D1.1: Needs, Acceptance and Readiness Report

Author(s):	Siem Buseyne, Glen Joris, Konstantinos Tsiakas, Tine van Daal, Leen Catrysse, Vincent Donche, Sven De Maeyer, Halszka Jarodzka, and David Gijbels
Editor(s):	Siem Buseyne and David Gijbels
Responsible Organisation:	Universiteit Antwerpen (UANTWERPEN)
Version-Status:	V1
Submission date:	31/03/2026
Dissemination level:	Public

Funded by
the European Union

Table of Contents

Note on Publication Status

Chapters 1, 2, 4, and 5 of this deliverable are based on research outputs that have been prepared as journal manuscripts and submitted for peer review. Upon acceptance, the corresponding journal articles will serve as the final, citable versions. The versions included in this deliverable represent pre-publication forms and may undergo revisions following the peer-review process.

Executive Summary

This report presents Deliverable 1.1 of the EYE-TEACH project and aims to provide an overview of teachers' needs, acceptance, and preparedness regarding AI-supported educational technologies that use behavioral data, including eye-tracking, to support reading comprehension and instructional decision-making. To address this aim, the report combines evidence from a mix of complementary research methods, including a systematic/scoping literature review, a qualitative interview study, and large-scale quantitative data collection based on survey and vignette methods. Together, these approaches offer both an in-depth and broader understanding of how teachers perceive such technologies, what conditions shape their acceptance, and to what extent they feel prepared to work with them, thereby directly addressing Objective 1.1 of the EYE-TEACH project (i.e., identifying teachers' needs, acceptance, and readiness regarding AI-assisted ET-analytics tools across relevant use cases).

Chapter 1 presents the scoping review, which maps the existing literature on teachers' perceptions of behavioral data-driven AI-supported educational technologies. The review shows that this body of research is still relatively limited and methodologically underdeveloped, with many studies relying on small-scale and mainly qualitative designs. At the same time, the chapter identifies several recurring themes that appear to shape teachers' perceptions and adoption of such systems, including trust, transparency, reliability, pedagogical usefulness, workload, and ethical concerns such as privacy and fairness. The review thereby sets the scene for the empirical work in the following chapters by identifying both the current state of knowledge and the main issues requiring further investigation.

Chapter 2 reports the qualitative interview study, in which teachers reflected on AI-supported eye-tracking applications for reading. Using a qualitative design, this chapter explores in greater depth how teachers perceive the opportunities, challenges, and conditions for using such technologies in practice. The findings show that teachers are interested in tools that can support differentiation, provide insight into students' reading processes, and foster reflection, but they also emphasize that these tools should support rather than replace teacher judgment. The chapter further highlights the importance of transparency, human oversight, responsible data use, and the need for teachers to develop not only technological knowledge, but also pedagogical, content-related, and ethical competencies.

Chapter 3 provides an overview of the study design that was used in the large-scale quantitative part of the project. This chapter explains how different configurations of teacher–AI collaboration were operationalized through vignettes that varied in system autonomy. It clarifies the methodological basis for the quantitative analyses reported in the next chapters and shows how the project examined teachers' responses to different forms of AI support, ranging from more teacher-controlled to more autonomous systems.

Chapter 4 presents findings from the large-scale quantitative data collection on teachers' preparedness to work with AI-supported educational technologies. Using survey data, this chapter examines differences in teachers' self-reported AI-related preparedness. The results indicate that preparedness is unevenly distributed. Teachers in more technology-oriented professional roles reported higher preparedness, while some differences also emerged by gender and country. By contrast, age, teaching experience, and educational level did not appear to show consistent unique effects when other variables were taken into account.

Chapter 5 reports the results of the vignette-based quantitative analyses on teachers' acceptance of systems with different levels of autonomy. The findings indicate that more automation does not automatically lead to more positive evaluations. Teachers did not respond most positively to the most autonomous systems. Instead, they showed the strongest preference and trust for configurations in which the system offered limited, teacher-controlled support. These findings suggest that teachers value AI-supported educational technologies most when they preserve teacher agency, pedagogical oversight, and interpretive control.

Finally, Chapter 6 integrates the findings across the preceding chapters and discusses their broader implications for the EYE-TEACH project. Taken together, the findings show that teachers are generally open to AI-supported educational technologies, particularly when these are designed according to a human-centered and hybrid-intelligence logic. Successful implementation depends not only on technical functionality, but also on trust, explainability, usability, ethical safeguards, and alignment with classroom practice. The most promising path for EYE-TEACH therefore lies not in replacing teachers with increasingly autonomous systems, but in developing teacher-centered AI support tools that are transparent, reliable, pedagogically meaningful, and aligned with teachers’ professional role.

1. Scoping Review: Setting the Scene

Authors: Siem Buseyne, Glen Joris, Konstantinos Tsiakas, Tine van Daal, Vincent Donche, Sven De Maeyer, Leen Catrysse, Halszka Jarodzka, and David Gijbels

Note: This chapter reflects work that has been prepared as a journal manuscript and submitted for peer review. Upon acceptance, the journal article will serve as the final, citable version.

1.1. Introduction

Artificial intelligence (AI) is increasingly being integrated into educational settings to enhance teaching and learning. AI-assisted educational technology has the potential to translate large amounts of data into real-time insights about student performance, personalize learning experiences, and support data-driven decision-making in the teaching–learning environment (Wang et al., 2024). Recent reviews have primarily focused on the role of large language models, such as ChatGPT, in computer-based learning tasks (e.g., Albadarin et al., 2024; Ali et al., 2024; Deng et al., 2025; Jin & Sercu, 2025; Zhang & Tur, 2024). However, a broader range of AI in education drawing on behavioral data in computer-based learning environments remains underexplored. Among others, these behavioral traces include keystrokes, mouse clicks, eye movements, and speech patterns (di Mitri et al., 2018). Such data are core to learning analytics, which refers to "the measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens, 2013, p. 1382). Over the past decade, learning analytics has shown that fine-grained behavioral traces can be productively fed into a variety of systems, such as teacher-facing dashboards (e.g., Wiley et al., 2023). When coupled with contemporary AI techniques, these data streams can increasingly drive classroom interventions, accelerating the field's capacity to understand and optimize learning at scale (e.g., Wang et al., 2024).

To capture this wider scope of applications, the current study adopts the term behavioral data-driven AI-assisted educational technology to describe systems that deploy AI methods to leverage multimodal data streams to model learner behavior, interactions, and knowledge states. Although such technologies hold significant promise, their effective adoption depends on important human factors, in which acceptance, readiness, preparedness, and self-efficacy are crucial for effective adoption (Cukurova, 2024; Molenaar, 2022). To date, however, no comprehensive review has examined these factors in relation to the broader spectrum of behavioral data-driven AI-assisted educational technology for computer-based learning tasks.

To address this gap, the present study adopts a scoping review methodology to map existing empirical research on educators' acceptance, preparedness, readiness, and use of behavioral data-driven AI-assisted educational technology for computer-based learning tasks in formal education.

1.2. Theoretical Framework

The theoretical framework presented here provides an overview of relevant theories for examining educators' acceptance, readiness, preparedness, and self-efficacy in relation to educational technology. In addition, it outlines how these theories have been applied in recent years to study AI use in educational contexts.

1.2.1. Technology Acceptance Models

Research on technology adoption offers useful theoretical foundations for understanding educators' use of AI-assisted educational technology. The Technology Acceptance Model (TAM; Davis, 1989), grounded in the Theory of Reasoned Action (Fishbein & Ajzen, 1975) and the Theory of Planned Behavior (TPB; Ajzen, 1991), posits that perceived usefulness (i.e., the belief that a technology enhances performance) and perceived ease of use (i.e., the belief that using it will be free of effort) shape attitudes and intentions toward adoption.

Building on these earlier models, the Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003) integrates elements from TAM, TPB, and other models to identify four key determinants of adoption: performance expectancy (i.e., the extent to which a person believes that using the system will enhance their job performance), effort expectancy (i.e., how easy one thinks it is to use the system), social influence (i.e., the extent to which a person feels triggered by influential others to use the system), and facilitating conditions (i.e. how much a person believes that there are adequate organizational and technical resources in place to support their use of the system).

Both TAM and UTAUT have been widely applied in educational contexts (e.g., Granić & Marangunić, 2019; Scherer & Teo, 2019). In a systematic review of 71 studies, Granić and Marangunić (2019) concluded that TAM and its numerous extensions have become the dominant paradigm in explaining technology acceptance in education. Perceived usefulness consistently emerged as the strongest determinant of adoption across a wide range of learning technologies. Complementing this, Scherer and Teo (2019), in a meta-analysis of 45 studies, demonstrated that TAM has proven effective in explaining both pre- and in-service teachers' behavioral intentions to use technology. However, they noted that TAM research has been conducted predominantly in Asian contexts, often with relatively small teacher samples, and was typically applied to technology in general rather than to specific tools.

Similarly, a systematic review by Xue et al. (2024) on UTAUT in higher education identified 162 studies and found a strong geographical bias, with most research conducted in Asia and North America. This review also highlighted an overreliance on student participants compared to educators. In terms of technologies, a growing emphasis on mobile learning tools and platforms was observed, reflecting a shift toward mobile-centric educational models accelerated by the COVID-19 pandemic.

In recent research, the frameworks have also been applied to the study of AI in education (Ali et al., 2024; Lu & Lin, 2025). For example, Choi et al. (2023) proposed an extended TAM model that incorporates teachers' pedagogical beliefs and perceived trust, highlighting the role of human and contextual factors in shaping teachers' acceptance of educational AI tools. For instance, Ali et al. (2024) conducted a meta-analysis of 30 empirical studies that explicitly applied TAM and UTAUT to the acceptance and use of AI and AI-based applications in education. Their analysis showed that both models significantly explained stakeholders' behavioral intentions toward AI adoption, but TAM emerged as the stronger predictor. The study covered diverse contexts and applications, with studies predominantly focusing on general AI adoption, acceptance or use, and the use of ChatGPT. However, research applying TAM or UTAUT to the specific context of behavioral data–driven AI-assisted educational technology remains limited.

1.2.2. Technology Readiness and Preparedness

Technology readiness offers another important lens to examine educators' perspective on adoption. Parasuraman's (2000) Technology Readiness Index (TRI) conceptualizes readiness as "people's propensity to embrace and use new technologies for accomplishing goals in home life and at work" (p. 308). The TRI distinguishes between enablers (optimism and innovativeness) and inhibitors (discomfort and insecurity), offering a more general psychological disposition toward technology use (Pangriya & Priya Singh, 2021; Porter & Donthu, 2006). As with the TAM, the TRI has been adopted across a broad range of contexts (e.g., Blut & Wang, 2020). To our knowledge, its use in educational settings is, however, limited. Summak et al. (2010) applied the TRI with primary school teachers in Turkey and found gender differences, with male teachers reporting higher readiness than females, while no differences were observed by age or subject area. A more recent study of Nouraldeen (2023), focusing on AI in education found that students' technology readiness is a positive predictor of their AI adoption. Gender played a key moderating role, with male students showing stronger links between technology readiness and AI adoption.

The related concept of technological preparedness has been used in education to denote educators' capacity to fulfill professional responsibilities in technology-rich environments (Arviv Elyashiv & Rozenberg, 2024; Gill, 2023; Muñiz, 2018). This capacity involves multiple dimensions, including the construct of self-efficacy, understood here as teachers' belief in their capability to use technology effectively. A central framework for conceptualizing and assessing such preparedness is the Technological Pedagogical Content Knowledge model (TPACK; Mishra & Koehler, 2006), derived from the Pedagogical Content Knowledge framework (PCK; Shulman, 1986). TPACK is one of the most frequently cited models in the educational technology research field (Davies & West, 2014; Kholid et al., 2023; Schmid et al., 2024). Reflecting its prominence, a review of reviews by Schmid et al. (2024) identified 21 systematic reviews and two meta-analyses focused on TPACK. More recently, TPACK has also been adapted to the context of AI in education (Ning et al., 2024; Yue et al., 2024) and adopted for research on teachers' professional development for AI in education (Dogan et al., 2025). However, the majority of these studies focus primarily on general applications of AI within TPACK or on TPACK frameworks tailored to generative AI (Lan et al., 2025).

1.2.3. Rationale for the Scoping Review

Together, the theories and frameworks presented above offer valuable perspectives for understanding educators' acceptance, preparedness, readiness, and use of educational technologies. However, as the recent reviews and meta-analyses discussed above show, most existing research has concentrated on AI acceptance in general terms or on specific generative AI applications such as ChatGPT for computer-based learning tasks. Much less is known about how educators perceive, prepare for, and adopt behavioral data-driven AI-assisted educational technology, leveraging learning analytics to support learners and educators. This gap underscores the need for a scoping review to map the current state of research in this emerging area and to identify directions for future inquiry.

1.2.4. Research Questions

Guided by the rationale presented above, this scoping review addresses the following primary research question (RQ): What is the current state of research on educators' acceptance, preparedness, and readiness for using behavioral data-driven AI-assisted educational technology for computer-based learning tasks? This overarching question is further specified into five sub-questions:

(RQ1) What does the existing literature reveal about the (a) publication trends and (b) methodological approaches to study educators' acceptance, preparedness, readiness, and use of these technologies?

(RQ2) What theoretical models and frameworks have been used to study educators' acceptance, preparedness, readiness, and use of these technologies?

(RQ3) How and for what purposes do teachers currently use these technologies?

(RQ4) What factors influence educators' use of these technologies?

(RQ5) What barriers, challenges, and benefits do educators perceive when using these technologies?

1.3. Method

This scoping review was conducted in accordance with the JBI scoping review methodology recommendations outlined in Peters et al. (2022) and the PRISMA-ScR checklist Tricco et al. (2018). This study was preregistered on OSF before data collection. This preregistration can be consulted via the following link: https://osf.io/32xfu/?view_only=baaa35e4783c4bed9a6178cec980caed.

The following sections give an overview of the planned methods used in this scoping review. Supplementary materials referred to in this manuscript can be found via https://osf.io/c7puf/?view_only=2295e11fa2cf4711ab7929fb6168bf31.

1.3.1. Search

1.3.1.1. Eligibility Criteria

The search strategy was developed using the Population–Concept–Context (PCC) framework to define eligibility criteria (Peters et al., 2022).

First, regarding the population, we included studies that focused on educators actively involved in teaching as well as studies targeting pre-service teachers. Studies exclusively examining other stakeholders not engaged in teaching, or those focusing only on students or learners without addressing the educator, were excluded.

Second, regarding concept, we included studies that investigated educators' acceptance, preparedness, readiness, and/or use of behavioral data-driven AI-assisted educational technology, as well as studies addressing factors influencing such use (e.g., infrastructure, technical support, system characteristics, ethical aspects) or involving related constructs (i.e., self-efficacy as low-order construct of preparedness). Studies that focused on educational technology tools not involving both behavioral data-driven and AI-assisted elements were excluded. As outlined hereunder in the Screening section, this was further specified after a first screening round.

Third, regarding context, we included studies conducted in formal education settings across all levels of education. Studies situated in non-formal or informal educational contexts were excluded.

Fourth, research types were also specified for the search. Eligible sources were empirical studies using quantitative, qualitative, or mixed methods. Non-empirical publications (e.g., literature reviews, theoretical or methodological papers, opinion pieces, or commentaries) were excluded. Furthermore, records were excluded if their title and abstract were not available in English. Full texts were included if available in Dutch, English, French, German, or Spanish. Eligible publication types comprised journal articles, conference papers and proceedings, dissertations and theses, and preprints. Other document types were excluded.

1.3.1.2. Database Search

To achieve comprehensive coverage across disciplines, the search strategy included major databases relevant to the study's focus, such as those in social sciences, psychology, and computer science. An overview of the databases per interface is provided in Table 1. To complement database searches, a backward citation search of included articles was also conducted. Furthermore, although not specified in the preregistration, a forward citation search was also performed for completeness.

Table 1

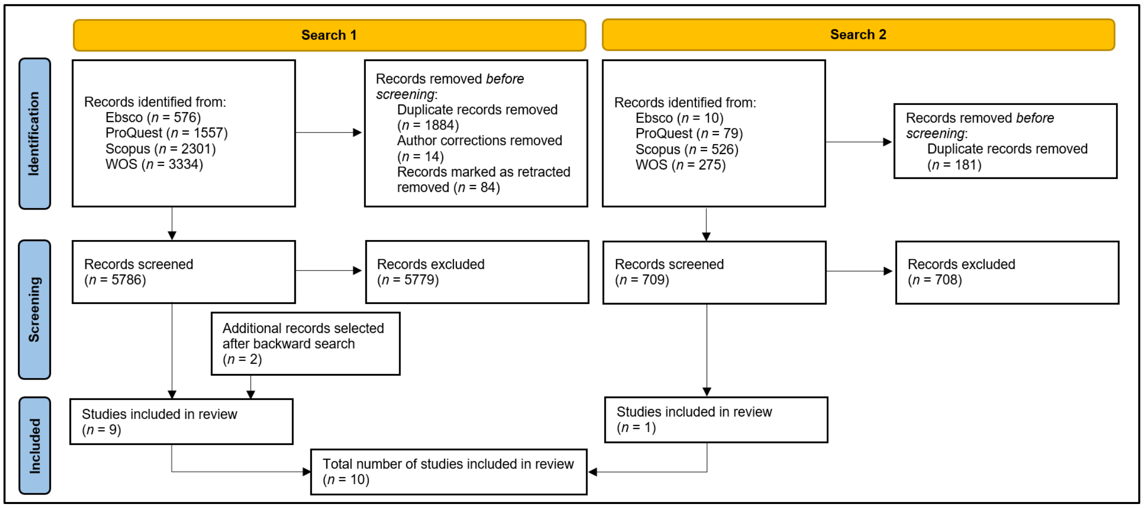
Overview of the Selected Databases per Interface.

Interface	*Database*
EBSCO	OpenDissertations, ERIC
Web of Science	Web of Science Core Collection, ProQuest™ Dissertations & Theses Citation Index, Preprint Citation Index, SciELO Citation Index
Scopus	Scopus
ProQuest	Sociological Abstracts, Social Services Abstracts, Publicly Available Content Database, APA PsycArticles, Computer Science Database, East & South Asia Database, East Europe & Central Europe Database, Education Database, Latin America & Iberia Database, Psychology Collection, Social Science Database, UK & Ireland Database

The exact query strings per database, presented in the next section, can be found via the OSF preregistration. As specified in the preregistration, the searches were conducted twice. The first search was completed on 1 April 2025, covering publications up to March 2025, and the second on 22 July 2025 to capture studies published after March 2025. This two-stage approach was adopted to account for the rapid developments in AI in education. Figure 1 provides an overview of the study selection process across both searches, thereby illustrating the additional contribution of the second search. An overview of the deduplicated identified records is provided in the supplementary materials.

Figure 1

PRISMA Flow Diagram of Study Selection Process



Note. This flow diagram is adapted from the PRISMA 2020 template (Page et al., 2021), available at https://www.prisma-statement.org/prisma-2020-flow-diagram.

1.3.2. Screening

1.3.2.1. Search 1

All retrieved records from the first search were first deduplicated in Microsoft Excel prior to screening. In addition, simple author corrections and records marked as retracted were also excluded prior to screening.

Screening was conducted in two stages: (a) title and abstract screening and (b) full-text screening. At both stages, studies were assessed against the eligibility criteria described in the Eligibility Criteria section. During title and abstract screening, records were classified as include, exclude, or unsure, with the latter carried forward for full-text assessment. To ensure consistency, 10% of the records were independently screened by two reviewers. Inter-rater reliability (IRR) was calculated based on the results of full text screening by two raters using Cohen's kappa (κ), with a threshold of κ ≥ .70 considered acceptable. If this threshold was reached, the remaining 90% of records could be screened by one reviewer. If IRR fell below the threshold, discrepancies had to be discussed, the screening procedure had to be refined, and another 10% of records needed to be re-screened until sufficient agreement was established.

Since sufficient consensus was not reached during the initial screening of 10% of the sources (κ = .13; n = 578), the two reviewers met to discuss the discrepancies. This discussion revealed that the criteria related to the concept under investigation required further refinement. As a result, the exclusion criteria related to the concept were further specified: (a) studies lacking a clear description of the application or tool, (b) studies addressing artificial intelligence in general without focusing on behavioral data-driven educational technologies, and (c) studies not involving the application of learning analytics, as defined in the introduction section.

In the subsequent 10% sample, sufficient consensus was reached at the full-text screening stage, with only one discrepancy between reviewers (ProQuest record 179; κ = .86). A total of seven articles were included from this set. A backward search was performed on the selected articles, and two articles were added consequently (i.e., Aslan et al., 2019; Wiedbusch et al. 2021). An additional forward search revealed no further articles.

1.3.2.2. Search 2

The second search conducted in July 2025 yielded 890 records. Following screening, one additional article met the eligibility criteria (Feldman-Maggor et al., 2025). A backward citation search was also performed, but no further relevant articles were identified.

1.3.3. Extraction

For the data extraction from the selected articles, a structured form was developed to capture bibliographic, methodological, and substantive information from each included study. Specifically, we recorded bibliographic details (title, author(s), year of publication, source type such as journal article, conference paper, or dissertation, journal or publisher, and DOI/URL) and study aims (stated aims, research questions, and hypotheses if available). Sample characteristics included educational level (e.g., early childhood, primary, secondary, post-secondary and short-cycle tertiary, higher education, formal adult education, or pre-service education), country and continent, sample size, gender distribution, and average age. Design characteristics comprised research setting (in vivo or in vitro), study design (qualitative, quantitative, or mixed methods), measures obtained from educators (e.g., questionnaires, interviews, observations, log data, or physiological measures), signal channels and modalities used by the AI-assisted system (as conceptualized by Di Mitri et al., 2018), and the main theoretical frameworks applied (e.g., TAM, UTAUT, TPACK). Furthermore, we extracted information on the categories of AI tools (based on Wang et al., 2024), subject areas if specified, and any key findings and recommendations. The coding table used for data extraction is provided in the supplementary materials.

Two reviewers independently piloted the extraction on six studies, thereby aligning with the preregistered minimum of 10% of included studies. Coding decisions from both reviewers were compared, and discrepancies were discussed to refine the extraction protocol.

Overall, sufficient agreement was achieved for most entities, including study aims and objectives, research questions and hypotheses, education level (κ = 1.00), country and continent, sample size and gender distribution, study design, subject (i.e., 100% agreement each), and key findings and recommendations. Near-perfect agreement was also reached for measures of educator data collection, theoretical frameworks, and categories of AI tools (83% agreement each). Minor uncertainty arose in coding age when only ranges (e.g., 40–49 years) were reported. This was recoded as missing data.

Insufficient agreement was observed for two study aspects. Research setting showed no agreement beyond chance (κ = 0.00) due to unclear initial definitions; this was refined so that in vivo referred to authentic classroom data and in vitro to simulated or laboratory data. For signal channels and modalities, the categories were expanded to include clickstream data and keystrokes, which were initially absent but relevant for the current study. Consequently, both variables were recoded, and the remaining studies were double-coded on these dimensions to ensure reliability. Perfect agreement was met after a second round of coding for these variables (i.e., 100% agreement each). The extraction of the remaining studies was carried out by a single reviewer.

1.3.4. Synthesis

Data synthesis was conducted in two complementary tiers, combining descriptive and thematic approaches in line with the preregistered protocol.

For RQ1 and RQ2, analysis was based primarily on coded data fields. Findings were synthesized through descriptive quantification of study characteristics, including publication years, outlets, geographic scope, and educational levels, as well as methodological approaches such as research setting, study design, data collection methods, and sample characteristics. For RQ2, theoretical models and frameworks were mapped by frequency and type, distinguishing between studies that applied established frameworks (e.g., TAM) and those without an explicit theoretical basis.

For RQ3, synthesis combined coded fields with additional qualitative extraction from the included articles. Contexts of technology use were described in terms of subject domains, learning objectives, educational levels, and instructional settings, while purposes were analyzed according to predefined categories (cf., Wang et al., 2024). This coding was complemented additional details such as differences in learning objectives, instructional formats, or the scale of implementation. Finally, we examined the relationship between subject domains, categories of AI systems, and the types of student data used, to highlight functional distinctions between data sources and their pedagogical applications.

For RQ4 (factors influencing educators' use), a thematic synthesis approach was applied. Open coding of study findings and author discussions was used to identify recurring factors affecting adoption, which were then grouped under broader thematic categories. RQ5 (barriers, challenges, and benefits) built on these themes by considering whether the identified factors were presented as enablers, barriers, or challenges in the included studies.

1.4. Results

1.4.1. Publication Trends and Methodological Approaches

To address RQ1, the results are organized into two parts. First, we examine publication trends, focusing on when and where the selected studies were published, their geographic scope, and the educational levels they addressed. Second, we analyze the methodological approaches employed, detailing the types of research settings, study designs, data collection methods, and participant samples. An overview of the included studies and their main characteristics is presented in Table 2. In what follows, educators will be further referred to as teachers, as they were referred to as such in all selected articles.

1.4.1.1. Publication Trends

The ten included studies were published between 2019 and 2025, indicating that research on this subject is a relatively recent development. The majority of publications appeared in the last five years, with most studies in 2021 (*n* = 3; Nazaretsky et al., 2021; Seo et al., 2021; Wiedbusch et al., 2021). Two studies were published in 2024 (Kim et al., 2024; Kumor et al., 2024) and 2025 (Feldman-Maggor et al., 2025; Rodrigues et al., 2025), suggesting that the field is continuing to grow. Only one study was published in 2022 (Nazaretsky et al., 2022) and another in 2023 (Mejia-Domenzain et al., 2023). The earliest study was Aslan et al. (2019).

The selected studies were disseminated through a range of publication outlets. Of the ten studies, six were published as peer-reviewed journal articles in British Journal of Educational Technology (Nazaretsky et al., 2022), Frontiers in Education (Wiedbusch et al., 2021), and International Journal of Educational Technology in Higher Education (Seo et al., 2021), International Journal of Artificial Intelligence in Education (Feldmann-Magor et al., 2025; Rodrigues et al., 2025), and TechTrends (Kumor et al., 2024). Three studies appeared in conference proceedings: Artificial Intelligence in Education 2023 (Mejia-Domenzain et al., 2023), CHI Conference on Human Factors in Computing Systems 2019 (Aslan et al., 2019), and European Conference on Technology Enhanced Learning (Nazaretsky, 2021). One was made available as a working paper on arXiv (Kim et al., 2024). This distribution across education, psychology, and computer science illustrates the multidisciplinary nature of the field.

Table 2

Overview of the Selected Studies and Main Characteristics

Reference	Source type	Educational level	Country	Study design	Sample size
Aslan et al. (2019)	Conference Paper	Secondary	Turkey	Mixed	37 students & 1 teacher
Feldman-Maggor et al. (2025)	Journal article	Secondary	NA	Mixed	4 teachers
Kim et al. (2024)	Working paper/ Preprint	Higher	South Korea	Qual.	6 teachers
Kumor et al. (2024)	Journal article	Secondary	US	Qual.	5 teachers
Mejia-Domenzain et al. (2023)	Conference Paper	Higher	Switzerland	Qual.	10 teachers
Nazaretsky et al. (2021)	Conference Paper	Primary & Secondary	NA	Qual.	16 teachers
Nazaretsky et al. (2022)	Journal article	Primary & Secondary	Israel	Qual.	6 teachers
Rodrigues et al. (2025)	Journal article	Primary	Brasil	Qual.	5 teachers
Seo et al. (2021)	Journal article	Higher	NA	Qual.	12 students & 11 teachers
Wiedbusch et al. (2021)	Journal article	Secondary	US	Mixed	104 teachers & 7 teachers

Note. A complete overview of all coded characteristics is provided in the supplementary materials.

The geographic focus of the included studies shows a relatively diverse spread across continents, although not all studies reported the country or broader region in which the study took place. Studies were conducted in North America (United States, *n* = 2; Kumor et al., 2024; Wiedbusch et al., 2021), Asia (Israel, *n* = 1, Nazaretsky et al., 2022; and South Korea, *n* = 1, Kim et al., 2024), South America (Brazil, *n* = 1, Rodrigues et al., 2025), and Europe (Switzerland, *n* = 1, Mejia-Domenzain et al., 2023). In addition, one study was conducted in Turkey (Aslan et al., 2019) and three publications did not specify a country or regional context (Feldman-Maggor et al., 2025; Nazaretsky et al., 2021; Seo et al., 2021).

The included studies were distributed across several educational levels, though not all predefined contexts were represented. Most studies were conducted in secondary education (*n* = 4; Aslan et al., 2019; Feldman-Maggor et al., 2025; Kumor et al., 2024; Wiedbusch et al., 2021) and higher education (*n* = 3; Kim et al., 2024; Mejia-Domenzain et al., 2023; Seo et al., 2021). Two studies spanned both primary and secondary education (Nazaretsky et al., 2021; Nazaretsky et al., 2022), and one focused exclusively on primary education (Rodrigues et al., 2025). Notably, no studies were identified in early childhood education, post-secondary or short-cycle tertiary education, formal adult education, or pre-service teacher education.

1.4.1.2. Methodological Approaches

Across the ten studies, both in vivo and in vitro approaches were represented, though research situated in authentic educational contexts was more prevalent. Specifically, six studies were conducted in vivo, presenting teachers with data drawn directly from real classroom practice (Aslan et al., 2019; Feldman-Maggor et al., 2025; Kumor et al., 2024; Mejia-Domenzain et al., 2023; Nazaretsky et al., 2021; Nazaretsky et al., 2022). The remaining four were carried out in vitro, using simulated or laboratory-based scenarios (Kim et al., 2024; Rodrigues et al., 2025; Seo et al., 2021; Wiedbusch et al., 2021). In addition, most studies employed a qualitative design (*n* = 7; Kim et al., 2024;

Kumor et al., 2024; Mejia-Domenzain et al., 2023; Nazaretsky et al., 2021; Nazaretsky et al., 2022 ; Rodrigues et al., 2025 ; Seo et al., 2021), typically relying on interviews to capture teachers' perspectives. The remaining studies used a mixed methods design (Aslan et al., 2019; Feldman-Maggor et al., 2025; Wiedbusch et al., 2021) and no studies with fully quantitative designs were identified. Interviews were used across all of the studies, supplemented in some cases by focus groups (Nazaretsky et al., 2021), questionnaires (Feldman-Maggor et al., 2025; Wiedbusch et al., 2021), or observational techniques (Aslan et al., 2019; Kim et al., 2024).

The reviewed studies varied considerably in terms of samples. Sample sizes ranged from studies with one teacher to larger surveys involving over 100 teachers. Most studies included fewer than 20 participants (Aslan et al., 2019; Kim et al., 2024; Kumor et al., 2024; Mejia-Domenzain et al., 2023 ; Nazaretsky et al., 2021; Nazaretsky et al., 2022; Rodrigues et al., 2025; Seo et al., 2021). Although most studies focused primarily on teachers as participants, some studies collected both data from students and instructors (Seo et al., 2021; Aslan et al., 2019). Where specified, the samples were often female-majority (Aslan et al., 2019; Feldman-Maggor et al., 2025; Rodrigues et al., 2025; Seo et al., 2021), though balanced male–female distributions were also observed. None of the studies explicitly included or reported on non-binary or other gender identities. Overall, age reporting was very limited across the included studies. Where reported, teacher participants were on average in their early to mid-40s (Kim et al., 2024; Rodrigues et al., 2025; Seo et al., 2021).

1.4.2. Theoretical Models and Frameworks

Regarding RQ2, the reviewed studies made only limited use of theoretical models and frameworks. The TAM, referred to by Kumor et al. (2024) and Nazaretsky et al. (2022) emerged as the most frequently applied framework. Other studies drew on more diverse perspectives, such as the five-factor model of learner–instructor interaction (Seo et al., 2021), self-regulated learning theory, information processing theory, and the cognitive theory of multimedia learning (Wiedbusch et al., 2021), or the model of trust in automation (Feldman-Maggor et al., 2025). In contrast, five publications did not specify a guiding framework and instead presented exploratory or design-oriented work (Aslan et al., 2019; Kim et al., 2024; Mejia-Domenzain et al., 2023; Nazaretsky et al., 2021; Rodrigues et al., 2025).

1.4.3. Contexts and Purposes of Technology Use

To address RQ3, we first outline the contexts of technology use, describing the subject areas, instructional settings, and educational levels in which teachers employ behavioral data-driven AI-assisted educational technology. We then turn to the purposes of technology use, examining how these tools are applied for adaptive learning, intelligent assessment, and learner profiling.

1.4.3.1. Context of Technology Use

In the studies examined, teachers employ behavioral data-driven AI-assisted educational technologies across a diverse set of teaching contexts. This diversity is evident not only in the types of courses and learning objectives but also in the classroom level, instructional setting, and the scale of deployment.

To start, when it comes to the diversity of courses and learning objectives, the included studies show that behavioral data-driven, AI-assisted education technologies are used across a wide range of domains, from high school mathematics (Kumor et al., 2021; Rodrigues et al., 2021; Meijia-Domenzain et al, 2023; Aslan et al., 2019) to biology (Nazaretsky et al., 2021), chemistry (Feldman-Maggor et al., 2022), physics (Nazaretsky et al., 2021), and English as a Foreign Language (EFL, Kim et al., 2022). While the majority of these domains are concentrated in hard science domains with well-defined rules and procedures, such as mathematics or chemistry, this result also shows that these technologies can also support learning in less structured, open-ended domains. For example, in the context of EFL writing courses in which Kim et al. (2022) conducted their research, teachers used AI dashboards to analyze the structure of students' essays, the coherence of their arguments, and the application of critical reasoning, enabling targeted feedback that goes beyond rule-based aspects of language, such as vocabulary or grammar correction.

Building upon the latter observation, it is clear that the learning objectives within these contexts can also vary considerably. In many courses, the focus is on subject-related knowledge and skills, such as mastering algebra, calculus, or chemical stoichiometry (Kumor et al., 2021; Feldman-Maggor et al., 2022). In these contexts, AI-assisted educational technology helps teachers track student progress in real time, identify knowledge gaps, and provide individualized practice opportunities tailored to each student's performance. For example, in the study of Feldman-Maggor et al. (2022), AI dashboards cluster students according to similarities in their responses to stoichiometry problems, supporting differentiated instruction and targeted feedback (Feldman-Maggor et al., 2022). In other contexts, the emphasis shifts toward higher-order cognitive skills, such as reasoning, argumentation, and conceptual understanding. This is, for instance, the case for the EFL courses previously discussed

(Kim et al., 2022), but also in biology and physics courses in which Nazaretsky et al. (2021) conducted their research. Students are expected to construct explanations of complex scientific processes and reason through abstract concepts. AI systems in these contexts automatically evaluate written responses, identify misconceptions, and assess the quality of students' reasoning, which enables teachers to provide targeted feedback to support conceptual understanding and scientific argumentation.

Furthermore, AI-assisted educational technologies demonstrate considerable adaptability across different educational levels, supporting learning from elementary and secondary school, as seen in basic numeracy (Rodrigues et al., 2021), high school mathematics (Kumor et al., 2021), and 11th-grade chemistry (Feldman-Maggor et al., 2022), up to university courses in subjects such as biology, computer science, psychology, and engineering (Seo et al., 2021; Aslan et al., 2021). This range illustrates their capacity to accommodate learners at different cognitive and developmental stages.

Additionally, these technologies are implemented in diverse instructional environments, including face-to-face classrooms, blended learning formats, and fully online settings. In the context of blended learning, for instance, flipped classroom dashboards have been developed and used to support self-regulated learning, giving teachers insights into how students manage planning and engagement across pre-class and in-class activities (Mejia-Domenzain et al., 2021). Similarly, in a study of Aslan et al. (2021), multimodal engagement systems have been deployed in online university classrooms, where they monitor participation through a combination of facial expression analysis and performance data, enabling teachers to identify disengaged students and intervene in real time.

The latter two studies also highlight differences in resource availability and the scale at which these technologies operate. Some systems are specifically designed for low-resource classrooms, as in the case of MathAIde, a mobile-based intelligent tutoring system that works with paper-based solutions and minimal internet requirements to support basic numeracy in underserved regions (Rodrigues et al., 2021). Others leverage advanced multimodal data capture, including eye-tracking and emotion recognition, to visualize students' cognitive and affective processes in real time, as demonstrated in the MetaDash project (Wiedbusch et al., 2021).

To end, the included studies show that AI-assisted educational technology can function at both the individual and group levels. Highly personalized systems, such as ALEKS in high school mathematics (Kumor et al., 2021) or ChatGPT-integrated dashboards for EFL writing (Kim et al., 2022), generate learning paths or feedback tailored to individual learners. At the same time, classroom-level tools provide insights that help teachers adjust collective instruction, as in chemistry classes where students are clustered according to their approaches to stoichiometry problems (Feldman-Maggor et al., 2022), or in online and blended environments where multimodal dashboards support whole-class monitoring (Aslan et al., 2021).

1.4.3.2. Purpose of Technology Use

Focusing on the technologies used in these studies and the main purposes for which they were developed, we can distinct three groups among the studies examined, largely aligning with the categories identified by Wang et al. (2024) in their systematic review on AI in education. We describe these three groups as (1) adaptive learning and personalized tutoring, (2) intelligent assessment and classroom management, and (3) profiling and prediction.

The first group is concerned with adaptive learning and personalized tutoring, in which AI technologies are used to guide students along individualized learning paths and allow practice at their own pace. In our study sample, this group is represented in three studies. For instance, in the study of Kumor et al. (2021) and in the context of high school mathematics, the ALEKS system is used by teachers to generate learning paths based on students' prior responses, number of attempts, and time spent on problems. In this approach, this system relies on algorithms that analyze a student's past performance to tailor instruction to individual needs. Another example is the "ITS Unplugged system" that makes use of AI algorithms to analyze individual student performance data to tailor the learning experience. Based on students' strengths and weaknesses in numeracy, the system can adapt exercises (Rodrigues et al., 2021).

The second group focuses on the use of technology for intelligent assessment and classroom management, including AI tools that automate evaluation and provide actionable insights for teachers. This group is the most frequently represented in our sample, appearing in five studies. For example, Kim et al. (2022) developed a dashboarding system that is linked with ChatGPT to analyze students' essays and chat interactions in real time, scoring work, detecting errors, and linking prompts to learning objectives, thereby enabling teachers to give targeted feedback. Another example is the study of Nazaretsky et al. (2021) in which natural language processing and machine learning systems are used to evaluate written responses, identify misconceptions, and assess reasoning quality. Other tools in this group cluster students based on response patterns and visualize these through interactive dashboards, facilitating differentiated instruction and focused feedback (e.g., Feldman-Maggor et al., 2022).

The third group, profiling and prediction, focuses on technology that is used for monitoring learner behavior and anticipating outcomes. This group appears in four studies in our sample. Examples of this group can be found in Meijia-Domenzain et al. (2023) where teachers used flipped classroom dashboards to visualize students' self-regulated learning profiles, highlighting patterns in planning, engagement, and strategy use. Another example is Aslan et al. (2021) where teachers employed real-time multimodal systems to capture behavioral and emotional engagement through facial expressions and performance data, allowing them to intervene when students appeared disengaged. Additional technologies in this cluster integrate eye-tracking, log files, emotion recognition, and reinforcement learning to provide real-time visualizations of cognitive, affective, metacognitive, and motivational processes (Wiedbusch et al., 2021).

Wang et al. (2024) also identified a fourth category in their systematic review: emerging AI-powered technologies, including educational robots and immersive VR/AR application that are used for creating new opportunities for interaction, engagement and experiential learning. However, this category did not appear in our study sample, aligning with Wang et al. (2024) who also noted that this group is less frequently reported.

Table 3

Overview of the studies with respect to Subject, AI category and Student Data Used

Reference	*Subject*	*Purpose of Use*	*Student Data*
Aslan et al. (2019)	Mathematics	profiling and prediction	clickstream and keystrokes
			multimodal
Feldman-Maggor et al. (2025)	Chemistry	intelligent assessment and management	clickstream and keystrokes
Kim et al. (2024)	English as a Foreign Language	intelligent assessment and management	clickstream and keystrokes
Kumor et al. (2024)	Mathematics	adaptive learning and personalized tutoring	clickstream and keystrokes
Mejia-Domenzain et al. (2023)	Mathematics	profiling and prediction	clickstream and keystrokes
Nazaretsky et al. (2021)	Physics and Biology	intelligent assessment and management	clickstream and keystrokes
Nazaretsky et al. (2022)	Biology	intelligent assessment and management	clickstream and keystrokes
Rodrigues et al. (2025)	Mathematics	adaptive learning and personalized tutoring	text and images
Seo et al. (2021)	Various	profiling and prediction	clickstream and keystrokes
		intelligent assessment and management	text and images
		adaptive learning and personalized tutoring	multimodal
Wiedbusch et al. (2021)	Various	profiling and prediction	multimodal

Looking further into the relationship between the subject (i.e., course), the type of AI system, and the data used from learners, from a first glance, we observe that most systems use clickstream and keystroke data to capture student responses and interactions within the learning platform (e.g., Aslan et al., 2019; Feldman-Maggor et al., 2025; Kim et al., 2024), while in Rodrigues et al. (2025), in addition to textual inputs, pictures are used to capture written responses. On the other hand, multimodal data is employed in fewer studies in order to infer students' underlying cognitive, affective, metacognitive, and motivational states (e.g., Seo et al., 2021; Wiedbusch et al., 2021). This indicates a clear functional distinction, where clickstream/keystroke data primarily supports the assessment of explicit performance, while multimodal data is leveraged for profiling more complex, internal learner processes.

1.4.4. Factors Affecting the Use of AI-Based educational technology

Research highlights several factors influencing teachers’ use and adoption of AI-based educational technology, particularly in relation to building trust and integrating these tools into practice. These factors are structured along seven themes: trust and perceptions of AI, reliability and system performance, explainability and transparency, pedagogical considerations, workload and practical benefits, information presentation and usability, and ethical concerns.

A first key consideration concerns trust and perceptions of AI-based educational technology (Nazaretsky et al., 2021; Nazaretsky et al., 2022; Wiebusch et al., 2021). Teachers’ engagement with AI recommendations depends on the degree to which they trust the system and perceive its recommendations as compatible with their beliefs about students (Nazaretsky et al., 2021). Reliability and system performance constitute a second central factor. Teachers evaluate whether AI-based educational technology can provide consistent, reliable and valid outputs regarding assessment (Kumor et al., 2024; Rodriques et al., 2025) and monitoring of task completion (Rodrigues et al., 2025).

A third factor is the explainability and transparency of AI systems. Teachers often require clarity regarding how system decisions are produced (Feldman-Maggor et al., 2025; Nazaretsky et al., 2021), and explainable AI features are regarded as critical for enhancing the interpretability of outputs (Feldman-Maggor et al., 2025; Nazaretsky et al., 2021). Pedagogical considerations is the fourth factor influencing adoption, as AI-based educational technology is expected to facilitate differentiation, accommodate student pacing, and enable the delivery of personalized learning content (Feldman-Maggor et al., 2025; Kumor et al., 2024; Rodrigues et al., 2025; Seo et al., 2021). Training and ongoing professional development are critical factors for effective adoption of AI-based educational technology by teachers (Nazaretsky et al., 2021; Rodrigues et al., 2025).

A fifth important factor affecting the use of AI based educational technology, is that it reduces the workload of teachers. Teachers see value in AI-based educational technology for easing real-time monitoring of students (Aslan et al, 2019; Kumor et al., 2024; Wiedbusch et al., 2021), providing learning analytics (Aslan et al., 2019, Rodrigues et al., 2025), and better detection of student behavior to become more aware of students' needs (Aslan et al., 2019; Seo et al., 2021). AI-based educational technology helps providing real-time feedback (Rodrigues et al., 2025), providing explanations to students (Kumor et al., 2024), and even answer simple student questions (Seo et al., 2021). Furthermore, it helps teachers with future lesson planning (Rodrigues et al., 2025; Wiedbusch et al., 2021).

Information presentation and usability constitute a sixth factor, as teachers vary in their preferences for raw versus highly interpreted (Wiedbusch et al, 2021). Simplified, textual reports and clear diagnoses are often preferred over complex figures and visualizations (Rodrigues et al., 2025). Finally, ethical considerations regarding the collection, storage, and use of sensitive student data, affect teachers' willingness to employ AI-based educational technology in their practice (Kim et al., 2025; Seo et al., 2021).

1.4.5. Perceived Barriers, Challenges and Benefits

Building on the mapping of the factors that shape teachers' engagement with AI-based educational technology in RQ4, RQ5 examines whether teachers perceive them as barriers or benefits, and under what conditions they shift from enabling to constraining.

1.4.5.1. Trust and perceptions of AI

When trust is high in AI recommendations, it will enable the use of AI-based educational technology (Nazaretsky et al., 2021; Nazaretsky et al., 2022; Wiebusch et al., 2021). However, also several barriers are identified in the reviewed studies. First, when teachers perceive AI as needing to be perfect or as an "oracle" that should provide the objective truth, they tend to dismiss its recommendations when it does not meet this expectation (Nazaretsky et al., 2021). System errors also act as a barrier further undermining teachers' trust (Nazaretsky et al., 2022). Furthermore, confirmation bias can be a barrier. Teachers often rely on prior knowledge and intuition, resisting AI recommendations that contradict their beliefs (Nazaretsky et al., 2021).

1.4.5.2. Reliability and System Performance

Even though reliable assessments are mentioned as an important factor for using AI based educational technology, reliability is also perceived as a barrier. Teachers emphasize that machines cannot account for contextual factors that may affect student behaviour (Aslan et al., 2019; Kim et al., 2025; Nazaretsky et al., 2022). They express a strong need for control over assignments, feedback, and assessment outputs, since rigid algorithms may misjudge answers (e.g., marking notation as incorrect; Kumor et al., 2024) or provide feedback lacking depth (Kumor et al., 2024; Rodrigues et al., 2025). Some teachers also indicated that they do not get enough feedback about the specifics of what students might be struggling with (Kumor et al., 2024; Rodrigues et al., 2025) or the issue that AI cannot clarify its answers, which risks student misinterpretation (Seo et al., 2021). This underlines the importance of maintaining a human-in-the-loop approach (Memarian & Doleck, 2024). Other reported issues regarding system performance include navigating and adapting to the tools (Kumor et al., 2024), system freezes (Kumor et al., 2024), large application sizes (Rodrigues et al., 2025), and reliance on internet connectivity (Rodrigues et al., 2025). Such challenges hinder seamless use in classroom contexts.

1.4.5.3. Explainability and Transparency

The more explainable and transparent AI-based educational technology are, the more teachers understand it, accept it and trust it (Feldman-Maggor et al., 2025;

Nazaretsky et al., 2021). More specifically, teachers indicated that Explainable AI features increase their understanding of the recommendations of AI-based educational technology and this understandability influences their acceptance of recommendations and helps validating the system performance (Feldman-Maggor et al., 2025). It also helps teachers more if explanations of how AI makes decisions is compared to those of human experts (Nazaretsky et al., 2021). In addition, domain-driven explanations are more effective in building trust than purely data-driven explanations (Feldman-Maggor et al., 2025).

1.4.5.4. Pedagogical Considerations

When teachers perceive AI-based educational technology as supportive for differentiation, student pacing, and providing personalized content it acts as an enabler for their adoption of such tools (Feldman-Maggor et al., 2025; Kumor et al., 2024; Rodrigues et al., 2025; Seo et al., 2021). However, at the same time concerns are raised about the impact of AI-based educational technology on student learning. Teachers worry that excessive support from AI could reduce opportunities for exploration and discovery (Seo et al., 2021), and critical thinking (Kim et al., 2025). Conflicts may also arise when AI-generated feedback or grading diverges from teacher feedback, potentially undermining student–teacher relationships (Seo et al., 2021). In addition, teachers also expressed concerns about the pedagogical adjustments required to effectively integrate AI-based educational technology into classroom routines (Nazaretsky et al., 2021).

1.4.5.5. Workload and Practical Benefit

Workload effects are a mixed factor. While AI-based educational technology can support monitoring and analytics, some teachers report increased workload from checking additional AI-generated information or handling plagiarism detection in the context of generative AI tools (Kim et al., 2025).

1.4.5.6. Ethical Considerations

Teachers express unease about collecting sensitive student data (Kim et al., 2025; Seo et al., 2021), especially regarding social interaction cues and engagement tracking (Seo et al., 2021). These privacy concerns may limit willingness to use AI-based educational technology extensively.

1.5. Discussion

1.5.1. Key Findings

This scoping review addressed five research questions. Concerning RQ1, which examined (a) publication trends and (b) methodological approaches to the study of educators' acceptance, preparedness, readiness, and use of behavioral data-driven AI-assisted educational technology, the findings indicate that this is still an emerging area, with all studies published in the past six years. The field is highly explorative, multidisciplinary and, despite the limited number of studies, already geographically widespread, with studies identified across multiple continents. Yet the literature is concentrated in technology-oriented journals across a broad set of fields, with a lack of representation in teacher education and teacher professional development outlets. At the same time, coverage of educational levels remains limited, as no studies were identified in contexts such as formal adult education, or pre-service teacher education. Methodologically, most studies were small-scale, qualitative, and interview-based, with little quantitative or large-scale work, which again reflects the exploratory stage of this field. In addition, the studies predominantly rely on self-report data, without triangulating findings with observations and additional data sources. Moreover, most publications appear to be stand-alone investigations rather than part of a systematic, iterative design program, such as the Learning Awareness Tools – User eXperience (LATUX) workflow described by Martinez-Maldonado et al. (2015) .

With regard to RQ2, this review found that theoretical grounding is limited. This finding resonates with previous work in the educational technology field which focused on learning analytics dashboards (Matcha et al., 2020). It also reflects the exploratory phase of this research, already identified under RQ1. Among the included studies, TAM (Davis, 1989) remained the main model for studying technology acceptance, confirming its importance in the field, as specified in, among others, the research of Valtonen et al. (2022). Other perspectives were only sporadically applied, and several studies proceeded without any explicit theoretical basis. Although such exploratory work can generate useful early insights, studies without clear underlying theory risk producing descriptive rather than explanatory insights. Future empirical research therefore needs to be more theory-driven while also contributing to the further development of theory. Moreover, reliance on TAM alone underscores the need to investigate additional factors beyond its core constructs to capture the complexity of educators' acceptance, preparedness, readiness, and use of behavioral data-driven AI-assisted educational technology.

Addressing the third research question (RQ3), this review found that, in studies examining educators' acceptance, preparedness, readiness, and use of behavioral data-driven AI-assisted educational technology, teachers currently employ these technologies in diverse ways and for distinct purposes. More specifically, the reviewed studies show applications across a wide range of subjects, educational levels, and instructional settings, from low-resource primary classrooms to advanced multimodal dashboards in higher education. These technologies were not limited to structured domains such as mathematics but were also applied in more open-ended contexts, such as EFL writing or scientific reasoning tasks. In terms of purpose, the results indicate that current applications emphasize individualization, diagnosis, and monitoring.

In response to RQ4, this review indicates that educators' use of behavioral data-driven AI-assisted educational technology is shaped by a common set of factors. These factors included trust and perceptions of AI, reliability and system performance, explainability and transparency, pedagogical considerations, workload and practical benefits, technical issues, information presentation and usability, and ethical concerns.

With respect to RQ5, the same factors were reported as barriers, challenges, or benefits, and in some cases more than one of these, depending on how they were experienced. Some factors were described primarily as benefits. Explainability and transparency consistently increased teachers' acceptance of AI-assisted educational technology, as systems that offered understandable recommendations were viewed as more trustworthy and useful. Pedagogical considerations were also perceived as beneficial when technologies supported differentiation, pacing, and personalization.

Other factors were experienced in more mixed ways. Trust facilitated adoption when present, but low levels of trust limited teachers' willingness to rely on system outputs. Perceptions of AI likewise created both opportunities and barriers: when teachers had high expectations of AI or held low tolerance for system errors, they tended to dismiss its recommendations. Reliability and system performance functioned similarly; reliable and valid assessments encouraged adoption, whereas technical failures or lack of control created barriers. Workload was also reported as a double-edged factor: teachers valued the potential of these systems to reduce demands by automating monitoring, feedback, and lesson planning, yet some also described increased workload due to verifying AI-generated outputs or handling plagiarism detection. Information presentation and usability fell into this category as well, with flexible and simple reporting formats facilitating use, while overly complex or rigid presentations hindered it.

Finally, certain factors were mentioned almost exclusively as barriers. Technical issues were reported to disrupt classroom use. Ethical concerns were consistently perceived as barriers that limited teachers' willingness to adopt these technologies.

1.5.2. Strengths and Limitations

This scoping review has several notable strengths. The study was preregistered and followed established methodological guidance by adhering to the JBI framework (Peters et al., 2022) and the PRISMA-ScR checklist (Tricco et al., 2018), thereby enhancing transparency and methodological rigor. The search strategy was comprehensive, covering multiple databases across disciplines and supplemented by both backward and forward citation searches. Conducting two searches within a four-month period further strengthened the review by accounting for the rapid pace of developments in the field of AI in education. Furthermore, screening and extraction procedures were also conducted systematically, with clearly defined eligibility criteria based on the PCC framework, refined when initial IRR was insufficient. As with the screening, data extraction was piloted and achieved strong levels of agreement across most variables, with problematic categories recoded and double-checked to ensure reliability.

At the same time, an important limitation must be acknowledged. The review identified only a small body of eligible studies. Although we opened the scope to include languages other than English, language restrictions may still have excluded relevant sources and specifically relevant grey literature.

1.5.3. Implications and Future Research Recommendations

1.5.3.1. Implications

The findings of this review carry several implications for both research and practice in educational technology. First, the emerging and exploratory character of this field suggests that stakeholders should temper expectations about the maturity of current evidence on educators' acceptance and use of behavioral data-driven AI-assisted educational technology. The prevalence of small-scale, qualitative work indicates that the field is still building its empirical base, making it important to treat current insights as preliminary.

Second, the limited theoretical grounding highlights a need for stronger integration of existing frameworks. Instead, future studies and practical implementations should incorporate additional perspectives, such as pedagogical theories, sociotechnical models, or ethics-oriented frameworks.

Third, the recurring dual role of factors such as trust, workload, and usability underscores that these technologies should be designed and implemented with sensitivity to context-specific conditions. The same feature may function as either a benefit or a barrier depending on context, which implies that professional development, system transparency, and iterative teacher involvement in design are key for successful adoption. Recent research on hybrid intelligence provides a useful perspective for addressing these issues, as it emphasizes human-AI collaboration rather than substitution (Cukurova, 2024; Dellermann et al., 2019). From this perspective, enhancing trust requires transparency and control, ensuring that teachers understand and trust the AI's contributions without being overwhelmed by technical details. Likewise, professional development could include explicit training on how to interpret AI-generated data and feedback, thereby strengthening teachers' competence and agency in using these tools effectively. Evidence shows that hybrid intelligence approaches can improve teacher satisfaction and performance by prioritizing explainability and teacher agency, while offloading routine tasks to AI systems (Bredeweg & Kragten, 2022; Yu et al., 2025). Furthermore, the development of AI competence skills is linked to emotional engagement in teacher training, highlighting the importance of supporting teachers not only cognitively but also affectively when preparing them for hybrid intelligent environments (Xiao et al., 2025). Together, these insights point to professional development as a critical working area to make them fully prepared and ready for using these technologies.

1.5.3.2. Future Research Recommendations

Several opportunities for future research emerge from this review. First, there is a need to expand methodological diversity and scale. To date, most studies have been small and exploratory. Future work should incorporate larger samples, mixed-method approaches, and longitudinal designs that can test causal mechanisms and capture changes in acceptance and use over time. In addition, vignette methodologies (e.g., Skilling & Stylianides, 2020) may be particularly valuable for experimentally examining use cases that are difficult to test in authentic practice, such as variations in visualization design or modes of human–AI interaction. Such approaches could shed light on how specific system features influence educators' acceptance, preparedness, readiness, and use of behavioral data–driven AI-assisted educational technologies. Another promising avenue is the use of more observational methods in classrooms. In particular, more in-depth case studies could provide a valuable research perspective for further unraveling the interrelated factors that seem to influence technological adoption, offering richer insights into how these factors interact in practice.

Second, there is a need to investigate cross-context differences. Future studies could examine how acceptance and use vary across educational levels, subject domains, and cultural or institutional contexts, in order to avoid overgeneralization from small or geographically narrow samples.

Third, future research would benefit from broadening its theoretical foundations. While the TAM has proven useful as a starting point, it has limitations in explaining the full range of factors shaping educators' acceptance and consequently their eventual use of AI-assisted educational technology. At the same time, being selective in the choice of conceptual lenses may help ensure greater theoretical coherence and contribute to cumulative theory building. In this regard, TAM could be further adapted to account for hybrid intelligence applications and the contextual factors highlighted in this review. Moreover, it is important to study these factors not only in relation to educational AI in general or generative AI applications such as ChatGPT, for which some prior research already exists (e.g., Dahri et al., 2024; Shao et al., 2024), but also more specifically in the context of behavioral data-driven AI-assisted systems as examined in this review.

A lack of theoretical grounding often leads to tools that are misaligned with pedagogical needs and practical workflows. Human-centered design moves beyond treating teachers as end-users or evaluators of a developed system and instead actively involves them throughout the design and development process (Alfredo et al., 2024; Topali et al., 2025). By actively involving teachers, human-centered design can directly address the dual role of factors like trust and usability; it can ensure that systems are built with transparency to mitigate privacy and trust concerns, designed to support teacher agency while reducing workload, and are reliable and pedagogically aligned to a specific classroom context. Ultimately, human-centered design approaches can support the creation of AI systems that are not only technically efficient but also pedagogically sound, widely trusted, and sustainably adopted.

Finally, ethical and practical barriers warrant closer investigation. More empirical work is needed on how teachers perceive and navigate concerns related to factors such as data privacy and transparency (e.g., Agarwal, 2024), as well as practical challenges such as workload and system reliability. Future studies should build on the set of factors identified in this review, examining how they interact and under what conditions they facilitate or hinder adoption. Addressing these questions will be crucial for designing systems that minimize barriers and align more closely with educators' professional needs, thereby improving adoption.

1.6. Data Availability Statement

Additional materials, next to the ones included in appendix, can be retrieved in the online repository via the following link: https://osf.io/c7puf/?view_only=2295e11fa2cf4711ab7929fb6168bf31.

1.7. References

Studies marked with an asterisk (*) are included in the scoping review.

Agarwal, B. (2024). Ethical framework for AI in education. *Proceedings of the Doctoral Consortium Of the Nineteenth European Conference on Technology Enhanced Learning*. https://ceur-ws.org/Vol-3927/paper5.pdf

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, *3*(1), 60. https://doi.org/10.1007/s44217-024-00138-2

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, *6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Ali, I., Warraich, N. F., & Butt, K. (2024). Acceptance and use of artificial intelligence and AI-based applications in education: A meta-analysis and future direction. *Information Development*, *0*(0), 1–16. https://doi.org/10.1177/02666669241257206

Arviv Elyashiv, R., & Rozenberg, K. (2024). Fostering early career teachers' preparedness, self-efficacy and professional commitment: The role of teacher education. *Teaching and Teacher Education*, *148*, 104691. https://doi.org/10.1016/j.tate.2024.104691

*Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Arslan Esme, A. (2019). Investigating the impact of a teal-time, multimodal student engagement analytics technology in authentic classrooms. In S. A. Brewster, G. Fitzpatrick, A. L. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM. https://doi.org/10.1145/3290605.3300534

Blut, M., & Wang, C. (2020). Technology readiness: A meta-analysis of conceptualizations of the construct and its impact on technology usage. In *Journal of the Academy of Marketing Science* (Vol. 48, Issue 4, pp. 649–669). Springer. https://doi.org/10.1007/s11747-019-00680-8

Bredeweg, B., & Kragten, M. (2022). Requirements and challenges for hybrid intelligence: A case-study in education. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.891630

Choi, S., Jang, Y., & Kim, H. (2023). Influence of pedagogical beliefs and perceived trust on teachers' acceptance of educational artificial intelligence tools. *International*

Journal of Human–Computer Interaction, *39*(4), 910–922. https://doi.org/10.1080/10447318.2022.2049145

Cukurova, M. (2024). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, 1–20. https://doi.org/10.1111/bjet.13514

Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Aldraiweesh, A., Alturki, U., Almutairy, S., Shutaleva, A., & Soomro, R. B. (2024). Extended TAM based acceptance of AI-Powered ChatGPT for supporting metacognitive self-regulated learning in education: A mixed-methods study. *Heliyon*, *10*(8), e29317. https://doi.org/10.1016/j.heliyon.2024.e29317

Davies, R. S., & West, R. E. (2014). Technology integration in schools. In *Handbook of Research on Educational Communications and Technology* (pp. 841–853). Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_68

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, *227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Dogan, S., Nalbantoglu, U. Y., Celik, I., & Agacli Dogan, N. (2025). Artificial intelligence professional development: A systematic review of TPACK, designs, and effects for teacher learning. *Professional Development in Education*, 1–28. https://doi.org/10.1080/19415257.2025.2454457

*Feldman-Maggor, Y., Cukurova, M., Kent, C., & Alexandron, G. (2025). The impact of explainable AI on teachers' trust and acceptance of AI EdTech recommendations:

The power of domain-specific explanations. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-025-00486-6

Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley.

Gill, A. J. G. (2023). 'Looking forward': Non-traditional students perceptions of their readiness and preparedness for the transition to work after graduation. *Research in Post-Compulsory Education*, *28*(1), 149–172. https://doi.org/10.1080/13596748.2023.2166697

Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. In *British Journal of Educational Technology* (Vol. 50, Issue 5, pp. 2572–2593). Blackwell Publishing Ltd. https://doi.org/10.1111/bjet.12864

Jin, Y., & Sercu, L. (2025). ChatGPT interventions in higher education: A systematic review of experimental studies. *Journal of Computer Assisted Learning*, *41*(4). https://doi.org/10.1111/jcal.70072

Kholid, M. N., Hendriyanto, A., Sahara, S., Muhaimin, L. H., Juandi, D., Sujadi, I., Kuncoro, K. S., & Adnan, M. (2023). A systematic literature review of Technological, Pedagogical and Content Knowledge (TPACK) in mathematics education: Future challenges for educational practice and research. *Cogent Education*, *10*(2). https://doi.org/10.1080/2331186X.2023.2269047

*Kim, M., Kim, S., Lee, S., Yoon, Y., Myung, J., Yoo, H., Lim, H., Han, J., Kim, Y., Ahn, S.-Y., Kim, J., Oh, A., Hong, H., & Lee, T. Y. (2024). *Designing prompt analytics dashboards to analyze student-ChatGPT interactions in EFL writing*. http://arxiv.org/abs/2405.19691

*Kumor, T., Uribe-Flórez, L., Trespalacios, J., & Yang, D. (2024). ALEKS in high school mathematics classrooms: Exploring teachers' perceptions and use of this tool. *TechTrends*, *68*(3), 506–519. https://doi.org/10.1007/s11528-024-00955-0

Lan, G., Feng, X., Du, S., Song, F., & Xiao, Q. (2025). Integrating ethical knowledge in generative AI education: Constructing the GenAI-TPACK framework for university teachers' professional development. *Education and Information Technologies*, *30*(11), 15621–15644. https://doi.org/10.1007/s10639-025-13427-6

Lu, W., & Lin, C. (2025). Meta-analysis of influencing factors on the use of artificial intelligence in education. *The Asia-Pacific Education Researcher*, *34*(2), 617–627. https://doi.org/10.1007/s40299-024-00883-w

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2015). The LATUX workflow: Designing and deploying awareness tools in technology-enabled learning settings. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, 16-20-March-2015*, 1–10. https://doi.org/10.1145/2723576.2723583

*Mejia-Domenzain, P., Laini, E., Neshaei, S. P., Wambsganss, T., & Käser, T. (2023). Visualizing self-regulated learner profiles in dashboards: Design insights from teachers. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. AIED 2023* (pp. 619–624). Springer. https://doi.org/10.1007/978-3-031-36336-8_96

Memarian, B., & Doleck, T. (2024). Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100053. https://doi.org/10.1016/j.chbah.2024.100053

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, *57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Muñiz, E. J. (2018). Career readiness education (CRED) program. *Journal of Higher Education Theory and Practice*, *18*(6).

*Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, *53*(4), 914–931. https://doi.org/10.1111/bjet.13232

*Nazaretsky, T., Cukurova, M., Ariely, M., & Alexandron, G. (2021). Confirmation bias and trust: Human factors that influence teachers' attitudes towards AI-based educational technology. *Sixteenth European Conference on Technology Enhanced Learning*. https://ceur-ws.org/Vol-3042/paper_4.pdf

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, *16*(3), 978. https://doi.org/10.3390/su16030978

Nouraldeen, R. M. (2023). The impact of technology readiness and use perceptions on students' adoption of artificial intelligence: The moderating role of gender. *Development and Learning in Organizations*, *37*(3), 7–10. https://doi.org/10.1108/DLO-07-2022-0133

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. In *The BMJ* (Vol. 372). BMJ Publishing Group. https://doi.org/10.1136/bmj.n71

Pangriya, R., & Priya Singh, A. (2021). The differencing views of technology readiness and acceptance model: A literature review. *Turkish Journal of Computer and Mathematics Education*, *12*(4), 1–22. https://doi.org/doi.org/10.17762/turcomat.v12i13.8310

Parasuraman, A. (2000). Technology readiness index (TRI): A multiple-item scale to measure readiness to embrace new technologies. *Journal of Service Research*, *2*(4), 307–320.

Peters, M. D. J., Godfrey, C., McInerney, P., Khalil, H., Larsen, P., Marnie, C., Pollock, D., Tricco, A. C., & Munn, Z. (2022). Best practice guidance and reporting items for the development of scoping review protocols. *JBI Evidence Synthesis*, *20*(4), 953–968. https://doi.org/10.11124/JBIES-21-00242

Porter, C. E., & Donthu, N. (2006). Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. *Journal of Business Research*, *59*(9), 999–1007. https://doi.org/10.1016/j.jbusres.2006.06.003

*Rodrigues, L., Guerino, G., Silva, T. E. V., Challco, G. C., Oliveira, L., da Penha, R. S., Melo, R. F., Vieira, T., Marinho, M., Macario, V., Bittencourt, I. I., Dermeval, D., & Isotani,

S. (2025). MathAIde: A qualitative study of teachers' perceptions of an ITS unplugged for underserved regions. *International Journal of Artificial Intelligence in Education*, *35*(1), 2–30. https://doi.org/10.1007/s40593-024-00397-y

Scherer, R., & Teo, T. (2019). Unpacking teachers' intentions to integrate technology: A meta-analysis. In *Educational Research Review* (Vol. 27, pp. 90–109). Elsevier Ltd. https://doi.org/10.1016/j.edurev.2019.03.001

Schmid, M., Brianza, E., Mok, S. Y., & Petko, D. (2024). Running in circles: A systematic review of reviews on technological pedagogical content knowledge (TPACK). *Computers & Education*, *214*, 105024. https://doi.org/10.1016/j.compedu.2024.105024

*Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education*, *18*(1), 54. https://doi.org/10.1186/s41239-021-00292-9

Shao, C., Nah, S., Makady, H., & McNealy, J. (2024). Understanding user attitudes towards AI-enabled technologies: An integrated model of self-efficacy, TAM, and AI ethics. *International Journal of Human–Computer Interaction*, 1–13. https://doi.org/10.1080/10447318.2024.2331858

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Siemens, G. (2013). Learning analytics. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Skilling, K., & Stylianides, G. J. (2020). Using vignettes in educational research: a framework for vignette construction. *International Journal of Research & Method in Education*, *43*(5), 541–556. https://doi.org/10.1080/1743727x.2019.1704243

Summak, M. S., Bağlıbel, M., & Samancıoğlu, M. (2010). Technology readiness of primary school teachers: A case study in Turkey. *Procedia - Social and Behavioral Sciences*, *2*(2), 2671–2675. https://doi.org/10.1016/j.sbspro.2010.03.393

Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M. J., Er, E., Khalil, M., & Akçapınar, G. (2025). Designing human-centered learning analytics and artificial intelligence in education solutions: A systematic literature review. *Behaviour & Information Technology*, *44*(5), 1071–1098. https://doi.org/10.1080/0144929X.2024.2345295

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Valtonen, T., López-Pernas, S., Saqr, M., Vartiainen, H., Sointu, E. T., & Tedre, M. (2022). The nature and building blocks of educational technology research. *Computers in Human Behavior*, *128*, 107123. https://doi.org/10.1016/j.chb.2021.107123

Venkatesh, Morris, Davis, & Davis. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167

*Wiedbusch, M. D., Kite, V., Yang, X., Park, S., Chi, M., Taub, M., & Azevedo, R. (2021). A theoretical and evidence-based conceptual design of MetaDash: An intelligent teacher dashboard to support teachers' decision making and students' self-regulated learning. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.570229

Wiley, K., Dimitriadis, Y., & Linn, M. (2024). A human-centred learning analytics approach for developing contextually scalable K-12 teacher dashboards. *British Journal of Educational Technology*, *55*(3), 845–885. https://doi.org/10.1111/bjet.13383

Xiao, J., Yang, Y., & Li, M. (2025). Empirical study on the feasibility of hybrid-flexible training model for developing teachers' artificial intelligence competence. *Education and Information Technologies*, *30*(12), 16835–16860. https://doi.org/10.1007/s10639-025-13460-5

Xue, L., Rashid, A. M., & Ouyang, S. (2024). The Unified Theory of Acceptance and Use of Technology (UTAUT) in higher education: A systematic review. In *SAGE Open* (Vol. 14, Issue 1). SAGE Publications Inc. https://doi.org/10.1177/21582440241229570

Yu, J., Yu, S., & Chen, L. (2025). Using hybrid intelligence to enhance peer feedback for promoting teacher reflection in video-based online learning. *British Journal of Educational Technology*, *56*(2), 569–594. https://doi.org/10.1111/bjet.13559

Yue, M., Jong, M. S.-Y., & Ng, D. T. K. (2024). Understanding K–12 teachers' technological pedagogical content knowledge readiness and attitudes toward artificial intelligence education. *Education and Information Technologies*, *29*(15), 19505–19536. https://doi.org/10.1007/s10639-024-12621-2

Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. *European Journal of Education*, *59*(2). https://doi.org/10.1111/ejed.12599

2. Educators' Perspectives on AI Systems Using Eye-Tracking to Support Reading

Authors: Siem Buseyne, Jolijn Verbergt, and David Gijbels

Note: This chapter reflects work that has been prepared as a journal manuscript and submitted for peer review. Upon acceptance, the journal article will serve as the final, citable version.

2.1. Introduction

Reading comprehension levels have been declining for years, with consistently disappointing results in both primary and secondary education (OECD, 2023; Tielemans et al., 2017). This downward trend is particularly worrying because the ability to construct meaning from text underpins students' academic success, lifelong learning, and active participation in society (Merchie et al., 2019). Strengthening reading comprehension is therefore not only an educational priority but also a societal necessity.

To understand and support students' reading comprehension, teachers currently rely on traditional methods such as observing reading behavior or asking students to verbalize their thinking while reading. However, these approaches are time-consuming if they need to be adopted for individual students, and not always feasible in everyday classroom practice (Merchie et al., 2019). As a result, many teachers continue to evaluate reading comprehension primarily through product-oriented assessments, for instance, by focusing on test results or final answers rather than on the reading process itself (Leslie & Caldwell, 2017). Yet, this approach provides limited information about how students engage with a text, where they encounter difficulties, and how their comprehension evolves over time. Moreover, research shows that teachers often struggle to accurately assess students' reading levels and to apply strategies tailored to their individual needs (Knoop-van Campen et al., 2021).

A promising way to address these challenges lies in approaches that capture and respond to the dynamics of the reading process through a collaborative interplay between teachers and technology. Recent advances in AI in education (AIEd) have opened new possibilities for such collaboration, supporting teachers not only in administrative and instructional tasks but also in monitoring and diagnosing learners' progress to identify suitable interventions (Chen et al., 2020; Wang et al., 2024). Central to these developments is the augmentation principle, which posits that AI should enhance rather than replace teachers' professional expertise (Cukurova et al., 2019). This view resonates strongly with the paradigm of Human-Centered AI (HCAI), which emphasizes that AI systems should "amplify, augment, and enhance human abilities, so as to empower people, build their self-efficacy, support creativity, recognize responsibility, and promote social connections" (Shneiderman, 2022, p. 1). It also aligns with the concept of hybrid intelligence, which envisions humans and AI combining their distinct but complementary capabilities to reach levels of performance unattainable by either alone (Dellermann et al., 2019; Kamar, 2016a, 2016b; Molenaar, 2022).

Within reading comprehension specifically, eye-tracking technology represents a particularly promising application (e.g., de-la-Peña, 2024). By recording learners' eye movements during reading, eye-tracking provides rich, fine-grained data about where students focus, hesitate, or revisit parts of a text (Carter & Luke, 2020; Mézière et al., 2023). Such data can make the invisible processes of reading visible and thereby support more adaptive instruction. However, these data are often high in volume and complexity, posing significant challenges for teachers to interpret in real time, particularly when they must attend to many students at once. Considering the notion of hybrid intelligence, in this context, AI could function as a collaborative partner that analyzes and translates complex eye-tracking data into pedagogically meaningful insights. Such systems could support teachers in different ways: providing real-time recommendations for instructional adjustments, suggesting targeted interventions for specific students, or automatically enacting certain adaptations.

New technologies only find their way into classrooms when teachers are prepared, have the necessary competencies, and are willing to adopt them (Ayanwale et al., 2022; Fakhar et al., 2024; Voogt et al., 2013). Considering literature on teachers' perceptions on AIEd is thus important when designing new technologies. Yet, existing research has predominantly examined AI as a broad concept. Most other studies have focused on generative AI tools, such as the use of ChatGPT. Far less is known about teachers' perceptions of systems that operate through hybrid intelligence and with varying levels of system- and teacher autonomy.

The present study addresses this gap by examining one specifically relevant context in which human-AI collaboration takes shape: AI-supported eye-tracking for reading comprehension. Specifically, the study investigates how teachers in compulsory education perceive the usefulness, usability, and their preparedness to work with such systems, as well as their preferred forms of collaboration. By focusing on this context, the study not only contributes to the emerging hybrid-intelligence field, but also informs the further design of such tools.

2.2. Theoretical Framework

The theoretical framework first introduces AIEd and outlines key frameworks that explain how AIEd can be conceptualized and shaped in educational contexts. Building on these perspectives, it then examines how emerging technologies that combine eye-tracking and AI can support reading comprehension in the classroom. Finally, it discusses factors influencing teachers' willingness to integrate such technologies into their practice, focusing in particular on the Technology Acceptance Model (TAM) and the Technological Pedagogical Content Knowledge (TPACK) framework.

2.2.1. Artificial Intelligence in Education

AIEd research brings together AI and the learning sciences with the aim of developing adaptive learning environments and AI-supported educational tools. The goal of AIEd is to make implicit learning computationally precise and explicit by uncovering how learning actually occurs (Luckin & Holmes, 2016). Feng and Law (2021) define AIEd as a multidisciplinary research area that combines approaches and technologies from fields like computer science and information science to tackle educational issues. The types of AIEd systems are also broad. Specifically, Wang et al. (2024) describe the following primary categories of AIEd applications: adaptive learning and personalized tutoring, profiling and prediction, intelligent assessment and management, and emerging products. In recent years, attention to the use of AIEd has been steadily increasing (Feng & Law, 2021). Despite this growing interest, however, the number of studies on AIEd is relatively low compared to other sectors, such as medicine and business (Cukurova et al., 2019).

Whereas AI was previously regarded primarily as a replacement for teachers, this view has shifted toward an augmentation perspective, in which AI is designed to support and enhance teachers' roles (Cukurova et al., 2019; Molenaar, 2022). Both the paradigms of hybrid intelligence and HCAI, introduced above, offer valuable lenses through which to understand emerging forms of collaboration between teachers and AI systems (Rokkones & Giannakos, 2025). A framework aligning with these paradigms is the Six Automations Model of Molenaar (2022). This framework outlines six levels of automation, with full human control (Level 0) and complete AI autonomy (Level 5) being the extremes. In between these extremes, the teachers' role gradually shifts.

Four intermediate levels can be identified where both the teacher and AI actively contribute. At the teacher support level (Level 1), the teacher remains in control of the learning environment while AI provides insights, such as through dashboards that analyze student behavior and performance, without taking direct action. In partial automation (Level 2), AI begins to take over specific, limited tasks like selecting exercises, although the teacher continues to oversee the broader learning process. Conditional automation (Level 3) involves AI managing and monitoring several tasks autonomously, while the teacher retains the ability to intervene when needed. Intelligent tutoring systems that offer detailed feedback are an example of this level. Finally, in high automation (Level 4), AI handles most instructional tasks independently and involves the teacher only in exceptional cases. The teacher adopts a more observational role and steps in only when necessary.

2.2.2. Reading Comprehension and Eye Tracking

A particularly promising domain in which hybrid intelligence could play a pivotal role is supporting educators in monitoring and enhancing students' reading comprehension. Reading comprehension is defined as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (Snow, 2002, p. 11). One of the most influential models of reading comprehension is the Construction–Integration Model proposed by Kintsch (1988), which posits that effective comprehension occurs when readers successfully progress through three distinct phases. In the first phase, or literal stage, the reader focuses on the explicit information in the text without yet engaging in interpretation. In the second phase, the reader establishes semantic connections among these elements, forming networks of ideas that enable an abstract understanding of the text. In the final phase, the reader integrates prior knowledge, previous experiences, and inferences with the textual information to achieve a deep level of comprehension. A proficient reader is therefore someone who not only constructs meaning from words and sentences but also integrates this meaning with relevant background knowledge (Kintsch, 1988). Interestingly, the concern over declining reading comprehension levels is partly due to the observation that many students remain stuck at the initial, literal level of understanding (de-la-Peña, 2024).

Assessing reading comprehension both for formative and summative purposes is inherently complex (Elleman & Oslund, 2019). In most classrooms, comprehension is measured by having students read a text and answer content questions, an approach which overall provides teachers with little insight into the underlying reading processes. Alternative methods also present challenges. For instance, think-aloud protocols are time-consuming and can impose additional cognitive load (Merchie et al., 2019). In response to these limitations, Merchie et al. (2019) advocate for further research into the added value of technology-enhanced elements within digital reading environments. This highlights the need to explore innovative solutions.

One promising approach lies in the use of eye-tracking technology, which provides fine-grained data on how readers engage with a text. Eye-tracking captures movements of the eyes during reading and has proven to be a strong predictor of reading comprehension (de-la-Peña, 2024). It allows for the identification of readers who experience difficulties with specific aspects of text processing, such as word recognition or the integration of information across sentences (Mézière et al., 2023).

During reading, the eyes make rapid jumps known as saccades and short pauses called fixations, sometimes moving backward to reread earlier words, i.e., regressions (de-la-Peña, 2024; Carter & Luke, 2020; Jarodzka & Brand-Gruwel, 2017). Efficient readers typically show short fixations, long saccades, and few regressions, while struggling readers display longer fixations and more regressions, often reflecting comprehension difficulties (de-la-Peña, 2024; Rayner, 1997). In addition to tracking visual attention, eye-tracking can also provide insights into readers' cognitive load and engagement (Catrysse et al., 2018; Zu et al., 2020).

Despite its potential, the richness and complexity of eye-tracking data make real-time interpretation challenging for teachers, particularly in classrooms with many students. To bridge this gap, artificial intelligence can function as a collaborative partner, analyzing, interpreting, and translating raw gaze data into pedagogically meaningful insights and eventually suggestions for actionable feedback (Burch et al., 2022). In this way, AI-supported eye-tracking systems exemplify the principles of hybrid intelligence, combining the analytical power of AI with teachers' pedagogical expertise to better understand and support students' reading comprehension.

2.2.3. Teachers' Readiness to Use New Technologies

Successful implementation of the technologies described in the previous section depends on multiple interrelated factors, including teachers' perceptions, acceptance, and knowledge (Ayanwale et al., 2022; Fakhar et al., 2024). One of the most widely used models to explain technology acceptance is the Technology Acceptance Model (TAM; Davis et al., 1989). TAM has proven highly robust and generalizable across contexts, user groups, and technologies, and has repeatedly been validated in research on educational technology (Scherer et al., 2019; Scherer & Teo, 2019; Venkatesh, 2000).

According to TAM, a user's intention to use a technology depends on their attitude toward it, which is shaped primarily by two key beliefs: perceived usefulness and perceived ease of use. Perceived usefulness is the extent to which an individual believes that using a system will improve their performance, whereas perceived ease of use reflects how effortless the system is considered to operate (Davis, 1989). Both factors are influenced by contextual and personal variables. The perceived usefulness of AIEd is often linked to its potential to provide teachers with deeper insights into students' cognitive and affective states. For example, Chounta et al. (2022) found that teachers especially valued AI systems capable of delivering real-time information about students' thoughts and attitudes. Such systems were seen as useful because they could enhance both teaching quality and learner support (Cukurova et al., 2023). Perceived ease of use, in turn, depends heavily on the classroom context and available infrastructure. Ertmer (2005) showed that adequate hardware, reliable software, and technical support are essential conditions for teachers' willingness to adopt ICT. When these are lacking, teachers often perceive the technology as cumbersome or impractical. Scherer et al. (2023) found that teachers' willingness to use technology increases with prior experience up to a point, after which it tends to decline, following an inverted U-shaped pattern. Other variables such as social influence, educational background, and motivation can also shape these perceptions (Venkatesh, 2000; Scherer et al., 2019). Demographic characteristics like gender and age have shown inconsistent relationships with technology adoption (Van Leeuwen et al., 2021). Additional factors such as ethical safeguards, trust, and transparency are likewise decisive for the adoption of AI (Cukurova, 2024; Cukurova et al., 2023; Nazaretsky et al., 2022). Finally, positive prior experiences with digital tools and high self-reported ICT competence further contribute to greater perceived ease of use (Scherer et al., 2015), whereas low ICT self-efficacy and concerns about learning outcomes can hinder adoption, even more so than technical barriers (Ertmer, 2005).

A widely recognized framework for mapping teachers' technological competence is the Technological Pedagogical Content Knowledge (TPACK) model (Mishra & Koehler, 2006). This framework outlines three core domains of teacher knowledge: technological knowledge, pedagogical knowledge, and content knowledge, as well as their intersections: technological pedagogical knowledge, technological content knowledge, and pedagogical content knowledge. At the center lies TPACK itself, representing the integrated application of all three domains to support effective technology use in the classroom (Koehler et al., 2013). The model has also been applied to the context of AIEd (Ning et al., 2024; Celik, 2023). Specifically, Celik (2023) expanded the framework into the Intelligent-TPACK model, introducing ethical knowledge as a fourth essential component. Based on Celik (2023) this ethical knowledge can be defined as teachers' ability to critically evaluate whether AI-based tools operate in fair, inclusive, transparent, and accountable ways for all learners. The model also emphasizes AI-specific technological knowledge, enabling teachers to critically evaluate AI-driven decisions and select tools aligned with pedagogical goals. Celik (2023) showed that teachers with greater knowledge of AI not only use it more confidently but also more responsibly and effectively.

2.3. Research Aim and Research Questions

Building on the theoretical perspectives outlined above, this study examines teachers' perceptions of AI-supported eye-tracking as a specific application of hybrid intelligence in the classroom. While eye-tracking can reveal detailed insights into students' reading processes and AI can translate these data into pedagogically meaningful feedback, the successful implementation of such systems depends on teachers' willingness and capacity to use them effectively. Understanding how teachers evaluate (a) the usefulness, usability, and required knowledge associated with these technologies, and (b) the system characteristics they consider desirable, is therefore essential for the meaningful design and integration of AI-supported eye-tracking tools in educational practice.

To examine these aspects, the study combines three complementary frameworks: (a) the TAM (Davis et al., 1989) to explain teachers' intentions and attitudes toward adoption, (b) the Intelligent-TPACK model (Celik, 2023) to identify the knowledge domains teachers consider essential for responsible use, and (c) Molenaar's (2022) Six Automations Model to conceptualize varying ways of collaboration and autonomy between teachers and AI systems.

Accordingly, this study investigates how teachers in primary and secondary education perceive and evaluate AI-supported eye-tracking tools for enhancing reading comprehension, guided by the following research questions (RQs):

RQ1: Which factors influence teachers' willingness to use technologies that combine eye-tracking and AI to enhance reading comprehension in the classroom?

RQ2: What types of knowledge do teachers consider essential for successfully implementing eye-tracking and AI technologies in their teaching practice?

RQ3: At which level of collaboration between humans and technology, as described in Molenaar's Six Automations Model, do teachers perceive the greatest added value and feasibility?

2.4. Method

This section first explains the choice of research design, followed by a description of the sample. Next, the data collection method is presented, including the development of the research instrument and the data collection procedure. Finally, the data analysis process is outlined. When writing this manuscript, we adhered to the Journal Article Reporting Standards for Qualitative Research (JARS–Qual; Levitt, 2019) to ensure methodological transparency and rigor.

2.4.1. Research Design

This study adopted a qualitative exploratory design, using semi-structured interviews supported by vignettes (see Appendix A). This approach was chosen to gain in-depth insight into a phenomenon for which limited prior research exists, allowing the collection of rich and nuanced data across diverse respondents (Cohen et al., 2002).

Vignettes are particularly appropriate for examining teachers' perceptions, as they provide realistic, thought-provoking scenarios that elicit participants' beliefs, reasoning, and professional insights (Skilling & Stylianides, 2020). As defined by Skilling and Stylianides (2020), vignettes are "written, visual, or oral stimuli, aligned with relevant research paradigms and methodologies, reflecting realistic and identifiable settings that resonate with participants for the purpose of provoking responses" (pp. 542-543). In the present study, vignettes were integrated into in-depth interviews to encourage reflection on different forms of teacher–AI collaboration. A structured interview protocol (see Appendix B) was developed to ensure consistency across interviews and was informed by insights from the theoretical framework.

2.4.2. Sample Composition and Characteristics

The sample consisted of twenty teachers from Flanders, Belgium: ten from primary education and ten from secondary education. In the Flemish context, primary school teachers typically teach multiple subjects to the same class, whereas secondary school teachers are subject specialists. For this reason, the secondary education group in this study included language teachers, as they most frequently address reading comprehension in their daily practice. Participants were recruited through convenience sampling.

To ensure diverse perspectives, variation in age and teaching experience was deliberately sought, as these factors may influence teachers' perceptions and use of technology in the classroom (Scherer et al., 2015). Participants ranged in age from 21 to 52 years ($M = 38$, $SD = 9$), with teaching experience varying between 0.5 and 28 years ($M = 13.5$, $SD = 9$). An overview of participant characteristics is provided in Table 1.

Table 1

Overview of Participants

Respondent	*Pseudonym*	*Level of Education*	*Gender*	*Age*	*Teaching Experience (years)*
1	Tom	Secondary	Male	48	25
2	Rachel	Secondary	Female	46	23
3	Sarah	Secondary	Female	44	21
4	Yanis	Secondary	Male	29	4
5	Amira	Secondary	Female	42	19
6	Nina	Secondary	Female	31	5
7	David	Primary	Male	39	4
8	Zoë	Secondary	Female	27	5
9	Emma	Primary	Female	37	16
10	Daniel	Secondary	Male	52	28
11	Helen	Primary	Female	48	25
12	Laura	Primary	Female	28	5
13	Noor	Primary	Female	29	8
14	Charlotte	Primary	Female	41	18
15	Mila	Primary	Female	29	7
16	Elise	Primary	Female	44	20
17	Hannah	Primary	Female	21	0.5
18	Liam	Primary	Male	39	2
19	Sophie	Secondary	Female	47	20
20	Megan	Secondary	Female	41	14

2.4.3. Data Collection

2.4.3.1. Initial Protocol Design

Molenaar's Six Automations Model served as the framework for constructing the vignettes. Of the six levels defined in this model, four involve active collaboration between humans and technology (Molenaar, 2022). As shown in the overview in Table 2, these four levels formed the basis for the vignette scenarios developed in this study. The vignette framework by Skilling and Stylianides (2020) was used to guide the vignette design. These authors proposed a method for constructing vignettes based on three key elements: conception, design, and administration. In addition, each vignette was written to a length of approximately 200 words, consistent with recommended guidelines (Jeffries & Maeder, 2005; Skilling & Stylianides, 2020).

Table 2

Overview of the Vignettes per Automation Level

Level	*Description*	*Vignette*
1	Teacher only	/
2	Teacher assistance	1
3	Partial automation	2
4	Conditional automation	3
5	High automation	4
6	Full automation	/

The hypothetical scenarios presented in the vignettes illustrated concrete forms of collaboration between teacher and technology, while remaining sufficiently abstract to invite respondents' own interpretations and insights (Jeffries & Maeder, 2005). The vignettes were structured along a developmental continuum in which the role of AI gradually increased and the role of the teacher correspondingly decreased across the four scenarios.

The vignettes were adapted to the contexts of primary and secondary education, with minor differences to reflect the realities of each teaching setting and ensure familiarity for participants (Skilling & Stylianides, 2020; Torres, 2009). Topics were tailored to each educational level: volcanoes for the fourth grade in primary school and climate change for secondary education. The complete set of eight vignettes used in the study can be found in Appendix A.

A consistent set of open-ended questions was used across all vignettes to facilitate comparison of participants' responses (Skilling & Stylianides, 2020). The semi-structured interview guide (see Appendix B) was developed deductively, with questions based on the conceptual framework and RQs of both collaborating researchers. Specifically for this study, the questions were designed to address concepts from the TAM and TPACK models.

2.4.3.2. Protocol Piloting and Refinement

To test and refine the interview protocol, two pilot studies were conducted with three teachers from primary and secondary education, selected through convenience sampling. The pilots were spaced one week apart to allow for iterative improvement based on initial findings. During these sessions, the duration required to discuss all four vignettes and the corresponding questions was recorded, along with the nature and depth of the participants' responses. Teachers also provided feedback on the realism and recognizability of the scenarios, as well as the clarity of the interview questions.

Based on this feedback, both the interview guide and vignettes were refined. The number of questions per vignette was reduced from sixteen to five by removing overlapping or redundant items. Each vignette was standardized to describe the same classroom context (i.e., same teacher and students), and an introductory paragraph was added to the interview guide to outline the hypothetical classroom setting, including grade level, lesson topic, and a brief explanation of the eye-tracker's functioning. This ensured that all participants shared a common understanding of the scenario.

2.4.3.3. Data Collection and Processing

This study received ethical approval from the Ethics Committee of the University of Antwerp. Interviews were conducted either face-to-face or online, depending on participants' preferences and practical feasibility. Prior to participation, each teacher received an information letter and signed an informed consent form. All interviews were audio-recorded with participants' permission and lasted on average 55 minutes (range: 42–60 minutes).

The recordings were transcribed verbatim, after which filler words, hesitations, and repetitions were removed to enhance readability. All personal identifiers (e.g., participants' names and school affiliations) were deleted. The resulting pseudonymized transcripts were used for subsequent data analysis.

2.4.4. Data Analysis

The data were analyzed using NVivo 1.7. A combined deductive–inductive coding approach was applied. A preliminary coding tree, derived from the theoretical frameworks of TAM and TPACK, provided the analytical structure. The parent codes per vignette were perceived usefulness, perceived ease of use, and required knowledge. Within these categories, subcodes were developed to capture more specific dimensions. For usefulness and ease of use, these subcodes were guided by the criteria list proposed by Davis (1989). A large proportion of teachers' statements could be categorized directly under these predefined codes.

Inductive coding was used to capture responses that did not fit the initial framework. These were first grouped under an "other factors" category. During a second round of coding, several inductive subcodes were integrated into the existing conceptual categories. Overlapping or weakly differentiated codes were merged to improve the internal coherence and clarity of the coding tree.

Where necessary, separate subcodes were developed per vignette. This approach enabled the identification of patterns, similarities, and differences between the various collaboration scenarios. The final version of the coding tree is presented in Table C1, Appendix C.

2.5. Results

2.5.1. Factors Influencing Teachers' Willingness to use the Technology

RQ1 explores the factors that influence teachers' willingness to use AI-supported eye-tracking technologies. The results are presented according to the key determinants of this willingness, followed by an analysis of how these factors vary across different levels of system automation.

2.5.1.1. Perceived Usefulness

One of the main perceived advantages of using eye-tracking technology to support reading comprehension is its ability to make invisible reading processes visible. Teachers often find it difficult to determine where students encounter difficulties while reading. This is especially true for quiet or less assertive students, who are less likely to ask for help. Eye-tracking was perceived as a valuable tool because it could provide teachers with direct insight into such challenges. For instance, Yanis, a 29-year-old secondary education teacher, mentioned with Vignette 1: "You can ask students: 'How well did you understand it? Where were the difficulties?' But I often notice that you get very little response when you just ask that question. . . . We can't look inside students' heads." He thinks that the technology "makes something that is currently hard to control, controllable for the first time."

Teachers also indicated that they found the technology as described useful for mapping students' reading behavior and engagement. Respondents teaching in large classes explained that they often have little awareness of who is genuinely reading and who is not. They believed that eye-tracking could help make students' engagement with the text visible. It could reveal whether students are "pretending to read" and whether they apply reading strategies independently (e.g., rereading a section or using contextual clues). For instance, Emma, 37-year-old primary education teacher, remarked at Vignette 1 that "you can actually find out who really does it and who doesn't." She also thought that "there should be some extra information linked to it, for example, when students stop reading during class. Because [she thinks] some of them just don't read at all."

Another frequently mentioned advantage among the respondents concerns the potential for differentiation in the classroom. Teachers indicated that this technology could help them better respond to the individual needs of their students. This was particularly emphasized by teachers working in large or heterogeneous classes, with considerable variation in reading proficiency or language background. By using this technology, teachers imagined being able to group students according to similar needs, such as reading pace or vocabulary knowledge.

At the higher levels of automation, where the technology takes over more control and makes decisions regarding students' reading processes (as in Vignette 3 and Vignette 4), teachers believed that such a system could free up time for them to provide individualized support. They envisioned the technology as a kind of learning pathway that students could follow independently. Students who read fluently could work autonomously, which would allow teachers to dedicate more time to those requiring additional guidance. Specifically, Laura, a 28-year-old primary education teacher, mentioned at Vignette 3 that "as a teacher, your hands are freed up more to guide other children. You know that the system is taking care of part of it for you. You can keep an eye on it through your dashboard." According to her, "for the students where things really go wrong you then have the time to work one-on-one or sit with them for a moment. So it's really a system that would allow you to work much more in a personalized way."

Some respondents also saw opportunities to use the technology in situations where the teacher is not present, for example, during study periods in secondary education or as part of homework in primary education. A few teachers mentioned the potential use of eye-tracking in team teaching settings, where one teacher could monitor the dashboard, while the other continues leading the instruction.

Furthermore, respondents emphasized the usefulness of the technology for teachers themselves. Some saw it as a tool to reflect on and improve their own teaching practices, believing that the system could provide valuable insights into the effectiveness of their instructional approaches. As David, a 39-year-old primary education teacher, explained when discussing Vignette 1:

> The thing about teaching in general is that it's so hard to quantify, you never really know: 'Am I doing well or not?' It's almost impossible to grasp. And

sometimes you think you're doing great, but it turns out later that you weren't at all. So the more information you get, the better you can adjust your teaching.

2.5.1.2. Perceived Ease of Use

The ease of use of the technology was assessed through the concepts of user-friendliness, interaction, flexibility, task performance capability, competence, and learning to use the technology (Davis, 1989). As with the previous category, substantial overlap was found among these concepts; therefore, the results are presented under three overarching themes: interaction with the dashboard, integration into classroom practice, and learning to work with the technology.

Interaction With the Dashboard. Most respondents expressed a predominantly positive view of the user-friendliness of AI-based eye-tracking technology. The dashboard and its interface design were identified as key elements influencing usability. Teachers emphasized that the dashboard should be visually clear, ideally featuring color codes or pop-up signals to make information easily interpretable. A system requiring constant navigation through multiple screens or menus was considered detrimental to usability.

Respondents stressed that the information displayed on the dashboard must be concise and easy to interpret. At the lower levels of automation (i.e., Vignette 1 and Vignette 2), teachers imagined the dashboard as relatively simple. However, as automation levels increased, so did concerns about its complexity. When discussing Vignette 3, Amira, a 42-year-old teacher in secondary education, compared an advanced dashboard to "a DJ's mixing console or a new electric car with way too many buttons and controls". In addition, teachers, especially those managing large classes, worried that an overload of information could become overwhelming and reduce the system's ease of use.

Integration into Classroom Practice. A recurring concern among respondents was how to integrate the proposed technology into their existing classroom routines. At the higher levels of automation, many expected the dashboard to demand additional attention, particularly in terms of monitoring alerts or maintaining an overview of the class. They anticipated a higher cognitive load, as they would need to divide their attention between the classroom and the dashboard. Teachers also pointed out that dashboard use could affect classroom dynamics. Some feared that they might become less engaged with students during lessons because their focus would shift toward monitoring data rather than interacting directly with learners. Several respondents emphasized that this technology should not be used in every lesson. Instead, they viewed it as a supplementary tool suitable for specific moments, such as screening sessions at the start of the school year or individual guidance during remedial instruction. A recurrent comment concerned the technical reliability of the system. While many teachers perceived the technology as relatively straightforward, they simultaneously expressed concerns about its fragility. For instance, issues with Wi-Fi connectivity or insufficiently charged devices could pose significant barriers. Respondents emphasized the need for a stable internet connection and adequate technical support to ensure smooth classroom implementation. Some also questioned whether the technology would be compatible with existing student management systems.

Training and Technical Support. At the lowest level of automation, many respondents indicated that they expected to master the proposed technology relatively quickly. In such cases, a brief introduction to the system's functionality was seen as sufficient to begin using it. Tom, a 48-year-old teacher in secondary education, for instance, mentioned at Vignette 1: "with some practice, it seems fairly manageable to me. Actually, quite easy, yes." As automation levels increased, differences between teachers became more pronounced. While some anticipated few difficulties, others expressed a clear need for additional guidance and support to feel confident and secure when using the system. For instance, Amira mentioned at Vignette 3: "I think someone who is very structured and good at multitasking will find it a straightforward tool. For me . . . I think it would help to have a colleague nearby so we could explore it together."

Teachers who felt competent to use the technology often referred to their previous experience with digital tools in their teaching practice. This experience appeared to strengthen their confidence in using new technologies. Nonetheless, even these teachers emphasized the importance of targeted professional development during implementation. Several respondents expressed a desire for a clear introduction explaining how the system works and providing insight into the underlying technology: how data are generated, processed, and displayed. They also wanted clarity on the pedagogical purpose of the tool: when and why it should be used, and how it could meaningfully complement their instruction.

Almost all respondents stated that they would prefer to experiment with the system themselves before applying it in real classroom contexts. Moreover, there was a strong call for a school-wide approach to the introduction of such technologies. Developing a shared vision within subject departments was seen as an important lever for successful implementation. Teachers also advocated for accessible training opportunities, such as workshops, and instructional videos. Several teachers also emphasized the importance of adequate technical support as a prerequisite for integrating such technologies effectively. As Amira pointed out when discussing Vignette 4, schools often operate with limited IT resources, making local peer support essential:

> Our IT colleagues are just two people for a school with 700 students and a lot of teachers. We all work with laptops and Chromebooks, so they can't possibly solve everything. That's why we have the expert pool; there are quite a few colleagues who are IT-skilled and can help others learn as well.

2.5.1.3. Other Factors

In addition to the factors described above, teachers identified several broader factors that influence their willingness to adopt AI-based eye-tracking technologies in the classroom. They are discussed in depth below.

The Role of the Teacher. The teacher's role emerged as a central theme in evaluating the usability of AI-supported technology in the classroom. Respondents noted that such technologies could reshape their role in multiple ways. As the vignettes described higher levels of automation, teachers expressed increasing concern about their autonomy and professional agency. For instance, in Vignette 4, Rachel, a 46-year-old teacher in secondary education expressed concerns, stating that "there would be resistance if something like that were fully implemented". She added "this is also not what we became teachers for. Yes, that's really just, well, supervising or something."

While most teachers were open to integrating the technology, they emphasized that this was conditional on maintaining an active role in the classroom instead of having a passive, observant role. They cautioned that as the system begins to make more decisions autonomously, part of the teacher's control and pedagogical discretion may be lost. As Liam, a 39-year-old primary education teacher, explained when discussing

Vignette 2, “I think that, over time, you might start clicking ‘yes’ too quickly and let the AI take over everything, gradually losing control over the children’s learning process, perhaps even out of convenience, just letting the technology do its job.”

Social Interaction. Another frequently mentioned concern was the potential loss of social interaction. Respondents feared that as the system operates more autonomously, the interaction between teachers and students could diminish. Tom noted when discussing Vignette 4 that “the danger, of course, is that over time you end up saying, ‘Here, just sit behind a screen!’, and that the interaction gets lost.”

The image of students working silently behind screens raised concerns about the classroom climate. Most teachers believed that reading improves through dialogue and the exchange of interpretations, emphasizing peer and teacher interaction as essential to comprehension. As Zoë, a 27-year-old teacher in secondary education explained, “That interaction with each other; I think it’s really important that they also learn from one another. In the last two [vignettes], I really feel that disappears; it just becomes ‘me and my computer’.”

Several respondents also expressed concern that technology might not account for students’ emotional states. They questioned whether the system could respond appropriately to students who might, for example, be having a difficult day. For instance, Megan, a 41-year-old secondary education teacher, mentioned the following when discussing Vignette 4: “Of course, the technology can respond to what it sees at the level of students’ eyes, but does it really know them personally? I don’t think so. And that sometimes matters”. She added that “what happened during recess, all those socio-emotional things [are factors] that AI can’t detect just from an eye tracker.”

Reading on a Screen. When putting themselves in their students’ place, several respondents also questioned the impact of prolonged screen reading. Many found paper-based reading more comfortable and natural. For instance, Laura noted when discussing Vignette 1: “reading a text on a computer, and I’m not sure if this is scientifically proven, but it just seems harder to me than reading from a book.”

Teachers noted that reading on screens could negatively affect motivation and concentration. Some observed that students were more easily distracted or read less deeply when texts were presented digitally. Others highlighted the physical and tactile differences between print and digital formats. For instance, Nina, a 31-year-old secondary education teacher finds that reading from paper is “much more interesting because you really engage deeply with a text: circling words, marking patterns”. According to her, “this system seems the complete opposite! There’s no interaction, no shared exploration; it feels lifeless.” Teachers also raised concerns about screen fatigue, especially since students already spend substantial time on digital devices outside school.

Finally, some respondents mentioned interface-related issues. One teacher noted that scrolling creates an unnatural reading experience, making it harder for students to grasp the overall structure of a text. According to this respondent, swiping would be preferable to scrolling for maintaining reading flow.

Trust in AI. Trust in AI emerged as another key factor. According to several participants, transparency in data handling was viewed as a fundamental requirement. Respondents emphasized the need for clear guarantees regarding privacy and data management, noting that it is often unclear to themselves, students, or parents what happens to the data collected by these systems. Concerns were raised about who has access to the data, how long it is stored, and for what purposes it may be used. Sophie, 47-year-old teacher in secondary education stated: “The eye-tracking movements are being monitored, who has access to that? Sure, we can use it in an educational context, but who’s to say that the platform won’t indirectly use all those children’s data? There definitely needs to be clarity.”

Teachers also questioned the reliability and interpretability of the system. Some expressed uncertainty about how AI makes decisions or generates notifications, particularly at higher automation levels. They wanted to understand the basis of the system’s reasoning to be able to trust its output. Aligning with this, Emma mentioned: “I always wonder how accurate those interpretations really are. You still have to interpret what the computer has seen. I still trust myself more than a machine.”

2.5.2. Knowledge Required to use the Technology

In relation to RQ2, this section explores which types of knowledge teachers consider essential for successfully integrating AI-based eye-tracking technologies into their classroom practice. The analysis of interview data was guided by the Intelligent-TPACK model (Celik, 2023). Each of the four AI-related components are discussed in what follows: technological knowledge, technological-pedagogical knowledge, technological-content knowledge, and ethical knowledge. The results are presented thematically by knowledge component, with attention to differences among respondents and across the four vignettes.

2.5.2.1. Technological Knowledge

Across all vignettes, respondents made remarks about the expected use and functioning of the technology itself, which were coded as technological knowledge. Most statements within this component occurred in relation to the vignette representing the lowest level of automation (i.e., Vignette 1). At this level, nearly all respondents described the system as relatively simple to use, considering themselves sufficiently competent to handle it. They referred to fundamental digital skills such as opening a program, consulting a dashboard, and interpreting simple visual information (e.g., color codes or graphs). In Vignette 2 (i.e., partial automation), respondents provided similar comments. However, at Vignette 3, expectations regarding the required technological knowledge became more divergent. Some participants assumed that the system described in this vignette would demand more technical understanding. Others believed that at the highest level of automation, less technological expertise would be needed. For instance, Charlotte, a 41-year-old primary education teacher, stated that “it’s basically just turning on your computer. Finally, fewer skills are needed than before. I just have to look at my screen until the computer tells me to say something.”

2.5.2.2. Technological-Pedagogical Knowledge

Beyond technical operation, many respondents also referred to technological pedagogical knowledge. The ability to interpret system signals and feedback was frequently emphasized. This was among others mentioned by Zoë: “You also need to be able to interpret it, of course. . . . I think it’s something you really need to learn.” Teachers expressed a need to understand what notifications mean, how reliable they are, and how to translate them into appropriate pedagogical actions. This was considered especially challenging in the higher automation vignettes, where the volume of data increased and the system made more autonomous decisions, making it harder to maintain classroom oversight. Specifically in Vignette 3, two respondents stressed the importance of maintaining strong classroom management while using the technology. They wanted to know how to balance technology use with retaining control and authority in the classroom.

2.5.2.3. Technological-Content Knowledge

Several respondents also emphasized the need for technological-content knowledge.

One teacher noted the need to understand the instructional content embedded in the system (e.g., explanatory videos or terminology), as teachers must be able to explain or contextualize these elements to students. A further respondent highlighted the importance of knowing how to interpret the technology's output correctly, ensuring that data-driven feedback is understood and acted upon appropriately.

2.5.2.4. Ethical Knowledge

The ethical knowledge referred to by the teachers mainly centered on aspects of transparency, fairness, inclusiveness, and accountability. First, teachers expressed a strong need for knowledge about the system's transparency, both in terms of how the technology produces its feedback and how student data are collected, stored, and used. For instance, Rachel mentioned: "I do wonder whether that is also stored for each individual student."

Fairness-related knowledge related to the accuracy and appropriateness of the system's interpretations. For example, Megan mentioned the need to know "whether it will be 100% correct. Is the information I'm receiving actually accurate?" Teachers wanted to know whether the AI system might misrepresent students' reading abilities or create misleading profiles, especially if the system generalizes too much. At Vignette 1, Nina raised that she wanted to know whether eventually, we will "end up creating an incorrect picture of a student's profile[.] Is it overly sensitive?"

Knowledge on the inclusiveness focused on whether the system could account for individual differences: variation in reading levels, student backgrounds, or socio-emotional states. As Megan mentioned: "Does AI know about the student's background?"

Regarding knowledge of accountability, teachers mainly stressed that, as it is for now, ultimate responsibility would still lie with the teacher. For instance, as mentioned by David: "If something goes wrong, then ultimately the responsibility lies with me and not with the system." Aligned with this, many teachers stressed that they wished to retain professional agency and not delegate key instructional decisions entirely to the system.

2.5.3. Preferred Level of Collaboration Between Teacher and Technology

At the end of each semi-structured interview, respondents were explicitly asked which level of collaboration between teacher and technology they considered most desirable. The twenty teachers each indicated their preference for one or a combination of the four vignettes. Preferences were distributed across all four vignettes. Two teachers preferred Vignette 1 (Sarah, Elise), and five selected Vignette 2 (Yanis, Amira, Emma, Sophie, Megan). Four respondents favored a combination of Vignettes 1 and 2 (Nina, Zoë, Noor, Charlotte). Two teachers chose Vignette 3 (Rachel, Mila), while three preferred a combination of Vignettes 2 and 3 (Tom, Daniel, Liam). One teacher indicated a mixed preference for Vignettes 1 and 4 (Hannah). Finally, three teachers favored Vignette 4 (David, Helen, Laura).

During the interviews, a clear pattern emerged: teachers with little or no experience with AI tended to prefer the lower levels of automation. They favored scenarios in which control remained primarily with the teacher and AI played a supportive role. In contrast, teachers who considered themselves more AI-proficient showed preferences that were more evenly distributed across the four vignettes.

2.5.3.1. Teacher Support - Vignette 1

Respondents who selected Vignette 1 primarily valued the active role of the teacher and the sense of control it affords. They emphasized the importance of being able to make didactic decisions based on their own classroom observations. While they saw a role for technology in signaling potential reading difficulties, they wanted to decide themselves how to act on such information.

By choosing the vignette with the lowest degree of automation, these respondents also associated their preference with a higher degree of teacher involvement. They believed that in this scenario, the teacher remains more attuned to students and that learning occurs in a more social and interactive environment. In their view, students would engage more with each other, thereby learning collaboratively and independently searching for solutions. As Zoë noted: “I think it’s really important that they also learn from one another and gain insights from each other. In the last two vignettes, I felt that really disappears. The interaction is gone.”

2.5.3.2. Partial Automation – Vignette 2

The majority of respondents viewed the partial automation level as the most desirable form of collaboration between teacher and technology. Similar to Vignette 1, teachers valued maintaining an active teaching role, but they also appreciated the balance between human and machine in this scenario. The system was seen as an assistant that helps teachers intervene more quickly and effectively, without diminishing their own role.

Respondents stressed their wish to retain autonomy in the classroom. They appreciated that the technology offered suggestions rather than directives, allowing them to accept or ignore AI input at their discretion and thus maintain control over classroom decisions. This was for instance mentioned by Amira: “I don’t have any problem with suggestions, as long as we still have the freedom to use them or not.”

Two respondents mentioned that this level of automation could be particularly useful in classes with students requiring additional support, such as newcomers or students with language difficulties. The adaptive differentiation capabilities of the system were seen as a way to ensure that the right type of assistance reaches the right student. Respondents also noted that the technology described in this vignette seemed easy and quick to use.

2.5.3.3. Conditional Automation – Vignette 3

Only a limited number of respondents preferred Vignette 3. Mila described this as the system described in Vignette 3 as the ideal balance between technological support and teacher autonomy, as it allows students to work independently without the teacher feeling out of control: “You could just stand behind the computers and watch how the students are doing, how things are going. But you could still step in right away if needed.” Other respondents valued the potential for differentiation at this level. They appreciated that the system could handle routine adjustments (e.g., in providing remedial or enrichment activities) while leaving the teacher free to intervene on a content or pedagogical level. In this way, the pedagogical responsibility remains with the teacher, despite increased technological involvement.

2.5.3.4. High Automation – Vignette 4

Finally, some respondents preferred the high automation level, in which the technology autonomously monitors and manages most tasks. The primary motivation for this choice was the potential for differentiation. Teachers believed that the system's automated adaptation could effectively address the diverse needs of students by offering support to weaker readers while simultaneously challenging more advanced ones. This was regarded as a major advantage, as it could enhance the efficiency of learning processes and allow teachers to intervene where truly necessary.

According to Yanis, "if that were in the long term", the fourth vignette "just seems like something that could greatly improve education, reading level, and reading comprehension". David viewed this level of automation as a valuable alternative to co-teaching. In his perspective, the technology would enable students to work independently while he could track their progress and evaluate their performance afterwards, providing individualized feedback based on concrete results.

2.6. Discussion and Conclusion

This study examined the extent to which teachers are willing to integrate hybrid AI-based eye-tracking technology into their classroom practice. Three research questions guided the study. The first examined which factors influence willingness, using the TAM (Davis et al., 1989).

Consistent with the TAM (Davis et al., 1989), perceived usefulness emerged as an important determinant of teachers' willingness to use AI-supported eye-tracking. Teachers particularly valued the technology's capacity to make otherwise invisible reading processes visible, thereby providing deeper insight into students' reading difficulties. This reflects earlier findings by Chounta et al. (2022), who reported that teachers appreciate systems that surface internal learning processes. Teachers also highlighted opportunities for differentiation, greater learner independence, and more targeted instructional support, and several respondents saw potential to evaluate and adjust their own teaching practice through the feedback provided. This suggests that teachers do not seek fully automated solutions, but rather human–AI hybrid intelligence systems in which AI augments their professional judgement instead of replacing it (Cukurova, 2024).

Perceived ease of use was mainly shaped by the intuitiveness of the dashboard and the volume of information displayed. Teachers expressed concern that larger classes would generate more alerts, potentially undermining usability. Practical integration into daily routines also surfaced as a significant issue, especially the challenge of dividing attention between the dashboard and students. The concerns resonate with previous work emphasizing the importance of reliable technology, sufficient technical support, and compatibility with existing systems (Ertmer, 2005; Scherer et al., 2015). Teachers generally expected that low-automation scenarios could be adopted quickly, whereas more advanced applications would require additional guidance. Notably, even teachers with strong digital competence emphasized the need for orientation, training, and school-wide coordination. Departmental support and a shared vision were seen as important conditions for successful implementation.

Several additional factors were highlighted. Teachers consistently highlighted their desire to maintain professional autonomy; they were open to AI as long as they retained control over key instructional decisions. Trust in AI was equally central: teachers wanted transparency about how the system generates recommendations, interprets data, and stores student information. This aligns with research identifying transparency, autonomy, and trust as critical to the adoption of AI in education (Cukurova et al., 2023) and with TAM extensions that treat perceived trust as a central factor (Choi et al., 2023).

Teachers also raised affective and relational concerns, such as the potential reduction of classroom interaction and possible negative effects of screen-based reading on motivation and comprehension. These concerns suggest that such technologies may be most appropriate when used in a targeted way rather than as a default mode of reading instruction. They also raise the question of whether, in future, similar systems could be developed for non-screen-based reading.

Finally, some participants pointed to broader ethical and societal issues, including ecological impact, and dependency on commercial technologies.

2.6.1. Teachers' Knowledge

The study shows that teachers consider technological (TK), technological-pedagogical (TPK), technological-content (TCK), and ethical knowledge essential for successful integration of AI-based eye-tracking technology. This resonates with TPACK's emphasis on the interplay of technology, pedagogy, and content. Strikingly, some components—general pedagogical knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK)—were scarcely mentioned spontaneously, suggesting that teachers primarily focus on functional and tech-pedagogical aspects. At the same time, the results underline the importance of Intelligent-TPACK, which explicitly incorporates ethical knowledge as necessary for integrating AIEd contexts.

Technological knowledge was seen as especially relevant at lower automation levels; views diverged at higher levels. Some teachers anticipated a greater need for technical skills, while others expected to intervene less actively. These differences indicate a need for targeted training on specific AI system operations, as also emphasized by Celik (2023). Technological-pedagogical knowledge likewise proved crucial: teachers wanted to understand when and how the technology is didactically meaningful and how to interpret system alerts. Technological-content knowledge surfaced in questions about how AI operates substantively and how tailored tasks are generated, pointing to the need to prepare teachers both technically and substantively. Finally, ethical knowledge was frequently cited—especially at higher automation—reflecting concerns about transparency, individual differences, and potential loss of pedagogical control. These concerns are consistent with Celik (2023), Celik et al. (2022), Cukurova (2024), and Nazaretsky et al. (2022), who foreground trust, critical reflection, and human oversight.

2.6.2. Preferred Collaboration Between Teacher and Technology

Most teachers preferred lower levels of automation in which technology is supportive while the teacher retains control—specifically vignettes 1 and 2 (teacher support and partial automation). Teachers valued autonomy, discretion over didactic decisions, and active engagement in the learning process. At higher automation (vignettes 3 and 4), benefits such as automatic differentiation and efficiency were acknowledged but more often accompanied by perceived risks—loss of control, diminished social interaction, and weaker alignment with classroom realities.

These findings reinforce prior insights that teachers wish to deploy AI in ways that enhance, rather than replace, their professional role (Cukurova, 2024; Molenaar, 2022). Teachers attach great importance to retaining control over the learning process and choosing how to act on system information. Even among those open to higher automation, the need for understanding system operation and the ability to adjust remained key. Willingness to engage with more automated technologies appears greater among teachers with more AI experience.

2.6.3. Limitations

Several limitations should be taken into account when interpreting the findings of this study. First, this study is based on hypothetical vignette scenarios instead of technology with which teachers could directly interact. While vignettes offer valuable insight into teachers' expectations and reasoning, they do not capture actual use or the situated challenges that may arise during real classroom implementation.

Participants were recruited mainly in one geographical area, with an overrepresentation of teachers working in urban contexts. It is important to note that the current study may not reflect perspectives from teachers in rural schools, schools with alternative pedagogical models, or special needs schools. Including a broader and more diverse sample in future studies could strengthen the transferability of the findings.

Third, although the Intelligent-TPACK framework guided the study, not all components emerged equally in teachers' accounts. General pedagogical knowledge, content knowledge, and pedagogical content knowledge were mentioned less frequently. This may reflect the structure of the vignettes or the interview focus, which foregrounded technological and hybrid-intelligence aspects. It remains unclear to what extent teachers see these other knowledge domains as relevant to the integration of AI-supported eye-tracking.

2.6.4. Implications

Despite these limitations, the findings offer several implications for research, design, and practice. The results highlight the importance of transparency, interpretability, and teacher control in hybrid-intelligent AIEd systems for teachers. Designers should ensure that dashboards present information clearly, avoid cognitive overload, and allow teachers to decide how to act on system recommendations. Integrating hybrid AIEd systems into existing school platforms may reduce practical barriers and support everyday use.

Next, professional development initiatives should therefore go beyond technical training to include: interpreting system feedback, understanding the pedagogical purposes of AIEd tools, evaluating the fairness and inclusiveness of AI outputs, and maintaining human oversight in hybrid-intelligence settings. Collaborative learning within subject teams may further foster a shared vision for implementation.

Third, the findings indicate that teachers prefer hybrid-intelligence systems that augment rather than replace their professional role. Future research on hybrid intelligence in education should therefore examine more systematically how different distributions of autonomy between humans and AI shape instructional decision-making, perceived professional agency, and classroom interaction. While the present study offers an initial contribution to this line of inquiry, a larger-scale quantitative investigation would be a valuable next step to validate and extend these insights across broader teacher populations.

2.6.5. Future Research Recommendations

Future research could build on this study in several ways. First, experimental or mixed-methods designs would allow researchers to observe how teachers actually use AI-supported eye-tracking systems in practice, including how specific hardware configurations or physical set-ups influence usability and acceptance. Second, school-level and meso-level factors warrant closer examination, such as ICT policies, availability of technical support, and teacher workload, as these conditions likely shape whether such technologies can be meaningfully integrated into classroom routines. Third, studies should investigate how AI-supported eye-tracking functions across diverse student populations, including multilingual learners and students with learning difficulties, to determine whether the technology adequately supports varied learning needs. Finally, longitudinal research could provide valuable insight into how teachers' perceptions evolve over time as they gain hands-on experience with hybrid-intelligence tools.

2.7. Conclusion

This study shows that teachers are willing to use AI-supported eye-tracking when it operates as a form of hybrid intelligence, i.e., when it augments rather than replaces their professional agency. Usefulness, ease of use, and transparency emerged as essential conditions for adoption, underscoring teachers' preference for systems that support rather than automate pedagogical decision-making. The study responds to the call by Merchie et al. (2019) to better understand how technology can be meaningfully integrated into digital reading environments. When developed in close collaboration with teachers, hybrid-intelligence applications such as AI-supported eye-tracking can contribute to more targeted support for diverse learners and deeper insight into the reading process.

2.8. References

Ayanwale, M. A., Sanusi, I. T., Adelana, O. P., Aruleba, K. D., & Oyelere, S. S. (2022). Teachers' readiness and intention to teach artificial intelligence in schools. *Computers and Education: Artificial Intelligence*, *3*, 100099. https://doi.org/10.1016/j.caeai.2022.100099

Burch, M., Haymoz, R., & Lindau, S. (2022). *The benefits and drawbacks of eye tracking for improving educational systems. Proceedings of the 2022 Symposium on Eye Tracking Research and Applications (ETRA '22)* (pp. 1–5). ACM. https://doi.org/10.1145/3517031.3529242

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Catrysse, L., Gijbels, D., & Donche, V. (2018). It is not only about the depth of processing: What if eye am not interested in the text? *Learning and Instruction*, *58*, 284–294. https://doi.org/10.1016/j.learninstruc.2018.07.009

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, 107468.

https://doi.org/10.1016/j.chb.2022.107468

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, *66*(4), 616–630. https://doi.org/10.1007/s11528-022-00715-y

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, *8*, 75264– 75278. https://doi.org/10.1109/access.2020.2988510

Choi, S., Jang, Y., & Kim, H. (2023). Influence of pedagogical beliefs and perceived trust on teachers' acceptance of educational artificial intelligence tools. *International Journal of Human–Computer Interaction*, *39*(4), 910–922. https://doi.org/10.1080/10447318.2022.2049145

Chounta, I.-A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education*, *32*(3), 725–755. https://doi.org/10.1007/s40593-021-00243-5

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. https://doi.org/10.4324/9780203224342

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, *56*(2), 469–488. https://doi.org/10.1111/bjet.13514

Cukurova, M., Kent, C., & Luckin, R. (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology*, *50*(6), 3032–3046. https://doi.org/10.1111/bjet.12829

Cukurova, M., Miao, X., & Brooker, R. (2023). Adoption of artificial intelligence in schools: Unveiling factors influencing teachers' engagement. *International Conference on Artificial Intelligence in Education*, 151–163. https://doi.org/10.1007/978-3-031-36272-9_13

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, *35*(8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982

de-la-Peña, C. (2024). Eye-tracking contribution on processing of (implicit) reading comprehension. *Journal of New Approaches in Educational Research*, *13*(1), 13. https://doi.org/10.1007/s44322-024-00013-w

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Elleman, A. M., & Oslund, E. L. (2019). Reading comprehension research: Implications for practice and policy. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 3–11. https://doi.org/10.1177/2372732218816339

Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development*, *53*(4), 25–39. https://doi.org/10.1007/bf02504683

Fakhar, H., Lamrabet, M., Echantoufi, N., El Khattabi, K., & Ajana, L. (2024). Artificial Intelligence from teachers' perspectives and understanding: Moroccan Study. *International Journal of Information and Education Technology*, *14*(6). https://doi.org/10.14569/ijacsa.2024.0150450

Feng, S., & Law, N. (2021). Mapping artificial intelligence in education research: A network-based keyword analysis. *International Journal of Artificial Intelligence in Education*, *31*(2), 277–303. https://doi.org/10.1007/s40593-021-00244-4

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: Towards a model of real-world reading. In *Journal of Computer Assisted Learning* (Vol. 33, Issue 3, pp. 193–201). Wiley Online Library. https://doi.org/10.1111/jcal.12189

Jeffries, C., & Maeder, D. W. (2005). Using vignettes to build and assess teacher understanding of instructional strategies. *Professional Educator*, *27*, 17–28. https://doi.org/10.1080/08878730.2011.552667

Kamar, E. (2016a). Directions in hybrid intelligence: Complementing AI Systems with human intelligence. *IJCAI*, 4070–4073.

Kamar, E. (2016b). Hybrid workplaces of the future. *XRDS: Crossroads, The ACM Magazine for Students*, *23*(2), 22–25. https://doi.org/10.1145/3013488

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163. https://doi.org/10.1037/0033-295x.95.2.163

Knoop-van Campen, C. A. N., Kok, E., van Doornik, R., de Vries, P., Immink, M., Jarodzka, H., & van Gog, T. (2021). How teachers interpret displays of students' gaze in reading comprehension assignments. *Frontline Learning Research*, *9*(4), 116–140. https://doi.org/10.14786/flr.v9i4.881

Koehler, M., Mishra, P., Akcaoglu, M., & Rosenberg, J. M. (2013). The technological pedagogical content knowledge framework for teachers and teacher educators. *ICT Integrated Teacher Education: A Resource Book*, 2–7. https://doi.org/10.1177/016146810610800610

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70. https://doi.org/10.1177/002205741319300303

Leslie, L., & Caldwell, J. S. (2017). Assessments of reading comprehension: Challenges and directions. In S. E. Israel (Ed.), *Handbook of research on reading comprehension* (2nd ed., p. 219–240). Guilford Press.

Levitt, H. M. (2019). Reporting standards for qualitative research in psychology: What are they, and why do we need them? In H. M. Levitt, *Reporting qualitative research in psychology: How to meet APA style journal article reporting standards* (pp. 3–18). American Psychological Association. https://doi.org/10.1037/0000121-001

Luckin, R., & Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. London: Pearson. https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf

Merchie, E., Gobyn, S., De Bruyne, E., De Smedt, F., Schiepers, M., Vanbuel, M., Ghesquière, P., Van den Branden, K., Van Keer, H. (2019). Effectieve, eigentijdse begrijpend leesdidactiek in het basisonderwijs: Wetenschappelijk eindrapport van een praktijkgerichte literatuurstudie [Effective, contemporary reading comprehension instruction in primary education: Final scientific report of a practice-oriented literature review]. Vlaamse Onderwijsraad.

Mézière, D. C., Yu, L., Reichle, E. D., Von Der Malsburg, T., & McArthur, G. (2023). Using eye-tracking measures to predict reading comprehension. *Reading Research Quarterly*, *58*(3), 425–449. https://doi.org/10.1002/rrq.498

Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1177/016146810610800610

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, *57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Nazaretsky, T., Cukurova, M., & Alexandron, G. (2022). An instrument for measuring teachers' trust in AI-based educational technology. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 56–66. https://doi.org/10.1145/3506860.3506866

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, *16*(3), 978. https://doi.org/10.3390/su16030978

OECD. (2023) , PISA 2022 results (Volume I): The state of learning and equity in education, PISA, OECD Publishing, Paris, https://doi.org/10.1787/53f23881-en

Rayner, K. (1997). Understanding eye movements in reading. *Scientific Studies of Reading*, *1*(4), 317–339. https://doi.org/10.1207/s1532799xssr0104_2

Rokkones, R. K., & Giannakos, M. (2025). Toward hybrid teaching intelligence: Investigating the potential of teacher–AI collaboration using large language models. *Behaviour & Information Technology*, 1–19. https://doi.org/10.1080/0144929X.2025.2564368

Scherer, R., Siddiq, F., & Teo, T. (2015). Becoming more specific: Measuring and modeling teachers' perceived usefulness of ICT in the context of teaching and learning. *Computers & Education*, *88*, 202–214. https://doi.org/10.1016/j.compedu.2015.05.005

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Scherer, R., & Teo, T. (2019). Unpacking teachers' intentions to integrate technology: A meta-analysis. *Educational Research Review*, *27*, 90–109. https://doi.org/10.1016/j.edurev.2019.03.001

Scherer, R., Siddiq, F., Howard, S. K., & Tondeur, J. (2023). The more experienced, the better prepared? New evidence on the relation between teachers' experience and their readiness for online teaching and learning. *Computers in Human Behavior*, *139*, 107530. https://doi.org/10.1016/j.chb.2022.107530

Shneiderman, B. (2022). Human-centered AI. *Proceedings of the 5th Workshop on Human Factors in Hypertext*, 1–2. https://doi.org/10.1145/3538882.3542790

Skilling, K., & Stylianides, G. J. (2020). Using vignettes in educational research: a framework for vignette construction. *International Journal of Research & Method in Education*, *43*(5), 541–556. https://doi.org/10.1080/1743727x.2019.1704243

Snow, C. E. (2002). Defining Comprehension. In *Reading for understanding: Toward and R&D program in reading comprehension*. RAND Education. https://www.videnomlaesning.dk/media/2526/reading-for-understanding.pdf

Tielemans, K., Vandenbroeck, M., Bellens, K., Van Damme, J., & De Fraine, B. (2017). *Het Vlaams lager onderwijs in PIRLS 2016. Begrijpend lezen in internationaal* *perspectief en in vergelijking met 2006.*

Torres, S. (2009). Vignette methodology and culture-relevance: Lessons learned through a project on successful aging with Iranian immigrants to Sweden. *Journal of Cross-Cultural Gerontology*, *24*, 93– 114. https://doi.org/10.1007/s10823-009-9095-9

van Leeuwen, A., Knoop-van Campen, C. A. N., Molenaar, I., & Rummel, N. (2021). How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K-12. *Journal of Learning Analytics*, *8*(2), 6–21. https://doi.org/10.18608/jla.2021.7325

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342– 365. https://doi.org/10.1287/isre.11.4.342.11872

Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge–a review of the literature. *Journal of Computer Assisted Learning*, *29*(2), 109– 121. https://doi.org/10.1111/j.1365-2729.2012.00487.x

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167

Zu, T., Hutson, J., Loschky, L. C., & Rebello, N. S. (2020). Using eye movements to measure intrinsic, extraneous, and germane load in a multimedia learning environment. *Journal of Educational Psychology*, *112*(7), 1338. https://doi.org/10.1037/edu0000441

3. Vignette-Based Survey Design and Data Collection

Authors: Siem Buseyne and David Gijbels

This chapter provides an overview of the design, data collection procedure, and sample characteristics of the vignette-based survey. This data collection served as a shared empirical basis for the quantitative chapters that follow, which focus respectively on teachers' preparedness and readiness for AI-supported educational technology (Chapter 4) and on the relationship between system autonomy and teachers' acceptance of AI-supported systems (Chapter 5).

3.1. Study Design

The study employed an online vignette-based survey design in which teachers were asked to reflect on a series of hypothetical scenarios describing the use of an AI-supported eye-tracking system to support reading activities in their teaching practice. Vignette methodology was chosen to systematically vary key system characteristics while holding the instructional context constant, allowing participants to evaluate different system configurations in a controlled yet ecologically meaningful way.

Across the vignettes, the instructional context remained identical: students engage in a digital reading task while an AI-supported eye-tracking system analyzes their reading behavior (e.g., reading pace, rereading, pauses) to support the teacher. What varied across scenarios was the degree of system autonomy, ranging from observation-only support to fully autonomous system behavior. This design enabled the examination of teachers' perceptions, preferences, and acceptance judgments across distinct levels of automation within a single coherent framework.

3.2. Procedure

Participants were recruited internationally by the different project partners via their respective networks and via social media. After providing informed consent, respondents first answered a set of background questions concerning their demographic characteristics, teaching context, and professional experience. Next, participants completed a set of self-report measures assessing their knowledge, skills, and beliefs regarding artificial intelligence in education, including an AI-adapted TPACK instrument and general AI-related self-efficacy and subjective norm items.

Following this preparedness section, participants were introduced to the vignette task. They were asked to keep a specific group of students in mind—preferably a group aged nine years or older—while reading the scenarios. Four vignettes were presented sequentially, each describing a different configuration of an AI-supported eye-tracking system used to support reading comprehension. After each vignette, participants evaluated the system using standardized acceptance measures and perceived system autonomy items, and they were asked to indicate design preferences regarding system feedback, explanations, and visualization.

At the end of the vignette part, participants ranked four system configurations in terms of (a) their willingness to use the system in future teaching and (b) their trust in the system. Several open-ended questions were included throughout the survey to allow participants to elaborate on their evaluations and preferences.

Because this is an online survey, some participant attrition occurred throughout its completion. As a result, the sample size differs slightly between analyses focusing on teachers' preparedness (Chapter 4) and those focusing on system acceptance and autonomy (Chapter 5). An overview of the sample characteristics is therefore provided in the respective chapters.

3.3. Vignette Manipulation: System Autonomy Levels

The four vignettes represented increasing levels of system autonomy, based on the model of Molenaar (2022):

- Teacher assistance: the system visualizes students' reading behavior on a dashboard without intervening or making decisions.
- Partial automation: the system provides small, predefined interventions in response to specific reading behaviors, while the teacher remains primarily responsible for instruction.
- Conditional automation: the system independently adjusts instructional elements and provides recommendations, while keeping the teacher informed and allowing final control.

- High automation: the system autonomously adapts content and support for students and alerts the teacher only in exceptional cases.

The design of the vignettes was also informed by a teacher workshop organised in collaboration with WP5 of the EYE-TEACH project, ensuring that the scenarios reflected authentic classroom needs and concerns. The vignette designs were further refined through several short iterative testing phases with teachers, allowing the descriptions to be adjusted for clarity, realism, and pedagogical relevance.

The ordering of the vignettes reflects a gradual shift from teacher-controlled to system-driven instructional decision-making and forms the conceptual backbone of the acceptance analyses reported in Chapter 5.

3.4. Ethical Considerations

The study was conducted in accordance with applicable ethical guidelines and data protection regulations. Ethical approval for this study was obtained from the Ethics Committee for the Social Sciences and Humanities (Approval No. SHW_2025_47_1). Participation was voluntary, responses were collected anonymously, and participants could withdraw at any time without consequences. All data are stored securely and are used exclusively for research purposes within the scope of the EYE-TEACH project and related scientific publications.

4. Individual Differences in Teachers' Perceived Preparedness for AI-Supported Educational Technologies

Authors: Siem Buseyne, Glen Joris, Konstantinos Tsiakas, Tine van Daal, Leen Catrysse, Vincent Donche, Sven De Maeyer, Halszka Jarodzka, and David Gijbels

Note: This chapter reflects work that has been prepared as a journal manuscript and submitted for peer review. Upon acceptance, the journal article will serve as the final, citable version.

4.1. Introduction

The successful adoption of AI-supported educational technologies does not solely depend on system characteristics, but also on teachers' preparedness to meaningfully integrate such technologies into their instructional practice (Dogan et al., 2025). Within research on technology integration, Technological Pedagogical Content Knowledge (TPACK) has emerged as one of the most widely used frameworks to conceptualize teachers' professional knowledge for teaching with technology (Koehler et al., 2014; Mishra, 2019; Mishra & Koehler, 2006; Schmidt et al., 2009). TPACK emphasizes that effective technology integration requires not only technological knowledge, but also an understanding of how technology interacts with pedagogical approaches and subject-matter content.

Originally conceptualized by Mishra and Koehler (2006), the TPACK framework builds on Shulman's (1986) notion of pedagogical content knowledge by explicitly incorporating technology as a core component of teachers' professional knowledge. The framework distinguishes three primary knowledge domains: Content Knowledge (CK), Pedagogical Knowledge (PK), and Technological Knowledge (TK). The model emphasizes that effective technology integration emerges from their dynamic interplay rather than from any single domain in isolation. Therefore, beyond these core domains, TPACK theorizes several intersecting knowledge components, including Pedagogical Content Knowledge (PCK; i.e., knowledge of how to teach specific subject matter), Technological Content Knowledge (TCK; i.e., knowledge of how technology can be used to represent the subject matter), and Technological

Pedagogical Knowledge (TPK; i.e., knowledge of how to apply technology to support teaching practices and pedagogical approaches), culminating in Technological Pedagogical Content Knowledge (TPACK) as an integrated form of knowledge required for teaching specific content with appropriate pedagogical strategies and technologies (Koehler et al., 2014; Mishra & Koehler, 2006; Scherer et al., 2017; Schmidt et al., 2009).

Over the past two decades, TPACK has been extensively applied in both pre-service and in-service teacher research (Scherer et al., 2017). At the same time, substantial methodological and conceptual challenges have been identified (Backfisch et al., 2025). Reviews and validation studies consistently show that TPACK is most often assessed using self-report instruments, which are efficient for large-scale data collection but raise concerns regarding construct validity, content specificity, and the distinction between knowledge, skills, and beliefs (e.g., Kopcha et al., 2014).

Despite these limitations, recent reviews underline that self-report TPACK measures remain highly valuable for describing teachers' perceived preparedness to engage with new educational technologies, especially in large and heterogeneous samples (Scherer et al., 2017). From this perspective, TPACK is best interpreted not as a precise measure of instructional quality, but as an indicator of teachers' perceived capability to integrate technology in pedagogically and contextually appropriate ways. Such perceptions are known to play a critical role in teachers' willingness to experiment with, adopt, and sustain the use of innovative technologies.

In the context of AI-supported systems, preparedness takes on particular importance. These systems require teachers to interpret multimodal data, align AI-generated insights with pedagogical goals, and retain professional agency in instructional decision-making. In response to the increasing prominence of artificial intelligence in education, the TPACK framework has recently been extended to explicitly account for AI-related knowledge and competencies. Building on the original TPACK model, Ning et al. (2024) propose an AI-TPACK framework that integrates teachers' understanding of AI technologies, their pedagogical affordances, and their subject-specific applications. This extension emphasizes teachers' ability to select, interpret, and pedagogically orchestrate AI systems, while remaining aware of their limitations and implications for teaching and learning.

Despite these conceptual advances, systematic insights into how AI-TPACK varies across different groups of teachers remain limited (Karataş & Ataç, 2025). Existing studies have primarily focused on validating AI-TPACK constructs or examining relationships between knowledge components, while paying less attention to how teachers' background characteristics (Celik, 2023; Ning et al., 2024), such as professional experience, teaching context, subject domain, or national setting, relate to their level of AI-TPACK. Yet, such differences are highly relevant from an educational policy and implementation perspective. Understanding whether preparedness for AI-supported systems differs between teacher groups and across contexts is essential for designing targeted professional development initiatives, ensuring equitable implementation, and aligning AI-in-education policies with teachers' actual preparedness.

Accordingly, the aim of this chapter is to provide an empirically grounded overview of teachers' preparedness for AI-supported educational technologies, using AI-related TPACK as an indicator of perceived preparedness. Specifically, the chapter aims to (a) describe overall levels of AI-related TPACK among participating teachers and (b) examine how teachers' preparedness varies across background characteristics and national contexts.

Guided by these aims, the following research question is addressed: How does teachers' preparedness in the technology-related dimensions of AI-TPACK (i.e., AI-TK, AI-TCK, AI-TPK, and AI-TPACK) vary as a function of contextual and teachers characteristics?

4.2. Literature

Empirical research examining AI-TPACK is gradually emerging (see Table 1), providing further insights into how teachers' characteristics vary across their understanding and proficiency with various AI-driven tools such as intelligent tutoring systems and adaptive learning systems. The available studies have explored a range of teacher characteristics, such as gender, age, and teaching position, but have typically examined these factors in isolation, without systematically considering how multiple background characteristics interact.

Karataş and Ataç (2025) examined the relation between AI-TPACK and traditional TPACK, and found that male teachers and those with previous AI experience scored higher in all AI-TPACK areas, particularly in TK. Prior AI experience has also been identified as an important factor for the development of technological knowledge in the context of AI (Chan & Tang, 2024; Li et al., 2025; Wang, 2024), showing the importance of AI-training and skill development in understanding how AI-systems work and are used in educational settings. Li et al. (2025) also highlighted the importance of other characteristics, such as teaching discipline and teaching position, noting that humanities teachers reported lower levels of technological knowledge regarding AI than science teachers. Additionally, An et al. (2025) and Chan and Tang (2024) pointed to the role of age, with younger teachers demonstrating a more advanced understanding of AI-TCK, and thus reflecting a stronger capacity to understand and critically engage with the ways in which AI technologies interact with subject-specific content.

Table 1

Literature Review AI-TPACK

Study	*Examined factors*
Karataş and Ataç (2025)	Gender, prior AI experience
Chan & Tang (2024)	Age, gender
Li et al., (2025)	Teaching disciplines, teaching positions, professional titles, skills training and support policies
Wang (2024)	Prior AI experience
An et al. (2025)	Age

However, as previously mentioned, research on how AI-related TPACK varies across different teacher populations remains limited, despite its importance for professional development initiatives and AI-in-education policies. Given these constraints, and considering that AI-TPACK builds directly on the broader TPACK framework, it is useful to draw on the more extensive empirical literature on TPACK in relation to other digital technologies. This body of work offers a richer evidence base and provides a more comprehensive understanding of how teachers' demographic and professional characteristics relate to their technology-related knowledge.

In the broader empirical literature on TPACK, gender emerges as one of the most consistently documented sources of variation. A large-scale meta-analysis by Scherer et al. (2026) examined gender disparities based on 915 effect sizes from 102 primary studies and 420 teacher samples, with a total of 681,745 teachers. Results show small differences favouring male teachers, particularly in TK and TPK. However, because the reported effects varied across studies, samples, countries, measures, and time, Scherer et al. (2026) emphasize the importance of examining additional moderating factors, including education level and country-level indicators (e.g., gender inequality and economic development), that shape the broader societal context in which teachers' digital self-beliefs develop.

Building on the notion that teachers' country of employment constitutes a relevant contextual characteristic that may influence TPACK, the empirical literature remains limited in the identification of cross-national differences in TPACK. As demonstrated by Padilla-Escorcia et al. (2025), most studies adopt a country-specific focus rather than a comparative perspective (due to interpretation issues in cross-country comparisons; see Castéra et al., 2020). Moreover, systematic reviews (e.g., Padilla-Escorcia et al., 2025) and most individual empirical studies do not examine cross-national differences in TPACK. An exception is the work of Castéra et al. (2020), who conducted a cross-national comparison of teacher educators across six European and Asian countries. Their findings show that the country teachers work in significantly influences TPACK, particularly in technology-related dimensions. They argue that national education systems, ICT infrastructure and teacher training traditions may likely explain these variations.

Next to gender and country of employment, age and teaching experience have also been identified as important factors influencing teachers' technological knowledge. Most dominantly, studies indicate a negative relationship between age and technological knowledge, with younger teachers demonstrating higher proficiency in digital technologies (e.g., Hamilton, 2020; Mahdum, 2015). This pattern likely reflects a cohort effect, as younger teachers have typically received initial training in contexts with greater access to digital technologies, whereas older teachers entered the profession when such tools were less prevalent.

A related factor alongside age is teaching experience. More experienced teachers may exhibit higher technological knowledge due to accumulated classroom practice and engagement in professional development initiatives (see e.g., Kumala et al., 2022;

Qiu et al., 2022). These contradictory findings suggest that the relationship between age, experience, and technological competence is complex, mediated by factors such as ongoing learning opportunities, institutional support, and individual motivation, which is also mentioned in the literature on AI-TPACK (Li et al., 2024).

Contradictory results can be found on the relationship between teachers' education level and how they score on technology-related dimensions of TPACK. Li et al., (2022), for instance, found that the seven dimensions of TPACK differed significantly according to teachers' highest acquired education level, where the higher the education level, the better the teachers' TPACK abilities. However, in research of Mahdum (2015), Plantedo (2023) and Hamilton (2013), teachers' education level was not significantly associated to TK when controlling for other factors such as age, gender and experience. The inconsistent association between education level and technology-related TPACK can be better understood in the context of the relatively recent integration of technology pedagogy into teacher education. Traditionally, programs focused more on subject content and general pedagogy, with limited attention to systematic technology integration. As such, higher degree attainment did not necessarily entail structured preparation for pedagogically meaningful technology use. Differences in TK may therefore reflect when and where teachers were trained, as well as their engagement in ongoing professional development, rather than degree level itself.

4.3. Research Hypotheses

Guided by prior work on TPACK and emerging evidence on AI-TPACK, we formulated hypotheses to examine whether teachers' background characteristics are associated with their perceived preparedness for AI-supported technologies. Given consistent findings for some predictors (e.g., gender, age), we state directional hypotheses where warranted and non-directional hypotheses where the literature is mixed. We test these hypotheses across four technology-related AI-TPACK dimensions (AI-TK, AI-TCK, AI-TPK, AI-TPACK)

H1. AI-related TPACK will differ across countries of employment.

H2. Age will be negatively associated with AI-related TPACK.

H3. Male teachers will report higher AI-related TPACK than female teachers.

H4. AI-related TPACK will differ across levels of educational degree.

H5. Teaching experience will be associated with AI-related TPACK.

H6. AI-related TPACK will differ across educational contexts.

H7. Teachers with additional professional roles (coach, IT coordinator, teacher trainer) will report higher AI-related TPACK.

4.4. Methods

4.4.1. Data Source and Sample

The analyses reported in this chapter draw on data from the vignette-based survey study administered in Qualtrics and described in Chapter 3. The focus is on respondents who completed the preparedness block, which included the AI-related TPACK instrument.

Participants' (*n* = 350) mean age was 43.44 years (*SD* = 10.45, range 22–68). Average teaching experience was 16.11 years (*SD* = 9.94, range 1–40). Regarding gender, 70.86% identified as women (*n* = 248), 28.00% as men (*n* = 98), and 1.14% as other/prefer not to say (*n* = 4). Participants' highest educational degree was Master's or equivalent for 53.43% (*n* = 187), Bachelor's or equivalent for 23.14% (*n* = 81), Doctoral or equivalent for 19.43% (*n* = 68), and Other for 4.00% (*n* = 14).

In terms of educational context, 42.29% worked in secondary education (*n* = 148), 26.86% in primary education (*n* = 94), 26.86% in higher education (*n* = 94), and 4.00% in other contexts (*n* = 14). Additional professional roles were reported by a minority of participants: 12.57% indicated a pedagogical coaching role (*n* = 44), 15.14% a teacher trainer role (*n* = 53), and 5.43% an IT coordination role (*n* = 19).

Participants were employed across nine country groups: Spain (30.00%, *n* = 105), Belgium (19.14%, *n* = 67), Finland (12.29%, *n* = 43), Netherlands (10.00%, *n* = 35), Malta (9.43%, *n* = 33), Germany (6.00%, *n* = 21), Italy (5.14%, *n* = 18), Poland (5.14%, *n* = 18), and Other countries (2.86%, *n* = 10).

4.4.2. Measures

4.4.2.1. AI-TPACK

Teachers' preparedness for AI-supported educational technologies was operationalized using an AI-TPACK self-report instrument adapted from Ning et al. (2024). Items were measured on a five-point Likert scale (1: Strongly disagree to 5: *Strongly agree*). The analyses focused exclusively on the technology-related dimensions of AI-TPACK: AI-TK, AI-TCK, AI-TPK, AI-TPACK. These dimensions reflect teachers' perceived ability to understand, interpret, and meaningfully apply AI technologies in educational settings, and are considered particularly relevant indicators of preparedness for AI-supported systems. Composite scale scores were computed by averaging item responses per dimension, with higher scores indicating higher perceived AI-related TPACK.

The technology-related AI-TPACK dimensions comprised 22 items across four subscales: AI-TK (5 items), AI-TCK (6 items), AI-TPK (6 items), and AI-TPACK (5 items) (Appendix A, Table A1). Because the original instrument was developed in Chinese, we used the published translated English wording as a basis and conducted a structured expert review to produce a clearer international English version. Educational technology and teacher education experts evaluated each item for (a) semantic clarity, (b) avoidance of double meanings, and (c) applicability across educational levels and national contexts. Revisions were restricted to wording and did not alter the target construct of any item. The final items (and a traceable overview of wording changes) are provided in Appendix A, Table A1.

4.4.2.2. Teacher Characteristics

Teacher characteristics were collected using self-report items. Participants reported their country of employment ("In which country are you currently teaching?"), age (open-ended numeric response), gender (woman, man, non-binary, or prefer not to say), and the educational context(s) in which they currently teach (multiple selection: primary, lower secondary, upper secondary, higher education, adult education, and special education). They also indicated their highest level of formal education completed (six-category item ranging from upper secondary to doctoral), years of teaching experience (open-ended numeric response; analyses used years working as a teacher/educator in total), and professional role(s) (multiple selection including teacher, school manager/director, pedagogical coach/mentor, IT coordinator, teacher trainer, and other).

For analysis, several variables were recoded to improve interpretability. Country of employment was recoded by pooling countries with fewer than 10 respondents (i.e., Angola, Canada, Chile, Grenada, Jordan, Portugal, Sweden, Switerland and the UK) into an "Other" category. Gender responses "Non-binary" and "Prefer not to say" were combined into a single category due to small cell sizes. Highest educational degree categories "Upper secondary education," "Post-secondary non-tertiary education," and "Short-cycle tertiary education" were combined into "Other," and the resulting variable was treated as an ordered factor (Other, Bachelor's, Master's, Doctoral). For educational context, the multi-select "Type(s) of education" item was transformed into a single categorical indicator reflecting participants' highest mainstream context (primary, secondary, higher education), while respondents selecting only adult education and/or special education were coded as "Other"; lower and upper secondary education were further collapsed into a single "Secondary education" category. Professional roles were operationalized as three binary indicators (i.e., pedagogical coach, teacher trainer, IT coordinator) derived from whether each role was selected in the multi-select roles item.

4.4.3. Data Analysis

Quantitative analyses were conducted using R (version 4.3.1; R Core Team, 2023). The code supporting the analyses is provided via the following link: https://osf.io/qu4ne/overview?view_only=8a9732f5c7cd4667b0201ba466d04059.

Descriptive statistics (means, standard deviations, and distributions) were first computed for each AI-related TPACK dimension to provide an overview of teachers' overall preparedness for AI-supported educational technologies. Internal reliability statistics (i.e., Cronbach's alpha and McDonald's omega) were calculated using the *psych* package (version 2.6.1; Revelle, 2026).

To address the research question, a single multiple multivariate regression analysis was performed. Teachers' AI-related TPACK dimensions were entered simultaneously as dependent variables, while country of employment, age, gender, educational degree, teaching experience, educational context, and professional roles were included as independent variables. This multivariate approach allows for the examination of associations between multiple correlated outcome variables and a set of predictors within a single analytical framework (Ganesh, 2010).

Continuous predictors were mean-centred prior to analysis to reduce potential multicollinearity. Multicollinearity diagnostics indicated no problematic levels of collinearity among the predictors. Multivariate significance was evaluated using Wilks' Lambda based on Type III sums of squares as implemented in the *car* package (version 3.1-2; Fox et al., 2001).

For predictors showing significant multivariate effects, follow-up univariate analyses were conducted to examine associations with individual AI-related TPACK dimensions. For the univariate models, effect sizes were reported as partial eta squared (ηp^2), computed from the Type III sums of squares within the model including all predictors. These were calculated using the *effectsize* package (version 0.8.5; Ben-Shachar et al., 2019). Following conventional guidelines for ηp^2, values of approximately .01, .06, and .14 were interpreted as small, medium, and large effects, respectively (Richardson, 2011).

To facilitate interpretation of categorical effects, estimated marginal means were calculated using the *emmeans* package (version 1.8.7; Lenth et al., 2017). Pairwise comparisons were adjusted for multiple testing using Tukey's honestly significant difference correction (e.g., Lane, 2010). The conventional *p* value reflects the probability of observing the obtained results assuming the null hypothesis is correct but it does not provide insight into how large or practically meaningful the observed effects are (Diener, 2010). Therefore, for significant categorical predictors from the univariate models, standardized mean differences, Cohen's (1988) *d* were computed for pairwise comparisons based on estimated marginal means using the *eff_size* function from the *emmeans* package. Effect sizes were calculated using the model-based residual standard deviation as the standardizer (i.e., the square root of the mean squared error from the corresponding linear model). Confidence intervals for Cohen's *d* were derived using the residual degrees of freedom of the respective model. For interpreting these effect sizes, we follow Cohen's (1988) widely used guidelines, which classify values of $d \approx 0.20$ as small, $d \approx 0.50$ as medium, and $d \approx 0.80$ as large.

4.5. Results

Results are presented in two steps, starting with descriptive statistics and followed by multivariate and univariate inferential analyses. Descriptive statistics for AI-related TPACK dimensions are provided in Table 2. A full overview of the output from analyses is provided via the following link:

https://osf.io/qu4ne/overview?view_only=8a9732f5c7cd4667b0201ba466d04059.

Table 2

Descriptive Statistics and Intercorrelations for AI-Related TPACK Dimensions

Dimension	*M*	*SD*	Omega	Alpha	AI-TK	AI-TCK	AI-TPK
AI-TK	3.34	0.85	.84	.83	–		
AI-TCK	3.24	0.91	0.90	0.90	0.88	–	
AI-TPK	2.94	0.95	0.90	0.90	0.80	0.82	–
AI-TPACK	3.23	0.95	0.88	0.88	0.84	0.86	0.87

Note. Values below the diagonal represent Pearson correlations. Reported internal consistency statistics (i.e., Omega and Alpha) are estimated across languages.

4.5.1. Multivariate Analysis

Multicollinearity was assessed using variance inflation factors. All GVIF-adjusted values were well below commonly accepted thresholds, indicating no problematic multicollinearity among the predictors.

A multivariate multiple regression analysis was conducted to examine whether teachers' background characteristics and professional roles were associated with the AI-related TPACK dimensions. Using Wilks' Lambda as the multivariate test statistic, significant overall multivariate effects were found for country of employment ($\Lambda = .79$, $F_{(32,\ 1196.45)} = 2.45$, $p < .001$), gender ($\Lambda = .93$, $F_{(8,\ 648.00)} = 3.12$, $p = .002$), educational degree ($\Lambda = .90$, $F_{(12,\ 857.51)} = 3.00$, $p < .001$), and having an IT coordination role ($\Lambda = .97$, $F_{(4,\ 324)} = 2.63$, $p = .03$). No significant multivariate effects were observed for age ($\Lambda = .99$, $F_{(4,\ 324)} = 1.017$, $p = .40$), teaching experience ($\Lambda = .99$, $F_{(4,\ 324)} = 0.49$, $p = .74$), educational context ($\Lambda = .95$, $F_{(12,857.51)} = 1.43$, $p = .15$), or holding a pedagogical coaching ($\Lambda = .98$, $F_{(4,\ 324)} = 1.82$, $p = .12$) or teacher trainer role ($\Lambda = .98$, $F_{(4,\ 324)} = 2.03$, $p = .09$).

4.5.2. Univariate Follow-up Analyses

Separate univariate linear models were estimated for each AI-related TPACK dimension, including all predictors entered in the multivariate model. For clarity of interpretation, Table 3 reports the univariate follow-up results only for those predictors that showed a significant multivariate effect, including F values and effect sizes expressed as partial eta squared.

The univariate analyses revealed a significant overall effect of country on three AI-related TPACK dimensions (i.e., AI-TK, AI-TCK, and AI-TPACK). However, none of the Tukey-adjusted univariate pairwise comparisons between individual countries reached statistical significance (i.e., all $p > .05$).

Univariate follow-up analyses revealed significant associations between gender and AI-TK, AI-TCK and AI-TPACK. After Tukey adjustment, men reported significantly higher AI-TCK than women ($t_{327} = 3.37$, mean difference = 0.35, $p = .002$, $d = 0.42$, 95% CI [0.17, 0.66]), and significantly higher overall AI-TPACK ($t_{327} = 2.56$, mean difference = 0.28, $p = .03$, d = 0.32, 95% CI [0.07, 0.56]). The difference between men and women on AI-TK approached statistical significance ($t_{327} = 2.32$, mean difference = 0.22, $p = .054$, $d = 0.29$, 95% CI [0.04, 0.53]). Comparisons involving the other/not specified gender category were not statistically significant.

For education level one significant univariate association was found for AI-TK. However, the Tukey-adjusted pairwise comparisons between degree levels were not statistically significant.

Consistent and robust univariate effects were observed for holding an IT coordination role. Teachers with an IT coordination role reported significantly higher AI-TK than teachers without such a role ($t_{327} = 3.08$, mean difference = 0.56, $p = .002$, $d = 0.75$, 95% CI [0.27, 1.24]). Similar effects were found for AI-TCK ($t_{327} = 2.70$, mean difference = 0.55, $p = .01$, $d = 0.66$, 95% CI [0.18, 1.14]), AI-TPK ($t_{327} = 2.31$, mean difference = 0.51, $p = .02$, $d = 0.56$, 95% CI [0.08, 1.05]), and overall AI-TPACK ($t_{327} = 2.91$, mean difference = 0.63, $p = .004$, $d = 0.71$, 95% CI [0.23, 1.19]). These findings indicate a strong association between technology-focused professional roles and teachers' preparedness for AI-supported educational technologies.

Table 3

Univariate Follow-Up Analyses for AI-Related TPACK Dimensions

Predictor	AI-TK		AI-TCK		AI-TPK		AI-TPACK	
	F	ηp^2	*F*	ηp^2	*F*	ηp^2	*F*	ηp^2
Country	2.45*	.06	2.60**	.06	1.76	.04	2.06*	.05
Gender	4.10*	.02	5.82**	.03	1.21	.01	3.35*	.02
Education	2.95*	.03	1.26	.01	1.12	.01	0.43	.004
IT role	9.51**	.03	7.29**	.02	5.31*	.02	8.44**	.03

Note. F values represent Type III tests from separate univariate linear models for each outcome. ηp^2 = partial eta squared. All models included country of employment, age (centered), gender, educational degree, work experience (centered), educational context, pedagogical coaching role, teacher trainer role, and IT coordination role as predictors. *$p < .05$. **$p < .01$.

4.6. Discussion

This study examined whether teachers' perceived preparedness for AI-supported educational technologies operationalized via the technology-related AI-TPACK dimensions (AI-TK, AI-TCK, AI-TPK, AI-TPACK) varies as a function of teachers' characteristics and contextual characteristics.

4.6.1. Summary of Main Findings

Drawing on emerging AI-TPACK research (e.g., Ning et al., 2024; Karataş & Ataç, 2025; Chan & Tang, 2024; Li et al., 2025; Wang, 2024; An et al., 2025) and broader TPACK literature (e.g., Scherer et al., 2017; Castéra et al., 2020; Padilla-Escorcia et al., 2025; Scherer et al., 2026), we formulated and tested hypotheses about cross-national, demographic, and role-based differences in perceived AI-related preparedness.

Consistent with H1, country of employment showed evidence of meaningful variation for several AI-TPACK dimensions. The magnitude of the country effects ranged from small to approximately medium, with the clearest differentiation observed for AI-TK and AI-TCK and somewhat smaller effects for overall AI-TPACK. The absence of clear country-to-country contrasts suggests that the observed variation may be diffuse rather than driven by a small set of pronounced national differences.

H2 was not supported, as age did not show a meaningful association with the AI-TPACK outcomes once other predictors were considered. This contrasts with some prior work reporting lower levels of technology-related knowledge or confidence among older teachers in both AI-TPACK and broader technology integration research (e.g., Chan & Tang, 2024; An et al., 2025; Hamilton, 2020; Mahdum, 2015). A possible interpretation is that AI is a relatively recent and rapidly diffusing technological domain in education, which may reduce the cohort-based differences often observed for more established educational technologies. Another explanation is that some earlier studies relied on single-predictor analyses (e.g., An et al., 2025) rather than models that account for multiple predictors.

H3 received partial support. Gender differences emerged primarily for AI-TCK and overall AI-TPACK, with effects in the small range by conventional benchmarks, and a similar small pattern for AI-TK. This result is broadly consistent with the wider TPACK literature, where gender differences, if present, tend to be small and most visible in technology-related dimensions (Scherer et al., 2026). The pattern also aligns with emerging AI-TPACK (Karataş & Ataç, 2025). It is important to note that these differences reflect variation in self-perceptions rather than differences in actual competence or enacted instructional practice. One possible contributing factor is that individual differences in affective traits can shape how respondents evaluate their own abilities. For instance, empirical and meta-analytic research has shown mean gender differences in neuroticism (i.e., women scoring higher on average than men), and neuroticism is typically linked to more negative self-evaluations (Jorm, 1987; Weisberg, 2011).

H4 was weakly supported. Educational degree showed limited differentiation across the AI-TPACK outcomes, with effects in the small range and most evident for AI-TK. This pattern is consistent with the mixed evidence in the broader TPACK literature, where some studies find degree-related differences (e.g., Li et al., 2022) but others report no meaningful associations once demographic and contextual covariates are taken into account (e.g., Mahdum, 2015; Plantedo, 2023; Hamilton, 2013). A plausible explanation is that degree level is an imperfect proxy for AI-specific preparation, given that systematic training in AI-supported pedagogy is still emerging and may vary substantially by program, institution, and other context-related aspects.

H5 was not supported. Teaching experience did not uniquely relate to AI-TPACK outcomes when considered alongside other predictors.

Similarly, H6 was not supported. Educational context (primary, secondary, higher education) did not meaningfully differentiate AI-TPACK outcomes.

H7 was partially supported, with the strongest and most consistent evidence observed for the IT coordination role. Teachers who reported an IT coordination role differed from those without such a role across all four AI-TPACK dimensions. The unique effects of IT role in the univariate models were small, yet the adjusted mean differences between IT coordinators and non-coordinators were consistently in the medium range. By contrast, pedagogical coaching and teacher trainer roles did not show comparable differentiation. This could indicate that AI has not yet been systematically integrated into many coaching and teacher training practices. As AI becomes more institutionalized in professional development ecosystems, these roles may become more predictive of teachers' perceived AI-related preparedness.

4.6.2. Limitations

Several limitations should be considered when interpreting the findings. First, although the study included teachers from multiple countries and the multilingual administration is an important strength, the sample sizes for some countries were relatively small. In addition, the participating countries were based in Europe. Future research would benefit from including a broader international sample, with stronger representation across both Western and non-Western or Eastern contexts. Such work could offer a more nuanced understanding of how national, cultural, and policy-related conditions shape teachers' perceived preparedness for AI-supported educational technologies.

Second, the study relied on translated versions of a self-report AI-TPACK instrument. Although careful translation and adaptation procedures were applied, cross-language administration may still introduce subtle differences. Moreover, self-report measures capture teachers' perceived preparedness rather than their actual competence or enacted classroom practice. While this was not the aim of the present research, it is important to consider that such measures cannot fully establish whether teachers are able to apply AI-supported technologies effectively in practice. Future research should therefore complement self-report instruments with additional sources of evidence, such as expert ratings, performance-based assessments, classroom observations, portfolio tasks, or scenario-based evaluations of teachers' decision making.

4.6.3. Implications

The findings have several practical implications. The absence of clear differences between pedagogical coaches and teachers without such a role suggests that AI-related expertise is not yet systematically embedded in pedagogical support functions. This points to a need for targeted professional development not only for teachers, but also for those responsible for supporting teachers' instructional practice. In this regard, train-the-trainer initiatives may be especially promising. Equipping pedagogical coaches, teacher trainers, and similar support professionals with stronger AI-related pedagogical competencies may be a key step toward building sustainable support structures within schools and teacher education settings.

More broadly, the results suggest that professional development initiatives should be differentiated rather than one-size-fits-all. Since teachers in IT coordination roles consistently reported higher preparedness, these professionals may serve as valuable internal change agents in schools. Policy and professional development efforts should focus on building teachers' confidence, pedagogical understanding, and critical capacity to use such tools in meaningful and responsible ways.

4.7. Conclusion

Taken together, the findings suggest that differences in perceived AI-related preparedness are most clearly associated with (a) technology-focused professional roles and (b) gender, with (c) country context contributing small-to-approximately-medium differentiation at the overall level but without clear country-to-country contrasts. Age, teaching experience, and educational context did not uniquely differentiate AI-TPACK outcomes in this multivariable framework.

4.8. References

An, X., Bai, J., Li, Y., Zhou, Y., Chai, C. S., & Hong, J.-C. (2025). Development of an artificial intelligence Technology Pedagogical Content Knowledge (AI-TPACK) scale for human and AI concurrently as teachers. *The Asia-Pacific Education Researcher*. https://doi.org/10.1007/s40299-025-01045-2

Backfisch, I., Scherer, R., Schneider, J., & Lachner, A. (2025). How valid, really? A meta-analysis of the validity evidence of Technological Pedagogical and Content Knowledge (TPACK) self-report assessments. *Educational Research Review*, *49*, 100718. https://doi.org/10.1016/j.edurev.2025.100718

Ben-Shachar, M. S., Makowski, D., Lüdecke, D., Patil, I., Wiernik, B. M., Thériault, R., & Waggoner, P. (2019). *effectsize: Indices of effect size*. https://doi.org/10.32614/CRAN.package.effectsize

Castéra, J., Marre, C. C., Yok, M. C. K., Sherab, K., Impedovo, M. A., Sarapuu, T., Pedregosa, A. D., Malik, S. K., & Armand, H. (2020). Self-reported TPACK of teacher educators across six countries in Asia and Europe. *Education and Information Technologies*, *25*(4), 3003–3019. https://doi.org/10.1007/s10639-020-10106-6

Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Chan, K. K.-W., & Tang, W. K.-W. (2024). Evaluating English teachers' artificial intelligence readiness and training needs with a TPACK-based model. *World Journal of English Language*, *15*(1), 129. https://doi.org/10.5430/wjel.v15n1p129

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Diener, M. J. (2010). Cohen's d. In *The Corsini Encyclopedia of Psychology* (pp. 1–1). Wiley. https://doi.org/10.1002/9780470479216.corpsy0200

Dogan, S., Nalbantoglu, U. Y., Celik, I., & Agacli Dogan, N. (2025). Artificial intelligence professional development: A systematic review of TPACK, designs, and effects for teacher learning. *Professional Development in Education*, 1–28. https://doi.org/10.1080/19415257.2025.2454457

Fox, J., Weisberg, S., & Price, B. (2001). car: Companion to applied regression. In *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.package.car

Ganesh, S. (2010). Multivariate linear regression. In *International Encyclopedia of Education* (pp. 324–331). Elsevier. https://doi.org/10.1016/B978-0-08-044894-7.01350-6

Hamilton, C. (2013). *An investigation into the relationships between the Technological Pedagogical Content Knowledge of university teacher education faculty and their age, rank, and gender*.

Karataş, F., & Ataç, B. A. (2025). When TPACK meets artificial intelligence: Analyzing TPACK and AI-TPACK components through structural equation modelling. *Education and Information Technologies*, *30*(7), 8979–9004. https://doi.org/10.1007/s10639-024-13164-2

Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The Technological Pedagogical Content Knowledge framework. In *Handbook of Research on Educational Communications and Technology* (pp. 101–111). Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_9

Kopcha, T. J., Ottenbreit-Leftwich, A., Jung, J., & Baser, D. (2014). Examining the TPACK framework through the convergent and discriminant validity of two measures. *Computers & Education*, *78*, 87–96. https://doi.org/10.1016/j.compedu.2014.05.003

Kumala, F. N., Ghufron, A., & Pujiastuti, P. (2022). Elementary school teachers' TPACK profile in science teaching based on demographic factors. *International Journal of Instruction*, *15*(4), 77–100. https://doi.org/10.29333/iji.2022.1545a

Lane, D. M. (2010). Tukey's Honestly Significant Difference (HSD). In *Encyclopedia of Research Design* (Vol. 0, pp. 1566–1570). SAGE Publications, Inc. https://doi.org/10.4135/9781412961288.n478

Lenth, R. v., & Piaskowski, J. (2017). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://doi.org/10.32614/CRAN.package.emmeans

Li, S., Liu, Y., & Su, Y.-S. (2022). Differential analysis of teachers' Technological Pedagogical Content Knowledge (TPACK) abilities according to teaching stages and educational levels. *Sustainability*, *14*(12), 7176. https://doi.org/10.3390/su14127176

Mahdum, M. (2015). Technological Pedagogical and Content Knowledge (TPACK) of English teachers in Pekanbaru, Riau, Indonesia. *Mediterranean Journal of Social Sciences*. https://doi.org/10.5901/mjss.2015.v6n5s1p168

Mishra, P. (2019). Considering contextual knowledge: The TPACK diagram gets an upgrade. *Journal of Digital Learning in Teacher Education*, *35*(2), 76–78. https://doi.org/10.1080/21532974.2019.1588611

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, *16*(3), 978. https://doi.org/10.3390/su16030978

Padilla-Escorcia, I. A., García-Rodríguez, M. L., & Aguilar-González, Á. (2025). Mathematics teachers' knowledge for teaching with digital technologies: A systematic review of studies from 2010 to 2025. *Education Sciences*, *15*(12), 1598. https://doi.org/10.3390/educsci15121598

Plantado, J. C. (2023). Usage of Classroom Technology and the Technological Pedagogical Content Knowledge (TPACK) of Mathematics Teachers. *Southeast Asian Mathematics Education Journal*, *13*(1), 43-56.

Qiu, C., He, H., Chen, G., & Xiong, M. (2022). Pre-service teachers' perceptions of technological pedagogical content knowledge in mainland China: A survey of teachers of Chinese as a second language. *Education and Information Technologies*, *27*(5), 6367–6391. https://doi.org/10.1007/s10639-022-10888-x

R Core Team. (2021). *R: A language and environment for statistical computing.* https://www.R-project.org/

Revelle, W. (2026). psych: Procedures for psychological, psychometric, and personality research. In *CRAN: Contributed Packages* (2.6.1). https://doi.org/10.32614/CRAN.package.psych

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Scherer, R., Tondeur, J., & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the

Technological, Pedagogical, and Content Knowledge (TPACK) model. *Computers & Education*, *112*, 1–17. https://doi.org/10.1016/j.compedu.2017.04.012

Scherer, R., Campos, D. G., Fütterer, T., & Siddiq, F. (2025). *Unraveling gender disparities in teachers' technological pedagogical and content knowledge: A large-scale meta-analytic review*. https://doi.org/10.31234/osf.io/gqm7a

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological Pedagogical Content Knowledge (TPACK). *Journal of Research on Technology in Education*, *42*(2), 123–149. https://doi.org/10.1080/15391523.2009.10782544

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Wang, K., Ruan, Q., Zhang, X., Fu, C., & Duan, B. (2024). Pre-service teachers' GenAI anxiety, technology self-efficacy, and TPACK: Their structural relations with behavioral intention to design GenAI-assisted teaching. *Behavioral Sciences*, *14*(5), 373. https://doi.org/10.3390/bs14050373

4.9. Appendices

4.9.1. Appendix A

Table A1 provides an overview of the original and modified wording of the AI-TPACK instrument items.

Table A1

Original and Modified Wording of the Technology-Related AI-TPACK Items (Adapted From Ning et al., 2024)

Subscale	*Item*	*Original wording (Ning et al., 2024)*	*Modified wording*
AI-TK	1	I am familiar with commonly encountered AI technologies in the educational environment.	I am familiar with commonly encountered AI technologies in the educational environment.
	2	I possess the capability to easily acquire AI technologies necessary for teaching.	I am able to easily learn to use new AI technologies for teaching.

Subscale	*Item*	*Original wording (Ning et al., 2024)*	*Modified wording*
	3	I frequently incorporate AI technologies in the pedagogical context.	I frequently use AI technologies in my teaching.
	4	I am proficient in using AI technologies to enhance the instructional process.	I am proficient in using AI technologies to enhance the instructional process.
	5	I am knowledgeable about using AI technologies for interactive teaching purposes.	I know how to use AI technologies for interactive teaching purposes.
AI-TCK	1	I am familiar with AI in specific academic domains, such as mathematical intelligent tutoring systems.	I am familiar with AI in specific subjects.
	2	I am capable of effortlessly using AI in specific academic domains.	I am capable of effortlessly using AI in specific subjects.
	3	I am proficient in using AI to update my knowledge base within the academic discipline.	I am proficient in using AI to update my knowledge base within my subjects.
	4	I can select appropriate AI tools based on the subject matter I am teaching.	I can select appropriate AI tools based on the subject matter I am teaching.
	5	I am adept at using AI to effectively enhance students comprehension of the material.	I am proficient in using AI to help students better understand the learning content.
	6	I can use AI to broaden the knowledge horizons of students.	I am able to use AI to expand students' knowledge.
AI-TPK	1	I am capable of using AI to enhance my pedagogical perspectives.	I am able to use AI to enhance my pedagogical approaches.
	2	I am able to apply appropriate AI in various teaching activities.	I am able to apply appropriate AI in various teaching activities.
	3	I have the capacity to select AI to sustain students motivation and interest.	I am able to use AI to keep students motivated and interested.

Subscale	*Item*	*Original wording (Ning et al., 2024)*	*Modified wording*
	4	I can apply AI to assess the learning outcomes of students.	I can apply AI to assess the learning outcomes of students.
	5	I am proficient in using AI to optimize classroom instructional management.	I am proficient in using AI to optimize classroom instructional management.
	6	I possess the ability to explain information derived from AI to provide real-time feedback.	I am able to provide real-time feedback using information from AI.
AI-TPACK	1	I am knowledgeable in integrating AI with educational content and teaching methods to improve classroom teaching efficiency and effectiveness.	I know how to integrate AI with subject content and teaching methods to make classroom instruction more effective.
	2	I am capable of selecting appropriate teaching methods and AI based on the educational content for instruction.	I am able to choose suitable teaching methods and AI tools based on the subject content I teach.
	3	I can use AI to create, simulate, and adapt scenarios that are in line with the educational content.	I can use AI to create, simulate, and adapt scenarios that are in line with the educational content.
	4	I can use personalized AI to select suitable teaching methods as well as guide students in practical learning.	I can use AI to choose suitable teaching methods and guide student learning.
	5	I will use AI for self-directed learning, further deepening my subject knowledge and understanding of educational pedagogical theories.	I use AI to deepen my own professional learning.

5. System Autonomy and Teachers' Acceptance

Chapter authors: Siem Buseyne, Glen Joris , Konstantinos Tsiakas, Tine van Daal, Leen Catrysse, Vincent Donche, Sven De Maeyer, Halszka Jarodzka, and David Gijbels

Note: This chapter reflects work that has been prepared as a journal manuscript and submitted for peer review. Upon acceptance, the journal article will serve as the final, citable version.

5.1. Introduction

Artificial intelligence in education (AIEd) has become a central focus in educational policy and research. AIEd, as shown by Wang et al. (2024), captures a wide range of systems. However, much of the recent attention has concentrated on interactive generative Artificial Intelligent (AI) chatbots such as ChatGPT (e.g., Ravšelj et al., 2025). Alongside these chatbot systems, however, a different class of AIEd systems remains relatively underexplored: behavioral data-driven AIEd that uses fine-grained trace data (e.g., clicks, keystrokes, gestures, physiological signals, or eye movements) to support teachers and students. Recent work has begun to conceptualize the diversity of such systems (e.g., Wang et al., 2024). An important theme in this literature is the degree of system and human autonomy, that is, the extent to which an AI system merely mirrors and visualizes data, generates alerts, or makes and enacts recommendations with limited human involvement (Cukurova, 2025; Molenaar, 2022). System autonomy is central in several conceptual AIEd frameworks. Molenaar (2022), for example, proposed the Six Levels of Automation model, which conceptualizes system configurations along a spectrum of decreasing teacher autonomy and increasing system autonomy. Similarly, Cukurova (2025) distinguishes between configurations that differ in how responsibilities and control are distributed between AI-driven automation and human decision-making. Within this line of work, hybrid intelligent systems are understood as socio-technical configurations in which humans and AI systems combine distinct but complementary capabilities to accomplish tasks at a level of performance that neither could achieve alone (Dellermann et al., 2019; Kamar, 2016).

Understanding how teachers evaluate hybrid intelligent systems across varying levels of autonomy is crucial, as teachers are central decision-makers in educational practice, and their acceptance and willingness to rely on AI systems ultimately shape whether and how such tools are implemented in classrooms. Yet, to our knowledge, despite its prominence in theoretical work, empirical research has not systematically examined how different levels of autonomy in hybrid intelligent systems affect teachers' technology acceptance and other perceptions. While the Technology Acceptance Model (TAM; Davis, 1989) and its extensions (e.g., Venkatesh et al., 2003) have been widely applied to study educational technology adoption, much of this empirical work predates contemporary systems and was developed largely outside AI-driven contexts. Only recently has TAM-based research begun to engage with AI-specific applications in education, most notably generative AI systems such as chatbots (e.g., Kong et al., 2024). This emerging body of work signals a shift in focus, but remains fragmented and does not yet account for key AI-specific characteristics, such as varying levels of system autonomy, particularly in hybrid intelligent systems.

The present study addresses this gap by investigating teachers' acceptance of the functionalities of a hybrid intelligence system, presented in vignettes that vary in level of autonomy. Building on TAM (Davis, 1989) and the hybrid-intelligence frameworks of Molenaar (2022), we examine how different autonomy configurations influence teachers' perceptions of the system functionalities. By adopting a vignette-based approach, the study allows teachers to evaluate different aspects of hybrid intelligent systems in concrete, context-rich scenarios, which has been shown to facilitate more nuanced insights into perceptions and concerns than abstract questionnaire items alone (Skilling & Stylianides, 2020).

5.2. Theoretical Framework

5.2.1. Technology Acceptance Model

TAM, first introduced by Davis (1989) posits that perceived usefulness and perceived ease of use predict behavioral intention to use technology, which in turn leads to actual use. Throughout the years, TAM has been adopted in various contexts and has also been adapted by several researchers. For instance, Venkatesh and Davis (2000) expanded TAM to TAM2 by including both social influence processes (e.g., subjective norm, image, voluntariness) and cognitive instrumental determinants (e.g., job relevance, output quality, result demonstrability) as key predictors of perceived usefulness and usage intentions. The Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003) complements TAM1 and TAM2 by unifying constructs from eight acceptance theories and reorganizing them into four central determinants of technology adoption (i.e., performance expectancy, effort expectancy, social influence, and facilitating conditions) while also introducing moderating effects of gender, age, experience, and voluntariness.

Although empirical research on teachers' acceptance of AIEd is still in its early stages, recent work demonstrates that TAM remains the most widely used model to explain teachers' AI adoption and perceptions towards AI-supported educational technologies (Xue et al., 2025). TAM has also been specifically adapted to the context of AIEd. For example, Kong et al. (2024) tailored TAM measures to reflect teachers' perceptions of generative AI in teaching, thereby increasing conceptual fit and measurement validity within AIEd contexts.

In addition to core TAM constructs, Kong et al. (2024) highlight trust as a promising avenue for future AIEd research. Trust in AI has been conceptualized as multi-faceted, encompassing perceptions of reliability, competence, fairness, and value alignment (Araujo et al., 2020). Within educational contexts specifically, trust has been identified as a central determinant of the willingness to adopt AI-powered systems (Nazaretsky et al., 2025). Overall, this research points to the need for a deeper understanding of how teachers perceive, interpret, and evaluate AI technologies, including potential psychological levers and barriers shaping their acceptance, in particular when system autonomy varies.

5.2.2. Teachers' perception of Hybrid Intelligence Systems

While several recent studies have applied TAM to investigate teachers' acceptance of generative AI tools (Kılıç & Çelik, 2025; Şimşek et al., 2025), this research has largely centered on text-generative systems, most notably chatbots. Much less is known about how teachers evaluate other categories of AIEd systems, particularly those that rely on behavioral, multimodal, or real-time learner data. This is notable because many emerging AIEd technologies differ substantially from chatbot systems, for instance in terms of functionality. One particularly promising example is the use of an eye-tracking data-driven AI system to support learners and teachers during reading tasks. Such systems can monitor students' reading behaviors and translate these data into actionable insights or targeted recommendations for teachers to intervene in an informed manner to better support learners. For a broader overview of different AIEd systems in terms of functionality, we refer the reader to Wang et al. (2024).

A substantial subset of these systems can be understood within the broader paradigm of hybrid-intelligent educational systems, in which humans and AI collaborate by combining their distinct but complementary capabilities to support instruction and learning (Cukurova, 2025; Dellermann et al., 2019; Kamar, 2016). A defining characteristic of such systems is their potential to operate at different levels of system autonomy. Molenaar (2022) highlighted system autonomy as a critical design dimension that shapes how teachers interact with AI. Drawing on insights from previous research, Molenaar (2022) proposed a Six Levels of Automation model that articulates how responsibility shifts from teacher to AI across increasingly automated system configurations. At the lowest level (L0), Teacher Only, all monitoring and instructional decisions remain under full human control. In Teacher Assistance (L1), AI supports the teacher by detecting and diagnosing learner behaviour and visualizing this information (e.g., through dashboards), but it does not act on learners directly. For example, the system may present patterns about students' reading pace and rereading behavior to the teacher via a dashboard, without taking any intervention initiatives by itself. In Partial Automation (L2), AI begins to take over specific, bounded instructional tasks, such as providing small, real-time micro-interventions (e.g., short explanations of difficult terms), while the teacher continues to supervise and regulate most instructional decisions. Conditional Automation (L3) represents a shift in monitoring responsibility, where AI not only executes a broader set of adaptive actions but also initiates alerts or recommendations when learners require support, making timely communication between AI and the teacher essential. In such configurations, the system may, for instance, adapt selected instructional elements and notify the teacher when a learner's progress deviates from expectations, while recommending possible follow-up actions.

In High Automation (L4), most instructional decisions and adaptive actions are performed autonomously by the AI system, with the teacher intervening primarily when signaled (by the system). Here, the system could autonomously guide learners through individualized reading paths by adjusting task difficulty, providing scaffolds or enrichment, and modifying activities based on engagement indicators, alerting the teacher only in exceptional cases. Finally, Full Automation (L5) reflects a hypothetical state in which the AI-system independently manages all instructional processes without human oversight.

Together, these levels reflect fundamentally different human-AI role configurations, expectations for transparency, and demands on teacher oversight. As system autonomy increases, so do potential benefits (e.g., increased responsiveness, efficiency), but also potential risks (e.g., reduced human control, trust concerns, opaqueness of system decisions). While hybrid intelligence systems can reduce cognitive workload by automating routine tasks, they must be carefully designed to prevent teacher over-reliance and ensure active cognitive engagement. Rather than assuming that greater automation necessarily enhances system effectiveness, for instance, by improving student learning outcomes or increasing instructional efficiency, it is essential to examine how autonomy configurations are evaluated by teachers. Investigating teachers' responses to varying degrees of delegated control enables us to assess whether system autonomy functions as a meaningful determinant of technology acceptance within hybrid-intelligent educational systems. Without insights into these perceptions, systems may be designed at autonomy levels that provoke resistance, misuse, overreliance, or frictions in terms of regulation of the learning and teaching process, undermining the effectiveness of these systems. Understanding the role of teachers' perceptions is therefore essential for designing and implementing hybrid-intelligent AIEd tools that are both pedagogically aligned and adoption-ready. From a value-sensitive design perspective, such insight is also needed to understand how human values are interpreted and prioritised in teaching practice when evaluating AI-supported educational systems (Shen et al., 2025). However, despite its importance, system autonomy remains unexamined in empirical studies of teacher acceptance, including those grounded in TAM.

5.2.3. Research Aims, Research Questions, and Hypotheses

To address the abovementioned gap, the present study systematically varies system autonomy across the four central levels (i.e., L1, L2, L3, and L4) of the Six Levels of Automation Model (Molenaar, 2022) presented in the previous section, and examines how teachers evaluate a hypothetical behavioral-data-driven AIEd tool operating under each configuration. We excluded the endpoints of full teacher control (L0) and full AI control (L5) as they fall outside the scope of hybrid intelligence interactions.

The first research question (RQ) focuses on how system autonomy shapes core technology-acceptance dimensions. Specifically, RQ1 asks: How do the four automation levels relate to participants' evaluations of the system in terms of (a) perceived ease of use, (b) perceived usefulness, and (c) attitude toward using the system?

Intention to use, a core TAM outcome, reflects an end-stage adoption judgment that may be harder to evaluate in absolute terms for these hypothetical scenarios. Likewise, trust, although not part of TAM, is a critical factor for autonomous AIEd systems and is inherently comparative across different autonomy configurations. Rather than relying solely on isolated item ratings, the vignette-based design allows teachers to evaluate multiple system configurations side by side, supporting more holistic and comparative judgments. To capture these evaluations, we therefore examined how teachers ranked the four system versions in terms of intention to use and trust. This comparative approach offers a complementary perspective to traditional survey-based assessments by reducing reliance on abstract judgments, mitigating socially desirable responding, and enabling the assessment of relative preferences grounded in rich, contextually embedded representations of AI-supported systems.

RQ2 therefore asks: How do the four automation levels differ in terms of teachers' (a) intention to use and (b) trust in the system, as reflected in their comparative rankings of the system types?

5.3. Method

The design, conduct, and reporting of this vignette experiment follow the APA Journal Article Reporting Standards for Quantitative Research (JARS-Quant; American Psychological Association, 2024). This study was approved by the The Ethics Committee for Social Sciences and Humanities (SHW_2025_47_1). Supplemental materials referred to in this study can be found via the following link: https://osf.io/f2g4s/overview?view_only=5c4f0f163d00403abd3532c0e7644395.

5.3.1. Participants and Recruitment

The target population consisted of educators working in formal education across nine European countries (i.e., Belgium, Finland, Germany, Italy, Malta, Poland, Sweden, Spain, and the Netherlands). The study was conducted as part of a broader multi-country European research initiative. The participating countries reflect the project's predefined sampling structure. Participants were recruited via social media, mailing lists of partner schools, universities, and educational institutions, teacher associations, and teacher-training programs. Snowball sampling was encouraged by inviting participants to forward the survey link within their professional networks. Participants did not receive any compensation for completing the survey.

A total of 350 respondents completed at least the first part of the questionnaire and correctly answered a bogus attention-check question, which served as the inclusion criterion for the present study. From this sample, 237 participants completed the full survey, including at least three out of four vignettes. An overview of the sample characteristics is provided in Table 1.

Table 1

Sample Characteristics

Variable	Category	*n*	%	Mean (SD)
Age (years)	–	349	–	43.44 (10.45)
Overall teaching experience (years)	–	350	–	16.11 (9.94)
Gender	Woman	248	71	–
	Man	98	28	–
	Non-binary / other	4	1	–
Country	Belgium	67	19	–
	Finland	43	12	–
	Germany	21	6	–
	Italy	18	5	–
	Malta	33	9	–
	Netherlands	35	10	–
	Poland	18	5	–
	Spain	105	30	–
	Other countries	10	3	–
User language	Dutch	98	28	–
	Spanish	110	31	–
	English	49	14	–
	Finnish	40	11	–
	German	21	6	–
	Italian	13	4	–
	Polish	17	5	–
	Swedish	2	1	–
Degree	Upper secondary	7	2	–
	Post-secondary non-tertiary	3	1	–
	Short-cycle tertiary	4	1	–
	Bachelor's or equivalent	81	23	–
	Master's or equivalent	187	53	–
	Doctoral or equivalent	68	19	–
Education Context	Primary education	94	27	–
	Lower secondary education	44	13	–
	Upper secondary education	104	30	–
	Higher education	94	27	–
	Other	14	4	–

Note. For education context, participants could indicate multiple options at the same time; this was recoded to the highest level.

5.3.2. Design and Procedure

At the start of the questionnaire, after providing informed consent, participants reported demographic and professional background information. An instructed-response item (i.e., bogus item) served as an attention check, and participants who failed this item (*n* = 24) were excluded from analyses.

For RQ1, a within-subjects (repeated-measures) vignette experiment was employed with one manipulated factor, Automation Level, comprising four levels (L1–L4). The four levels corresponded to the central levels of the Six Levels of Automation Model (Molenaar, 2022). In L1 (teacher assistance), the system functioned as an observation-only dashboard. In L2 (partial automation), the system provided small, real-time micro-interventions while the teacher retained primary responsibility. In L3 (conditional automation), the system adjusted selected instructional elements and recommended teacher interventions. In L4 (high automation), the system created fully autonomous individualized learning paths and alerted the teacher only in exceptional cases. The reading comprehension context was intentionally chosen and framed in a generic manner to remain applicable across educational levels and instructional settings. Full vignette texts are provided in the supplemental materials. To illustrate the design and content of the experimental manipulation, Vignette 2, representing Level 2, is presented below:

> As students read, the system detects individual difficulties and offers small, real-time support actions for specific tasks. For example, Robin sees a brief explanation of a difficult term after rereading it several times. Alexis receives a hint to continue reading and come back later if needed, when pausing too long at a particular sentence. However, the system does not make further interventions, such as adjusting the difficulty of the reading materials. The dashboard can show you which interventions have been triggered for which students, and when they occurred. As the teacher, you remain primarily responsible for instruction and can use this information to guide your follow-up. However, the system does respond directly to certain student behaviors. The system operates automatically in the background, while you maintain an overview via the dashboard.

Before viewing the vignettes, participants were asked to anchor their responses by first envisioning a specific group of students, preferably aged nine years or older, and to indicate the group's average age, as well as a topic from a predefined list of reading or text-based subjects. This procedure is consistent with recommendations for vignette construction that emphasize realistic and meaningful contextualization to support valid interpretation of hypothetical scenarios (e.g., Skilling & Stylianides, 2020).

Qualtrics randomly selected three of the four vignettes for each participant and randomized their presentation order. We limited the number of vignettes to three to prevent the questionnaire from becoming too long and to minimize redundant repeated measures. After each vignette, participants completed a fixed set of questionnaire items described in the Measures section. After completing the three randomized vignettes, participants had the option to view and respond to the remaining vignette. As a result of this design, the number of responses per vignette was comparable but not identical, ranging from n = 214 to n = 231 across the four automation levels.

At the end of the survey, participants completed two drag-and-drop ranking tasks comparing the four automation levels in terms of their behavioral intention to use it and their trust towards it (1 = highest, 4 = lowest). These ranking tasks were shown to all participants, enabling full comparative ordinal outcomes.

The survey instrument was pilot-tested and iteratively refined using think-aloud interviews with one participant each from primary, secondary, and higher education. After finalization, the instrument was translated from English into Dutch, Finnish, French, German, Italian, Polish, Spanish, and Swedish by trained translators to align with the linguistic backgrounds of participants in the targeted countries. The translations were reviewed by native-speaking researchers to verify clarity and equivalence.

5.3.3. Measures

For each vignette, technology acceptance was assessed using items adapted from Kong et al. (2024). Perceived usefulness (three items), perceived ease of use (three items), and attitude toward using the system (three items) were each measured on 7-point Likert-type scales. Perceived system autonomy was assessed with four items on a 7-point Likert-type scale adapted from McGrath et al. (2025). Table 2 presents an overview of all multi-item Likert-type scale measures included in the study.

Additional questions were added about how such tools should be designed to support teaching. These items fall outside the scope of the present study and were therefore not included in the analyses. Finally, behavioral intention to use and trust across the four automation levels were assessed with the two ranking tasks described earlier.

Table 2

Internal Consistency of Multi-Item Scales

Domain	Scale	*k*	Response format	Mc Donald ω	Cronbach α
AI TAM (per vignette; Kong et al., 2024)	Perceived Usefulness	3	7-point Likert	.94	.94
	Perceived Ease of Use	3	7-point Likert	.81	.81
	Attitude Toward Using	3	7-point Likert	.93	.93
Perceived System Autonomy (per vignette; McGrath et al., 2025)		4	7-point Likert	.91	.91

Note. k = number of items. For repeated measures per vignette, average values of internal consistency levels are reported.

Internal consistency was examined for all multi-item scales using McDonald's ω, which does not rely on the strict assumption of tau-equivalence (Béland et al., 2018). For comparability with prior work, Cronbach's α was also reported. Estimates were calculated on the final analytic sample after excluding participants who failed the attention check. As shown in Table 2, general AI acceptance scales, vignette-specific TAM measures, and the perceived system autonomy scale all demonstrated at least acceptable internal consistency (i.e., > .70).

5.3.4. Data Analysis

Quantitative analyses were performed in R (version 4.3.1; R Core Team, 2023). The code used for these analyses can be retrieved via the following link: https://osf.io/f2g4s/overview?view_only=5c4f0f163d00403abd3532c0e7644395.

Internal reliability measures (i.e., McDonald's ω and Cronbach's α) for multi-item scales were calculated using the *psych* package (version 2.3.6; Revelle, 2024). In what follows an overview of the analyses is given per hypothesis.

5.3.4.1. Manipulation Check

Before examining the main acceptance outcomes, we first verified whether the vignette conditions successfully elicited the intended differences in perceived system autonomy (i.e., manipulation check). Establishing this distinction is essential to ensure that participants meaningfully differentiated between the automation levels embedded in the vignettes. To this end, we conducted a linear mixed-effects analysis with perceived system autonomy as the dependent variable and automation level as a fixed effect. The model included a random intercept for participants to account for the repeated-measures structure of the data. Automation level was treated as a categorical predictor with levels ordered from lowest to highest autonomy. The model was estimated using the *lme4* package (version 1.1-34; Bates et al., 2015), and Type III tests of fixed effects were obtained using the *car* package (version 3.1-2; Fox et al., 2001). Estimated marginal means for each automation level were computed using the *emmeans* package (version 1.8.7; Lenth & Piaskowski, 2017), and pairwise comparisons among the four levels were performed with Tukey-adjusted p-values to control for multiple testing. Using the same package, via the *eff_size* function, Cohen's (1988) *d* standardized effect sizes for pairwise contrasts were computed based on the estimated marginal means, with the pooled standard deviation derived from the sum of the participant-level random intercept variance and the residual variance of the mixed-effects model.

5.3.4.2. Hypothesis 1

Following the manipulation check, we examined whether differences in system autonomy translated into differences in teachers' acceptance-related evaluations of the system. To this end, we estimated three separate linear mixed-effects models corresponding to the three TAM-based outcome variables: perceived ease of use (H1a), perceived usefulness (H1b), and attitude toward using the system (H1c). In each model, Automation Level (four levels; within-subjects factor) was specified as the primary predictor.

Because each participant evaluated multiple vignette conditions, all models included a random intercept for participants to account for the repeated-measures structure of the data. In addition to Automation Level, Country was included as a covariate to adjust for potential cross-national differences in perceived relevance or pedagogical applicability related to differences in educational contexts and policy environments.

All models followed the same mixed-effects modeling framework and estimation procedures as the manipulation check outlined above. Where appropriate, estimated marginal means, pairwise contrasts, and effect sizes were computed to facilitate interpretation of differences between autonomy configurations.

5.3.4.3. Hypothesis 2

To address RQ2 and H2, participants' (a) intention to use and (b) trust were examined using two drag-and-drop ranking tasks, in which the four automation levels were ranked from 1 (highest) to 4 (lowest). These tasks were designed to elicit explicit comparative judgments, allowing participants to directly express relative preferences among the system configurations rather than evaluating each system in isolation.

The resulting ranking data were analyzed using Plackett-Luce models, attributed to Plackett (1975) and Luce (1959). These models are specifically designed for complete ranking data and estimate the relative worth of each item based on how frequently it is ranked above competing alternatives. Separate Plackett-Luce models were fitted for intention to use and trust. Rankings were converted into ranking objects and analyzed using the *PlackettLuce* package (version 0.4.4; Turner et al., 2025) . Model parameters represent relative preferences on a log-worth scale.

5.4. Results

5.4.1. Manipulation Check

As a manipulation check, linear mixed-effects model was estimated with perceived system autonomy as the dependent variable, automation level (vignette) as a fixed effect, and participant as a random intercept to account for the repeated-measures design. The model revealed a significant main effect of automation level, $\chi^2(3) = 210.16$, $p < .001$, indicating that perceived system autonomy differed across the four automation levels.

Perceived system autonomy increased monotonically with higher levels of automation. Pairwise comparisons and their effect sizes are reported in Table 3. Vignette 4 was perceived as significantly more autonomous than all other system versions (all $p < .001$), with a large effect between Vignette 1 and 4. Differences between Vignette 2 and Vignette 3 were small and did not reach statistical significance after Tukey adjustment, whereas all other pairwise contrasts were rather moderate and statistically significant. Overall, these findings confirm that, as intended, participants distinguished between the intended automation levels, thereby supporting the manipulation check.

Table 3

Pairwise Comparisons of Perceived System Autonomy across Vignettes

Contrast	*ΔM*	SE	*p* (Tukey)	Cohen's *d*	95% CI for *d*
V1 – V2	−0.74	0.11	< .001	−0.52	[−.67, −.36]
V1 – V3	−1.00	0.11	< .001	−0.70	[−.85, −.54]
V1 – V4	−1.57	0.11	< .001	−1.10	[−1.26, −.94]
V2 – V3	−0.26	0.11	.08	−0.18	[−.33, −.03]
V2 – V4	−0.84	0.11	< .001	−0.59	[−.74, −.43]
V3 – V4	−0.58	0.11	< .001	−0.40	[−.55, −.25]

Note. *ΔM* reflects differences in estimated marginal means. Effect sizes are reported as Cohen's *d* based on the model residual variance. Positive values indicate higher perceived autonomy for the first vignette in the contrast.

5.4.2. Technology acceptance dimensions

To examine H1, three linear mixed-effects models were estimated with perceived ease of use, perceived usefulness, and attitude toward using as dependent variables. In each model, automation level (vignette) was included as a fixed effect, country was entered as a control variable, and participant was specified as a random intercept to account for the repeated-measures design.

For perceived ease of use, the analysis revealed a significant main effect of automation level, $\chi^2(3) = 10.08$, $p = .02$. Pairwise comparisons, reported in Table 4, indicated that Vignette 1 was perceived as easier to use than Vignette 2, with a small-to-moderate effect size. None of the remaining pairwise contrasts reached statistical significance after Tukey adjustment, and perceived ease of use did not increase consistently across higher levels of automation.[1] Effect sizes for the remaining comparisons were small.

For perceived usefulness, the main effect of automation level was not statistically significant, $\chi^2(3) = 5.91$, $p = .12$. Pairwise comparisons showed no significant differences between automation levels. Corresponding effect sizes were small.

Similarly, for attitude toward using, the effect of automation level was not significant, $\chi^2(3) = 6.20$, $p = .10$. None of the pairwise comparisons between automation levels reached statistical significance, and all effect sizes were small.

---

[1] As a robustness check, we re-estimated all models for H1 after excluding responses from countries with fewer than ten participants to reduce the potential influence of small national subsamples. Results were highly similar to those obtained in the full sample, with consistent effects of automation level and comparable pairwise contrasts. Full results are reported in the Supplementary Materials.

Table 4

Pairwise Comparisons of Technology Acceptance Dimensions Across Automation Levels

Contrast	ΔM	SE	*p* (Tukey)	Cohen's *d*	95% CI for *d*
			Perceived ease of use		
V1 – V2	0.21	0.07	.01	0.17	[.06, .27]
V1 – V3	0.16	0.07	.11	0.12	[-.02, .23]
V1 – V4	0.14	0.07	.17	0.11	[-.01, .21]
V2 – V3	-0.06	0.07	.84	-0.04	[-.15, .06]
V2 – V4	-0.07	0.07	.71	-0.06	[-.16, .05]
V3 – V4	-0.02	0.07	.99	-0.01	[-.11, .09]
			Perceived usefulness		
V1 – V2	0.11	0.09	.59	0.08	[-.04, .20]
V1 – V3	-0.004	0.09	.99	0.003	[-.12, .12]
V1 – V4	-0.17	0.09	.21	0.12	[-.002, .24]
V2 – V3	-0.12	0.09	.54	-0.08	[-.20, .04]
V2 – V4	0.06	0.09	.91	-0.04	[-.08, .16]
V3 – V4	0.18	0.09	.18	0.12	[-.003, .24]
			Attitude toward using		
V1 – V2	0.18	0.09	.17	0.12	[-.01, .24]
V1 – V3	0.02	0.09	.99	0.01	[-.10, .13]
V1 – V4	0.14	0.08	.38	0.09	[-.02, .21]
V2 – V3	-0.16	0.08	.25	-0.11	[-.22, .01]
V2 – V4	-0.04	0.08	.96	-0.03	[-.14, .08]
V3 – V4	0.11	0.08	.52	0.08	[-.03, .19]

Note. *ΔM* reflects differences in estimated marginal means. Effect sizes are reported as Cohen's d based on the model residual variance. Positive values indicate higher perceived autonomy for the first vignette in the contrast.

5.4.3. Rankings of System Configurations

For H2, participants' rankings of the four system configurations in terms of intention to use were analyzed using a Plackett–Luce model. Relative to the observation-only system (System 1), the system that allowed small, teacher-controlled interventions (System 2) was significantly more likely to be ranked higher (β = 1.15, z = 9.37, p < .001). The system that independently adjusted instructional elements while keeping teachers informed (System 3) was also more likely to be ranked higher than the observation-only system (β = 0.58, z = 4.66, p < .001). In contrast, the fully autonomous system that alerted teachers only in rare cases (System 4) did not differ significantly from the observation-only system (β = 0.13, z = 1.09, p = .28). Overall, the model-implied preference ordering indicated that System 2 was most strongly preferred, followed by System 3, whereas Systems 1 and 4 were least preferred and did not differ reliably from one another.

Participants' rankings of the four system configurations in terms of trust were analyzed using the same approach. Using the observation-only system (System 1) as the reference, the system allowing small, teacher-controlled interventions (System 2) was significantly ranked as more trustworthy (β = 0.33, z = 2.85, p = .004). In contrast, the system that independently adjusted instructional elements while keeping teachers informed (System 3) was significantly less likely to be ranked as trustworthy than the observation-only system (β = –0.50, z = –4.02, p < .001). The fully autonomous system that alerted teachers only in rare cases (System 4) was least trusted, showing a strong and significant negative difference relative to the observation-only system (β = −0.86, z = −6.69, p < .001). Overall, the model-implied trust ordering indicated that System 2 was most trusted, followed by System 1, whereas Systems 3 and 4 were trusted significantly less, with System 4 ranked lowest.

5.5. Discussion

The present study aimed to examine how different levels of system autonomy are related to teachers' acceptance of a behavioral data-driven AIEd tool. Building on the Six Levels of Automation Model (Molenaar, 2022), we employed a within-subjects vignette study in which we presented teachers with different vignettes describing a hypothetical AI system designed to support reading comprehension, with each vignette corresponding to a different system autonomy level. Teachers' acceptance was assessed using the technology acceptance dimensions: perceived ease of use, perceived usefulness, and attitude toward using the system. We also included two comparative judgements for intention to use and trust in the AI systems. This approach allowed us to investigate not only whether system autonomy affects absolute acceptance perceptions, but also how it shapes relative preferences and trust in a human-AI collaborative context.

5.5.1. Interpretation of Results

Across the TAM-dimensions, differences in automation had relatively little impact on teachers' acceptance. Neither perceived usefulness nor attitude toward using the system differed significantly across the four levels of automation. Effect sizes were also found to be generally small. Only for perceived ease of use, a significant difference was observed between two vignettes: the system that provided small, real-time micro-interventions while the teacher retained primary responsibility (L2) was perceived slightly less easy to use than the observation-only dashboard system (L1). This suggests that a modest increase in system autonomy may create tension for teachers, as they may need to navigate the overlap between their own instructional control and the system's emerging agency. Although teachers still retain primary control at L2, the presence of micro-interventions may create a sense of shared control that can be experienced as intrusive, making the system feel less easy to use compared to the non-intrusive L1 dashboard, which leaves decision-making entirely with the teacher. While L1 supports teacher control through passive observation, L2 may be perceived as a system that requires teachers to process and respond to system actions alongside their instructional responsibilities, potentially increasing perceived complexity despite preserving ultimate teacher control. However, this effect was only limited for this system automation level, and does not indicate a systematic trend.

By contrast, ranking-based measures on intention to use and trust revealed a clear and more consistent pattern of preferences. When participants were required to directly compare system configurations and autonomy levels, the system that allowed limited, teacher-controlled interventions (L2) was consistently ranked as highest for both intention to use and trust. Systems with higher levels of autonomous behavior, particularly those that acted with minimal teacher involvement (i.e., L4), were ranked lower, especially on trust. These findings indicate that teachers are sensitive to differences in autonomy, but such distinctions are most apparent when they are required to make explicit trade-offs between system configurations. This also contributes to the literature on value-sensitive AI design in education (Shen et al., 2025), by suggesting that teachers' perceptions of AI systems are partly grounded in how such systems align with human values in teaching practice.

5.5.2. Strengths and Limitations

The present study has several methodological and conceptual strengths. First, it addresses a clear empirical gap in the AIEd literature by systematically examining system autonomy as a design dimension in teachers' acceptance of behavioral data-driven AI systems. While autonomy is central in recent conceptual frameworks of hybrid intelligence, it has rarely been operationalized and tested empirically. By explicitly varying autonomy across multiple system configurations based on Molenaar's (2022) model, the study offers a theory-informed and structured approach to studying human–AI collaboration in education.

Second, the within-subjects vignette design constitutes an important strength. Presenting multiple autonomy configurations to the same participants allowed for direct comparisons while considering stable individual differences in attitudes toward AI. The use of rich, contextually embedded vignettes likely supported more realistic and nuanced evaluations than abstract questionnaire items alone, particularly for complex systems that teachers may not yet encounter in daily practice. The manipulation check further indicates that participants meaningfully distinguished between the intended autonomy levels, although the distinction between Levels 2 and 3 appeared less pronounced, as this comparison was only marginally significant.

Third, the study combines traditional TAM-based measures with comparative ranking tasks, which proved to be particularly informative. While differences across autonomy levels were limited on absolute Likert-type acceptance ratings, the ranking tasks for intention to use and trust revealed clear and theoretically meaningful preference patterns. The results highlight the added value of moving beyond isolated acceptance judgments toward designs that explicitly elicit relative preferences.

Fourth, the study draws on a large and diverse international sample of teachers across multiple educational levels and European countries, increasing the robustness and generalizability of the findings. The careful translation and review process further strengthens the internal validity of the cross-national data collection.

Despite these strengths, several limitations should be acknowledged. First, while the vignette approach enabled systematic manipulation of system autonomy, it necessarily simplifies complex instructional realities. Real AIEd systems may combine multiple autonomy levels simultaneously or allow teachers to dynamically adjust autonomy during use. As such, the four vignettes presented here should be interpreted as useful prototypes rather than being representative of all possible human–AI arrangements.

Second, the study focused primarily on acceptance-related outcomes, drawing on TAM and complementary constructs such as trust. Although these constructs are critical for understanding adoption readiness, they do not yet capture actual instructional impact or concrete pedagogical alignment.

Finally, it is worth noting that the study did not explicitly model individual differences or contextual moderators that may shape teachers' evaluations of autonomous AIEd systems. For example, perceptions may vary depending on teachers' prior experience with AI, their broader teaching orientations, or self-beliefs about their ability to work effectively with emerging technologies. Investigating how such factors influence responses to different autonomy configurations represents an important avenue for future research.

Taken together, the strengths of this study lie in its theory-driven operationalization of system autonomy, its innovative use of comparative evaluation methods, and its broad international scope. At the same time, its limitations underscore the need for complementary longitudinal, design-based, and intervention studies that examine how teachers' acceptance and trust in autonomous AIEd systems develop through sustained, real-world use.

5.6. Data Availability Statement

Additional materials, next to the ones included in appendix, can be retrieved in the online repository via the following link:

https://osf.io/f2g4s/overview?view_only=5c4f0f163d00403abd3532c0e7644395.

5.7. References

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Béland, S., Cousineau, D., & Loye, N. (2018). Utiliser le coefficient omega de McDonald à la place de l'alpha de Cronbach. *McGill Journal of Education*, *52*(3), 791–804. https://doi.org/10.7202/1050915ar

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, *56*(2), 469–488. https://doi.org/10.1111/bjet.13514

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Fox, J., Weisberg, S., & Price, B. (2001). *car: Companion to Applied Regression*. https://doi.org/10.32614/CRAN.package.car

Kamar, E. (2016). Hybrid workplaces of the future. *XRDS: Crossroads, The ACM Magazine for Students*, *23*(2), 22–25. https://doi.org/10.1145/3013488

Kılıç, E., & Çelik, H. E. (2025). Factors affecting acceptance of ChatGPT-4o by English language instructors: The extended TAM approach. *Computers and Education: Artificial Intelligence*, *9*, 100452. https://doi.org/10.1016/j.caeai.2025.100452

Kong, S. C., Yang, Y., & Hou, C. (2024). Examining teachers' behavioural intention of using generative artificial intelligence tools for teaching and learning based on the extended technology acceptance model. *Computers and Education: Artificial Intelligence*, *7*, 100328. https://doi.org/10.1016/j.caeai.2024.100328

Lenth, R. V., & Piaskowski, J. (2017). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://doi.org/10.32614/CRAN.package.emmeans

Luce, R. D. (1959). *Individual choice behavior.* John Wiley.

McGrath, M. J., Lack, O., Tisch, J., & Duenser, A. (2025). Measuring trust in artificial intelligence: validation of an established scale and its short form. *Frontiers in Artificial Intelligence*, *8*. https://doi.org/10.3389/frai.2025.1582880

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, *57*(4), 632–645. https://doi.org/10.1111/ejed.12527

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2025). The critical role of trust in adopting AI-powered educational technology for learning: An instrument for measuring student perceptions. *Computers and Education: Artificial Intelligence*, *8*. https://doi.org/10.1016/j.caeai.2025.100368

Plackett, R. L. (1975). The analysis of permutations. *Applied Statistics*, *24*(2), 193. https://doi.org/10.2307/2346567

R Core Team. (2023). *R: A language and environment for statistical computing.* https://www.R-project.org/

Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., A. Iahad, N., Abdulla, A. A., Akopyan, A., Aldana Segura, M. W., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydın, F., Badran, A., Balbontín-Alvarado, R., ben Saad, H., … Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLOS ONE*, *20*(2), e0315011. https://doi.org/10.1371/journal.pone.0315011

Revelle, W. (2024). *psych: Procedures for Psychological, Psychometric, and Personality Research* (2.4.1). https://cran.r-project.org/web/packages/psych/index.html

Shen, Y., Le, H., Wang, Z., & Wang, Q. (2025). Teachers' perceptions of value-sensitive AI in education: A case study of AI tutor. *Teaching and Teacher Education*, *168*, 105253. https://doi.org/10.1016/j.tate.2025.105253

Şimşek, A. S., Cengiz, G. Ş. T., & Bal, M. (2025). Extending the TAM framework: Exploring learning motivation and agility in educational adoption of generative AI. *Education and Information Technologies*, *30*(15), 20913–20942. https://doi.org/10.1007/s10639-025-13591-9

Skilling, K., & Stylianides, G. J. (2020). Using vignettes in educational research: A framework for vignette construction. *International Journal of Research & Method in Education*, *43*(5), 541–556. https://doi.org/10.1080/1743727X.2019.1704243

Turner, H., Kosmidis, I., & Firth, D. (2025). *PlackettLuce: Plackett-Luce Models for Rankings* (0.4.4). https://doi.org/10.32614/CRAN.package.PlackettLuce

Venkatesh, Morris, Davis, & Davis. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167

Xue, L., Ghazali, N., & Mahat, J. (2025). Artificial intelligence (AI) adoption among teachers: A systematic review and agenda for future research. *International Journal of Technology in Education*, *8*(3), 802–824. https://doi.org/10.46328/ijte.1191

6. Synthesis and Implications

Chapter Authors: Siem Buseyne & David Gijbels

This final section brings together the key insights from the preceding chapters and considers their implications for the further development, implementation, and governance of AI-supported educational technologies in general, and the EYE-TEACH project in particular, thereby synthesizing first evidence needed to address Objective O1.1 on teachers' needs, acceptance, and readiness regarding AI-assisted ET-analytics tools.

Taken together, the findings across the studies in this report provide a coherent picture of how teachers perceive AI-supported educational technologies that rely on behavioral data, and more specifically how they respond to the idea of AI-supported eye-tracking for reading comprehension. Across the scoping review, the qualitative study, the preparedness analysis, and the vignette-based technology acceptance study, results show that teachers are not categorically opposed to AI-supported educational technology, but their openness is clearly conditional. Acceptance depends less on the mere presence of AI and more on how the technology is designed, how much control teachers retain, how transparent and trustworthy the system is, and whether it fits with teachers' pedagogical goals and classroom realities.

A first overarching conclusion is that the findings strongly support an augmentation-oriented rather than a replacement-oriented view of AI in education. Across the different studies, teachers valued configurations in which AI supports professional judgment rather than takes over pedagogical decision-making. This aligns with broader perspectives on human-centered AI and hybrid intelligence, which emphasize that AI should enhance rather than displace human expertise. In the context of EYE-TEACH, this means that the development of AI-supported tools should prioritize teacher agency, preserve meaningful teacher–student interaction, and avoid over-automation of core educational processes. In turn, this also suggests that policy frameworks and funding schemes should explicitly encourage teacher-in-the-loop designs and treat human-centered or hybrid-intelligence principles as important criteria in the development and evaluation of AI-supported educational technologies.

A second conclusion is that trust, transparency, and explainability are not secondary concerns, but central conditions for adoption. Teachers' willingness to engage with behavioral data-driven AI tools depends not only on whether these systems are technically functional, but also on whether teachers can understand how the systems work, how outputs are generated, what the data are used for, and where the limits of the system lie. Questions of privacy, accountability, fairness, and data governance therefore need to be treated as integral parts of design and implementation rather than as issues to be addressed afterward. This points to the need for clear standards and guidelines at broader policy levels, including strong attention to explainability, child-appropriate privacy protections, and transparent communication about data use, in line with wider regulatory developments such as the GDPR and the AI Act.

A third key insight is that pedagogical fit is essential. Teachers did not evaluate AI-supported eye-tracking only in terms of efficiency or innovation, but through the lens of educational value. They recognized the potential of such tools to support diagnosis, differentiation, and formative feedback, but also stressed that reading instruction is a socially and pedagogically rich activity that should not be reduced to individual screen-based interaction. The findings therefore suggest that AI-supported eye-tracking is likely to be most valuable when used in targeted, formative, and pedagogically embedded ways, rather than as a stand-alone or fully automated instructional solution. This has implications not only for design, but also for educational policy and implementation: behavioral data-driven AI should primarily support diagnosis, monitoring, and formative feedback, while safeguarding teacher autonomy, peer interaction, and opportunities for discussion and guided instruction.

A fourth takeaway is that teacher preparedness to use behavioral data-driven AI cannot be assumed. Successful implementation requires not only well-designed tools, but also sustained investment in professional learning. Support should go beyond technical training and include attention to pedagogical integration, interpretation of dashboards and system outputs, ethical reflection, and critical understanding of what AI can and cannot do in educational settings. This implies a need for professional development frameworks, as well as school-level opportunities for collaborative learning, peer exchange, and joint reflection around concrete tools and use cases. The findings also point to a need for train-the-trainer initiatives, since no heightened levels of preparedness were observed among pedagogical coaches or related roles.

Overall, the integrated findings suggest that the most promising path forward for EYE-TEACH lies in developing AI-supported tools that make complex behavioral data pedagogically usable while keeping teachers meaningfully in the loop. Such tools are most likely to be accepted when they are transparent, reliable, ethically governed, and clearly supportive of teachers' professional role. In that sense, the implications of this report extend beyond EYE-TEACH itself: they point toward a broader vision of AI in education in which innovation is guided not by automation alone, but by human-centered design, pedagogical relevance, and responsible implementation.