

AI-ED-SAT: design and validation of a questionnaire for self-assessment of teaching skills in educational AI

AI-ED-SAT: diseño y validación de un cuestionario para autoevaluar competencias docentes en IA educativa



 Andresa Sartor-Harada - *Universidad Internacional de La Rioja, UNIR (Spain)*

 Juliana Azevedo Gomes - *Universidad Internacional de La Rioja, UNIR (Spain)*

ABSTRACT

The accelerated integration of Artificial Intelligence (AI) in education poses new challenges for teacher training and assessment. This study presents the design, validation, and psychometric analysis of the AI-ED-SAT questionnaire, a self-assessment tool designed for teachers to diagnose their level of preparedness in the pedagogical, ethical, and curricular use of AI. The AI-ED-SAT questionnaire was validated within the Design-Based Research approach. It comprises four dimensions aligned with UNESCO's (2024a, 2024b) current AI competency frameworks: conceptual understanding of AI, pedagogical use of intelligent tools, ethical and critical reflection, and curricular integration. Its construction was based on an exhaustive theoretical review and an iterative validation process using the Delphi method with 12 experts in educational technology, AI, and teacher training. Subsequently, in a pilot test involving a sample of 128 teachers from various educational levels, the reliability (Cronbach's $\alpha = 0.93$), content validity, and construct validity were analyzed. The results of the exploratory factor analysis (EFA) confirmed the grouping of items into four theoretical factors, while the confirmatory factor analysis (CFA) showed excellent fit indices (CFI = 0.96; RMSEA = 0.045). The AI-ED-SAT is a robust and up-to-date tool that is useful in both research and teacher training programs. Its self-reflective approach helps strengthen teachers' critical literacy and professional agency in the face of the challenges posed by AI in education.

Keywords: artificial intelligence; education; teacher competencies; questionnaire; instrument validation.

RESUMEN

La integración acelerada de la Inteligencia Artificial (IA) en el ámbito educativo plantea nuevos retos para la formación y evaluación de competencias docentes. Este estudio presenta el diseño, validación y análisis psicométrico del cuestionario AI-ED-SAT, una herramienta de autoevaluación dirigida a docentes para diagnosticar su grado de preparación en el uso pedagógico, ético y curricular de la IA. Desarrollado bajo un enfoque de investigación basado en diseño (Design-Based Research), el instrumento se estructuró en torno a cuatro dimensiones alineadas con los actuales marcos de competencias en IA de la UNESCO (2024a, 2024b): comprensión conceptual de la IA, uso pedagógico de herramientas inteligentes, reflexión ética y crítica e integración curricular. Su construcción se apoyó en una revisión teórica exhaustiva y un proceso iterativo de validación mediante el método Delphi con 12 expertos en tecnología educativa, IA y formación docente. Posteriormente, en una prueba piloto a una muestra de 128 docentes de distintos niveles educativos, se analizó su fiabilidad (α de Cronbach = 0.93), validez de contenido y validez de constructo. Los resultados del análisis factorial exploratorio (AFE) confirmaron la agrupación de los ítems en cuatro factores teóricos, mientras que el análisis factorial confirmatorio (AFC) mostró índices de ajuste excelentes (CFI = 0.96; RMSEA = 0.045). El AI-ED-SAT se configura como una herramienta robusta y actualizada, útil tanto en la investigación como en programas de formación docente. Su enfoque autorreflexivo contribuye a fortalecer la alfabetización crítica y la agencia profesional del profesorado frente a los desafíos de la IA en la educación.

Palabras clave: inteligencia artificial; educación; competencias docentes; cuestionario; validación de instrumento.

INTRODUCTION

The emergence of Artificial Intelligence (AI) in educational ecosystems is transforming not only teaching and learning methodologies, but also the ethical, pedagogical, and organizational frameworks that govern teaching practice. While technological advances have enhanced the development of adaptive virtual environments, intelligent tutoring systems, and automated assessment processes, their implementation in school contexts requires specific, situated, and reflective teacher preparation (Holmes et al., 2019; UNESCO, 2021; UNESCO, 2024b).

Beyond access to digital devices or platforms, education professionals must develop critical AI competencies that enable them to understand how AI works, assess its social and ethical implications, and design pedagogical practices consistent with the principles of equity, inclusion, and educational value. In this context, AI teaching competencies should be understood as a multidimensional set of knowledge, skills, and attitudes that go beyond traditional digital literacy (Celik, 2023; Chen et al., 2020).

However, the literature reveals a shortage of specific, validated instruments that enable the assessment of these skills in relation to AI in a rigorous, contextualized, and formative manner. Most existing models focus on generic indicators of digital competence, which limits their ability to address the particular challenges posed by the integration of AI in education (Zawacki-Richter et al., 2019). This gap hinders both empirical research and the design of teacher training programs that cater to new technological demands.

In response to this need, this study presents the design, validation, and psychometric analysis of the AI-ED-SAT (Artificial Intelligence in Education – Self-Assessment Tool), a self-assessment questionnaire aimed at teachers at different educational levels, with the purpose of diagnosing their level of competence in the use of AI in the school environment. The instrument is structured around four key dimensions—conceptual understanding of AI, pedagogical use of tools, ethics and considerations in the classroom, and curriculum integration—and has been developed using a mixed methodological approach, combining theoretical review, expert judgment validation (Delphi method), and exploratory and confirmatory factor analyses. Its objective is twofold: to provide a rigorous tool for educational research and, at the same time, to serve as a training resource that promotes professional reflection and continuous improvement among teachers in the face of the challenges posed by AI in education.

Although the questionnaire was developed entirely in Spanish, it was decided to retain an English acronym to facilitate its visibility and adoption in future international research. This decision responds to the growing presence of English as the *lingua franca* in the field of educational AI, which may favor its inclusion in academic databases, global collaboration networks, and translation and cross-cultural adaptation processes.

Education and emerging technologies: the new teaching paradigm

The advancement of emerging technologies, particularly AI, is profoundly reshaping the contemporary educational landscape. Recent research highlights the urgent need to redefine teaching skills in light of this transformation (Celik, 2023; Holmes et al., 2019; UNESCO, 2023). Far from being limited to technical mastery, these competencies include a critical understanding of the impact of technology on

teaching and learning processes, the ethical use of data, ensuring inclusion, and designing personalized and meaningful experiences.

The digital transformation has brought about significant structural changes in education, affecting not only the tools available and pedagogical dynamics but also regulatory frameworks and the professional roles of teachers. In this context, technologies such as augmented reality, big data, blockchain, and AI itself not only open up new teaching possibilities but also impose the need for a profound reconfiguration of teachers' professional competencies (Bo, 2024; UNESCO, 2024b).

This new paradigm requires teachers to perform more complex functions: designing learning experiences, acting as critical mediators between technology and the curriculum, and managing digital knowledge ethically. Consequently, technological integration in the classroom cannot be reduced to instrumental or technical use but requires a contextualized pedagogical approach capable of assessing its social, ethical, and educational implications and adapting them to the needs and trajectories of students (Cabero-Almenara et al., 2020).

Various international initiatives, such as the DigCompEdu framework (Ghomi & Redecker, 2019) and UNESCO's guidelines for teacher competence in ICT, as well as those currently specific to the use of AI, have established benchmarks for the development of digital competences in teachers. However, these frameworks tend to take a generalist perspective, focusing on cross-cutting skills and established technologies, without addressing in sufficient depth the specific challenges posed by emerging technologies, particularly AI. This limitation underscores the need to develop models that integrate specific dimensions of critical appropriation of these technologies, aiming to promote new digital literacies that position teachers as active agents of transformation within contemporary educational ecosystems.

Although DigCompEdu (Ghomi & Redecker, 2019) is a widely recognized framework for assessing teachers' digital competence, its generalist orientation limits its ability to address the specificities of Artificial Intelligence in depth. This instrument describes cross-cutting areas of competence applicable to various digital technologies, without differentiating between established and emerging tools, or explicitly integrating ethical and curricular dimensions specific to AI. In contrast, AI-ED-SAT focuses exclusively on the educational use of Artificial Intelligence, incorporating indicators linked to the conceptual understanding of algorithms, contextualized pedagogical application, ethical reflection, and strategic curricular integration. Thus, the AI-ED-SAT does not seek to replace frameworks such as DigCompEdu, but rather to complement them, offering a level of detail adapted to the unique challenges and opportunities presented by AI in educational settings.

Artificial Intelligence in education: potential, risks, and controversies

The inclusion of Artificial Intelligence into education represents one of the most significant transformations in recent years, with direct implications for teaching practices, pedagogical models, and teaching-learning processes. Its presence manifests itself in multiple forms: intelligent tutoring systems, adaptive platforms, learning analytics, chatbots, recommendation engines, and automated feedback tools (Holmes et al., 2019; Zawacki-Richter et al., 2019).

This technological expansion offers significant opportunities to personalize learning, optimize assessment processes, and alleviate teachers' administrative tasks. However, it also poses substantial challenges that cannot be ignored. These include

algorithmic biases, the opacity of automated systems, privacy violations, and the risk of delegating fundamental pedagogical decisions to systems that operate with logic alien to human judgment (Chesterman, 2021; Ferrante, 2021; Huang, 2023; Selwyn, 2019).

In this sense, AI cannot be conceived solely as a functional tool, but rather as a complex socio-technical phenomenon that involves political, pedagogical, and ethical decisions (Flores-Vivar & García-Peñalvo, 2023). Its implementation in education requires teachers not only to acquire operational skills, but also to develop a critical literacy that enables them to understand how algorithms work, identify their potential biases, and assess their social and educational implications (García Peñalvo et al., 2024; Lane et al., 2023; Luckin, 2017).

A recent and relevant ethical reference is the Safe AI Education Manifesto (Alier Forment et al., 2024), which sets out clear recommendations for the responsible use of AI in educational contexts. Its fundamental principles include transparency in systems, accountability, protection of student privacy, promotion of equity, and fostering critical literacy in the face of algorithmic technologies. These principles reinforce the ethical dimension of the AI-ED-SAT, particularly in items related to bias identification, automated decision-making, and educational impacts, thereby consolidating the instrument as a tool that integrates updated ethical principles relevant to contemporary teaching practice.

Despite this outlook, teacher training in AI remains in its incipient stages. Existing proposals tend to focus on instrumental aspects, without incorporating a comprehensive view that considers both its pedagogical and ethical dimensions (Delgado et al., 2024; Holmes & Tuomi, 2022). Given this gap, it is essential to design valid and reliable instruments that allow for the evaluation not only of teachers' familiarity with AI, but also of their ability to integrate it in a critical, contextualized, and pedagogically meaningful way. To this end, it is necessary to articulate rigorous methodological processes that combine expert judgment with robust statistical analysis, as proposed by Sampieri (2018) and Muñiz and Fonseca-Pedrero (2019).

Teaching skills in AI: beyond digital literacy

The emergence of Artificial Intelligence (AI) in educational settings has consolidated its position as a disruptive technology, with applications ranging from personalized learning to the automation of assessment and administrative tasks (Holmes et al., 2019; Owan et al., 2023). Its implementation poses significant challenges that require teachers to have a deep understanding of its technical foundations and to critically reflect on the associated risks, such as algorithmic biases, opacity in decision-making, technological dependence, and inequalities in access (Celik, 2023; Sharma et al., 2024).

In this context, developing specific AI skills is a strategic priority for education systems. Unlike generic digital skills, which focus on managing digital resources or virtual environments, teaching skills in AI requires a multidimensional understanding that integrates technical, ethical, pedagogical, and curricular aspects. Various conceptual documents have begun to develop this new competency profile, including UNESCO's (2024a, 2024b) AI competency frameworks for teachers and students, UNESCO's AI and the Future of Teaching document (UNESCO, 2021), the update of the DigCompEdu framework to the AI context (Caena & Redecker, 2019), and the studies by Chen et al. (2020), Chiu et al. (2024), and Ng et al. (2023).

Khreisat et al. (2024) propose a structure based on four interrelated areas:

1. Understanding the principles and functioning of AI.
2. Ability to use AI tools for educational purposes.
3. Critical knowledge of its ethical and social implications.
4. Curriculum integration from an interdisciplinary approach.

This proposal reflects a broader conception of the role of teachers, which transcends the instrumental use of technology. Rather than merely operating tools, teachers must position themselves as reflective and critical agents, capable of evaluating the pedagogical meaning, social consequences, and curricular relevance of AI in their educational practices. As Holmes et al. (2019), Zawacki-Richter et al. (2019), and Ayuso del Puerto and Gutiérrez Esteban (2022) warn, many existing frameworks remain descriptive or eminently technical, which limits their ability to guide truly transformative professional development.

From this perspective, it is necessary to create assessment tools that respond to this complexity. Such tools must extend beyond measuring declarative knowledge to encompass dimensions such as pedagogical competence, critical thinking, and curricular integration. The AI-ED-SAT instrument, developed within the framework of this study, follows this logic, as it is structured around four key dimensions: conceptual understanding, pedagogical application, ethical dimension, and curricular integration. This approach aims to capture the complexity of teacher professional development in the era of AI, acknowledging the need for critical literacy that enables educators to navigate the challenges posed by emerging technologies proactively.

Assessment of teaching skills: development and validation of instruments

In a context of educational transformation driven by emerging technologies, it is crucial to have valid and reliable instruments for diagnosing the level of teaching competence in relation to artificial intelligence. In this sense, self-assessment is a formative and reflective strategy that enables the identification of strengths and areas for improvement from a situated perspective (Panadero et al., 2019; Rachbauer et al., 2025).

Although there are numerous instruments for measuring teacher digital competence (Cabero-Almenara et al., 2020; Ghomi & Redecker, 2019), most focus on generic technical skills, without specifically addressing the pedagogical, ethical, and curricular complexities posed by AI. This limitation has created an urgent need for assessment models that capture not only teachers' technological familiarity but also their ability to integrate AI in a critical, contextualized, and pedagogically meaningful way (Celik, 2023; Delgado et al., 2024).

The construction of instruments with these characteristics requires a rigorous methodological approach that combines theoretical, empirical, and practical criteria. In recent years, a mixed methodology has been established that includes systematic literature review for item development, validation through expert judgment (such as the Delphi method), and psychometric analysis (content and construct validity, internal reliability) (Diefes-Dux et al., 2010; Muñoz & Fonseca-Pedrero, 2019; Sampieri, 2018). In this process, participatory and interactive approaches, such as Design-Based Research (DBR), have proven particularly useful in ensuring the semantic and pedagogical relevance of the instrument.

Self-assessment has also established itself as an effective tool for encouraging professional reflection, providing data for the design of personalized training plans, and promoting teacher agency (Borge et al., 2005). Unlike external assessment models, it places teachers at the center of the diagnosis and improvement process.

The AI-ED-SAT instrument aligns with this framework, as it is designed to assess teaching competencies related to AI in educational contexts comprehensively. Its development is based on a critical review of the specialized literature and recent reference frameworks that address the pedagogical, ethical, and curricular implications of AI (Celik, 2023; Holmes et al., 2019; Zawacki-Richter et al., 2019).

Based on this analysis, four key dimensions were identified that make up a competency profile geared toward a critical and formative appropriation of AI:

1. **Conceptual understanding of AI**
This includes basic knowledge of algorithms, machine learning, and adaptive systems, with a focus on their educational applications. This dimension seeks to position teachers as informed actors in the face of dominant technological discourses (Holmes et al., 2019).
2. **Pedagogical use of AI tools**
This assesses the ability to apply AI tools for clear educational purposes, taking into account the level of the students and the academic context. It encompasses activity design, personalized learning, and automated assessment, extending beyond mere technical use (Chen et al., 2020).
3. **Ethics and critical considerations regarding AI**
This encompasses the identification of risks such as algorithmic biases, privacy issues, and decision automation. This dimension promotes critical literacy, positioning the teacher as an ethical mediator (Celik, 2023).
4. **Curricular integration of AI**
Measures the ability to incorporate AI into teaching and curriculum planning through interdisciplinary projects, active methodologies, and adaptation to new technological scenarios (Rachbauer et al., 2025).

It should be noted that, although some competencies may appear to overlap—particularly between the pedagogical use of AI tools and their curricular integration—the two are distinguished by their level of application: while the second dimension refers to educational use in the classroom (microplanning), the fourth focuses on the strategic incorporation of AI into general curriculum planning (macroplanning), including interdisciplinary approaches and institutional alignment. This differentiation avoids ambiguities and captures different levels of teacher appropriation.

These dimensions are articulated in a complementary manner, enabling a holistic assessment of the teaching role within the digital ecosystem. Unlike fragmented or overly technical models, the AI-ED-SAT proposes an integrated framework where pedagogical action, conceptual knowledge, curricular vision, and ethical commitment are intertwined in the same assessment device.

Figure 1
Diagram of the instrument

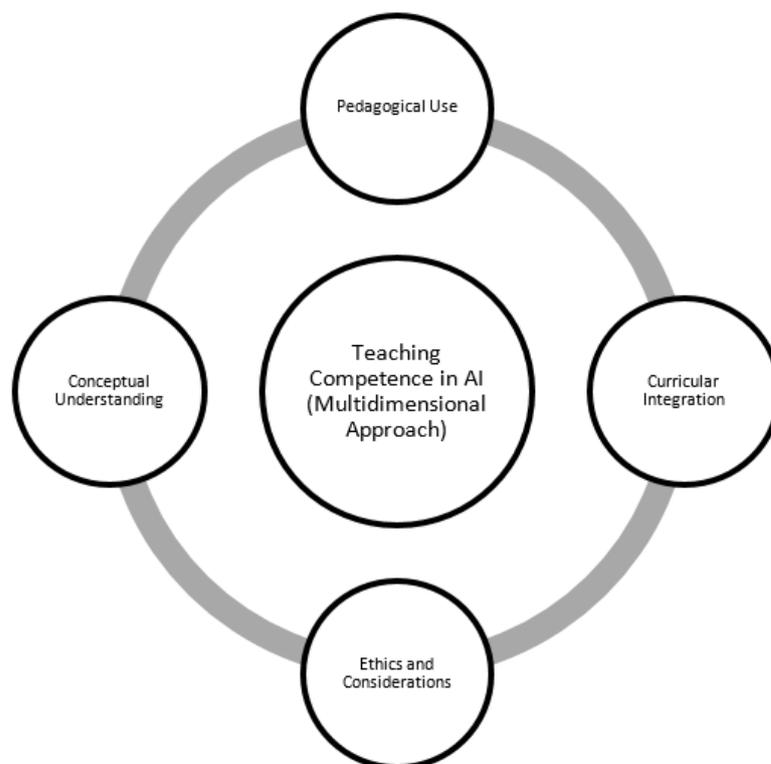


Figure 1 presents a schematic representation of the AI-ED-SAT model, highlighting the interrelationship between its dimensions and its theoretical foundation. The graphic order of the dimensions –Pedagogical Use, Conceptual Understanding, Curriculum Integration, and Ethics and Considerations – does not imply a hierarchy, but rather reflects a circular, non-linear arrangement that emphasizes their interdependence. The dimensions of the model do not follow a predetermined order of acquisition but rather form a multidimensional framework that can be developed in parallel, complementary, or combined ways, depending on the teacher's profile, previous experience, and educational context.

In short, the incorporation of AI in the educational field requires a comprehensive understanding on the part of teachers, combining conceptual mastery with critical, ethical, and pedagogical appropriation. Despite advances in the theoretical definition of these competencies, a lack of specific, valid, and reliable instruments remains, hindering the rigorous and contextualized diagnosis of these competencies.

The AI-ED-SAT is conceived as a response to this need. Its methodological design, based on a combination of qualitative and quantitative approaches, including Design-Based Research, expert judgment, and factor analysis, seeks to ensure the theoretical and psychometric soundness of the instrument, as well as its educational usefulness in improving teaching practice.

METHODOLOGY

Methodological design of the instrument

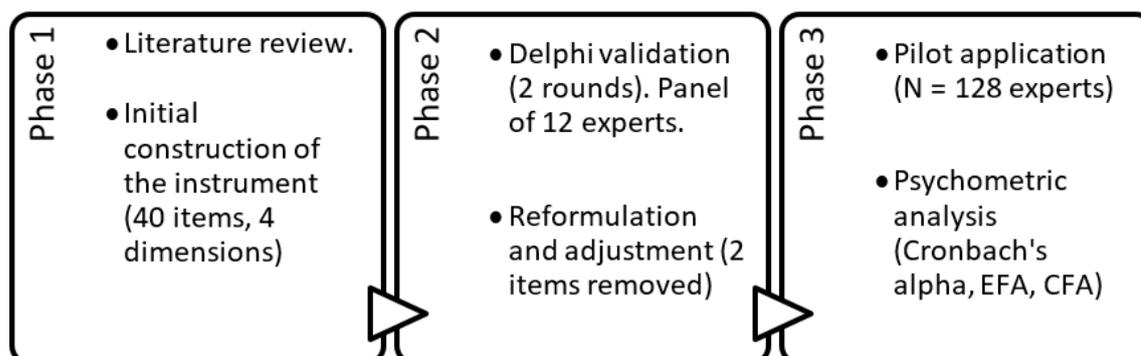
This study is part of a quantitative, descriptive, and correlational approach aimed at designing, validating, and conducting psychometric analysis of a teacher self-assessment instrument on the use of Artificial Intelligence (AI) in education. Methodologically, the principles of Design-Based Research (DBR) (Brown, 1992) were integrated with current guidelines for the validation of educational instruments (Muñiz & Fonseca-Pedrero, 2019; Sampieri, 2018).

The process was structured in three main phases:

1. Construction of the questionnaire, based on a systematic review of recent literature.
2. Content validation through expert judgment, using the Delphi method.
3. Pilot application and psychometric analysis of the instrument.

Figure 2

Stages of the development and validation process



A pilot test was carried out as an intervention, in which the instrument was applied to individuals with characteristics similar to those of the target sample. The purpose of this test was to adjust those items that might require modifications in terms of operability, wording, or other aspects, as well as to optimize the appearance of the instrument (Hernández-Sampieri et al., 2006).

The literature review on teaching competencies in the use of Artificial Intelligence (AI) in education led to the configuration of the first structure of the self-assessment questionnaire around two fundamental dimensions that cover the spectrum of analysis of teaching practice: (1) the pedagogical integration of AI in the classroom and (2) teaching competency in AI and its impact on teaching.

The first draft was constructed using a process based on the design-based research (DBR) methodology, as described by Brown (1992). The process was structured in three phases: in the first, a review of the literature on AI in education and teaching competencies in digital environments was carried out; in the second, the dimensions were defined and the items that would make up the questionnaire were developed; and in the third, the instrument was validated using the Delphi method

with a panel of experts and learning (Chen et al., 2020; Holmes et al., 2019; Luckin, 2017; UNESCO, 2024b).

The instrument, called AI-ED-SAT (Artificial Intelligence in Education – Self-Assessment Tool), was designed to assess the level of teacher competence in four key dimensions:

1. General knowledge of AI and its educational application
2. Pedagogical use of AI tools
3. Ethics and considerations regarding AI in the classroom
4. Curricular integration of AI

Initially, 40 items were developed, written as statements and distributed evenly across the four dimensions (10 items per dimension). Responses were structured on a 5-point Likert scale:

1 = Strongly disagree | 2 = Disagree | 3 = Neither agree nor disagree | 4 = Agree | 5 = Strongly agree

The decision to include a balanced number of 10 items per dimension was based on the criteria of structural symmetry and comparability between dimensions, seeking to ensure a balanced psychometric analysis. Although the nature of the dimensions could have allowed for an extension or reduction of items, an initial equitable distribution was chosen as a starting point for validation, in line with common practices in similar instruments (Muñiz & Fonseca-Pedrero, 2019). Subsequently, the validity analysis enabled the instrument to be refined, resulting in a final version comprising 38 items.

This scale allows for the capture of differentiated levels of self-assessment without overloading decision-making and is suitable for factor analysis. The validation of the instrument sought to ensure that the indicators adequately reflected the integration of AI in education and the teaching skills necessary for its effective implementation.

The final version of the questionnaire, comprising 38 items distributed across four dimensions, is available in full in Appendix 1. The inclusion of the complete instrument provides a detailed overview of the indicators used to assess teaching competence in AI, facilitating its possible replication in subsequent studies or training programs.

Instrument content validation: the Delphi method

The second phase of the study aimed to validate the content of the preliminary draft of the questionnaire that had been previously developed. To this end, the Delphi method was used, widely recognized in the field of Education Sciences for its ability to clarify complex issues through a structured process of consultation among experts organized in a panel (Reguant-Álvarez & Torrado-Fonseca, 2016). This iterative technique fosters informed consensus through successive rounds of evaluation, enabling a progressive convergence of informed opinions (Linstone & Turoff, 2002; Reguant-Álvarez & Torrado-Fonseca, 2016).

The instruments used for data collection consisted of questionnaires designed to assess specific dimensions and items, including sections for qualitative observations.

The collection process was organized with a flexible yet limited timeframe to facilitate the effective participation of a panel of experts, who were distributed geographically and had varying availability.

At this stage, the panel of experts was formed, prioritizing the representativeness and professional diversity of the participants over the absolute number of participants. To this end, selection criteria were defined to ensure a balance between up-to-date knowledge on the integration of Artificial Intelligence (AI) in educational contexts and the teaching skills required for its effective application.

Although the panel included a higher proportion of university educators, this choice was made in response to the need for experts with a systematic view of teacher development and experience in continuing education, which is crucial for assessing cross-cutting skills, such as those related to AI. However, the presence of primary and secondary school professionals with direct experience in applying AI technologies in the classroom was ensured, as well as accredited specialists in artificial intelligence with scientific output and participation in educational innovation projects.

Based on the preliminary structure of the instrument, the central criterion was to include professionals from three key profiles: (1) teachers from different educational levels with experience in the pedagogical use of emerging technologies; (2) specialists in AI applied to education; and (3) teacher trainers in the field of educational technology. This strategy enabled the integration of a multidisciplinary perspective, consistent with the study's objectives.

Although there is no normative consensus on the ideal size of Delphi panels, various authors suggest practical guidelines. Generally, it is considered that a minimum of ten participants ensures the stability of the consensus, as samples of fewer than seven may compromise representativeness (Reguant-Álvarez & Torrado-Fonseca, 2016). In line with this premise, Skulmoski et al. (2007) argue that, for relatively homogeneous panels, a sample of between ten and fifteen experts can be methodologically sound.

Twelve experts who met the established criteria were invited to participate in the study. The invitation included a detailed description of their professional profile, information about their teaching and research roles, and a self-assessment of their suitability for the study's purpose.

All the experts accepted the invitation, forming a panel with a high level of specialization. Seventy-five percent of the group held doctoral degrees, comprising seven university trainers in teacher training programs, three basic education teachers (from primary and secondary education) with direct experience in using AI in the classroom, and two specialists with combined experience in academic research and teacher training in educational AI. Both experts have published indexed works on AI in education, served as advisors on technology integration projects in educational institutions, and collaborate in international networks focused on academic innovation and digital ethics.

In addition, a balanced gender distribution was sought in the composition of the panel, which included seven women (58%) and five men (42%), in line with the representativeness and diversity criteria established in the methodological design.

Common inclusion criteria included:

- A minimum of five years' experience in educational technology, teacher training, or applied research on AI.
- Participation in innovation initiatives or digital skills training projects.

- Representation of different levels of the education system (basic and higher education).

Table 1*Composition of the Delphi expert panel*

Category	Number
Total number of experts	12
With doctoral degrees	9
University trainers	7
Primary and secondary school teachers	3
Specialists in AI and teacher training	2

The researchers in charge supervised data collection during the validation phase, which was conducted through an iterative process of consulting experts, utilizing email as the primary channel of communication. The procedure was structured in two successive rounds, which was considered adequate to facilitate the convergence of opinions and reach a methodologically sound consensus (Linstone & Turoff, 2002). From the outset, participants were informed about this structure and the degree of involvement required, with the aim of ensuring transparency and commitment on the part of the panel (Reguant-Álvarez & Torrado-Fonseca, 2016).

The primary purpose of the first round was to evaluate the quantitative and qualitative aspects of the items in the instrument, with a focus on two fundamental dimensions: conceptual relevance and linguistic clarity. At the quantitative level, the experts were asked to rate each item on a five-point Likert scale, where one represented "not relevant" or "unclear," and five indicated "very relevant" or "very clear." At the same time, space was provided for open comments, which allowed for the collection of qualitative suggestions aimed at improving the wording, accuracy, or appropriateness of the items. This mixed approach, based on previous work (Gallant & Luthy, 2020; Grand-Guillaume-Perrenoud et al., 2023; Paulin et al., 2024; Ramírez, 2019; Ramírez-Montoya & Lugo-Ocando, 2020), enriches the instrument from an interpretive perspective.

Based on the analysis of this first phase, six items were reformulated based on qualitative observations, and two were eliminated because they were considered redundant or conceptually weak. These adjustments led to the design of the form for the second round, which focused on re-evaluating the modified items. Again, a five-point Likert scale was used to assess their relevance and clarity after the modifications. The objective of this second iteration was to confirm the validity of the changes introduced and consolidate a final version that reflected the panel's consensus. The results obtained enabled the process to be closed with a definitive structure comprising 38 items, distributed across four theoretical dimensions.

The results of both rounds were analysed using a combination of quantitative techniques:

- Descriptive statistics: calculation of mean, standard deviation, and percentiles for each item.
- Kendall's coefficient of concordance (W): used to measure the degree of agreement among experts.
- Acceptance criterion based on the 80th percentile: established as the minimum threshold for considering items accepted, according to Mauri et al. (2007).

This analytical approach enabled the accurate identification of items that achieved high levels of consensus, as well as those that required adjustments or elimination.

Pilot study and psychometric analyses

Once the expert judgment validation process was complete, the revised version of the instrument underwent a pilot study with a sample of 128 teachers from various educational contexts in Spain and Latin America. The selection was intentional, based on criteria of accessibility and institutional diversity.

Sample profile:

- Primary education: 34%
- Secondary education: 31%
- Higher education: 35%
- Average teaching experience: 9.6 years
- Gender distribution: 69% women, 31% men

The application was carried out online. The sample consisted of teachers from seven Spanish-speaking countries: Spain ($n = 49$), Mexico ($n = 27$), Argentina ($n = 18$), Colombia ($n = 12$), Chile ($n = 9$), Peru ($n = 7$), and Uruguay ($n = 6$). This distribution was based on criteria of institutional accessibility and geographical diversity, ensuring a balanced representation of different educational contexts. In terms of academic level, the proportion of participants remained similar across all countries, with representation from all three levels: primary, secondary, and tertiary education.

All participants signed a digital informed consent form before accessing the questionnaire. The data collected were processed using IBM SPSS v28 and AMOS v24 software and underwent three types of statistical analysis:

a) Internal consistency

The internal reliability of the questionnaire was estimated using Cronbach's alpha coefficient, calculated individually for each dimension and for the instrument as a whole. This analysis made it possible to evaluate the internal consistency of the items grouped around the different theoretical constructs:

- General knowledge about AI: $\alpha = 0.88$
- Pedagogical use of AI tools: $\alpha = 0.90$
- Ethics and considerations: $\alpha = 0.87$
- Curricular integration: $\alpha = 0.91$
- Total (AI-ED-SAT): $\alpha = 0.93$

These values demonstrate high internal consistency across all dimensions, exceeding the threshold of 0.70 recommended by specialized literature.

In addition to Cronbach's alpha coefficient, McDonald's omega coefficients were calculated, which exceeded the threshold of 0.85 in all cases (total $\omega = 0.94$). This metric, considered a more robust estimate of reliability in scales with multidimensional

structures (McDonald, 1999), reinforces the evidence of the instrument's high internal consistency. Although no temporal reliability analyses (test–retest) or factor invariance tests by subgroups (educational level, gender, or country) were applied in this phase, due to the cross-sectional nature of the design and the sample size, the relevance of these techniques for future validation research with longitudinal designs or larger samples is recognized.

Although the number of participants can be considered adequate in relation to the total number of items in the instrument, various methodological studies (Costello & Osborne, 2005; Lloret-Segura et al., 2014; MacCallum et al., 1999) agree that sample adequacy does not depend solely on absolute size, but on other factors such as item communality, the number of expected factors, and the robustness of statistical indicators. Along the same lines, Hogarty et al. (2005) point out that the quality of factorial solutions is more influenced by the average communality and overdetermination of factors than by the sample size itself. In this case, the high internal consistency of the instrument ($\alpha = 0.93$), the KMO sample adequacy index = 0.91, the significance of Bartlett's sphericity test, and factor loadings greater than 0.60 in all items allow us to consider the sample of 128 teachers as methodologically sound for the factor analysis performed.

b) Exploratory factor analysis (EFA)

In order to identify the underlying structure of the questionnaire, an EFA was performed using the maximum likelihood method and Varimax rotation. The adequacy of the sample was confirmed by:

- KMO index = 0.91
- Bartlett's sphericity test: $\chi^2 (df = 703) = 3982.23, p < .001$

Both indicators showed excellent conditions for the application of factor analysis. Factor loadings were greater than 0.60 for all retained items, and the total variance explained reached 67.8%, distributed evenly among the four proposed dimensions.

In accordance with methodological recommendations and to provide greater transparency in the results, Annex 2 includes the complete matrix of factor loadings, communalities, and specific variances corresponding to the 38 items retained in the exploratory factor analysis. This information enables a detailed examination of the factor structure of the instrument and supports the robustness of the model obtained.

c) Confirmatory factor analysis (CFA)

To verify the empirical adequacy of the theoretical model derived from the EFA, a CFA was performed using a two-factor model, evaluating the quality of the fit using the following indicators:

- $\chi^2/df = 2.34$
- CFI = 0.96
- TLI = 0.95
- RMSEA = 0.045

All the values obtained indicate an excellent fit of the factorial model, confirming the structural validity of the AI-ED-SAT instrument as a tool for teacher self-assessment on the use of Artificial Intelligence in educational settings.

Although it is methodologically recommended to apply EFA and CFA on independent samples, several studies (Brown, 2015; Worthington & Whittaker, 2006; Lloret-Segura et al., 2014) recognize that in applied research contexts with limited samples, it is acceptable to use both procedures on the same sample, provided that the statistical results are robust and the limitation is made explicit. In this study, this strategy was chosen for exploratory and initial validation purposes, recognizing the need for future replication in broader contexts.

RESULTS

Validation by expert judgment (Delphi method) and concordance analysis

During the first round of the Delphi method, the 40 items that made up the questionnaire were evaluated by the panel of experts in relation to two fundamental aspects: conceptual relevance and linguistic clarity. Although most items received high ratings, six were reformulated and two were eliminated because they did not meet the minimum threshold of the 80th percentile in the relevance dimension.

In the second round, the experts reevaluated the adjusted items. The results were analyzed using Kendall's coefficient of agreement (W), whose values showed a statistically significant level of agreement in both dimensions evaluated. This finding supports the existence of consensus among experts, despite the diversity of their profiles and approaches.

Table 2

Expert agreement – Kendall's coefficient

Aspect evaluated	Kendall's W	Chi-square	df	Sig. (p)
Relevance	0.27	84.44	26	< .001
Clarity	0.24	74.53	26	< .001

To ensure content validity, the approach of Mauri et al. (2007) was applied, adapting the cutoff point to the 80th percentile instead of the arithmetic mean. Additionally, a conditional relationship between relevance and clarity was established: the clarity of an item was only evaluated if it had previously been considered relevant. In this way, conceptually weak items were avoided from a formal perspective. The concordance coefficients obtained reflect an adequate consensus, which justifies the incorporation of the reformulated items and the exclusion of those that did not meet the established criteria.

Although Kendall's coefficients are moderate ($W = 0.27$ for relevance and $W = 0.24$ for clarity), these values are consistent with the diverse nature of the expert panel and the complementary qualitative approach of the Delphi method. Previous studies have highlighted that intermediate levels of agreement may be methodologically acceptable in heterogeneous panels, particularly when combined with qualitative analyses that enrich interpretation and facilitate informed adjustments to the items (Grand-Guillaume-Perrenoud et al., 2023; Reguant-Álvarez & Torrado-Fonseca, 2016). Therefore, the magnitude of W reflects a reasonable balance between diversity

of perspectives and sufficient consensus to support the content validity of the instrument.

Internal consistency

Internal reliability analysis was performed on the pilot sample ($N = 128$ teachers) using Cronbach's alpha coefficient, both for the individual dimensions and for the instrument as a whole. All coefficients greatly exceeded the reference value of 0.85, indicating high internal consistency between the items in each dimension and in the questionnaire as a whole:

Table 3
Internal consistency (Cronbach's alpha)

Dimension	Cronbach's Alpha
General knowledge about AI	0.88
Pedagogical use of AI	0.9
Ethics and considerations	0.87
Curricular integration of AI	0.91
Total AI-ED-SAT	0.93

In addition to the internal consistency analysis, descriptive statistics were calculated for the scores obtained in the pilot test. The means per dimension ranged from 3.2 to 4.1 (on a 1-5 Likert scale), indicating a generally positive perception among teachers. The dimension with the highest score was "Pedagogical use of AI" ($M = 4.1$, $SD = 0.6$), followed by "General knowledge about AI" ($M = 3.8$, $SD = 0.7$). The dimensions "Curricular integration" and "Ethics and considerations" had means of 3.5 ($SD = 0.8$) and 3.2 ($SD = 0.9$), respectively, suggesting areas with greater scope for professional development among teachers.

This result supports the internal consistency of the instrument and confirms that the items are grouped logically and consistently within their respective constructs.

Exploratory factor analysis (EFA)

To explore the underlying structure of the questionnaire, exploratory factor analysis (EFA) was performed using the maximum likelihood method with Varimax rotation. The sample adequacy was excellent ($KMO = 0.91$) and Bartlett's sphericity test was significant ($\chi^2 (df = 703) = 3982.23$, $p < .001$), which justified the application of the factorial model.

Considering the geographical diversity of the sample, special attention was paid to possible lexical and usage differences between Spain and Latin America. During the design of the items, idiomatic expressions and regional technical terms were avoided, prioritizing clear, neutral, and accessible language. Both the Delphi phase and the pilot test included spaces for linguistic observations, which were analyzed qualitatively and, where necessary, led to the reformulation of potentially ambiguous terms. This care contributed to improving the semantic validity of the instrument and its applicability in diverse Spanish-speaking educational contexts.

The EFA revealed four main factors, in line with the proposed theoretical structure, which together explained 67.8% of the total variance. Factor loadings

remained above 0.60 in all dimensions, confirming an adequate level of item saturation in their respective factors.

During this process, the 40 original items were used. For the presentation in Table 4, a reduced set of 10 representative items (2-3 per dimension) was selected to illustrate the most prominent factor loadings concisely and facilitate readability. The complete matrix, which includes the 38 final items after wording adjustments and statistical refinement, along with their specific communalities and variances, is presented in Appendix 2.

Table 4
Factor loadings of selected items (EFA)

Item	Factor loading (EFA)
Item 1	0.72
Item 2	0.81
Item 3	0.75
Item 4	0.77
Item 5	0.8
Item 6	0.69
Item 7	0.73
Item 8	0.78
Item 9	0.74
Item 10	0.82

These results show that the selected items maintain a clear and consistent structure in each dimension and are representative of the factorial pattern obtained for the instrument as a whole.

Confirmatory factor analysis (CFA)

To empirically confirm the structural validity of the questionnaire, a confirmatory factor analysis (CFA) was conducted on the theoretical model derived from the EFA. The model evaluated was a two-factor model, which allowed for the verification of both the general common variance and the specific variance attributable to each dimension.

The fit indices obtained were:

- $\chi^2/df = 2.34$
- CFI = 0.96
- TLI = 0.95
- RMSEA = 0.045

These values fall within the ranges established as indicators of excellent fit (Hu & Bentler, 1999), confirming that the structure of the AI-ED-SAT instrument adequately reflects the proposed theoretical model. Specifically, the CFI (Comparative Fit Index) and TLI (Tucker-Lewis Index) exceed the threshold of 0.95, while the RMSEA remains below 0.06, indicating the model's validity from a multidimensional perspective.

This analysis provides solid evidence that the questionnaire has a robust factorial structure, capable of accurately capturing the four theoretical components

proposed: knowledge about AI, pedagogical use, ethics and considerations, and curricular integration.

In line with the Delphi method-based methodological process, and with the aim of ensuring the clarity, relevance, and theoretical consistency of the items, the instrument was refined after qualitative and quantitative analysis of the validation rounds. In this process, two items were eliminated because they did not meet the established relevance threshold, and six more were reformulated due to issues of clarity, ambiguity, or suitability for the teaching profile. Table 5 presents a summary of the actions taken on the items during this expert review phase.

Table 5
Items eliminated and reformulated after the Delphi process

Original item	Action taken	Justification	New version (if applicable)
Item 5 – “I understand the mathematical and logical foundations of AI.”	Removed	Low level of relevance for the general teaching profile; excessively technical content; did not exceed the 80th percentile.	—
Item 12 – “I have designed learning activities that incorporate AI.”	Removed	Uncommon level of implementation in the average teaching profile; low relevance as perceived by experts.	—
Item 4 – “I am familiar with terms such as machine learning and neural networks.”	Reformulated	Technical language lacking sufficient clarity; more accessible wording suggested.	I am familiar with general terms related to AI, such as “machine learning” or “neural networks.”
Item 7 – “I can explain to other teachers how basic AI works.”	Reformulated	Ambiguity in the term “basic AI”; need for greater precision.	I can explain the fundamental concepts of AI to other teachers in a simple way.
Item 16 – “I use chatbots or virtual assistants to answer students' questions.”	Reformulated	Technical ambiguity and diversity of interpretation; imprecise wording.	I use tools such as chatbots or virtual assistants to help answer students' questions.
Item 27 – “I identify the social impacts of AI in education.”	Reformulated	Item too broad; specifying the focus is recommended.	I identify how AI can affect access, equity, and social interaction in educational contexts.
Item 33 – “I am familiar with curriculum frameworks that incorporate AI in education.”	Reformulated	Ambiguity about the availability of frameworks; context not easily generalizable.	I am familiar with examples of curriculum proposals or policies that promote the inclusion of AI in education.
Item 36 – “I use AI-based simulations to reinforce concepts in the classroom.”	Reformulated	Unclear terminology; more practical and concrete wording is suggested.	I use AI-supported simulations or interactive tools to reinforce student learning.

DISCUSSION

The results obtained in the validation of the AI-ED-SAT instrument confirm the conceptual and psychometric soundness of its design, allowing it to be considered a robust tool for assessing teacher competence in the use of Artificial Intelligence (AI) in educational settings. In line with the reference frameworks proposed by international organizations such as UNESCO (2021) and the OECD (2022), the questionnaire presents a comprehensive, up-to-date, and contextualized approach to the training requirements of teachers in response to the rise of AI in education.

The high internal consistency obtained in all dimensions (Cronbach's α ranging from 0.87 to 0.93) supports the internal coherence of the items around well-defined constructs. Likewise, both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) provided clear empirical evidence of the structural validity of the proposed model, with fit indices (CFI = 0.96; RMSEA = 0.045) that far exceed the minimum standards required in the psychometric literature (Hu & Bentler, 1999). These findings support the conclusion that the AI-ED-SAT assesses differentiated and complementary dimensions of teaching knowledge and practice in relation to AI.

Beyond psychometric indicators, the results underscore the relevance of the AI-ED-SAT as a diagnostic and formative tool in contexts where teacher literacy in AI is becoming increasingly strategic. As Celik (2023) warns, the development of teaching competencies in AI must go beyond the technical management of tools, also encompassing the capacity for critical analysis, meaningful pedagogical design, and curricular adaptation to new technological scenarios. The instrument addresses this need by integrating four key dimensions: conceptual understanding, pedagogical use, an ethical approach, and curricular integration. This segmentation is not only consistent with previous studies (Chen et al., 2020; Zawacki-Richter et al., 2019) but also allows for differentiated and specific feedback, which is highly valuable for guiding teacher professional development processes.

Potential of AI-ED-SAT for university teacher training in digital and hybrid environments

The growing incorporation of digital and hybrid modalities in higher education requires university teacher training to integrate specific skills for teaching mediated by emerging technologies, including Artificial Intelligence. AI-ED-SAT, with its four-dimensional structure and self-reflective nature, offers an operational framework for diagnosing and developing these skills in diverse academic contexts. In digital environments, its results can guide the selection of AI-based resources, strategies, and pedagogical approaches, promoting the adaptation of content and methodologies to virtuality. In hybrid modalities, the instrument facilitates the identification of practices that coherently integrate the potential of AI in both physical classrooms and virtual environments, promoting pedagogical coherence and digital inclusion. Its application in university professional development programs also enables the design of personalized, evidence-based training itineraries, enhancing adaptive learning and the continuous updating of teachers in response to technological advances.

One of the distinctive strengths of the tool lies in its self-assessment format, which encourages personal reflection and positions teachers as active protagonists in their continuing education. Unlike assessment models focused on generic digital competencies, the AI-ED-SAT focuses on a specific technology—AI—and its

contextualized pedagogical use, which represents an innovative and necessary contribution in the face of the rapid integration of AI systems into educational platforms, virtual learning environments, and automated assessment devices (Baltazar, 2023; Owan et al., 2023).

From a methodological perspective, the combination of the Design-Based Research (DBR) approach with the Delphi technique provided rigor, flexibility, and contextual sensitivity in the development of the instrument. This approach enabled iterative and consensual development, articulating the theoretical foundation through practical validation by a panel of experts. Previous research has supported the effectiveness of this methodological combination in developing complex educational instruments (Diefes-Dux et al., 2010; Rachbauer et al., 2025), especially when a balance between technical soundness and pedagogical appropriateness is required.

These findings are consistent with previous studies that have identified greater teacher proficiency in the technical use of digital tools compared to more complex ethical or curricular dimensions (Celik, 2023; Chen et al., 2020). Compared to similar instruments focused on general digital competencies, the AI-ED-SAT presents slightly higher means on items related to the specific pedagogical application of AI. This difference can be attributed to the instrument's contextualized approach, which is designed specifically to capture teachers' perceptions of the use of artificial intelligence in real educational contexts.

Despite the promising results, it is essential to acknowledge certain methodological limitations that affect the interpretation of the findings. First, the pilot sample was selected for convenience, which introduces potential bias and limits the generalization of the results to other educational contexts. Second, the instrument was validated exclusively in Spanish-speaking countries, which restricts its cross-cultural applicability. These limitations should be taken into account when evaluating the validity and utility of the questionnaire in other educational settings.

In this sense, the results should be understood as a preliminary validation of the instrument. It is recommended that the study be replicated in larger and more diverse samples, particularly in higher education contexts and virtual learning environments, to strengthen the generalizability of the findings and refine the instrument's sensitivity to different educational realities.

CONCLUSIONS

The development and validation of the AI-ED-SAT questionnaire represent a significant contribution to the field of educational innovation and teacher training, offering a rigorous, relevant, and up-to-date tool for self-assessing professional competencies related to the use of Artificial Intelligence (AI) in school contexts. Its design is based on a robust theoretical approach, a participatory construction informed by expert consensus, and a comprehensive psychometric validation process, ensuring its usefulness for both research and training improvement purposes.

Among the main contributions of the instrument are:

- Its structure in four key dimensions—conceptual understanding, pedagogical application, ethical approach, and curricular integration—comprehensively reflects the current challenges facing teaching practice in the face of advances in AI.

- Its formative and self-reflective nature allows teachers to identify strengths, areas for improvement, and professional development needs autonomously and in context.
- Its versatility for use in various training contexts, both face-to-face and virtual, makes it an adaptable tool for initial and continuing training programs, with the potential to personalize training itineraries.

In addition to its immediate contributions, the AI-ED-SAT opens up various lines of future research and development, among which the following are recommended:

- Expand its empirical validation with larger, more representative, and culturally diverse samples.
- Perform factorial invariance analyses to verify the stability of its structure at different educational levels, geographic regions, or groups of teachers with heterogeneous profiles.
- Develop linguistically and culturally adapted versions for non-Spanish-speaking contexts, considering both technical translation and semantic adaptation.
- Integrate the questionnaire into digital teacher training platforms, where the results can be linked to adaptive learning paths, promoting automated formative feedback.

In a global scenario where AI is redefining not only teaching and learning methods but also the ethical, social, and political frameworks that underpin educational decision-making, it is becoming increasingly urgent to provide teachers with tools that facilitate their critical literacy, professional autonomy, and transformative agency. In this sense, the AI-ED-SAT aspires to be more than just a measurement tool; it seeks to become a catalyst for reflection, empowerment, and ethical commitment among teachers in the face of the technological challenges of the 21st century.

REFERENCES

- Alíer Forment, M., García Peñalvo, F. J., Casañ Guerrero, M. J., Pereira, J. A., & Llorens-Largo, F. (2024, 8 October). *Safe AI in Education Manifesto* (Version 0.4.0). Safe AI in Education Manifesto. <https://manifesto.safeaieducation.org>
- Ayuso del Puerto, D., & Gutiérrez Esteban, P. (2022). La inteligencia artificial como recurso educativo durante la formación inicial del profesorado. *RIED-Revista Iberoamericana de Educación a Distancia*, 25(2), 347-362. <https://doi.org/10.5944/ried.25.2.32332>
- Baltazar, C. (2023). Herramientas de IA aplicables a la educación. *Technology Rain Journal*, 2(2), e15-e15. <https://doi.org/10.55204/trj.v2i2.e15>
- Bo, N. S. W. (2024). OECD digital education outlook 2023: Towards an effective education ecosystem. *Hungarian Educational Research Journal*, 15(2), 284-289. <https://doi.org/10.1556/063.2024.00340>
- Borge, R., García, J., Oliver, R., & Salomón, L. (2005). *Competencias y diseño de la evaluación continua y final en el Espacio Europeo de Educación Superior*. Dirección General de Universidades, MEC.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings.

- Journal of the Learning Sciences*, 2(2), 141-178.
https://doi.org/10.1207/s15327809jls02_02_2
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.
- Cabero-Almenara, J., Barroso-Osuna, J., Palacios-Rodríguez, A., & Llorente-Cejudo, C. (2020). Marcos de competencias digitales para docentes universitarios: Su evaluación a través del coeficiente competencia experta. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 23(3), 17-34.
<https://doi.org/10.6018/reifop.414501>
- Caena, F., & Redecker, C. (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (DigCompEdu). *European Journal of Education*, 54(3), 356-369.
<https://doi.org/10.1111/ejed.12345>
- 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>
- 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>
- Chesterman, S. (2021). Through a glass, darkly: Artificial intelligence and the problem of opacity. *The American Journal of Comparative Law*, 69(2), 271-294.
<https://doi.org/10.1093/ajcl/avabo12>
- Chiu, T. K., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, 100171.
<https://doi.org/10.1016/j.caeo.2024.100171>
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1), 1-9.
<https://doi.org/10.7275/jyj1-4868>
- Delgado, N., Carrasco, L. C., de la Maza, M. S., & Etxabe-Urbieta, J. M. (2024). Aplicación de la inteligencia artificial (IA) en educación: Los beneficios y limitaciones de la IA percibidos por el profesorado de educación primaria, educación secundaria y educación superior. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 27(1), 207-224.
<https://doi.org/10.6018/reifop.577211>
- Diefes-Dux, H. A., Zawojewski, J. S., & Hjalmanson, M. A. (2010). Using educational research in the design of evaluation tools for open-ended problems. *International Journal of Engineering Education*, 26(4), 807-819.
- Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos: ¿Por qué deberían importarnos? *Nueva Sociedad*, (294), 27-36.
<https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>
- Flores Vivar, J. M., & García Peñalvo, F. J. (2023). Reflexiones sobre la ética, potencialidades y retos de la inteligencia artificial en el marco de la Educación de Calidad (ODS4). *Comunicar: Revista Científica de Comunicación y Educación*, (74), 37-47.
<https://doi.org/10.3916/C74-2023-03>
- Gallant, D. J., & Luthy, N. (2020). Mixed methods research in designing an instrument for consumer-oriented evaluation. *Journal of MultiDisciplinary Evaluation*, 16(34), 21-43.
<https://doi.org/10.56645/jmde.v16i34.583>
- García Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 9-39.
<https://doi.org/10.5944/ried.27.1.37716>
- Ghomi, M., & Redecker, C. (2019, May). Digital competence of educators (DigCompEdu): Development and evaluation of a self-assessment

- instrument for teachers' digital competence. In *Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019)* (Vol. 1, pp. 541-548). SCITEPRESS. <https://doi.org/10.5220/0007679005410548>
- Grand-Guillaume-Perrenoud, J. A., Geese, F., Uhlmann, K., Blasimann, A., Wagner, F. L., Neubauer, F. B., Huwendiek, S., Hahn, S., & Schmitt, K.-U. (2023). Mixed methods instrument validation: Evaluation procedures for practitioners developed from the validation of the Swiss Instrument for Evaluating Interprofessional Collaboration. *BMC Health Services Research*, 23, 83. <https://doi.org/10.1186/s12913-023-09040-3>
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2006). *Metodología de la investigación* (4th ed.). McGraw-Hill Interamericana.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communalities, and overdetermination. *Educational and Psychological Measurement*, 65(2), 202-226. <https://doi.org/10.1177/0013164404267287>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570. <https://doi.org/10.1111/ejed.12533>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, 16(2), 2577-2587. <https://doi.org/10.15354/sief.23.re202>
- Khreisat, M. N., Khilani, D., Rusho, M. A., Karkkulainen, E. A., Tabuena, A. C., & Uberas, A. D. (2024). Ethical implications of AI integration in educational decision making: Systematic review. *Educational Administration: Theory and Practice*, 30(5), 8521-8527. <https://doi.org/10.53555/kuev.v30i5.4406>
- Lane, M., Williams, M., & Broecke, S. (2023). The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers (*OECD Social, Employment and Migration Working Papers*, No. 288). OECD Publishing. <https://doi.org/10.1787/ea0a0fe1-en>
- Linstone, H. A., & Turoff, M. (Eds.). (2002). *The Delphi method: Techniques and applications*. Addison-Wesley.
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, 1(3), 0028. <https://doi.org/10.1038/s41562-016-0028>
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99. <https://doi.org/10.1037/1082-989X.4.1.84>
- Mauri, T., Coll, C., & Onrubia, J. (2007). La evaluación de la calidad de los procesos de innovación docente universitaria: Una perspectiva constructivista. *Revista de Docencia Universitaria*, 5(1). <https://doi.org/10.4995/redu.2007.6290>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410601087>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16.

- <https://doi.org/10.7334/psicothema2018.291>
- Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, 71(1), 137-161. <https://doi.org/10.1007/s11423-023-10203-6>
- OECD. (2022). *Trends shaping education 2022*. OECD Publishing. <https://doi.org/10.1787/6ae8771a-en>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Basse, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self- and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education*, 34, 535-557. <https://doi.org/10.1007/s10212-018-0407-8>
- Paulin, A. M., Barriga-Arceo, F. D., Mendiola, M. S., & González, A. M. (2024). Evidencias de validez de un instrumento para evaluar la competencia digital docente en educación médica. *Investigación en Educación Médica*, 13(51), 82-92. <https://doi.org/10.22201/fm.20075057e.2024.51.23584>
- Rachbauer, T., Graup, J., & Rutter, E. (2025). Digital literacy and artificial intelligence literacy in teacher training. *Forum for Education Studies*, 3(1), 1842. <https://doi.org/10.59400/fesi1842>
- Ramírez, J. L. M. (2019). El proceso de elaboración y validación de un instrumento de medición documental. *Acción y Reflexión Educativa*, 44, 50-63. https://revistas.up.ac.pa/index.php/accion_reflexion_educativa/article/view/673
- Ramírez-Montoya, M. S., & Lugo-Ocando, J. (2020). Revisión sistemática de métodos mixtos en el marco de la innovación educativa. *Comunicar: Revista Científica de Comunicación y Educación*, 28(65), 9-20. <https://doi.org/10.3916/C65-2020-01>
- Reguant-Álvarez, M., & Torrado-Fonseca, M. (2016). El método Delphi. *REIRE. Revista d'Innovació i Recerca en Educació*, 9(1), 87-102. <https://doi.org/10.1344/reire2016.9.1916>
- Sampieri, R. H. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill México.
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. John Wiley & Sons.
- Sharma, D. M., Ramana, K. V., Jothilakshmi, R., Verma, R., Maheswari, B. U., & Boopathi, S. (2024). Integrating generative AI into K-12 curriculums and pedagogies in India: Opportunities and challenges. In M. S. P. Subathra & G. V. S. N. R. V. Prasad (Eds.), *Facilitating global collaboration and knowledge sharing in higher education with generative AI* (pp. 133-161). IGI Global. <https://doi.org/10.4018/979-8-3693-0487-7.ch006>
- Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6(1), 1-21. <https://doi.org/10.28945/199>
- UNESCO. (2021). *Artificial intelligence and the futures of learning: Towards an ethical and inclusive approach*. UNESCO. <https://www.unesco.org/en/digital-education/ai-future-learning>
- UNESCO. (2023). *Generative AI and the future of education*. UNESCO. <https://doi.org/10.54675/HOXG8740>
- UNESCO. (2024a). *AI competency framework for students*. UNESCO. <https://doi.org/10.54675/JKJB9835>
- UNESCO. (2024b). *AI competency framework for teachers*. UNESCO. <https://doi.org/10.54675/ZJTE2084>
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806-838.

<https://doi.org/10.1177/0011000006288127>

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher

education. *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>

APPENDIX 1

AI-ED-SAT Tool: First Version

Dimension 1: General knowledge of AI and its educational application

1. I am familiar with the basic concepts of AI and its applications in different sectors.
2. I understand how AI can influence education and learning.
3. I can identify AI tools that are applicable in the field of education.
4. I am familiar with terms such as machine learning and neural networks.
5. I understand the mathematical and logical foundations of AI.
6. I can distinguish between different types of AI and their applications in education.
7. I can explain to other teachers how basic AI works.
8. I know examples of AI used in learning platforms.
9. I can identify the advantages and limitations of AI in the educational context.
10. I have researched current trends in AI applied to education.

Dimension 2: Pedagogical use of AI tools

11. I use AI tools to improve my teaching practice.
12. I have designed learning activities that incorporate AI.
13. I am familiar with AI platforms that can support personalized teaching.
14. I evaluate the impact of AI tools on my students' learning.
15. I apply AI to the generation of teaching materials.
16. I use chatbots or virtual assistants to answer students' questions.
17. I implement AI for automatic task assessment.
18. I explore AI tools that promote adaptive learning.
19. I develop strategies to integrate AI into collaborative activities.
20. I participate in training on new AI tools in education.

Dimension 3: Ethics and considerations regarding AI in the classroom

21. I am aware of potential biases in AI systems.
22. I promote the ethical and responsible use of AI in the classroom.
23. I teach my students about privacy and security in the use of AI.
24. I reflect on the ethical challenges of automation in education.
25. I analyze how AI can influence decision-making.
26. I discuss with students the issues of transparency in AI algorithms.
27. I identify the social impacts of AI in education.
28. I propose practices to ensure the equitable use of AI.
29. I research AI regulations and standards in education.
30. I facilitate spaces for reflection on the impact of AI on society.

Dimension 4: Curricular integration of AI

31. I have integrated AI-related activities into my lesson plans.
32. I design teaching strategies that include AI as a learning tool.
33. I am familiar with curriculum frameworks that incorporate AI into education.

34. I participate in training on AI applied to teaching.
35. I develop interdisciplinary projects that include AI.
36. I use AI-based simulations to reinforce concepts in the classroom.
37. I coordinate activities with other teachers to integrate AI into the curriculum.
38. I research active methodologies that use AI in the classroom.
39. I design complete teaching units focused on AI.
40. I promote AI literacy among my students.

APPENDIX 2

Complete matrix of factor loadings, communalities (h^2), and specific variances for the AI-ED-SAT (EFA).

Item	Factor loading	Communal (h^2)	Specific variance ($1 - h^2$)
Item 1	0.72	0.52	0.48
Item 2	0.81	0.66	0.34
Item 3	0.75	0.56	0.44
Item 4	0.77	0.59	0.41
Item 5	0.80	0.64	0.36
Item 6	0.69	0.48	0.52
Item 7	0.73	0.53	0.47
Item 8	0.78	0.61	0.39
Item 9	0.74	0.55	0.45
Item 10	0.82	0.67	0.33
Item 11	0.76	0.58	0.42
Item 12	0.71	0.50	0.50
Item 13	0.79	0.62	0.38
Item 14	0.70	0.49	0.51
Item 15	0.73	0.53	0.47
Item 16	0.68	0.46	0.54
Item 17	0.75	0.56	0.44
Item 18	0.77	0.59	0.41
Item 19	0.72	0.52	0.48
Item 20	0.69	0.48	0.52
Item 21	0.70	0.49	0.51
Item 22	0.76	0.58	0.42
Item 23	0.74	0.55	0.45
Item 24	0.79	0.62	0.38
Item 25	0.71	0.50	0.50
Item 26	0.73	0.53	0.47
Item 27	0.69	0.48	0.52
Item 28	0.77	0.59	0.41
Item 29	0.75	0.56	0.44
Item 30	0.80	0.64	0.36
Item 31	0.78	0.61	0.39
Item 32	0.73	0.53	0.47
Item 33	0.76	0.58	0.42
Item 34	0.74	0.55	0.45
Item 35	0.72	0.52	0.48
Item 36	0.70	0.49	0.51
Item 37	0.69	0.48	0.52
Item 38	0.81	0.66	0.34

Note: The values follow the reported psychometric pattern, with high loadings and solid explained variance (67.8%).

Date of reception: 1 June 2025

Date of acceptance: 21 August 2025

Date of approval for layout: 23 September 2025

Date of publication in OnlineFirst: 15 October 2025

Date of publication: 1 January 2026