From End-Users to Co-Designers: Lessons from Teachers

Martina Galletti
Sony Computer Science Laboratories - Paris , France
University of Rome "La Sapienza, Italy
martina.galletti@sony.com

Valeria Cesaroni
University of Perugia, Italy
valeria.cesaroni@dottorandi.unipg.it

Abstract

This study presents a teacher-centred evaluation of an AI-powered reading comprehension tool, developed to support learners with language-based difficulties for English and Italian. Drawing on the Social Acceptance of Technology (SAT) framework, we investigate technical usability and the pedagogical, ethical, and contextual dimensions of AI integration in classrooms. We explore how teachers perceive the platform's alignment with inclusive pedagogies, instructional workflows, and professional values through a mixed-methods approach, including questionnaires and focus groups with educators. Findings revealed a shift from initial curiosity to critical, practice-informed reflection, with trust, transparency, and adaptability emerging as central concerns. The study contributes a replicable evaluation framework and highlights the importance of engaging teachers as co-designers in developing educational technologies.

1 Introduction

As Natural Language Processing (NLP) continues to advance, its applications in education are expanding rapidly—from intelligent tutoring systems to automated writing feedback and reading support (Su et al., 2023; Özer, 2024; Zawacki-Richter et al., 2019). These AI-powered tools promise to transform instruction (Maity and Deroy, 2024), yet a key question remains: How do they perform in real classrooms with real teachers and students? Do they align with the practical realities and pedagogical expectations of educators, ensuring both usability and instructional relevance? (Cesaroni et al., 2024) Many systems are developed in controlled settings with limited educator input (Celik et al., 2022; Cukurova and Luckin, 2018; Luckin and Cukurova, 2019), often overlooking pedagogical realities and learning science principles (Luckin and Cukurova, 2019; Cesaroni et al., 2025).

Addressing this disconnect requires greater attention to the roles educators play, not just as passive users, but as active contributors throughout the AI development cycle. Indeed, teachers have already played various roles in educational AI research (Celik et al., 2022). They served as models for AI training through classroom data (Su et al., 2014; Kelly et al., 2018), shared professional development information to improve predictive systems (Alzahrani and Alzahrani, 2025; Yoo and Rho, 2020), and provided student data to support AI-driven interventions (Bonneton-Botté et al., 2020; Nikiforos et al., 2020). They have also validated AI outputs by grading work and defining evaluation criteria (Huang et al., 2010; Yuan et al., 2020), influenced pedagogical alignment through instructional material selection (Dalvean and Enkhbayar, 2018; Fitzgerald et al., 2015), and in some cases, offered technical feedback on system design (Burstein et al., 2004). Despite these contributions, their role as evaluators who shape AI integration in classroom contexts remains largely underexamined.

Building on this foundation, the paper presents an evaluation framework of an AI-powered reading comprehension interface using a framework that places educators at the centre of AI integration. Although applied to a single interface in this study, the framework is generalizable to the evaluation of AI technologies across diverse educational contexts. This framework draws on the SAT model to examine the pedagogical, ethical, and practical dimensions of AI adoption in education. Through a mixed-methods approach involving questionnaires and focus groups, we not only assess how teachers perceive the system but also explore how their insights can shape more effective, inclusive, and ethically grounded AI implementation in real classroom settings.

Our findings highlight the value of participatory design, showing that teachers act as co-designers and evaluators, not just users. Their acceptance of

Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications, pages 505–516

AI tools relies on alignment with pedagogical values, transparency, and autonomy. While they saw promise in promoting inclusion and differentiated instruction, they also pointed to needed improvements in clarity, layout, and customisation. These insights call for ethically grounded, teacher-centred approaches and further research through long-term classroom use, evaluation and broader educator involvement.

2 Background

2.1 Existing Reading Comprehension Interfaces

Reading comprehension interfaces aim to support users in understanding and engaging with complex textual material. Unlike general reading tools, these systems are designed to go beyond passive reading by incorporating interactive features such as question answering, summarisation, sentence simplification, and semantic annotations. Existing tools focus on general users and surface-level comprehension, lacking therapeutic intent, multilingual support, and personalisation.

One of the most notable efforts in this domain is the Semantic Reader Project (Lo et al., 2023), which augments scientific documents with context-aware explanations, definitions, and citation-level summaries to help readers quickly identify core ideas. Similarly, systems like SciReader (Head et al., 2021) employ semantic highlighting, definitions on hover, and automatic summarisation to assist users, particularly researchers, in navigating dense academic material. However, these tools are not tailored to students with reading comprehension deficits or learning disorders.

Another promising direction involves gaze-driven sentence simplification interfaces, such as the work of Higasa et al. (2023), which are particularly relevant for language learners or readers with cognitive impairments. These systems use real-time eye-tracking data to detect reading difficulty and apply NLP techniques to simplify complex sentences. However, while useful as assistive technologies, they do not provide structured activities aimed at rehabilitating underlying comprehension deficits.

Complementing these assistive tools are educational systems like 3D Readers (3dR) and CACSR (Kim et al., 2006), which take a more interactive and instructional approach to enhancing reading comprehension. 3D Readers allow users to engage with texts through either verbal strategies (such as question generation) or visual strategies (like manipulating images), with immediate feedback provided to support learning (Johnson-Glenberg, 2007). Similarly, CACSR offers personalised instruction using techniques like visual imagery, graphic organisers, mnemonics, self-questioning, and summarization (Stetter and Hughes, 2011), also incorporating real-time feedback to support continuous assessment (Kim et al., 2006). Despite their effectiveness in educational settings, these systems are not designed with therapeutic goals in mind.

Moreover, existing systems are designed for English-language users. There appears to be only one known system available in Italian that supports integrated telerehabilitation: RIDInet [1]. The platform offers activity modules like the Cloze Application, which trains reading comprehension through multiple-choice tasks. However, RIDInet does not offer targeted exercises for developing word-level literal understanding, nor does it support the integration of prior knowledge with new textual input.

2.2 Assessment of AI-Powered Educational Technologies

Evaluating AI-powered educational tools poses a methodological challenge due to the lack of frameworks integrating pedagogical, psychological, and social dimensions of technology adoption. Existing models, such as the Technology Acceptance Model (TAM) (Davis, 1989) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003) offer robust tools for analysing perceptions of usefulness, usability, and behavioural intention. However, these frameworks emphasise generic constructs (e.g., efficiency, ease of use), but often neglect education-specific factors such as alignment with instructional goals, teacher–student dynamics, and pedagogical adaptability.

The Technological Pedagogical Content Knowledge (TPACK) framework (Koehler and Mishra, 2009) addresses the integration of technology into pedagogy, but primarily with a formative intent. It delineates the competencies required to design learning experiences that effectively combine technological tools with pedagogical strategies and disciplinary knowledge. However, it lacks evaluative tools for real-world adoption, omitting con-

[1] https://www.anastasis.it/ridinet/

cerns like ethical implications, institutional fit, and teacher autonomy. For instance, a teacher may possess TPACK proficiency in using an NLP tool yet refrain from adopting it due to ethical concerns (e.g., algorithmic bias) or practical constraints (e.g., misalignment with classroom workflows) — factors that lie outside the scope of TPACK.

To address this gap, this study adopts a mixed-methods approach guided by the SAT framework (Occhipinti et al., 2023) to explore the use of AI tools in educational settings. Unlike traditional models that primarily assess individual user experience or usability, the SAT framework views technology as part of a broader socio-technical system. By focusing on four interrelated dimensions (User Experience, Value Impact, and Trust), SAT enables an assessment that extends beyond subjective usability to encompass ethical, cultural, and contextual factors. In the context of schools, the study uses SAT to design questionnaires and focus groups that examine how a software for the teaching of reading comprehension aligns with pedagogical values, affects teacher relationships and institutional structures, and impacts trust. Special emphasis is placed on the Value Impact and Trust dimensions, which help uncover educators' perspectives on issues like inclusion, transparency, autonomy, and coherence with teaching practices.

3 System Description

In this paper, we present a novel and enhanced version of ARTIS (Galletti et al., 2023, 2024). ARTIS is a web-based educational tool designed to support reading comprehension for primary school students, with a particular focus on learners with reading difficulties or language-based learning disorders up to 11 years old (Galletti et al., 2023). The system integrates a multimodal approach to text comprehension by combining visual, auditory, and interactive components. The interface supports multilingual content (Italian and English), making it adaptable for bilingual contexts or second-language learners. While some of the assistive features were previously introduced in Galletti et al. (2024), this version introduces new rehabilitative features, an enhanced administrative dashboard, as well as updated design and graphics.

The design of the platform's features is grounded in the psycholinguistic model of reading comprehension proposed by Kintsch and van Dijk (Kintsch and Van Dijk, 1978; Galletti et al., 2023). This model outlines three levels of text comprehension. First, there is surface representation, which involves recognizing words and grammar (i.e. lexical and morphosyntactic understanding). Second, there is propositional representation, where readers connect ideas into meaningful sequences and structures. Finally, there is the mental model construction, where readers combine what the text says with their background knowledge.

Following Kintsch and Van Dijk's model, our interface includes different modules and exercises targeting different comprehension levels: lexical understanding, propositional structuring, and mental model integration, each progressively supporting deeper text processing. In the next subsection, we describe each module and the algorithms behind its core functionalities [2].

3.1 Assistive Features

Upon logging into the platform, students can access a digital library interface that displays a collection of illustrated literary and informational texts. Each title is visually represented with a stylised image and a short textual excerpt to support browsing and engagement. Once a text is selected, the reading interface presents the full passage along with assistive features such as read-aloud audio, using the Google text-to-speech API, synchronized text highlighting, and pace controls (e.g., play, pause, and speed adjustment). These supports are designed to aid comprehension while still requiring the child to engage actively with the text.

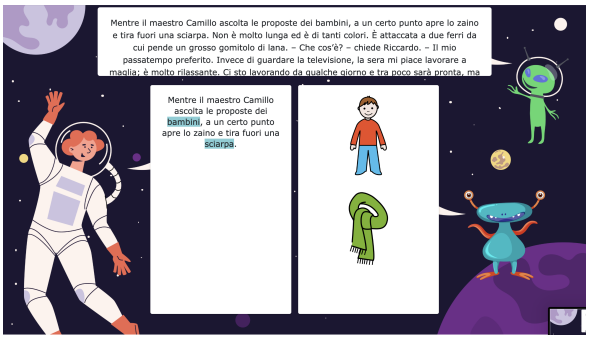

Figure 1: Keywords like *"bambini"* ("children") and *"sciarpa"* ("scarf") are highlighted and paired with pictograms to support understanding.

In a subsequent step, the text is presented sentence by sentence thanks to the spaCy Sentencizer [3] and key terms within the passage are visually highlighted and linked to pictograms that

[2] A recorded demonstration of our proof of concept is available at this link.

[3] https://spacy.io/api/sentencizer

illustrate their meaning, as in Figure 1. Secondly, keywords are extracted using a fine-tuned version of Keybert (Grootendorst, 2020). To ensure accuracy and prevent misleading outputs, the extracted keywords were manually reviewed by speech and language therapists as described in Galletti et al. (2023). Once the keywords were extracted from the sentences, after lemmatisation, we used the Arasaac API [4] to link them to pictograms.

In a third step, unfamiliar or specific terms are also supported with definitions and example sentences drawn directly from the source passage, as in Figure 2. These terms are selected either manually by the operator or automatically extracted as detailed in Galletti et al. (2023). Users can interact with the words to hear their definitions, view pictograms illustrating their meanings, and see them embedded in the text, supporting multisensory learning and strengthening decoding skills. For definitions, we used *gpt-3.5-turbo-0125*[5].

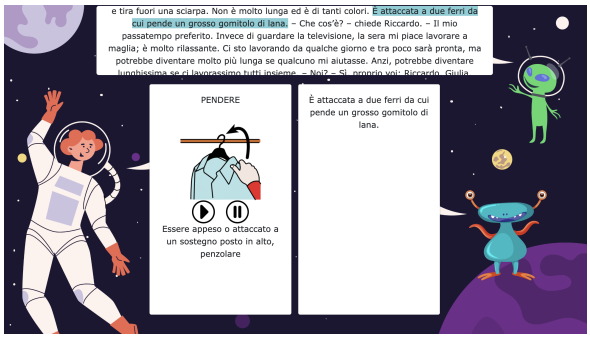

Figure 2: Vocabulary's support a not common Italian verb *"pendere"* ("to hang" or "to dangle"). The left panel shows the pictogram associated with it, a definition, and audio playback buttons to support its comprehension.

3.2 Rehabilitative Features

A variety of comprehension and language-focused exercises are included to deepen semantic processing and support inference-making skills. These activities comprise: (1) *"Leggi e rispondi"* ("Read and Respond"), where students answer comprehension questions generated by *gpt-3.5-turbo-0125* and manually validated by speech and language therapists; and (2) *"Trova le parole chiave"* ("Find the Keywords"), which engages learners in identifying key terms within the text. Keywords are generated using a fine-tuned version of KeyBert, as described in Galletti et al. (2023); and (3) *"Trova la rete semantica"* ("Build the Semantic Network"), which prompts students to connect words with semantically related concepts, as illustrated in Figure 3. Both the related terms and distractors for this task were generated using `gpt-3.5-turbo-0125`. To generate related and unrelated words, we used two prompts: one asked for a list of synonyms with definitions, and the other for non-synonyms that are semantically unrelated, both formatted as JSON arrays with appropriate keys.

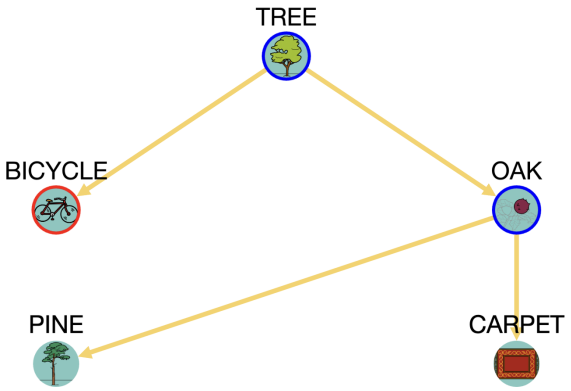

Figure 3: Example of a semantic network exercise. The user starts from the word "TREE" and must choose which of the two presented options is semantically related. If the correct option is selected, the network expands and presents two new options.

3.3 Administrative Dashboard

Finally, the interface includes an administrative dashboard that gives therapists full control over all aspects of the content generated by the AI algorithms. This dashboard allows for the management of texts, such as editing existing content, inserting new texts, or regenerating AI-related components, as well as the organisation and customisation of exercises. When a new text is added, the dashboard displays a preview of each AI-generated component—such as keywords, pictograms, sentences, and questions—for validation. This enables the educator to verify or adjust these associations before the material is presented to the learner. Additionally, the dashboard supports the enrollment of students and the assignment of personalised texts, avatars and exercises, helping to tailor the learning experience to each user. It also collects the data and makes it available for downloading to the therapist.

[4] https://arasaac.org/

[5] https://platform.openai.com/docs/models/gpt-3-5-turbo

Dimension	Description
Pedagogical Appropriateness	Evaluates alignment with inclusive and AAC (Augmentative and Alternative Communication)-based pedagogy, focusing on scaffolding, shared meaning-making, and student autonomy.
Inclusive Potential	Assesses support for diverse learners and compatibility with Universal Design for Learning (UDL) and cooperative learning strategies.
Teacher Readiness	Explores how hands-on experience influenced openness to AI, highlighting competence gaps and training needs.
Trust	Investigates confidence in AI-driven features and the balance between automation and teacher control, including transparency and ethical oversight.
Expectation Shift	Compares pre- and post-use attitudes to identify shifts in teachers' perceptions of AI in education.

Table 1: The five key dimensions which guided the focus group.

4 Methods

The evaluation framework is a two-phase approach: a questionnaire which assessed teachers' general attitudes and readiness toward AI, followed by a focus group conducted after hands-on interaction with the platform. This structure allowed us to compare abstract views of AI with teachers' hands-on experiences, highlighting how their perceptions align with pedagogical values, ethical concerns, and practical adoption barriers. This section outlines the questionnaire and focus group design; results are presented in Section 5 followed by their discussion in Section 6.

4.1 Questionnaire's Design

The questionnaire explored teachers' knowledge, perceptions, and attitudes toward digital and AI technologies in education. Its structure follows the four dimensions of the SAT framework: user experience, social disruptiveness, value alignment, and trust. As part of our contribution, the questionnaire is openly available to the community at this link, while the detailed results are available here.

Following an initial section collecting teacher details such as years of experience and subject area, Section 2 explored teachers' general approach to technology use. It included items on comfort with technology in personal and professional contexts, habits for staying updated on tools, and frequency of using digital resources for planning. Participants also rated the importance of technology in teaching and their interest in new tools. The section ended with a multiple-choice item on perceived barriers to tech integration, based on established research (Ertmer et al., 2012).

Section 3 focused on teachers' awareness and use of digital tools designed to support reading comprehension, drawing on research into the educational technology ecosystem (Tondeur et al., 2017). Participants first have to identify any relevant software or platforms they know, such as simplified reading tools, text-to-speech apps, or concept mapping software. They then need to indicate which they had used in teaching or planning, and report any reading comprehension technologies available at their schools.

Section 4 explores teachers' perceptions of AI in education, assessing their knowledge, expectations, trust, ethical concerns, and professional agency—defined as the capacity to shape one's practice within institutional and technological constraints (Biesta et al., 2015; Toom et al., 2015). A mixed-format design with Likert-scale and open-ended items enables a mixed-methods analysis, aligning with best practices in educational technology research (Ponce and Pagán-Maldonado, 2015). This last section is divided into three subsection focusing respectively on (A) *Trust and Risk/Benefit Perception* - using items adapted from the Propensity to Trust in AI scale (Mcknight et al., 2011), (B) *Ethical Awareness*, drew on critical AI literacy frameworks (Veldhuis et al., 2024) and (C) *Teacher Agency and Involvement*, with items informed by the TPACK framework (Mishra and Koehler, 2006) and research on teacher agency (Leijen et al., 2024). This subsection examines the perception of students' interest in AI, teachers' views on the importance of ethical and pedagogical training, and expectations for future integration and involvement in AI-related decisions.

4.2 Focus Group's Design

The focus group was intentionally designed to capture authentic, practice-informed insights from educators by combining experiential use of the platform with structured group reflection. Following the initial questionnaire, participating teachers engaged in a one-hour, hands-on session with the ARTIS platform. To preserve the ecological validity of the study, no prior exposure or formal training was provided. Instead, participants received minimal onboarding and quick-start instructions, allowing for natural, intuitive engagement with the interface.

The exploratory session encouraged teachers to

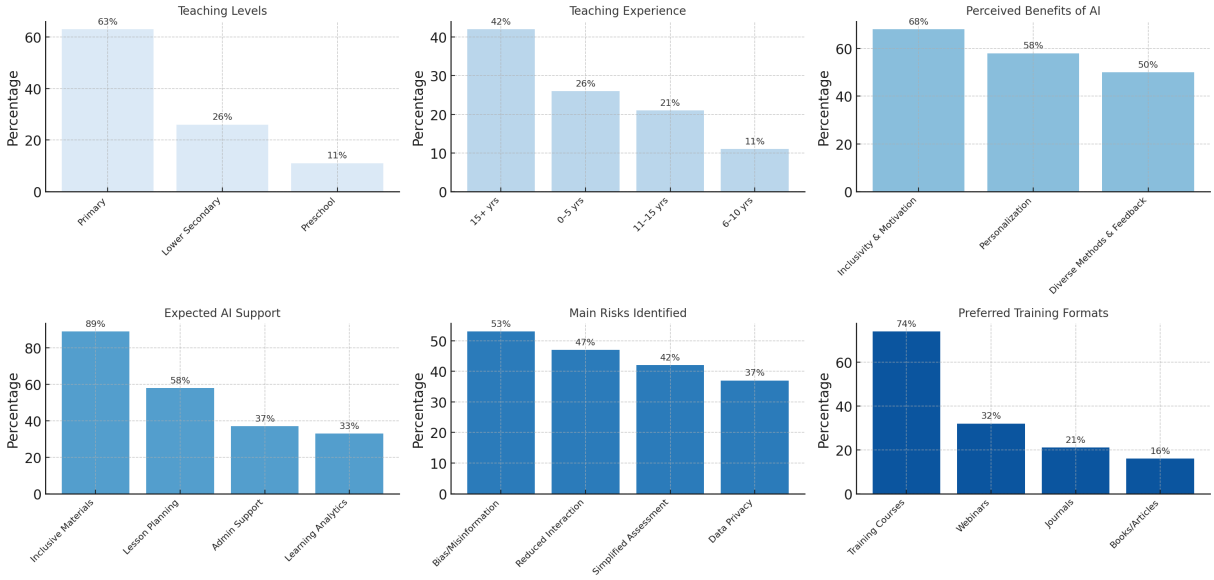

Figure 4: Summary of questionnaire responses from 19 teachers on AI in education, including teaching demographics, perceived benefits and risks, expected support, and preferred training formats.

freely navigate the platform, selecting from a variety of texts. Participants were asked to engage with materials spanning multiple educational levels to ensure a range of instructional contexts were represented. Additionally, a subset of four teachers was invited to interact with English-language texts, enabling the evaluation of bilingual and second-language accessibility features.

Immediately after the interaction phase, a structured focus group gathered in-depth reflections on five key dimensions of AI acceptance and pedagogical fit : (I) *Pedagogical Appropriateness*, (II) *Inclusive Potential*, (III) *Teacher Readiness*, (IV) *Trust*, and (V) *Reconfiguration of Expectations* —based on sociotechnical, ethical, and inclusive education frameworks and guided the analysis of teacher responses. Precise details on this key dimension can be found in Table 1.

5 Results

We recruited 19 teachers from the *Istituto Comprensivo di Narni Scalo (Italy)* to participate in this study. The sample included 12 primary school teachers, 5 lower secondary teachers, and 2 preschool educators. Although the ARTIS platform is primarily designed for literacy development and may have limited direct applicability in early childhood education, preschool teachers were intentionally included to examine how attitudes toward AI integration vary across educational levels. This inclusive approach captured diverse perspectives and enabled comparison of teachers' readiness for AI across contexts.

5.1 Questionnaire Results

When asked about digital tools used to support reading comprehension, teachers cited a mix of general productivity platforms, such as Google Workspace and Canva, alongside more specialized tools like Genially, Popplet, and Arasaac for augmentative and alternative communication. A few teachers had explored newer AI-powered tools such as NotebookLM and Napkin.AI. Participants were optimistic about AI's educational benefits, linking it to inclusivity, student motivation, personalized learning, diverse instructional formats, and instant feedback. Teachers saw AI as useful for creating inclusive materials, supporting planning and content design, reducing administrative tasks, and improving the use of learning analytics to guide instruction.

At the same time, teachers were mindful of the risks associated with AI in education. Concerns included the possibility of algorithmic bias, misinformation, and diminished teacher autonomy. Others raised ethical issues such as the reduction of student-teacher interaction, oversimplified forms of assessment, and the potential misuse of student data. In terms of professional development, participants showed a clear preference for structured training opportunities focused on the ethical and social dimensions of AI. Webinars, journals, and

other professional resources were also seen as helpful, though less commonly preferred. Overall, the findings suggested a cautiously optimistic attitude among teachers. While they see promise in the pedagogical affordances of AI, especially about inclusion and personalization, they also emphasize the need for thoughtful implementation, transparency, and support through well-designed training. Aggregated results are shown in Figure 4 and detailed results are available here.

5.2 Focus group results

The following paragraphs present findings from the focus group discussions, organized according to the five evaluative dimensions illustrated in Table 1.

Pedagogical Appropriateness Participants raised thoughtful concerns about the platform's alignment with inclusive pedagogical practices. A key point of critique centered on the sequencing of support features: currently, visual and linguistic scaffolds, such as keywords and pictograms, are presented before students actively engage with the text through the extercise. While well-intentioned, this design was perceived by many as potentially limiting student autonomy and interpretive effort. Several educators proposed reversing this order, suggesting that scaffolds introduced during or after initial engagement would better support active meaning-making.

Inclusive Potential Teachers generally recognized the platform's value in supporting differentiated instruction, especially for students with language-based or cognitive challenges. However, concerns were raised about the semantic precision and clarity of the pictograms, as well as the overall visual layout, both of which were seen as crucial to accessibility. Importantly, educators emphasized that the platform should not replace teacher-student interaction but instead enhance it, particularly through collaborative practices like co-selecting keywords and interpreting texts.

Teacher Readiness While initial questionnaire responses reflected a generally positive orientation toward AI in education, the post-use discussions revealed more grounded, experience-based perspectives. Teachers emphasized the importance of training that goes beyond technical operation to include pedagogical integration. They acknowledged the platform's potential to support differentiated learning and streamline resources, but stressed that its success would depend on its adaptability to real classroom contexts and its alignment with established instructional workflows.

Trust Trust in the platform emerged as closely tied to the degree of teacher agency and system transparency. However, this did not translate into a blanket rejection of the technology. Rather, educators identified specific areas for improvement, calling for enhanced user control and clearer communication about how AI-driven choices are made.

Expectations Shift Educators moved from abstract curiosity and cautious optimism to a more critical, practice-informed perspective. Their experiences prompted a clear set of priorities for the future development of AI in education: (1) Flexibility over rigidity – AI tools must be adaptable to diverse classroom contexts; (2) Transparency over opacity – teachers need to understand and shape how AI-driven decisions are made; (3) Support over substitution – technology should amplify, not replace, human interaction and pedagogical creativity. While initial enthusiasm was tempered by practical limitations, participants remained confident in the potential of AI-supported learning environments—particularly when such tools are designed to complement teacher expertise and foster meaningful student engagement. Importantly, the findings underscore the value of involving educators not merely as users, but as co-designers and evaluators in the development process.

6 Discussion

This study contributes to a growing body of research on the integration of AI in education by offering a practice-informed, teacher-centered perspective grounded in the SAT framework (Occhipinti et al., 2023). Our findings show how teachers' acceptance of AI tools is not static nor solely based on usability, but shaped dynamically through hands-on engagement, educational values, and pedagogical alignment. Methodologically, this study aligns with recent calls for participatory and iterative approaches to AI design in education (Luckin and Cukurova, 2019; Mouta et al., 2024).

Rather than viewing acceptance as a fixed variable to be assessed retrospectively (Celik et al., 2022), our approach positions teachers as formative agents, whose experiences, critiques, and creativity are integral to the ethical and effective development of technology. In line with Zawacki-Richter

et al. (2019), who underline the scarcity of qualitative studies that capture educators' voices in AI research, our work emphasizes the importance of interpretive approaches that explore how teachers make sense of AI tools in concrete pedagogical settings. By combining questionnaires with in-depth focus groups, we were able to reveal not only general trends in acceptance but also the nuanced ways in which teachers negotiated the role of AI in their practice. Specifically, participants moved from general skepticism to targeted suggestions, such as reversing scaffold sequencing or refining pictogram clarity, demonstrating a shift from rejection to co-design.

While 89% of teachers viewed AI as potentially beneficial for inclusive education (questionnaire), the focus group exposed significant caveats. Teachers emphasized that inclusion cannot be achieved through technical affordances alone but requires alignment with pedagogical routines and accessibility standards. These findings echo critiques of UDL frameworks when implemented in a top-down, compliance-oriented manner (Edyburn, 2010). Instead, participants advocated for adaptive interfaces, customizable visuals, and collaborative practices, such as co-selection of keywords, that preserve student-teacher interaction and interpretive autonomy. This underscores the need to reconceptualize inclusivity as a dynamic co-construction rather than a static feature.

Our findings also nuance assumptions in the literature about AI adoption motives. Their insistence on retaining control over scaffolding and keyword selection reflects a broader commitment to maintaining instructional intentionality. Similarly, the shift from interest in formal AI training (74%) to a focus on pedagogical and ethical guidance suggests that professional development should go beyond technical skills to include critical and ethical perspectives (Perrotta and Selwyn, 2020). By engaging teachers not as end-users but as evaluators and co-evaluators, this study contributes a replicable model for socio-technical evaluation and advances the debate on how educational technologies can be made aligned with the realities of classroom practice.

Finally, these insights are also shaped by the cultural and institutional context. The participating teachers, embedded within the Italian education system, approached AI-mediated feedback through pedagogical norms distinct from those observed in studies conducted in other countries. For example, while Anglo-American frameworks often emphasize data-driven personalization and performance metrics (Shum and Luckin, 2019; Selwyn, 2019), Italian educators tended to prioritize dialogic, relational approaches to learning and a strong emphasis on formative assessment as a collective rather than individualistic practice (Moretti et al., 2015; Pastore, 2020). Teachers highlighted the importance of maintaining pedagogical intentionality and student-teacher interaction, reflecting a broader educational tradition in Italy that values interpretive autonomy and humanistic principles (Viteritti, 2009). These values shaped how teachers perceived AI tools—not simply as assistive technologies, but as agents that must harmonize with existing curricular structures, ethical responsibilities, and institutional logics. This cultural lens helps explain why some technological affordances, such as automated scaffolding or visual simplifications, were met with ambivalence unless they could be flexibly adapted to local pedagogical aims. Cross-cultural comparisons are thus essential to avoid universalist assumptions in AI design and to ensure that integration strategies remain sensitive to educational diversity (Zhang, 2025).

6.1 Actionable steps

The findings confirm that when engaged early in the development process, teacher expertise plays a pivotal role in surfacing abstract concerns and translating them into actionable design feedback (UNESCO, 2021). Rather than perceiving resistance to AI as rooted in negative attitudes, the study highlights the value of participatory engagement, where teachers act as co-designers. In particular, initial concerns centred around algorithmic opacity and perceived lack of control (Mcknight et al., 2011), mirroring broader critiques of AI as black-box systems that conceal decision-making logic (Burrell, 2016). Participants consistently called for increased transparency, interpretability, and human oversight—features that contribute to what is often termed "calibrated trust" (Zhang et al., 2020), where users remain critically engaged while feeling empowered to understand and shape system outcomes.

The next phase of interface development started to operationalize the teachers' inputs into concrete design interventions. These included (I) co-designed interface modifications with customizable elements (e.g., reversible scaffold sequencing, teacher-defined, configurable visual aids), (2) im-

plementing adjustable transparency layers (e.g., explainable feedback rationales), (3) develop teacher-facing toggles for AI assistance, allowing instructors to choose when and how AI contributes during a session (e.g., real-time suggestion, post-activity reflection), (4) include a cultural/contextual alignment layer in system design documentation: capture assumptions embedded in educational norms (e.g., Italian vs. others), dsign localized variants where necessary and plan for comparative studies across national systems to test transferability.

7 Conclusions & Future Work

This study is among the first to systematically evaluate the ethical and pedagogical acceptability of AI in real educational contexts using a framework explicitly oriented toward sociotechnical reflection. It highlights the importance of involving teachers not just as end-users, but as active co-designers and evaluators—positioning them as key contributors in shaping how AI tools are developed and integrated into educational settings. Rather than viewing acceptance as a fixed variable to be assessed retrospectively (Celik et al., 2022), our approach positions teachers as formative agents, whose experiences, critiques, and creativity are integral to the ethical and effective development of technology.

Using the SAT framework, we captured nuanced perceptions of an AI-supported reading platform, revealing that teachers' acceptance depends on more than functionality: it is conditional on alignment with pedagogical values, transparency, and support for professional autonomy. In sum, the question is not only whether AI works, but whether it works with and for teachers, in alignment with the values and practices that define education. The shift from abstract optimism to context-sensitive critique underscores the importance of participatory, ethically grounded approaches to AI design.

Future work should also explore long-term classroom use to track evolving practices, involve a more diverse sample of educators, in a second school, and examine how training and co-design processes influence ethical and pedagogical alignment of AI technologies, further validating the SAT model for educational settings. Moreover, insights from the current evaluation framework will inform the next development cycle of the ARTIS interface, ensuring that future iterations are responsive to teacher feedback.

Limitations

While the study offers valuable insights and contributes meaningfully to the design of educational AI tools, certain limitations also highlight important avenues for future exploration. The teacher sample, though rich in contextual relevance, was relatively small and specific, which may affect the broader applicability of the findings. However, this focused scope allowed for in-depth engagement and formative feedback that can directly inform future iterations. Participants interacted with a prototype version of ARTIS, and some of their observations likely reflect temporary interface or usability elements rather than underlying structural challenges, providing useful direction for refinement.

Moreover, the study's temporal scope was limited to a single session, offering a snapshot of initial impressions rather than longitudinal insights. Nonetheless, this approach effectively captured early responses and surfaced key priorities for longer-term implementation. Similarly, although emerging needs for teacher training were identified during the focus group, the study did not incorporate formal training programs. This opens a promising path for future research to investigate how targeted support mechanisms influence adoption and pedagogical integration over time.

An additional consideration relates to the use of the SAT framework. While SAT served as a valuable and critically-informed structure for guiding the study, it remains a relatively new model, especially in educational contexts. In this study, we adapted a combination of existing scales and custom-developed items to reflect the SAT's four dimensions. While this tailoring ensured contextual relevance, it may reduce replication and comparability across studies. Advancing this work will require the development and validation of standardized SAT-based instruments to foster wider methodological consistency.

Despite these limitations, the study yields important design implications. Educational AI tools should enable flexibility and personalization, promote transparency to build trust, and support—rather than supplant—educators' pedagogical creativity.

References

3D Readers Software. https://www.sbir.gov/node/334161. Accessed: 10/04/2025.

Abdulaziz Alzahrani and Amal Alzahrani. 2025. Comprendiendo la adopción de chatgpt en universidades: el impacto del tpack y utaut2 en los docentes understanding chatgpt adoption in universities: the impact of faculty tpack and utaut2. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(1).

Gert Biesta, Mark Priestley, and Sarah Robinson. 2015. The role of beliefs in teacher agency. *Teachers and teaching*, 21(6):624–640.

N. Bonneton-Botté, G. Beucher, and F. Ollivier. 2020. Can tablet apps support the learning of handwriting? an investigation of learning outcomes in kindergarten classroom. *Computers & Education*, 144:103706.

Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society*, 3(1):2053951715622512.

J. Burstein, M. Chodorow, and C. Leacock. 2004. Automated essay evaluation: The criterion online writing service. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 23–26. Association for Computational Linguistics.

Ismail Celik, Muhterem Dindar, Hanni Muukkonen, and Sanna Järvelä. 2022. The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66(4):616–630.

Valeria Cesaroni, Martina Galletti, Eleonora Pasqua, Daniele Nardi, and 1 others. 2024. Towards trustworthy ai in inclusive education: A co-creation approach rooted in ecological frameworks. In *Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI,*.

Valeria Cesaroni, Eleonora Pasqua, Piercosma Bisconti, and Martina Galletti. 2025. A participatory strategy for ai ethics in education and rehabilitation grounded in the capability approach. *arXiv preprint arXiv:2505.15466*.

Mutlu Cukurova and Rose Luckin. 2018. Measuring the impact of emerging technologies in education: A pragmatic approach.

Michael Dalvean and Galbadrakh Enkhbayar. 2018. Assessing the readability of fiction: A corpus analysis and readability ranking of 200 english fiction texts. *Linguistic Research*, 35:137–170.

Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340.

Dave L Edyburn. 2010. Would you recognize universal design for learning if you saw it? ten propositions for new directions for the second decade of udl. *Learning Disability Quarterly*, 33(1):33–41.

Peggy A Ertmer, Anne T Ottenbreit-Leftwich, Olgun Sadik, Emine Sendurur, and Polat Sendurur. 2012. Teacher beliefs and technology integration practices: A critical relationship. *Computers & education*, 59(2):423–435.

Jill Fitzgerald, Jeff Elmore, Heather Koons, Elfrieda H Hiebert, Kimberly Bowen, Eleanor E Sanford-Moore, and A Jackson Stenner. 2015. Important text characteristics for early-grades text complexity. *Journal of Educational Psychology*, 107(1):4.

Martina Galletti, Eleonora Pasqua, Francesca Bianchi, Manuela Calanca, Francesca Padovani, Daniele Nardi, and Donatella Tomaiuoli. 2023. A reading comprehension interface for students with learning disorders. In *Companion Publication of the 25th International Conference on Multimodal Interaction*, pages 282–287.

Martina Galletti, Eleonora Pasqua, Manuela Calanca, Caterina Marchesi, Donatella Tomaiuoli, Daniele Nardi, and 1 others. 2024. Artis: a digital interface to promote the rehabiliatation of text comprehension difficulties through artificial intelligence. In *Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy*.

Maarten Grootendorst. 2020. Keybert: Minimal keyword extraction with bert.

Andrew Head, Kyle Lo, Waleed Ammar, and 1 others. 2021. Augmenting scientific papers with just-in-time, position-sensitive definitions of terms and symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

Taichi Higasa, Keitaro Tanaka, Qi Feng, and Shigeo Morishima. 2023. Gaze-driven sentence simplification for language learners: Enhancing comprehension and readability. *arXiv preprint arXiv:2310.00355*.

R. H. Huang, J. M. Spector, and J. F. Yang. 2010. Ict literacy and the development of digital learning resources. *Educational Technology Research and Development*, 58(2):191–204.

Mina C Johnson-Glenberg. 2007. *Web-based reading comprehension instruction: Three studies of 3D-readers*. Erlbaum.

Sean Kelly, Andrew M Olney, Patrick Donnelly, Martin Nystrand, and Sidney K D'Mello. 2018. Automatically measuring question authenticity in real-world classrooms. *Educational Researcher*, 47(7):451–464.

Ae-Hwa Kim, Sharon Vaughn, Janette K Klingner, Althea L Woodruff, Colleen Klein Reutebuch, and Kamiar Kouzekanani. 2006. Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. *Remedial and special education*, 27(4):235–249.

Walter Kintsch and Teun A Van Dijk. 1978. Toward a model of text comprehension and production. *Psychological review*, 85(5):363.

Matthew Koehler and Punya Mishra. 2009. What is technological pedagogical content knowledge (tpack)? *Contemporary issues in technology and teacher education*, 9(1):60–70.

Äli Leijen, Margus Pedaste, and Aleksandar Baucal. 2024. A new psychometrically validated questionnaire for assessing teacher agency in eight dimensions across pre-service and in-service teachers. In *Frontiers in Education*, volume 9, page 1336401. Frontiers Media SA.

Kyle Lo, Joseph Chee Chang, Andrew Head, and 1 others. 2023. The semantic reader project: Augmenting scholarly documents through ai-powered interactive reading interfaces. *arXiv preprint arXiv:2303.14334*.

Rosemary Luckin and Mutlu Cukurova. 2019. Designing educational technologies in the age of ai: A learning sciences-driven approach. *British Journal of Educational Technology*, 50(6):2824–2838.

Subhankar Maity and Aniket Deroy. 2024. Generative ai and its impact on personalized intelligent tutoring systems. *arXiv preprint arXiv:2410.10650*.

D Harrison Mcknight, Michelle Carter, Jason Bennett Thatcher, and Paul F Clay. 2011. Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS)*, 2(2):1–25.

Punya Mishra and Matthew J Koehler. 2006. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers college record*, 108(6):1017–1054.

Giovanni Moretti, Arianna Giuliani, and ARIANNA LODOVICA Morini. 2015. Flexible and dialogic instructional strategies and formative feedback: an observational research on the efficacy of assessment practices in italian high schools. In *ICERI2015 Proceedings*, pages 8229–8236. IATED.

Ana Mouta, Ana María Pinto-Llorente, and Eva María Torrecilla-Sánchez. 2024. Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 34(3):1166–1205.

S. Nikiforos, S. Tzanavaris, and K. L. Kermanidis. 2020. Automatic detection of learners' aggressive behavior in a virtual learning community using machine learning techniques. *Education and Information Technologies*, 25(4):3293–3313.

Carmela Occhipinti, Antonio Carnevale, Luigi Briguglio, Andrea Iannone, and Piercosma Bisconti. 2023. Sat: a methodology to assess the social acceptance of innovative ai-based technologies. *Journal of Information, Communication and Ethics in Society*, 21(1):94–111.

Mahmut Özer. 2024. Potential benefits and risks of artificial intelligence in education. *Bartın University Journal of Faculty of Education*, 13(2):232–244.

Serafina Pastore. 2020. How do italian teacher trainees conceive assessment?. *International Journal of Instruction*, 13(4):215–230.

Carlo Perrotta and Neil Selwyn. 2020. Deep learning goes to school: Toward a relational understanding of ai in education. *Learning, Media and Technology*, 45(3):251–269.

Omar A Ponce and Nellie Pagán-Maldonado. 2015. Mixed methods research in education: Capturing the complexity of the profession. *International journal of educational excellence*, 1(1):111–135.

Neil Selwyn. 2019. *Should robots replace teachers?: AI and the future of education*. John Wiley & Sons.

SJ Buckingham Shum and Rosemary Luckin. 2019. Learning analytics and ai: Politics, pedagogy and practices. *British journal of educational technology*, 50(6):2785–2793.

Maria Earman Stetter and Marie Tejero Hughes. 2011. Computer assisted instruction to promote comprehension in students with learning disabilities. *International Journal of Special Education*, 26(1):88–100.

Jiahong Su, Davy Tsz Kit Ng, and Samuel Kai Wah Chu. 2023. Artificial intelligence (ai) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, 4:100124.

Yen-Ning Su, Chia-Cheng Hsu, Hsin-Chin Chen, Kuo-Kuang Huang, and Yueh-Min Huang. 2014. Developing a sensor-based learning concentration detection system. *Engineering Computations*, 31(2):216–230.

Jo Tondeur, Johan Van Braak, Peggy A Ertmer, and Anne Ottenbreit-Leftwich. 2017. Understanding the relationship between teachers' pedagogical beliefs and technology use in education: A systematic review of qualitative evidence. *Educational technology research and development*, 65:555–575.

Auli Toom, Kirsi Pyhältö, and Frances O'Connell Rust. 2015. Teachers' professional agency in contradictory times. *Teachers and teaching*, 21(6):615–623.

UNESCO. 2021. Ai and education: Guidance for policy-makers.

Annemiek Veldhuis, Priscilla Y Lo, Sadhbh Kenny, and Alissa N Antle. 2024. Critical artificial intelligence literacy: A scoping review and framework synthesis. *International Journal of Child-Computer Interaction*, page 100708.

Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478.

Assunta Viteritti. 2009. A cinderella or a princess? the italian school between practices and reforms. *Italian Journal of Sociology of Education*, 1(Italian Journal of Sociology of Education 1/3):10–32.

Jin Eun Yoo and Minjeong Rho. 2020. Exploration of predictors for korean teacher job satisfaction via a machine learning technique, group mnet. *Frontiers in psychology*, 11:441.

H. Yuan, Y. Wang, and Y. Liu. 2020. Automated essay scoring based on two-stage learning. *IEEE Access*, 8:21906–21915.

Olaf Zawacki-Richter, Victoria I Marín, Melissa Bond, and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International journal of educational technology in higher education*, 16(1):1–27.

Li Zhang. 2025. Educational technology in a cross-cultural perspective: Applications and challenges of generative ai tools in language education in sino-foreign cooperative programs. *Journal of Humanities, Arts and Social Science*, 9(3).

Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 295–305.