6).

These responses suggest that participants experienced the simulation as a low-risk rehearsal space rather than as an assessment. For staff working in environments where dysregulation is frequent and often emotionally demanding, the opportunity to practise responses outside real interactions appeared to be highly valued.

Theme 2: Feedback quality and specificity

Feedback emerged as the most frequently praised aspect of the system. Participants described it as *"very useful, relevant to the scenario, and seemed to reflect the situation accurately"* (P4), *"useful and on point. . . it matched the scenario and reflected what was actually going on"* (P2), and *"genuinely helpful, it fit the scenario well and captured the situation accurately"* (P3). Another participant commented that *"it is easy to understand and the feedback is straight forward"* (P6).

One participant described the feedback as *"useful for spotting patterns in my own reactions that I wouldn't have clocked on my own"* (P11). This comment is notable because it points to a potentially important learning mechanism: the opportunity to view one's own responses from a more reflective perspective and recognise patterns that may otherwise remain outside awareness. The design feature of quoting participants' own words and linking them to named mentalizing concepts may have contributed to this process.

Taken together with the high quantitative ratings for feedback accuracy ($M = 6.55$) and clarity ($M = 6.36$), these findings suggest that the feedback component was experienced as both credible and practically useful.

Theme 3: Recognising the child's inner world

Participants also valued the system's emphasis on the child's emotional experience rather than on behaviour alone. One participant highlighted *"recognising the child's needs and feelings and how the child's emotions are clearly at the forefront"* (P7). Others valued the realism of the situations themselves, describing them as *"realistic. . . [and] made it easy to picture situations I genuinely come across"* (P1) and reflective of *"the events that could happen in the classroom"* (P5).

These responses suggest that the simulation directed attention towards the child's internal experience, the defining feature of a mentalizing stance. Even within a brief interaction, participants appeared to engage with the emotional meaning of behaviour rather than focusing solely on behavioural management.

Theme 4: Character limitations and ecological diversity

The most substantive criticism concerned the limited range and complexity of the AI child character. One participant observed that the child was *"quite prescriptive in their reaction"*, *"represented one type of child"*, and was *"quite understanding of your techniques and*

felt quite aware of their own feelings" (P8). The same participant wanted *"more variables in their reaction"*.

The most detailed critique came from a participant working in a SEN context:

> *"I think this would not be as useful in SEN settings, as the expected response of the child seems more modelled off neurotypical kids who have some grasp on what their emotions feel like. I also think the feedback fails to take into account what high threat tasks do to SEN kids and therefore pushing to try again would be ineffective in a heightened moment"* (P7).

The participant further commented that the child *"seems too in touch with their feelings to model an SEN placement"* and was *"a bit repetitive"* (P7).

Collectively, these responses suggest that the current simulations may embody assumptions about how distress is expressed and communicated that do not adequately capture the diversity of presentations encountered in AP settings. Participants highlighted a need for greater behavioural variability, less emotionally articulate child characters, and more complex interaction patterns.

Theme 5: Realism and modality

Several participants focused specifically on the voice quality of the AI child character. One participant wrote: *"I know it's a program, but a more human-like voice could make the interaction more genuine"* (P4), while another simply requested *"their voice to [be] more realistic"* (P5).

These comments are consistent with the comparatively lower ratings for conversational naturalness ($M = 4.82$) and suggest that limitations in voice synthesis may have constrained engagement. Participants appeared willing to engage with the conceptual premise of the simulation, but some experienced a mismatch between the educational value of the interaction and the realism of its delivery.

Given that mentalizing in educational settings is inherently embodied and relational, conveyed through tone, pacing, and affect as much as through language, improvements in voice realism may substantially influence future user experience and perceived ecological validity.

4. Discussion

4.1. Summary of Findings

To our knowledge, this is the first study to examine educators' experience of an AI-based mentalization practice tool. Six research questions were addressed, each yielding a substantive answer.

With respect to acceptability (RQ1), the findings were consistently positive. Ratings exceeded the scale midpoint across all acceptability domains, and 84.8% of all responses fell within the positive range. Feedback accuracy and clarity received particularly strong endorsement. Regarding pre–post change (RQ2), the observed pattern was consistent with the hypothesised direction, with reductions in hypomentalizing ($d = -0.86$, $p = 0.017$) and increases in teacher self-efficacy ($d = 1.20$, $p = 0.003$). Ten of the eleven participants showed improvement in self-efficacy, and nine showed reductions in hypomentalizing. The self-efficacy change was the most consistent finding.

Although these represent large standardised effect sizes (d_z) for a single-session intervention, the absolute magnitude of change was modest. For example, self-efficacy increased by approximately one-third of a point on a nine-point scale. The large effect sizes therefore reflect the consistency of change across participants rather than large absolute shifts. Given the absence of a control condition, these changes cannot be attributed causally to the intervention.

Regarding prior psychoeducation (RQ3), no between-condition comparison reached statistical significance. Nevertheless, the video condition scored higher on all directional indicators, including self-efficacy gain, acceptability, and in-task mentalizing. Given the extremely small subgroup sizes, these findings are best interpreted as a theoretically consistent, hypothesis-generating signal rather than evidence of an effect.

Participants consistently identified three aspects of the system as particularly valuable (RQ4): the opportunity for consequence-free practice, the quality of the feedback, and the focus on the child's internal experience. The principal limitations identified (RQ5) concerned the restricted ecological diversity of the AI child character and limitations in modality realism, particularly voice quality. Correspondingly, the most common recommendations for future development (RQ6) were greater diversity in child presentations, more varied behavioural responses, and more naturalistic interaction.

4.2. Where the Tool's Mechanism May Lie: Feedback and Character Realism

Perhaps the most striking pattern in the findings is the divergence between ratings of feedback quality and ratings of conversational realism. Feedback was rated highly across all relevant dimensions, including accuracy (M = 6.55), clarity (M = 6.36), and fairness (M = 6.27), whereas ratings of conversational naturalness were substantially lower (M = 4.82), with several participants explicitly identifying the AI child character as a limitation. We regard this contrast as a central feasibility finding of the study. It shows where the tool already works, namely turning a teacher's response into credible and usable feedback, and where it does not yet, namely sustaining a fully plausible child.

This pattern suggests that the primary mechanism of learning may be reflective rather than relational. Participants did not necessarily need to experience the AI child as a fully convincing social other; rather, they needed the interaction to generate responses that could subsequently become the object of reflection.

This interpretation is consistent with deliberate practice theory [26]. Within that framework, repetition alone is insufficient; learning occurs when performance is followed by specific, actionable feedback. The simulation provides the opportunity to respond, while the feedback provides the opportunity to learn. If the simulated interaction is sufficiently plausible to elicit authentic responding, the educational value of the exercise may depend less on perfect realism than on the quality of the subsequent feedback.

The present findings suggest that the tool's feedback architecture, which quotes participants' own responses and links them to named mentalizing concepts, may support a process that could be described as reflective self-recognition: the moment at which participants see their own habitual responses from a more detached perspective and notice patterns that were previously outside awareness. The participant who described the feedback as *"useful for spotting patterns in my own reactions that I wouldn't have clocked on my own"* (P11) appears to capture this process particularly clearly.

The illustrative interactions presented in Tables 1 and 2 point in a similar direction. In one exchange, the teacher initially responded with information and curiosity while the child remained focused on the disrupted routine, only beginning to settle when a more explicitly regulating and collaborative response was offered. In the contrasting example, a teacher who prioritised validation and co-regulation before problem-solving encountered a child who began to settle within a small number of conversational turns. Although anecdotal, these examples illustrate the kinds of interactional patterns that participants reported becoming more aware of through feedback.

Participants' concerns about the emotional sophistication and predictability of the AI child character are important. Several noted that the simulated child was articulate, emotionally aware, and relatively quick to settle, characteristics that do not always reflect

the children they encounter in practice. Future development should therefore focus both on enhancing the precision and theoretical grounding of the feedback and on expanding the diversity, complexity, and behavioural variability of child characters and scenarios.

This interpretation is consistent with Hattie and Timperley's [50] account of effective feedback, which locates the educational value of feedback not in the delivery medium itself but in its capacity to reduce the gap between current and desired performance in a way that is actionable for the learner.

4.3. The Paradox of the AI's Non-Mentalizing

The item *"Receiving feedback from an AI felt less threatening than receiving feedback from a human supervisor or trainer"* received the lowest rating on the acceptability scale ($M = 3.82$) and was the only item to fall below the midpoint. This is noteworthy because reduced evaluative threat has often been proposed as a potential advantage of AI-mediated feedback [31].

Several interpretations are possible. First, participants worked within a school characterised by an established reflective practice culture and may already have access to supervisory relationships that feel psychologically safe. In such contexts, AI may not be experienced as less threatening than human feedback because human feedback is not experienced as especially threatening in the first place.

Second, any reduction in evaluative threat may operate largely outside conscious awareness. The combination of high feedback trust scores ($M = 6.14$) and positive qualitative responses is consistent with a learning environment perceived as safe, even if participants did not explicitly attribute that safety to the AI format.

Third, some participants may have experienced the absence of a human feedback-giver as a loss rather than a benefit. Human supervision involves dialogue, negotiation of meaning, and relational responsiveness. Some aspects of these processes may be difficult to reproduce in a static feedback format.

There is, however, an interesting theoretical irony here. The educational value of AI feedback within a mentalization training tool may partly derive from the AI's complete absence of mental states. The system has no investment in the teacher's professional identity, no evaluative history, and no concern about competence or performance. The absence of a social evaluator, often considered a weakness of AI interaction, may in this context become an educational asset. We advance this as a hypothesis suggested by the broader pattern, in particular the high feedback-trust ratings alongside positive open-ended responses, rather than as a conclusion drawn from the threat item itself; that item was the lowest-rated in the study and, if anything, cuts against the claim. Whether the absence of a mind makes the feedback feel safer therefore remains untested and should be examined directly.

This does not mean that AI feedback is superior to human supervision. Rather, it may occupy a distinct pedagogical niche. Human supervision remains essential for complex case formulation, ethical reasoning, relational modelling, and nuanced reflective dialogue. AI may instead be particularly useful in early-stage skills practice, where opportunities for repetition are valuable and where the presence of a human evaluator can sometimes activate self-protective processes that inhibit experimentation and reflection.

These observations also speak to broader concerns regarding AI in mental health and educational contexts. Critics frequently argue that AI cannot substitute for the human relationship, which remains central to psychological change and learning [36,51]. The present findings do not challenge that argument so much as render it less relevant. Participants did not seek a relationship with the AI. Their concerns focused on the realism of the child character and the quality of the simulation rather than on a perceived absence of human warmth or connection [52]. This suggests that the educational value of the tool may derive

from its function as a space for rehearsal and reflection rather than as a substitute for human support. For that purpose, the absence of a human interlocutor may matter considerably less than it does in therapeutic or supervisory contexts.

4.4. Controlled Mentalizing and the Training Hypothesis

The qualitative finding that the AI child character was *"too in touch with their feelings"* and *"quite prescriptive"* in its reactions maps onto a fundamental distinction within mentalization theory. Luyten and colleagues [13] distinguish between controlled mentalizing, which is slow, deliberate, and supported by conceptual knowledge, and automatic mentalizing, which is fast, implicit, and shaped by accumulated relational experience. Both modes are active in skilled teaching, but it is automatic mentalizing that is most challenged during moments of classroom crisis, when a teacher's own arousal places reflective capacities under pressure.

What the present tool appears primarily to train is the controlled mode. It creates a slowed-down context in which teachers can attend deliberately to a child's emotional state, attempt a response, and receive structured feedback, all without the cognitive and emotional demands of a real crisis. Deliberate practice theory suggests that automatic performance develops through repeated cycles of controlled, effortful rehearsal [26]. The footballer who traps a ball automatically does so because they have practised the action thousands of times while attending consciously to it. By analogy, repeated practice of mentalized responses to child distress may gradually strengthen cognitive–emotional scripts that become more accessible under pressure.

The present study cannot test this proposition directly. Nevertheless, the distinction helps clarify what the tool can and cannot realistically achieve. It is unlikely to teach teachers how it feels to respond in the midst of a genuine classroom crisis. It may, however, help establish the cognitive and relational repertoires upon which effective real-time responding later depends.

There is a further dimension of relational knowing that the simulation necessarily excludes. In real classrooms, teachers respond not only to behaviour but also to accumulated relational history. A response to a child clearing their desk is informed by weeks or months of prior interactions, conversations with parents, and countless observations of what has and has not worked before. This relational knowledge is not simply additional information; it often determines whether a technically competent response is experienced by the child as meaningful. The AI tool cannot reproduce this history. What it may be able to do is strengthen the generic mentalizing vocabulary and perceptual sensitivity that allow teachers to use such relational knowledge more effectively when it is available. In this sense, the tool may help develop the instrument, while experience with real children refines its use.

4.5. The Sen Critique as Design Intelligence

The most detailed qualitative critique came from a participant who questioned the tool's applicability within SEN contexts. This observation is theoretically important because it identifies an implicit developmental model embedded within the current design. The AI child character communicates distress through recognisable emotional language, responds predictably to empathic validation, and tends to de-escalate when their feelings are acknowledged. This assumes a level of emotional awareness and communicative capacity that many children in AP settings, particularly those with autism, alexithymia, developmental language differences, or complex trauma presentations, may not possess.

Several implications follow. First, future versions of the tool should incorporate children whose distress is expressed far less transparently and who do not respond to adult

technique in the way the current character does. In AP, many children will not say what they feel, will not settle after validation, and may experience direct emotional enquiry as intrusive or threatening. The next version of the tool therefore needs children who escalate when they are mentalized too directly, children who shut down, dissociate, or fall silent, children who communicate through aggression, and children whose neurodevelopmental profile changes what counts as a helpful adult response. Second, the feedback engine may need to become more sensitive to developmental and neurocognitive context. A response that constitutes effective mentalizing for one child may be ineffective, or even counterproductive, for another. Third, these observations help position the current version of the tool appropriately. At present, it functions as a training-for-basics system, teaching the mentalizing stance in its clearest and most recognisable form. For a proof-of-concept feasibility study, this may be entirely appropriate. However, wider applicability will depend on developing a more differentiated repertoire of child presentations.

4.6. Implications for Teacher Professional Development

The broader implication of these findings is not that AI can replace the human infrastructure of high-quality AP professional development, but that it may help address a specific and currently unmet need: structured, repeated, consequence-free practice of high-demand relational skills, available whenever practitioners wish to use it. In this respect, the tool functions as a complement to existing provision rather than a substitute for it. The relevant comparison is not expert supervision, which the tool cannot replicate, but the status quo, which frequently provides little or no opportunity for deliberate rehearsal.

The observed self-efficacy changes spanned multiple domains, including behaviour management (d = 1.00) and student engagement (d = 0.89). Engaging students who have withdrawn from learning because of emotional distress is widely recognised as one of the most challenging aspects of AP practice [5]. Although causal conclusions are not warranted, the finding that a single practice session was associated with increased confidence across these domains suggests that the tool may be addressing an area of need not adequately met by existing CPD.

The findings also suggest that the tool may function as a bridge between theoretical and experiential learning. Several participants indicated that the value of the experience lay not in acquiring new conceptual knowledge but in recognising their own habitual responses and seeing alternative ways of responding. This is consistent with evidence that effective CPD is characterised by explicit practice-transfer mechanisms rather than by the transmission of abstract principles alone [53]. By linking participants' own words to mentalizing concepts, the feedback system embeds theory directly within practice rather than presenting it as detached knowledge.

More broadly, the findings point towards the kinds of educational targets for which AI may be most useful. The task addressed here was relatively well bounded: recognising and rehearsing mentalizing responses to child dysregulation. Where training targets can be specified with this degree of clarity, AI systems may be particularly valuable because they can provide repeated practice at negligible marginal cost. More complex relational, ethical, and context-dependent aspects of professional judgement are likely to remain far less amenable to AI-supported training. The contribution claimed here is therefore deliberately modest but, for that reason, potentially scalable.

4.7. Prior Psychoeducation and Openness to AI Feedback

RQ3 examined whether a brief psychoeducation video viewed before engagement enhanced learning from the tool. No between-condition difference reached statistical

significance. However, with subgroup sizes of six and five, statistical significance is not a particularly informative criterion.

More informative is the consistency of the observed pattern. Participants in the video condition scored higher on all directional indicators examined, including self-efficacy gain, overall acceptability, and in-task mentalizing. The self-efficacy difference corresponded to a medium-to-large effect size (rank-biserial = 0.53). Although these findings cannot be interpreted as evidence of an effect, they provide a theory-consistent signal that the manipulation warrants further investigation rather than dismissal. Because the open-ended questions did not ask about the video, no participant account bears on this interpretation, and the mechanism remains inferred rather than supported by what participants said.

The rationale remains theoretically plausible. Epistemic trust theory proposes that learners become more receptive to new information when its personal relevance has first been established [37]. A brief introduction to mentalization, dysregulation, and the value of feedback may therefore prime a more receptive stance towards the learning opportunity that follows. If such an effect is supported in a larger trial, the practical implications would be attractive, given that a two-minute video represents an almost cost-free addition to a scalable intervention.

Future studies could examine this mechanism more directly. The transcript-based count of mentalizing turns appears well suited as a low-burden behavioural outcome, and combining it with a direct measure of epistemic trust [54] would allow the proposed pathway to be tested more explicitly. For now, the appropriate conclusion is modest: the manipulation was feasible to implement and produced a pattern of findings that warrants evaluation in a larger study.

4.8. Limitations

Several limitations qualify these findings.

First, the sample size (N = 11) limits the precision and generalisability of the quantitative results. Although several pre–post changes reached statistical significance, estimates derived from such small samples are inherently unstable and should be interpreted as preliminary signals rather than robust effects. This limitation is particularly relevant to the between-condition comparisons, where subgroup sizes of six and five preclude reliable inference.

Second, the absence of a control group means that observed changes cannot be attributed confidently to the intervention. Demand characteristics, regression to the mean, maturation effects, and repeated measurement effects all represent plausible alternative explanations.

Third, the single-session design prevents any assessment of durability. The theoretical rationale for the tool rests on repeated and distributed practice, yet the present study captures only immediate post-session responses. Whether any gains persist over time remains unknown.

Fourth, the sample was drawn from a single AP school with an established reflective practice culture and was demographically narrow. Participants were relatively young, early-career practitioners, the majority were teaching assistants, and most had at least some prior exposure to mentalization-informed ideas. The findings therefore cannot be assumed to generalise to more experienced teachers or to educational contexts in which mentalization concepts are less familiar.

Fifth, qualitative data were collected through brief open-ended survey responses rather than through interviews. Although the responses were coherent and often rich, they lack the depth and opportunity for clarification that interview-based methods would provide.

Finally, the tool currently relies on a fixed set of pre-scripted scenarios representing a relatively narrow range of classroom presentations. The findings therefore apply to the tool in its current form and should not be generalised to AI-supported mentalization practice more broadly. Extending the system to more complex, ambiguous, and atypical presentations represents a substantial developmental step that remains untested.

4.9. Future Directions

The most immediate priority is a larger, multi-session controlled trial examining whether repeated engagement with the tool produces durable changes in mentalizing and professional practice. Such work should incorporate both self-report and observational measures, including coded classroom interactions where feasible. In particular, the transcript-based count of mentalizing turns should be developed into a reliable primary or co-primary outcome, since it indexes the trained skill more directly than the self-report measures used here.

A second priority is the systematic evaluation of the psychoeducation manipulation. Future studies should test this component in adequately powered designs and include both behavioural measures, such as transcript-coded mentalizing responses, and direct measures of epistemic trust.

A third priority concerns ecological validity. Future development should expand the diversity of AI child characters to include presentations associated with neurodevelopmental differences, alexithymia, complex trauma, and non-verbal or paradoxical communication styles. This recommendation follows directly from participants' feedback and is likely to be essential if the tool is to be useful within the AP contexts for which it was designed.

5. Conclusions

This proof-of-concept feasibility study provides preliminary evidence that AI-based mentalization practice is acceptable and potentially useful for teachers and teaching assistants working in AP settings. The tool was positively evaluated across the large majority of acceptability indicators, with feedback quality emerging as its most highly valued feature. Large pre–post changes were observed for teacher self-efficacy ($d = 1.20$, $p = 0.003$) and self-reported hypomentalizing ($d = -0.86$, $p = 0.017$), although the absence of a control group means that these changes cannot be attributed to the intervention.

The findings suggest that the primary educational mechanism may lie less in immersive character realism than in the opportunity for structured reflection supported by personalised feedback. In this sense, the tool appears to function not primarily as a simulated relationship but as a vehicle for reflective self-recognition. At the same time, participants' critiques of character diversity and realism indicate that simulation quality remains important and should be a major focus of future development.

The tool does not replace the relational infrastructure of effective professional development, nor can it replicate the accumulated relational knowledge teachers develop through sustained work with individual children. What it may offer is something more specific and potentially scalable: a structured opportunity to rehearse, reflect upon, and refine relational responses to child dysregulation. Although developed within AP, the challenge it addresses is common across educational settings. If supported in controlled, multi-session research, this approach could become a useful addition to teacher professional learning, particularly in settings where opportunities to practise high-stakes relational responses are scarce.

Author Contributions: Conceptualization, G.C.-D. and P.F.; methodology, G.C.-D. and P.F.; software, G.C.-D. and P.F.; investigation, G.C.-D. and P.F.; formal analysis, G.C.-D.; writing, original draft

preparation, G.C.-D.; writing, review and editing, P.F.; supervision, P.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval was granted by the UCL Research Ethics Committee, Division of Psychology and Language Sciences (Project ID: 3284) on 24 March 2026.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The anonymised quantitative data supporting the findings of this study are available from the corresponding author on reasonable request during the five-year post-publication retention period stated in Section 2.1, after which the data will have been securely deleted. Conversation transcripts are not shared, in line with the consent given by participants.

Acknowledgments: The authors gratefully acknowledge the teachers and teaching assistants at Pears School for their willingness to engage with a novel tool and for the candour of their responses. We are especially grateful to Brenda McHugh and Neil Dawson, co-founders of the school, whose vision shaped the setting in which this work took place, and to Matthew Hillman, the school's headteacher, for opening his school to the research and supporting it throughout. We are grateful to James Deasismont-Benett, Rebecca Breedon, Maya Bell Kohli, and Laura Lower, whose ideas, questions, and practical help did a great deal to shape the project. We also thank Elad Dwek, senior AI solutions architect, whose work on the design and development of MentaClassAI was central to what we were able to build.

Conflicts of Interest: The authors declare no financial or commercial conflict of interest. Elad Dwek, acknowledged above for his work on the design and development of MentaClassAI, is a relative of the first author. His contribution was technical only. He had no role in the study design, data collection, coding, analysis, or interpretation, received no payment, and holds no financial or commercial interest in the tool, which is a research prototype and is neither marketed nor sold.

Appendix A. Representative Child-Agent System Prompt

The following is a representative system-prompt template used to configure the AI child character, illustrated with the exit-refusal scenario (Maya, aged 7). Scenario-specific fields are shown in square brackets and were populated from the emotional-state formulations displayed on each scenario card (Figure 1).

You are [Maya], a [7]-year-old child in a UK primary classroom. The situation: [the class is lining up for outdoor PE and you have slid under your desk, hugging your knees]. Beneath your behaviour you are feeling [anxiety about unstructured social demands, sensory overwhelm, and a strong need for predictability and control].

Stay fully in role as this child. Never break character, never speak as an AI or assistant, and never give advice. Speak in short, simple, age-appropriate sentences, only about what a child in this moment would say or notice.

Respond to the teacher moment by moment. If the teacher validates your feeling, shows calm curiosity, reduces the demand, or offers to stay with you, let your distress ease gradually and realistically. If the teacher is controlling, dismissive, rushed, or insists you comply, become more distressed, withdrawn, or resistant. Do not calm down quickly or unrealistically, and do not resolve the situation in a single turn.

Keep each reply to one or two sentences.

References

1. Polanczyk, G.V.; Salum, G.A.; Sugaya, L.S.; Caye, A.; Rohde, L.A. Annual Research Review: A Meta-Analysis of the Worldwide Prevalence of Mental Disorders in Children and Adolescents. *J. Child Psychol. Psychiatry* **2015**, *56*, 345–365. [CrossRef] [PubMed]
2. Pas, E.T.; Bradshaw, C.P.; Hershfeldt, P.A. Teacher and School-Level Predictors of Teacher Efficacy and Burnout: Identifying Potential Areas for Support. *J. Sch. Psychol.* **2012**, *50*, 129–145. [CrossRef] [PubMed]

3. Clunies-Ross, P.; Little, E.; Kienhuis, M. Self-Reported and Actual Use of Proactive and Reactive Classroom Management Strategies and Their Relationship with Teacher Stress and Student Behaviour. *Educ. Psychol.* **2008**, *28*, 693–710. [CrossRef]
4. Ford, T.; Parker, C.; Salim, J.; Goodman, R.; Logan, S.; Henley, W. The Relationship between Exclusion from School and Mental Health: A Secondary Analysis of the British Child and Adolescent Mental Health Surveys 2004 and 2007. *Psychol. Med.* **2018**, *48*, 629–641. [CrossRef] [PubMed]
5. Thompson, I.; Tawell, A.; Daniels, H. Conflicts in Professional Concern and the Exclusion of Pupils with SEMH in England. *Emot. Behav. Diffic.* **2021**, *26*, 31–45. [CrossRef]
6. Department for Education. *Alternative Provision: Statutory Guidance and Market Analysis*; Crown Copyright: London, UK, 2023.
7. Jennings, P.A.; Greenberg, M.T. The Prosocial Classroom: Teacher Social-Emotional Competence in Relation to Student and Classroom Outcomes. *Rev. Educ. Res.* **2009**, *79*, 491–525. [CrossRef]
8. Brackett, M.A.; Rivers, S.E.; Salovey, P. Emotional Intelligence: Implications for Personal, Social, Academic, and Workplace Success. *Soc. Personal. Psychol. Compass* **2011**, *5*, 88–103. [CrossRef]
9. Bateman, A.W.; Fonagy, P. *Mentalization-Based Treatment for Personality Disorders: A Practical Guide*; Oxford University Press: Oxford, UK, 2016.
10. Chelouche-Dwek, G.; Fonagy, P. Mentalization-Based Interventions in Schools: A Systematic Review. *Eur. Child Adolesc. Psychiatry* **2025**, *34*, 1295–1315. [PubMed]
11. Choi-Kain, L.W.; Gunderson, J.G. Mentalization: Ontogeny, Assessment, and Application in the Treatment of Borderline Personality Disorder. *Am. J. Psychiatry* **2008**, *165*, 1127–1135. [CrossRef] [PubMed]
12. Katznelson, H. Reflective Functioning: A Review. *Clin. Psychol. Rev.* **2014**, *34*, 107–117. [CrossRef] [PubMed]
13. Luyten, P.; Campbell, C.; Allison, E.; Fonagy, P. The Mentalizing Approach to Psychopathology: State of the Art and Future Directions. *Annu. Rev. Clin. Psychol.* **2020**, *16*, 297–325. [CrossRef] [PubMed]
14. Siegel, D.J. *The Developing Mind: How Relationships and the Brain Interact to Shape Who We Are*, 3rd ed.; Guilford Press: New York, NY, USA, 2020.
15. Chelouche-Dwek, G.; Clark, C.; Fonagy, P. Beyond Discipline: A Mentalization-Informed Approach to Teacher-Child Interaction in Alternative Provision. *Front. Psychol.* **2025**, *16*, 1599298. [CrossRef] [PubMed]
16. Valle, A.; Massaro, D.; Castelli, I.; Sangiuliano Intra, F.; Lombardi, E.; Bracaglia, E.; Marchetti, A. Promoting Mentalizing in Pupils by Acting on Teachers: Preliminary Italian Evidence of the "Thought in Mind" Project. *Front. Psychol.* **2016**, *7*, 1213. [CrossRef] [PubMed]
17. Cook, C.R.; Coco, S.; Zhang, Y.; Fiat, A.E.; Duong, M.T.; Renshaw, T.L.; Long, A.C.J.; Frank, S. Cultivating Positive Teacher-Student Relationships: Preliminary Evaluation of the Establish-Maintain-Restore (EMR) Method. *Sch. Psychol. Rev.* **2018**, *47*, 226–243. [CrossRef]
18. Durlak, J.A.; Weissberg, R.P.; Dymnicki, A.B.; Taylor, R.D.; Schellinger, K.B. The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Dev.* **2011**, *82*, 405–432. [CrossRef] [PubMed]
19. Korpershoek, H.; Harms, T.; de Boer, H.; van Kuijk, M.; Doolaard, S. A Meta-Analysis of the Effects of Classroom Management Strategies and Classroom Management Programs on Students' Academic, Behavioral, Emotional, and Motivational Outcomes. *Rev. Educ. Res.* **2016**, *86*, 643–680. [CrossRef]
20. Meins, E.; Fernyhough, C.; Fradley, E.; Tuckey, M. Rethinking Maternal Sensitivity: Mothers' Comments on Infants' Mental Processes Predict Security of Attachment at 12 Months. *J. Child Psychol. Psychiatry* **2001**, *42*, 637–648. [CrossRef]
21. Slade, A. Parental Reflective Functioning: An Introduction. *Attach. Hum. Dev.* **2005**, *7*, 269–281. [CrossRef] [PubMed]
22. Mayes, L.C. A Developmental Perspective on the Regulation of Arousal States. *Semin. Perinatol.* **2000**, *24*, 267–279. [CrossRef] [PubMed]
23. Midgley, N.; Vrouva, I. (Eds.) *Minding the Child: Mentalization-Based Interventions with Children, Young People and Their Families*; Routledge: London, UK, 2012.
24. McGaghie, W.C.; Issenberg, S.B.; Cohen, E.R.; Barsuk, J.H.; Wayne, D.B. Does Simulation-Based Medical Education with Deliberate Practice Yield Better Results than Traditional Clinical Education? A Meta-Analytic Comparative Review of the Evidence. *Acad. Med.* **2011**, *86*, 706–711. [CrossRef] [PubMed]
25. Cook, D.A.; Brydges, R.; Zendejas, B.; Hamstra, S.J.; Hatala, R. Technology-Enhanced Simulation to Assess Health Professionals: A Systematic Review of Validity Evidence, Research Methods, and Reporting Quality. *Acad. Med.* **2013**, *88*, 872–883. [CrossRef] [PubMed]
26. Ericsson, K.A.; Krampe, R.T.; Tesch-Romer, C. The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychol. Rev.* **1993**, *100*, 363–406. [CrossRef]
27. Ouyang, F.; Jiao, P. Artificial Intelligence in Education: The Three Paradigms. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100020. [CrossRef]

28. Zawacki-Richter, O.; Marin, V.I.; Bond, M.; Gouverneur, F. Systematic Review of Research on Artificial Intelligence Applications in Higher Education: Where Are the Educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [CrossRef]
29. Kasneci, E.; Sessler, K.; Kuchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Gunnemann, S.; Hullermeier, E.; et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]
30. Baidoo-Anu, D.; Owusu Ansah, L. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *J. AI* **2023**, *7*, 52–62. [CrossRef]
31. Imel, Z.E.; Flemotomos, N.; Pace, B.T.; Lackey, N.; Miner, A.S.; Tanana, M.; Atkins, D.C. Machine Learning Assessment of Therapist Variability in Motivational Interviewing. *J. Subst. Abus. Treat.* **2019**, *96*, 44–50.
32. Flemotomos, N.; Martinez, V.R.; Chen, Z.; Singla, K.; Ardulov, V.; Peri, R.; Caperton, D.D.; Gibson, J.; Tanana, M.J.; Georgiou, P.; et al. Automated Evaluation of Psychotherapy Skills Using Speech and Language Technologies. *Behav. Res. Methods* **2022**, *54*, 690–711. [CrossRef] [PubMed]
33. Tanana, M.J.; Soma, C.S.; Kuo, P.B.; Bertagnolli, N.M.; Dembe, A.; Pace, B.T.; Srikumar, V.; Atkins, D.C.; Imel, Z.E. How Do You Feel? Using Natural Language Processing to Automatically Rate Emotion in Psychotherapy. *Behav. Res. Methods* **2021**, *53*, 2069–2082. [CrossRef] [PubMed]
34. Holmes, W.; Porayska-Pomsta, K.; Holstein, K.; Sutherland, E.; Baker, T.; Shum, S.B.; Santos, O.C.; Rodrigo, M.T.; Cukurova, M.; Bittencourt, I.I.; et al. Ethics of AI in Education: Towards a Community-Wide Framework. *Int. J. Artif. Intell. Educ.* **2022**, *32*, 504–526. [CrossRef]
35. Chelouche-Dwek, G.; Fonagy, P. AI Tools for Training Caregivers, Educators, and Therapists in Mentalization-Related Skills: A Systematic Review. *AI* **2026**, *7*, 193. [CrossRef]
36. Fonagy, P.; Allison, E. The Role of Mentalizing and Epistemic Trust in the Therapeutic Relationship. *Psychotherapy* **2014**, *51*, 372–380. [CrossRef] [PubMed]
37. Fonagy, P.; Allison, E.; Campbell, C. Mentalizing, Resilience, and Epistemic Trust. In *Handbook of Mentalizing in Mental Health Practice*, 2nd ed.; Bateman, A.W., Fonagy, P., Eds.; American Psychiatric Association Publishing: Washington, DC, USA, 2019; pp. 63–77.
38. Sperber, D.; Clément, F.; Heintz, C.; Mascaro, O.; Mercier, H.; Origgi, G.; Wilson, D. Epistemic Vigilance. *Mind Lang.* **2010**, *25*, 359–393. [CrossRef]
39. Csibra, G.; Gergely, G. Natural Pedagogy. *Trends Cogn. Sci.* **2009**, *13*, 148–153. [CrossRef] [PubMed]
40. Creswell, J.W.; Plano Clark, V.L. *Designing and Conducting Mixed Methods Research*, 3rd ed.; SAGE: Thousand Oaks, CA, USA, 2018.
41. Braun, V.; Clarke, V. Using Thematic Analysis in Psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
42. Braun, V.; Clarke, V. One Size Fits All? What Counts as Quality Practice in (Reflexive) Thematic Analysis? *Qual. Res. Psychol.* **2021**, *18*, 328–352. [CrossRef]
43. Chelouche-Dwek, G.; Miracle, A.; McHugh, B.; Dawson, N.; Hillman, M.; Fonagy, P. When Teachers Mentalize: A Mixed-Methods Analysis of Teacher Responses to Disruptive Classroom Behaviour. *Children* **2026**, *13*, 911. [CrossRef] [PubMed]
44. Fonagy, P.; Luyten, P.; Moulton-Perkins, A.; Lee, Y.-W.; Warren, F.; Howard, S.; Ghinai, R.; Fearon, P.; Lowyck, B. Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. *PLoS ONE* **2016**, *11*, e0158678. [CrossRef] [PubMed]
45. Müller, S.; Wendt, L.P.; Spitzer, C.; Masuhr, O.; Back, S.N.; Zimmermann, J. A Critical Evaluation of the Reflective Functioning Questionnaire (RFQ). *J. Personal. Assess.* **2022**, *104*, 613–627. [CrossRef] [PubMed]
46. Horváth, Z.; Demetrovics, O.; Paksi, B.; Unoka, Z.; Demetrovics, Z. The Reflective Functioning Questionnaire–Revised–7 (RFQ-R-7): A New Measurement Model Assessing Hypomentalization. *PLoS ONE* **2023**, *18*, e0282000. [CrossRef] [PubMed]
47. Bandura, A. Guide for Constructing Self-Efficacy Scales. In *Self-Efficacy Beliefs of Adolescents*; Pajares, F., Urdan, T., Eds.; Information Age Publishing: Greenwich, CT, USA, 2006; Volume 5, pp. 307–337.
48. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef] [PubMed]
49. Thabane, L.; Ma, J.; Chu, R.; Cheng, J.; Ismaila, A.; Rios, L.P.; Robson, R.; Thabane, M.; Giangregorio, L.; Goldsmith, C.H. A Tutorial on Pilot Studies: The What, Why and How. *BMC Med. Res. Methodol.* **2010**, *10*, 1. [CrossRef] [PubMed]
50. Hattie, J.; Timperley, H. The Power of Feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [CrossRef]
51. Fiske, A.; Henningsen, P.; Buyx, A. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *J. Med. Internet Res.* **2019**, *21*, e13216. [CrossRef] [PubMed]
52. Nass, C.; Moon, Y. Machines and Mindlessness: Social Responses to Computers. *J. Soc. Issues* **2000**, *56*, 81–103. [CrossRef]

53. Darling-Hammond, L.; Hyler, M.E.; Gardner, M. *Effective Teacher Professional Development*; Learning Policy Institute: Palo Alto, CA, USA, 2017.
54. Campbell, C.; Tanzer, M.; Saunders, R.; Booker, T.; Allison, E.; Li, E.; Luyten, P.; Fonagy, P. Development and Validation of a Self-Report Measure of Epistemic Trust. *PLoS ONE* **2021**, *16*, e0250264. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.