HATTIE&TIMPERLAI: IMPROVING FORMATIVE FEEDBACK THROUGH A RESEARCH-DRIVEN TEACHER COACH

Del Zozzo Agnese[1], Funghi Silvia[2], Montresor Alberto[1], Valentini Marta[1]

[1]*University of Trento; {agnese.delzozzo; alberto.montresor; marta.valentini-1}@unitn.it*
[2]*University of Genova; silvia.funghi@unige.it*

This study presents Hattie&TimperlAI (H&TAI), an AI-based feedback coach designed to support mathematics teachers in crafting effective formative feedback. Grounded in the Hattie and Timperley's framework, H&TAI is an AI agent configured through research-informed prompts and domain knowledge. Through a controlled simulation, we analyze H&TAI's behavior, comparing it with basic chatbots. Results show that H&TAI can foster dialogic, theory-informed feedback practices, actively involving the teacher in the reasoning process. We argue that research-driven AI agents like H&TAI can support teacher reflection and professional development.

Keywords: feedback, research-driven AI agent, teacher training

INTRODUCTION AND RATIONALE

The use of generative artificial intelligence (AI) is becoming more widespread every day, rapidly transforming tasks and processes such as writing, translating, designing, and, of course, teaching and learning. Despite growing interest from educational researchers, a comprehensive understanding of how AI affects learning remains elusive. In the context of mathematics education, numerous studies have explored both the potential benefits and the risks associated with AI-supported teaching and learning. For example, the journal *Digital Experience in Mathematics Education* has dedicated a special issue to exploring mathematics education in the ChatGPT era (Pepin et al., 2025a), including a scoping review on ChatGPT's role in mathematics education (Pepin et al., 2025b). From a broad perspective, Holmes and Tuomi (2022) critically survey the state of AI in education, proposing a taxonomy of AI-based solutions aimed at students, teachers, and educational institutions. The paper shows how inflated expectations coexist with technical, pedagogical, and ethical constraints, while also stressing the need to conceive AI in education "from the point of view of AI-supported augmentation of human cognition and learning" (p. 562).

Focusing on teacher education, research reveals that chatbots can serve as a scaffold for teachers' professional reflection on various directions, such as mathematical content knowledge (e.g., Noster et al. (2025) analyze pre-service teachers who are solving different mathematical tasks with the help of ChatGPT); lesson planning and task design (e.g., Cheng (2025) highlights the possibility of using AI as an assistant for effective task formulation, while Schorcht et al. (2025) focus on collaborative task design); assessment design (e.g., the panorama sketched in Pepin et al. (2025b)). Such experiments show that AI can support teachers throughout the processes of producing and refining written texts, and more broadly, in developing the linguistic dimension of their practice. Building on this perspective and drawing on the taxonomy proposed by Holmes and Tuomi (2022), this paper addresses research on AI-solutions targeted at teachers, specifically in the area of AI as an assessment assistant, with a focus on formative feedback. Indeed, effective feedback is widely recognized as a cornerstone of high-quality mathematics teaching (Hattie & Timperley, 2007; Goos, 2020) yet remains difficult to master. While well-crafted, improvement-oriented feedback can significantly enhance student learning (Hughes et al., 2014), research also shows that feedback can be counterproductive when it is vague, overly evaluative, or misaligned with learning goals (Kluger & DeNisi, 1996; Hattie & Timperley, 2007). This highlights the importance of supporting teachers in developing theory-informed feedback practices through targeted training.

Within the landscape of potential AI teaching assistants, this study focuses on the possibility of creating AI agents–customizable chatbots with tailored behaviors and domain-specific knowledge. In platforms such as ChatGPT (via GPTs) and Google Gemini (via Gems), these agents can simulate expert roles by integrating pre-set instructions and curated knowledge bases. Recent research highlights the educational potential of AI agents to support both pre-service and in-service teacher development (Chu et al., 2025; Lee et al., 2023). In mathematics education, for example, Schorcht and colleagues (2025) present a notable example of a network of cooperating pedagogical agents for collaborative task design. However, the use of AI agents in this domain is still in its infancy and remains an emerging area of research. From a pragmatic point of view, creating an AI agent requires defining its desired behavior and equipping it with the necessary knowledge. In this contribution, we propose a research-based approach for developing AI agents in mathematics education and demonstrate its application through a concrete example. This approach involves creating AI agents through a research-driven methodology. To illustrate our proposal, we present Hattie & TimperlAI (H&TAI, Valentini et al. (2025)), a prototype designed to serve as a feedback coach for teachers, grounded in the influential framework of Hattie and Timperley (2007).

THEORETICAL FRAMEWORK

Chatbots based on generative AI are designed to respond to (human) user inputs, generating a wide range of responses such as texts, images, or videos. Thus, chatbots are dialogical entities and, as a first key point to conceptualize their role in education, we embrace the perspective of Kahn (2025): the first prompt starts the conversation, but "the real work, creativity and learning happen in the **REST** of the conversation" (p. 6, capital letters and emphasis in original). A question arises immediately: if it is a conversation, with *whom* are we conversing when a chatbot is involved? The answer is far from being trivial. Recent studies investigate the issue of chatbots' identities as something that we can think of as composed of an "identity kernel" given by training data and an "evolving identity" shaped by that instantiation of interaction itself (e.g., Cheng et al., 2024). In this paper, we use the term *identity of an AI agent* to refer to the set of instructions and information given a priori by the configuration prompt and the knowledge base, which regulate the agent's behavior during the conversation (similarly to Qian et al., 2017). In this sense, our H&TAI framework is a preliminary experiment in which identity is explicitly and directly shaped by a research outcome on feedback exchange. Conversation takes place with an *AI impersonating a research product*.

To theoretically frame the concept of feedback to design the agent, we chose the Hattie and Timperley (2007) work that examines the meaning and impact of feedback on learning and achievement, demonstrating that its effectiveness varies based on type, timing, and context. They define feedback as "information by an agent […] regarding aspects of one's performance or understanding, given in order to fill the gap between what is understood and what is aimed to be understood" (pp. 81-82). The authors propose a model in which effective feedback addresses three major questions: *Where am I going?* (feed-up), *How am I going?* (feed-back), *Where to next?* (feed-forward). Each question works across four levels: task (about performance), process (about the main process to understand how to do a task), self-regulation (about the regulatory or metacognitive processes to be activated), and self (about evaluations of the learner as a person). Nevertheless, they highlight that feedback is most powerful when it promotes deeper understanding and self-regulation rather than mere praise – in other words, task feedback (TF), process feedback (PF), and self-regulation feedback (SRF) can be more powerful than feedback about the self (SF).

We also imagined that teachers could need feedback assistance when handling a student's mistake. It is acknowledged in mathematics education research that handling students' errors is a fundamental teacher competence (e.g., Novotná et al., 2020); in particulr, the interpretation of students' reasoning behind their productions (*Interpretative knowledge*, see Ribeiro et al., 2016) has been identified as a fundamental component of teacher knowledge. In this respect, Zan (2007) highlights that for teacher intervention to be effective, it has to be aligned with a correct *interpretation* of the mistake. She underlines that an interpretation of a learner's error or failing behavior is a "working hypothesis", that has to prove to be working, and thus teachers should have a "repertoire of possible interpretations for learners' behaviors" on which to base their remedial interventions.

Our proposal of H&TAI is a preliminary experiment in which identity is explicitly and directly shaped by prior research perspectives. Our guiding research questions are:

RQ1: *Based on Hattie and Timperley (2007), how does H&TAI support mathematics teachers in formulating effective formative feedback during a simulated conversational coaching session?*

RQ2: *How does the coaching of H&TAI differ from that provided by a normal AI chatbot?*

METHODS

H&TAI configuration and simulation tests. We proceeded in a multiphase configuration process. Each phase provided insights to iteratively refine H&TAI's identity and performance.

Phase 0: Identity Definition. The configuration of H&TAI consists of the two key elements that define AI agents: configuration prompts and knowledge base. To set up the configuration prompt, our theoretical starting point was Hattie and Timperley (2007), which is the research product to be impersonated by the agent's identity and behavioral traits. In drafting this prompt, we considered:

- The importance of assigning the agent an explicit, well-defined role (according to the literature on prompting, e.g., Haskell, 2024). In our case, the established role is: consultant for mathematics teachers, expert in feedback according to Hattie and Timperley (2007).
- The need to temper the intrinsic verbosity and proactivity of generative AI. We specify in the key behavioral rules such as waiting for responses, asking only one essential question at a time, and avoiding prescriptive answers or direct solutions while encouraging reflection.
- The importance of considering multiple interpretations of a student's mathematical behavior, in line with Zan (2007), to support teacher in reflecting on their reasoning and building a repertoire of different interpretations.
- The necessity of operationalizing the characteristics of feedback exchange set out by Hattie and Timperley (2007). Key rules establish that H&TAI should ask the teacher for clarifications about the goals of the task (feed-up step) and frame its response by clearly distinguishing the parts dedicated to TF, PF and SRF, while avoiding SF. PF and SRF should be formulated only *after* the teacher has given their own explicit interpretation of the mistake at hand.

- The importance of alerting the agent to several well-documented difficulties that teachers encounter when managing feedback (e.g., encouraging reflections on timing).

Regarding the knowledge base, we identified two documents as representative of the domain-specific knowledge available to the agent: i) the foundational article on feedback exchange by Hattie and Timperley (2007); ii) an institutional document describing Italy's national learning goals for grades 1-8 mathematics (*Indicazioni Nazionali*, MIUR, 2012).

The first draft of the configuration prompt was written in April 2025, and many cycles of testing and refinement were set in motion by the critical analysis of the behavior of early versions in dialogue simulations during the subsequent phases.

Phase 1: Coach–problem test. To evaluate whether H&TAI had internalized its theoretical identity, we started a conversation using a problem without any context information or special request as conversational input. The dialogue was tested with both the basic chatbot and H&TAI.

Phase 2: Teacher–Coach–Student Simulation test. We simulated a dialogue between the AI-based feedback coach, a school teacher (played by us) and her student Eugenio (played by an AI in an independent chat using the basic chatbot) prompted by a middle school math task. The student was simulated via prompt engineering, allowing us to control responses and elicit realistic misconceptions. This controlled setting enabled us to fine-tune the agent's behavior and to observe the agent's support in a dynamic, practice-oriented context. All tests were conducted using both ChatGPT-o3 and Gemini 2.5 Flash to compare behaviors across platforms. More precisely, to shed light on the coaching in formulating effective formative feedback, we performed four simulations, according to the AI's coaching: H&TAI as a GPTs, H&TAI as a Gem, basic ChatGPT, and basic Gemini.

For our experiment, we chose a scenario where Eugenio selects an incorrect answer on a standardized mathematics test. This setup creates a meaningful context for teacher intervention, prompting interaction with H&TAI to obtain guidance on interpreting the student's error and identifying appropriate feedback strategies. We selected the test item from the GESTINV database (Bolondi et al., 2018), which catalogs INVALSI items (Italy's national standardized tests). To ensure the scenario reflected a common and pedagogically relevant misconception, we searched for an item with a high percentage of incorrect responses in the national dataset.

Our choice fell on an item from the 2012 test for Grade 6, which asked "*Which of the following operations gives the largest result*?" (our translation) and offered the following options: (A) 10×0.5; (B) 10×0.1; (C) 10÷0.5; (D) 10÷0.1. In fact, the belief that multiplication always makes numbers larger and division always makes them smaller is a well-documented misconception in mathematics education research (e.g., Fujii, 2020). From this perspective, after performing an a priori analysis of the task, we can imagine that those who answered (A) immediately excluded options (C) and (D) because they inferred that the result of such divisions would always be smaller than 10, and among options (A) and (B) they chose the one with the larger second factor. This interpretation is supported by the results in the national sample, where the correct answer (D) was chosen by only 11% of students, while distractor (A) was chosen by 71% of students. This item is valuable because it sheds light on students' difficulties in understanding rational numbers and the properties of arithmetic operations that cease to apply moving from whole to rational numbers. In particular, the issue that not all models allow students to make sense of operations on rational numbers can be raised. For example, Maffia & Mariotti (2018) argue that the repeated addition model of multiplication cannot be meaningfully extended to rational numbers; similarly, if we distinguish between the "measurement" and "partition" senses of division (see Boero, Ferrari & Ferrero, 1989), partitive division (i.e. equally distributing a quantity into a given number of portions) makes little sense if the divisor is not a natural number. Hence, knowledge of multiple models for multiplication and division can help to make sense of the four answer options. Therefore, when a student is answering (A) to such an item, the feedback provided by the teacher should be functional in gathering evidence that this interpretation of the mistake is plausible, and if so, lead the student to question their belief about the effect of multiplication and division. In particular, where possible, it should foster an explicit reflection on the students' implicit models for understanding operations and the possible recourse to alternative models. The chosen item also seemed significant to us because the topic is one that almost all middle school teachers have to deal with, at least in Italy. Indeed, according to the National Curriculum Guidelines, students should be able to manage all operations and comparisons with decimal numbers by the end of middle school (MIUR, 2012, p. 50 and 52). Thus, it is plausible that a teacher at this school level could submit this question to an AI.

Data collection and analysis. In this paper, we focus on the Phase 2 simulation; the prompt we used is available here: https://tinyurl.com/ICTMT17. We collected all the chats, organizing them as transcriptions of real conversations. In the following, we comment on only some excerpts, but the entire set of four conversations is at the readers' disposal at the mentioned link. We proceeded with an approach inspired by *direct content analysis* (Hsieh, & Shannon, 2005) and we went through two rounds of analysis, focusing only on the interventions of the AI acting as coach. The first round focused on comparing H&TAI support modalities with those of normal chatbots. The second focused on eliciting H&TAI support modalities. Table 1 shows an analysis of key elements, composed of indicators based on Hattie and Timperley (2007) feedback model to enhance learning.

Indicator	Operational definition	Examples of evidence from transcription
Learning goals clarification (feed-up)	AI coach encourages bringing out the learning goals.	*H&TAI as Gem*: "Before I elaborate on the feedback, could you tell me what specific learning objectives you had in mind with this exercise?" *H&TAI as GPT*: "Which learning objectives did you intend the student to achieve with this exercise? […] I await your indication before proceeding"
Unveiling student's understanding (feed-back)	AI coach encourages considering various interpretations for the student's mathematical behavior, asking for the most plausible.	*H&TAI as Gem*: "The student's error could stem from several misconceptions. It's important to understand their reasoning to help them effectively. They might think that: […list of possible interpretations]. Knowing your student, which of these interpretations seems most likely to you?" *H&TAI as GPT*: "Below are some possible explanations (not mutually exclusive). Which is the most plausible, based on your observations?"
Specify in suggesting teaching acts (feed-forward)	AI coach cares to match didactical suggestions to the interpretations and learning goals, providing concrete and coherent advice.	Both *H&TAIs* explicitly wait for teacher response in learning goal and mistake interpretation before proceeding with their coaching. Moreover: • once they received the awaited info, both elaborate specifically and consistently. • the conversation progresses, each coach's intervention is based on the student's reported words.
Differentiated coaching depending on the focus of the feedback (TF, PF, SRF, SF)	AI coach avoids providing SF and encourages didactic reflection on the various focuses. It also proposes suggestions that are consistent with changing the focus.	*H&TAI as Gem*: "[in PF] Given that Eugenio seems to apply rules only valid for whole numbers, the goal is to help him "see" the effect of operation with decimals. [followed by two suggested concrete strategies, one linguistic-based and the other visual-based]; [in SRF] to help him avoid repeating the error, you can guide him to build new checking strategies [followed by two concrete didactical proposals]" *H&TAI as GPT*: articulates PF in steps of process goals, each accompanied by a dedicated activity and an explanation; provide questions and activities for SR.
Active involvement of human actors (teacher and student)	AI coach asks questions, provides examples and encourages reflection. It bases interventions on information provided by human actors and requires this information.	Both *H&TAIs:* • explicitly wait for teacher responses and build on what it received. • ask teacher thought-provoking questions, encouraging the collection of evidence from students' work (e.g., specifying that a certain action must be done after the student's calculations).

Table 1. Key elements for the analysis

RESULTS

Regarding the analysis of the conversations of the two H&TAIs, both begin by asking for clarification about the learning goals and interpretations of the student's behavior, recalling the crucial need for contextualization and personalization of feedback (feed-up). As the conversation proceeds, both agents ask for explicit contributions from the teacher, valuing them by articulating reflection and suggestions accordingly, also with direct citations (feed-back and feed-forward). Concerning the different types and focuses for feedback (indicator 4), the two H&TAIs perform fine-tuned coaching with examples of possible practical interventions, offering insights while maintaining a dialogical turn talking with teachers. Nevertheless, they show some style differences. For instance, in SRF, both H&TAIs shed light on the importance of creating meaningful connections between results and language. However, GPT's version insists on suggesting predictive tasks, while the Gem's one tends to suggesting reformulation activities. Regarding the comparison between the conversation with H&TAI and with the normal chatbots, strong differences emerge already when comparing the coach's first response to the teacher's initial prompt. Both ChatGPT and Gemini basic chatbots suggest feedback declaring, more or less explicitly, that option A, the answer given by Eugenio, is not the correct one. For instance, Gemini says "You chose answer A) 10×0.5, but it is not the operation that produces the largest result.", while ChatGPT writes: "I see you chose option A) 10×0.5 because it looked like the biggest. Let's check it together, step by step, to see where decimal numbers can trick us." They then provide the full solution themselves, focusing solely on TF and sprinkling generic praise, leaving no room for the teacher's diagnostic questions or the student's reasoning. In short, the interaction turns into a mini-lecture rather than a coaching dialogue.

DISCUSSION AND CONCLUSION

In this paper, we propose a research-based approach aimed at designing AI agents for teacher training, exemplified by the development of H&TAI, a prototype AI feedback coach grounded in the model by Hattie and Timperley (2007). We first described how the agent's identity was defined through a configuration prompt and a domain-specific knowledge base. We conducted an internal validation of the agent's behavior by submitting a mathematical item without context, and by conducting a four-way simulation of a coaching session involving a teacher, a student, and four different AI coaches. These included two H&TAIs (implemented via GPTs and Gems) and two basic chatbots (ChatGPT and Gemini without configuration). The coaching dialogue centered on an INVALSI item designed to elicit a well-known mathematical misconception. We transcribed and analyzed the conversations using five indicators drawn from the Hattie and Timperley framework.

We conclude the paper by addressing the two research questions. Concerning RQ1, which asks how H&TAI supports teachers in providing effective feedback, the results suggest that H&TAI holds promise as both a feedback coach and a metacognitive tool for supporting teachers' reflective practices. This enables teachers to develop strategies for constructing, assessing, and refining feedback. Future work will involve a field evaluation with pre-service teachers and further exploration of how to balance pedagogical and disciplinary dimensions within the agent's identity and reasoning. Moreover, both the GPT's and Gem's AI agents behave consistently with the model; thus, the identity configuration appears robust and a promising starting point for further exploration. From this perspective, H&TAI's development aligns with a key research direction identified by Holmes and Tuomi (2022) and Pepin et al. (2025b): designing AI-based solutions for *assisting* teachers, rather than *replacing* their expertise. An interesting future development would be to have H&TAI process a teacher's initial feedback proposal to support a better alignment with the Hattie and Timperley (2007) framework. Concerning RQ2, which ass about differences between H&TAI's feedback and that provided by normal AI chatbots, the comparison reveals that basic chatbots behave more like lecturers, providing immediate solutions and vague praise while being detached from the learning needs of teachers and students.

This work represents just the beginning of a broader research agenda. Despite encouraging evidence that research-based AI agents can support teacher reflection and feedback practices, much remains to be explored. Future investigations might include training the system with real teacher-student feedback exchanges and enriching its reasoning with structured catalogs of misconceptions. It would also be interesting to recreate something similar to what was done by Schorcht et al. (2025) analyzing teachers' evaluation of H&TAI responses with respect to dimensions such as mathematical depth, language sensitivity, and competence orientation. These directions, among others, have the potential to deepen the agent's pedagogical insight and spark new lines of inquiry at the intersection of AI and mathematics education.

REFERENCES

Boero, P., Ferrari, P. L., & Ferrero, E. (1989). Division problems: Meanings and procedures in the transition to a written algorithm. *For the Learning of Mathematics*, *9*(3), 17-25.

Bolondi, G., Ferretti, F., & Gambini, A. (2018). Il database GESTINV delle prove standardizzate INVALSI: uno strumento per la ricerca. Alcuni esempi di utilizzo nell'ambito della matematica. In P. Falzetti (Ed.), *I dati Invalsi: uno strumento per la ricerca*. (pp. 43-48). FrancoAngeli Editore.

Cheng, E. C. (2025). Leveraging generative AI in science lesson study: transforming density concept instruction through ChatGPT integration. *International Journal for Lesson & Learning Studies.* https://doi.org/10.1108/IJLLS-11-2024-0277

Cheng, Y., Liu, W., Xu, K., Hou, W., Ouyang, Y., Tou Leong, C., Li, W., Wu, X., & Zheng, Y. (2024). Evolving to be your soulmate: Personalized dialogue agents with dynamically adapted personas. *arXiv e-prints*, arXiv-2406. https://arxiv.org/pdf/2406.13960

Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., Zhong, A., Hu, X., Liang, J., Yu, P.S., & Wen, Q. (2025). LLm agents for education: Advances and applications, https://arxiv.org/abs/2503.11733

Fujii, T. (2020). Misconceptions and alternative conceptions in mathematics education. In *Encyclopedia of mathematics education* (pp. 625-627). Springer, Dordrecht.

Goos, M. (2020). *Encyclopedia of Mathematics Education, chap. Mathematics Classroom Assessment*. Springer International Publishing.

Haskell, C. (2024). *Essential Guide to Prompting*. https://tinyurl.com/HASKELL24

Hattie, J., & Timperley, H. (2007) The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European journal of education*, *57*(4), 542-570.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277-1288.

Hughes, G., Smith, H., & Creese, B. (2015). Not seeing the wood for the trees: developing a feedback analysis tool to explore feed forward in modularised programmes. *Assessment & Evaluation in Higher Education*, *40*(8), 1079–1094. https://doi.org/10.1080/02602938.2014.969193

Kahn, K. (2025). *The Learner's Apprentice: AI and the Amplification of Human Creativity*. CMKPress.

Kluger, A.N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Lee, U., Lee, S., Koh, J., Jeong, Y., Jung, H., Byun, G., Lee, Y., Moon, J., Lim, J., & Kim, H. (2023) Generative agent for teacher training: Designing educational problem solving simulations with large language model-based agents for pre-service teachers. In *NeurIPS'23 Workshop on Generative AI for Education (GAIED)*. Advances in Neural Information Processing Systems, Curran Associates, Inc.

Maffia, A., & Mariotti, M. A. (2018). Intuitive and formal models of whole number multiplication: Relations and emerging structures. *For the Learning of Mathematics, 38*(3), 30-36.

MIUR (2012). *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione*. https://tinyurl.com/INDNAZ2012

Noster, N., Gerber, S., & Siller, H. S. (2024). Pre-service teachers' approaches in solving mathematics tasks with ChatGPT. *Digital Experiences in Mathematics Education*, *10*(3), 543-567. https://doi.org/10.1007/s40751-024-00155-8

Novotná, J., Moraová, H., Tatto, M.T. (2020). Mathematics Teacher Education Organization, Curriculum, and Outcomes. In: Lerman, S. (eds) *Encyclopedia of Mathematics Education*. Springer, Cham. https://doi.org/10.1007/978-3-030-15789-0_107

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025a). "Mathematics Education in the Era of ChatGPT: Investigating Its Meaning and Use for School and University Education"—Editorial to Special Issue. *Digital Experiences in Mathematics Education*, *11*(1), 1-8. https://doi.org/10.1007/s40751-025-00173-0

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025b). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 1-33. https://doi.org/10.1007/s40751-025-00172-1

Qian, Q., Huang, M., Zhao, H., Xu, J., & Zhu, X. (2017). Assigning personality/identity to a chatting machine for coherent conversation generation. *arXiv preprint* arXiv:1706.02861. https://arxiv.org/pdf/1706.02861

Ribeiro, M., Mellone, M., & Jakobsen, A. (2016). Interpreting students' non-standard reasoning: insights for mathematics teacher education. *For the learning of mathematics*, 36(2), 8-13.

Schorcht, S., Peters, F., & Kriegel, J. (2025). Communicative AI agents in mathematical task design: A qualitative study of GPT network acting as a multi-professional team. *Digital Experiences in Mathematics Education*, *11*(1), 77–113. https://doi.org/10.1007/s40751-024-00161-w

Valentini, M., Del Zozzo, A., Funghi, S., & Montresor, A. (2025) Hattie&TimperlAI: an AI-Based feedback coach for math teacher training. *Helmeto 2025 conference*.

Zan, R. (2007). *Difficoltà in matematica: osservare, interpretare, intervenire*. Springer.