For citation / Atıf için:
Temur, S. (2026). Using Artificial Intelligence in measurement and evaluation processes in education. *Uluslararası Sosyal Bilimler ve Eğitim Dergisi – USBED 8(*14), 1-32. https://doi.org/10.5281/zenodo.17559801, https://dergipark.org.tr/tr/pub/usbed

Using Artificial Intelligence in measurement and evaluation processes in education

Süleyman TEMUR
Ph.D. (c) Niğde Ömer Halisemir Üniversitesi, Eğitim Fakültesi, Sosyal Bilgiler Eğitimi, 51000, Niğde, Türkiye
E-mail: temursuleyman19@gmail.com *ORCID: 0000-0002-5203-6553*

Makale Türü / Article Type:	Araştırma Makalesi / Research Article
Gönderilme Tarihi / Submission Date:	08/09/2025
Revizyon Tarihleri / Revision Dates:	12/10/2025 (Minor r.)
Kabul Tarihi / Accepted Date:	08/11/2025

Etik Beyan / Ethics Statement

✓ The author(s) declare(s) that the study is not subject to ethics committee approval.

Araştırmacıların çalışmaya katkısı / Researchers' contribution to the study

Author contribution: Wrote the article, collected the data, and analyzed/reported the results (100%).

Çıkar çatışması / Conflict of interest

The author(s) declare(s) that there is no potential conflict of interest in this study.

Benzerlik / Similarity

This study was scanned in the Turnitin program. The final similarity rate is 5 %.

Using Artificial Intelligencein measurement and evaluation processes in education

Abstract

In recent years, the potential benefits of Artificial Intelligence(AI) in education have gained significant momentum. Foremost among these gains are its contributions to the measurement and evaluation processes that form the foundation of education. In this context, this study examines the potential benefits, application areas, risks, and recommendations that can be developed for the use of AI in educational measurement and evaluation, employing a systematic literature review method. The research findings reveal that AI has the potential to offer significant contributions to students (in-depth performance insights, individualized education, identification of at-risk students, increased engagement and motivation), educators/administrators (time savings, automated grading, increased efficiency, objective evaluation), and measurement-evaluation processes (validity/reliability analysis, automated item generation, test improvement, data-driven decision-making). However, it is emphasized that ethical violations that may arise with the use of AI should also be considered. Specifically, the importance of adopting a controlled approach under teacher guidance in educational measurement and evaluation processes is one of the key findings of the study. In this context, various recommendations have been presented in technical, pedagogical, ethical, and legal dimensions for the effective and safe use of AI in educational measurement and evaluation processes. These recommendations aim to maximize the benefits from the innovative opportunities offered by AI while minimizing potential risks.

Keywords: Artificial Intelligence, Education, Measurement and evaluation, Potential benefits, Risks.

EXTENDED ABSTRACT

Introduction

This study aims to comprehensively examine the potential benefits and risks of integrating Artificial Intelligenceinto measurement and evaluation processes in education, using a systematic literature review method. The research highlights the opportunities offered by Artificial Intelligenceto students, educators, and measurement-evaluation processes, such as in-depth performance analysis, personalized education, and automated evaluation, while also drawing attention to potential risks, including ethical violations, data security concerns, and bias. In this respect, addressing the pedagogical and ethical dimensions of Artificial Intelligencetogether, thereby filling a gap in the existing literature, constitutes the unique and valuable aspect of the study. The findings emphasize the importance of controlled use of Artificial Intelligenceunder teacher guidance, offering concrete recommendations on the technical, pedagogical, ethical, and legal dimensions for effectively and safely implementing this technology. These recommendations make significant contributions to the field by aiming to maximize the innovative opportunities offered by Artificial Intelligencewhile minimizing potential risks.

Conceptual and Theoretical Framework

The rapid proliferation of digital technologies in education has necessitated the restructuring of traditional learning and assessment processes. This transformation has profoundly impacted measurement and evaluation approaches, particularly revealing the inadequacies of traditional tools in providing holistic performance assessments and effective feedback. In this context, Artificial Intelligence(AI) stands out as an integrable solution for measurement and evaluation processes in education. In the literature, academics and education stakeholders widely agree on AI's potential to enhance assessment quality and efficiency, emphasizing its capabilities such as automated scoring, instant feedback, and in-depth data analysis. The primary objective of this study is to comprehensively examine

the potential uses, benefits, and possible ethical risks of AI in the field of measurement and evaluation in education. In this regard, it aims to make a significant contribution to the literature by offering solutions to overcome the limitations of existing approaches and by providing a holistic perspective on the pedagogical and ethical dimensions of AI use. Finally, the research offers concrete recommendations across various dimensions for the responsible and effective use of AI in education.

Concepts

1. Artificial Intelligence(AI) is a branch of science that enables computer systems to imitate human-like abilities such as learning, problem-solving, and decision-making.
2. Measurement and Evaluation: Measurement and evaluation are processes that involve quantifying observations (measurement) to determine the characteristics of a particular situation or object, and then assigning meaning to these results to reach a judgment (evaluation).

Method

This research aims to thoroughly examine the potential benefits, current application areas, possible risks, and methods for mitigating these risks associated with using Artificial Intelligencein educational assessment and evaluation processes, through a systematic literature review. To this end, a detailed search strategy was employed using leading academic databases, including Web of Science, Scopus, ERIC, ResearchGate, and TR Dizin. Articles, theses, and conference papers obtained with predefined keyword combinations were meticulously filtered according to predefined inclusion and exclusion criteria, guided by the PRISMA flow diagram. Two independent researchers reviewed full texts of selected studies, and data extraction was systematically performed under themes such as potential benefits, risks, ethical issues, and future recommendations. The collected data were analyzed using a thematic coding method, aiming to comprehensively address the pedagogical and ethical dimensions of Artificial Intelligence's role in this field.

Findings

The findings of this study reveal that integrating Artificial Intelligenceinto measurement and evaluation processes in education offers significant benefits and potential risks. The data obtained indicates that AI contributes to these processes in three main dimensions: the student, the educator, and the measurement-evaluation processes themselves. Among the student-focused benefits, in-depth performance analysis and individualized learning strategies are particularly notable. In contrast, educator and administrator-focused benefits include time savings, automated scoring, and the possibility of more objective evaluations. Process-focused benefits encompass aspects such as facilitating validity and reliability analyses, automating item generation, and enhancing data-driven decision-making processes. However, the examinations conducted also reveal various risks associated with the use of Artificial Intelligence. These risks include critical issues such as algorithmic bias, lack of transparency, reliability problems, accountability concerns, and data security vulnerabilities. The findings emphasize the need to ensure that this technology is used relatively, reliably, and ethically, while benefiting from the efficiency and personalization opportunities offered by Artificial Intelligence.

Conclusion, Discussion, and Recommendations

This study reveals that the integration of Artificial Intelligenceinto educational assessment and evaluation processes offers multifaceted benefits, including personalized feedback and performance analysis for students, time savings and objective evaluation for educators, and validity-reliability analyses for assessment processes.

However, it is emphasized that significant ethical and practical risks, including algorithmic bias, data privacy concerns, and a lack of transparency, must also be meticulously addressed. Consistent with other studies in the literature, Artificial Intelligenceshould be seen not as a tool to replace teachers, but as an assistant that strengthens their professional roles and supports processes. In this context, the importance of training aimed at developing teachers' AI literacy skills for the effective and responsible use of Artificial Intelligencehas been highlighted. Recommendations suggest that AI technologies should be developed and implemented with an approach that aligns with pedagogical goals, adheres to ethical principles, and prioritizes the human factor.

INTRODUCTION

The integration of digital technologies into current educational processes has necessitated a radical restructuring of these processes (Senkivska, 2022). In this context, the use of technology in education offers a wide range of advantages, from solving problems faced by disadvantaged groups to diversifying and personalizing instruction, from accelerating and facilitating learning processes to developing creative thinking and ensuring remote accessibility to education (Günüç, 2017). Traditional, teacher-centered, rote-learning-based, and mostly multiple-choice test-characterized education systems have given way to modern educational paradigms that center the learner and emphasize the application and transfer of knowledge. This transformation is closely related to the capacity of developing technologies to perform basic operations via computers (Karakaya et al., 2024). This change has also profoundly affected assessment methods in education. As a fundamental component of the modern education system, assessment serves multiple purposes, including facilitating student learning, determining instructional outcomes, and informing educational policies (Tan et al., 2025). This is because the acquisition of knowledge requires testing; whether knowledge has been learned is determined and evaluated through measurement tools (Yeşilyurt, 2025). Furthermore, assessment is the backbone of quality education (Sok & Heng, 2023).

Parallel to the rapid development in education and training processes, the evolution of measurement processes into an assessment system supported by innovative technologies has become an inevitable necessity. The traditional measurement and evaluation methods widely applied today appear to be a hindrance in obtaining healthy results in a system where innovative technologies are integrated into education and training processes to the extent of creating a paradigm shift (Bulut & Akyıldız, 2024). However, traditional measurement and evaluation tools may be insufficient in holistically assessing student performance and providing effective feedback (Gao & Grisham-Brown, 2011). Furthermore, in traditional approaches, validity issues can sometimes arise in teacher-made tests (Effendi & Mayuni, 2022). For these reasons, it is crucial to revise measurement and evaluation processes in line with the requirements of the current age. Especially considering the expansion of Artificial Intelligence(AI)-based tools in education with various applications in different disciplines (Naidu & Sevnarayan, 2023; Temur,

2024a), the potential use of AI in the field of measurement and evaluation is also increasing (Bulut et al., 2024; Owan et al., 2023; Şahin et al., 2024). In this context, an AI transformation is also underway in measurement and evaluation research in education (Ayva et al., 2024). Indeed, one of the significant reflections of the rapid spread of innovative technologies in the educational world is observed in the field of measurement and evaluation (Bulut & Akyıldız, 2024). Recent developments in the field of AI have led to the emergence of automated essay scoring systems that promise efficient, reliable, and objective evaluation of written texts (İnceçay et al., 2024). Moreover, the elements that need to be considered when preparing instructional materials and especially measurement tools, such as "economy, practicality, and individual needs," have begun to be frequently emphasized as the development of technology and the active use of AI in education and training have started, and "AI" has entered the scene as a facilitating tool (Yeşilyurt, 2025).

Academics, in their research, consistently highlight AI's role as a valuable tool in measurement and evaluation, emphasizing its capacity for diverse assessment methods, immediate and comprehensive feedback, and robust data analysis capabilities (Güler et al., 2025; Korkmaz et al., 2025; Ölçer et al., 2025). They also identify specific benefits for autonomous evaluation, enhanced integrity, assignment checking, and automated assessment systems (Khan et al., 2024; Kuzembayeva et al., 2025), noting that AI improves evaluation quality and ease of use (Tovar & Gutiérrez Ocegueda, 2025). Similarly, teachers across various disciplines recognize AI's potential in question preparation, data analysis, efficient and rapid evaluation, and identifying student deficiencies (Seyrek et al., 2024; Şenel, 2024). Preschool and science teachers, in particular, emphasize AI's utility in developmental evaluation, objective measurement through precise data analysis, and multi-criteria evaluation opportunities, preventing measurement errors (Küçükkara et al., 2024; Sanca, 2022). Furthermore, school administrators and information technology teachers concur that AI automates tasks like data analysis and feedback, making evaluations more objective and accelerating performance assessment (Kaya, 2024; Danış, 2025). Prospective teachers also foresee the benefits of AI in individualized student assessment and addressing knowledge gaps (Çam et al., 2021). Collectively, these perspectives underscore a broad consensus among educational stakeholders regarding the significance of AI in measurement and evaluation, particularly its potential in data analysis, automated scoring, feedback provision, and process acceleration.

The primary objective of this study, in light of the available information, is to comprehensively explore the potential applications, benefits, and potential ethical implications of AI in educational measurement and evaluation. Furthermore, examining the impact of AI use on measurement and evaluation processes, which are among the most critical components of

education, in parallel with its increasing prevalence in education, is also an important objective of this research. Moreover, in line with the findings obtained, recommendations have been presented for a more responsible and effective use of AI in educational measurement and evaluation processes.

METHODOLOGY

This study aims to thoroughly examine the potential benefits, current application areas, potential risks, and recommendations for minimizing these risks associated with AI use in educational measurement and evaluation processes, employing a systematic literature review method. Systematic literature reviews are a scientific method that involves carefully selecting and analyzing previously published scientific works according to meticulously defined criteria to find answers to specific research questions (Karasar, 2023). In this context, the research aims to comprehensively analyze the relevant literature through a meticulously designed, multi-step process.

Databases and Search Strategy

To ensure access to current and authoritative sources in the literature for the literature review, the leading scientific databases were used. These databases and the search strategy used are as follows:

- Web of Science (WoS)
- Scopus
- ERIC
- ResearchGate
- TR Dizin

Detailed searches were conducted for each database using the following keyword combinations:

- ("Artificial Intelligencein education", "AI education", "education AI") and ("assessment evaluation", "evaluation assessment", "exam", "test", "grading")
- ("Artificial Intelligence", "AI") AND ("assessment in education", "evaluation in education")
- ("potential benefits", "opportunities", "advantages", "effectiveness") or ("potential risks", "ethical issues", "ethical dilemmas", "disadvantages", "limitations")
- ("future perspectives", "future trends", "recommendations")

Although no restrictions were applied regarding the publication year during the search process, priority was given to current literature, especially from the last 6 years (2018-2024), to reflect the rapid developments in AI and assessment-evaluation in education. In terms of language

limitations, the focus was on English and Turkish publications, which constitute a significant portion of academic production.

Inclusion and Exclusion Criteria

Inclusion Criteria:

- Studies directly addressing AI applications in educational assessment and evaluation processes.
- Studies examining the potential benefits, risks, and ethical implications of AI applications in assessment and evaluation.
- Articles published in peer-reviewed journals, theses, conference papers, and book chapters.
- Studies with a clearly stated research design and findings.

Exclusion Criteria:

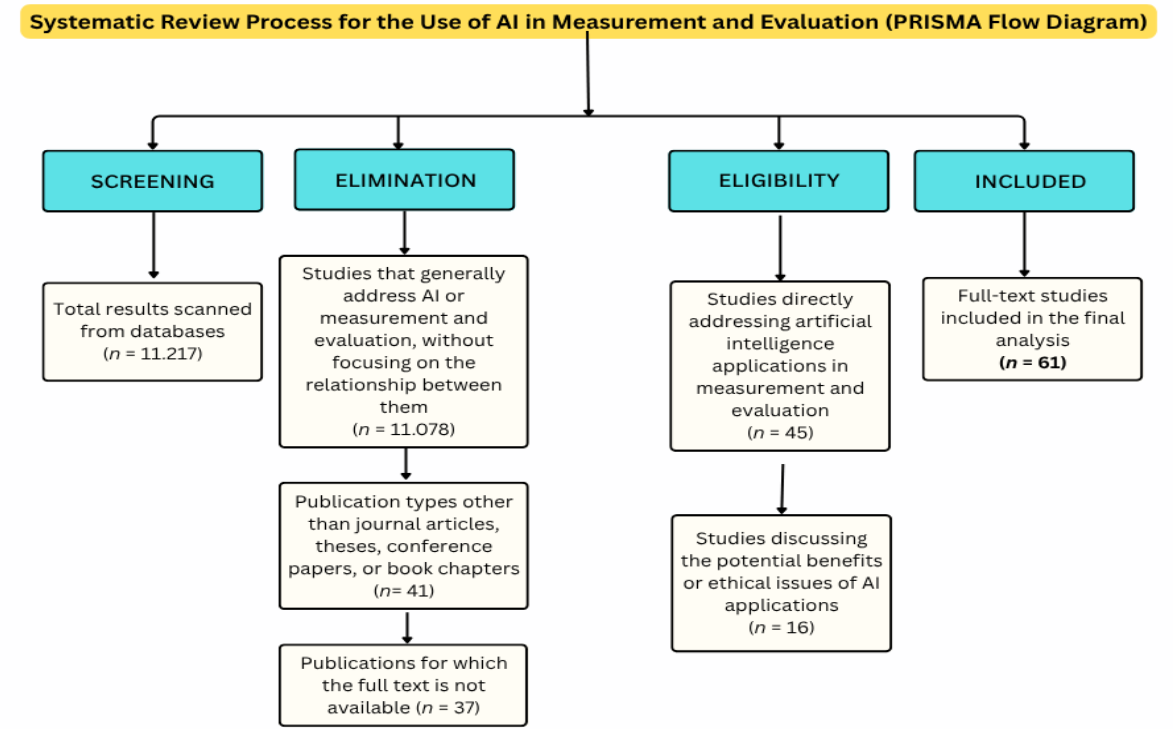
- Studies that generally address AI or assessment and evaluation, but do not focus on the relationship between the two.
- Publications for which the full text is not accessible.
- Publication types other than journal articles, theses, conference papers, or book chapters (e.g., opinion pieces, news articles, promotional materials, blog posts).

Study Selection Process and Data Extraction

The selection of acquired sources was carried out through an objective and repeatable process. All studies identified as a result of the search were first pre-screened based on inclusion and exclusion criteria by examining their titles and abstracts. The full texts of studies deemed suitable after pre-screening were obtained and read in detail to make the final inclusion decision. Two independent researchers conducted this process to reduce potential biases and ensure consistency.

The study selection process and stages were visualized using the “PRISMA Flow Diagram” (Figure 1). This diagram presents the entire process step-by-step, from the initial number of studies obtained from databases to screening, eligibility assessment, and the final number of included studies.

Figure 1. Systematic Review Process For The Use Of AI İn Measurement And Evaluation (PRISMA Flow Diagram)



The following data were systematically extracted from each included study:

- Study Aim/Research Questions
- Method Used
- Main Findings
- Potential Benefits of AI in Educational Measurement and Assessment Processes
- Potential Risks and Ethical Issues Related to AI in Educational Measurement and Assessment Processes
- Future Recommendations

The obtained sources were first filtered by examining their titles and abstracts in line with the research objectives. Then, the full texts of the sources deemed suitable according to the determined criteria were accessed and analyzed in detail. In this analysis process, key themes such as AI's contributions to measurement and evaluation processes (automatic grading, individualized feedback, item analysis, etc.), the current situation in application areas, potential risks (ethical issues, data security, bias, etc.), and strategies developed or proposed to reduce these risks were identified. The systematic review and analysis of the literature aimed to gain a deep understanding of the current knowledge base of AI in educational measurement and evaluation, the perspectives of different researchers on this topic, and future research and

application areas. This approach ensured that the study's findings were based on solid foundations and offered a comprehensive perspective.

Data Analysis

The collected literature was analyzed in detail using the thematic coding method. In this analysis process, predetermined main categories, including potential benefits, potential risks, and ethical issues related to the use of AI in measurement and evaluation processes in education, served as a basis. Thematic coding is a qualitative analysis method that enables the systematic identification and classification of recurring themes, concepts, and patterns within textual data (Baltacı, 2019).

As a result of the analysis, the number of included studies, their type (e.g., experimental, qualitative, mixed methods), publication year range, and central themes were summarized in tables. These tables allowed for the transparent and understandable presentation of the study's findings. The reviewed literature played a critical role in establishing the conceptual and analytical framework necessary for evaluating the use of AI in measurement and evaluation processes in education. This framework aims to make a significant contribution to the literature by addressing the role of AI in measurement and evaluation, considering both its pedagogical and ethical dimensions.

FINDINGS

Definition of AI and Its Use in Education

AI systems are tools that can be trained to simulate the human brain and perform routine tasks using large datasets, which McCarthy (2007, p. 2) defined as "the science and engineering of making intelligent machines" (Bengio et al., 2021). In general, AI can be characterized as computerized systems capable of performing tasks typically associated with human intelligence, such as learning, problem-solving, and achieving goals under specific or uncertain variable conditions (Dalton, 2024).

In education, AI refers to the application of intelligent computer systems to perform functions that typically demand human cognitive abilities, such as personalized learning, data analysis, and performance evaluation (Temur, 2024a). Within this context, educators can leverage AI's capabilities to analyze extensive student data, identify patterns and trends, and glean insights into individual learning needs (Çavuş, 2024). Lameras and Arnab (2022) point out that AI in education, a field of interest for approximately three decades, encompasses the design, implementation, and evaluation of tools, pedagogical models, teaching strategies, and frameworks, as well as the ethical implications and necessary teacher competencies for its practical use. Recent national and international studies (Ilham et al., 2024; Temur, 2025a;

Yadav & Anubha, 2024) consistently demonstrate a significant surge in research dedicated to the application of AI within the educational domain.

Sharma and Kumar (2023) emphasize the potential of AI to revolutionize learning environments, stating that this technology has garnered significant interest from educators, researchers, and policymakers. Temur (2024a) classified the potential application areas emerging from the integration of AI into educational processes as adaptive learning, instructional evaluation, virtual classrooms, smart campuses, and intelligent teaching robots. Furthermore, the impact of AI in education is supported by studies conducted in various fields such as language teaching, music education, museum education, and mathematics education (Çangal et al., 2025; Keleş, 2024; Nayıroğlu & Tutak, 2024; Özdemir, 2024; Schuster, 1986; Türker, 2024). Additionally, potential benefits that AI can provide in the field of education include personalized learning experiences (Popenici & Kerr, 2017), development of problem-solving skills (Cui & Wang, 2017), automated assessment and feedback (Ndukwe et al., 2019), accessibility and inclusivity (Cardona et al., 2023), interactive learning tools (Ayala-Pazmiño, 2023), improved teaching methods (Çelik at al., 2022), academic achievement and motivation (Peras at al., 2023), curriculum design (Alqahtani et al., 2023), and facilitation of administrative tasks (Qin at al., 2020).

In line with this information, the study provides a detailed examination of the potential benefits of integrating AI into educational assessment and evaluation processes. The study presents the advantages that AI offers to students, teachers, and administrators, as well as its effectiveness in this field, by compiling current literature findings. Furthermore, the possibilities offered by AI-based solutions have been evaluated across a broad spectrum. Accordingly, many benefits have been emphasized, from contributing to teachers' professional time management to the capacity for more accurate and objective evaluation of student performance. Additionally, specific application areas of AI, such as test development, scoring, providing feedback, and creating individualized education strategies, have been addressed. Finally, various tools and methods that AI can apply in educational measurement and evaluation are listed, and findings on the effectiveness of this technology are expressed in light of current studies in the literature.

Potential Benefits of AI Use in Measurement and Evaluation Processes in Education

AI-powered data analysis tools surpass traditional assessment approaches, enabling educators and administrators to gain in-depth insights into student performance, identify trends, predict future academic outcomes, and suggest targeted interventions for students in need of support (Bulut et al., 2024). The rapid development of AI technologies has brought significant innovations and improvements to assessment and evaluation processes in schools (Boduroğlu & Yiğiter, 2024). Indeed, AI, utilizing advanced technologies such as data analytics, machine

learning, and natural language processing, enables the validity and reliability analysis of exams and tests (Dumlu et al., 2024).

The integration of AI into assessment and evaluation processes offers significant advantages not only for students but also for teachers and administrators (Baker et al., 2016). This is because assessment and evaluation, a critical part of the teaching and learning process, is an area where teachers spend almost one-third of their professional time (Stiggins, 2014). In this context, AI-based solutions provide teachers with the opportunity to assess student performance more accurately, quickly, and objectively, while also enabling the development of individualized educational strategies tailored to students' needs (Bittencourt et al., 2020). Huang et al. (2023a) stated that AI systems support educators by saving time in identifying and designing assessment tools/methods appropriate for the curriculum, goals, and objectives. Furthermore, Boduroğlu and Yiğiter (2024) stated that AI has potential benefits in areas such as test content validity, item difficulty analysis, item discrimination analysis, test time management, language and comprehensibility control, bias analysis, test internal consistency, rubric creation, automatic scoring, and feedback provision. Çavuş (2024) also stated that specific areas where AI can be effectively applied in measurement and evaluation processes in education include grading, predictive data analysis, administrative analyses, similarity detection, data visualization, and the design of various assessment tools.

Studies on the Effectiveness of AI in Educational Measurement and Evaluation

Numerous recent academic studies support the potential of AI in measurement and evaluation. These studies reveal the innovations that AI offers compared to traditional methods, the benefits it provides, and some of the challenges it encounters.

Research indicates that AI can enhance the effectiveness of distractors in assessment tools, thereby improving item discrimination (Ceylan & Gökçe, 2024). In automated essay scoring, AI-based language models like ChatGPT-4.0 have shown comparable reliability to human evaluators, promising to alleviate educator workload and ensure consistent evaluations in foreign language writing assessments (İnceçay et al., 2024). Furthermore, AI detection tools show promise in measuring psychomotor skills, with findings suggesting that their data aligns closely with expert observations (Kucam & Malakcıoğlu, 2024). AI also significantly supports teachers in refining the psychometric properties of their assessments, contributing to enhanced test validity and reliability through diverse analyses and feedback (Boduroğlu & Yiğiter, 2024). Beyond specific tool development, AI shows broad applicability across educational levels. In preschool education, AI-supported models can personalize instruction and utilize data-driven measurements, thereby optimizing teaching processes and addressing individual student needs (Güven & Yumugan, 2024). The academic community generally concurs on AI's valuable role

in measurement and evaluation processes, with detailed examinations highlighting the benefits of tools like ChatGPT in these contexts (Bulut & Akyıldız, 2024). Despite these advancements, challenges remain. Studies on AI-generated exam content reveal inconsistencies in question appropriateness, length, and difficulty, underscoring the necessity of human oversight and expert guidance to ensure quality and effectiveness (Yeşilyurt, 2025). Overall, while AI offers substantial advantages in automated scoring, feedback provision, and psychometric improvement, its practical and reliable integration into educational measurement and evaluation processes necessitates continuous human expertise and supervision.

When these studies are evaluated collectively, it becomes apparent that AI has significant potential for application in the field of measurement and evaluation, offering distinct advantages, particularly in areas such as automated scoring, providing feedback, and enhancing the psychometric properties of tests. However, it is also emphasized that human expertise and continuous supervision are required for AI to be used effectively and reliably.

Within the framework of this information, specific application areas of AI in the measurement and evaluation field of education, along with the potential advantages these applications can provide, are presented below.

Specific Application Areas and Potential Advantages of AI in Educational Measurement and Evaluation

- Automated Grading: Teachers can save significant time by utilizing automated grading systems, allowing them to focus more on core tasks such as lesson planning and providing student support (Adiguzel et al., 2023; Whitmer et al., 2023).
- Learning Analytics Tools: Educators can effectively utilize learning analytics tools to monitor student progress comprehensively. These tools enable them to identify students who are at risk and facilitate data-driven decision-making aimed at improving overall student learning outcomes (Önder et al., 2023).
- In-depth Performance Insights: AI can transform assessment and evaluation in education by enabling students to gain in-depth performance insights (Çavuş & Kuzilek, 2024). AI can identify students' conceptual or procedural deficiencies through detailed error analysis and observe problem-solving approaches by tracking learning processes. This enables personalized feedback tailored to each student, while also identifying learning styles to facilitate the personalization of course materials (Eaton, 2023). Thanks to AI's predictive analytics capabilities, students' future difficulties can be anticipated, laying the groundwork for proactive interventions. All of this can help educators understand how students learn, creating more effective and personalized learning environments.

- Various Digital Tools: AI-powered tools such as formative assessment tools (Mentimeter, Nearpod, and Socrative), digital portfolios (Google Sites and WordPress), data visualization tools (Google Data Studio, Infogram, and Tableau), gamification tools (Classcraft, Kahoot!, and Quizlet), and digital assessment tools (Edulastic, ExamView, and Google Forms) can be effectively used in assessment and evaluation processes (Çavuş, 2024; Owan et al., 2023).
- The use of large language models in educational assessment and evaluation encompasses automated item generation, assessment tool development, assessment tool review and validity checking, automated scoring, formative assessment, and analysis of test results (Çüm, 2024).
- Scoring of Open-Ended Items: AI offers significant potential in scoring open-ended items through advanced "natural language processing" (NLP) and "machine learning" (ML) techniques (Beiting-Parrish & Whitmer, 2023). Studies in the literature (Alers et al., 2024; Jukiewicz, 2024; Poole & Coss, 2024; Yiğiter & Boduroğlu, 2025) have generally shown a strong correlation between student responses and AI and human scores. However, studies on AI performance in scoring handwritten open-ended items appear to be limited, and the issue has not yet been fully clarified.
- Automated Item Generation: Today, open-ended items can be generated with AI tools (Karakaya, 2024). A literature review by Kurdi et al. (2019) found that studies on automated item generation have increased since the mid-2010s.
- Cost-Effectiveness and Objectivity: AI-powered assessment tools have the potential to offer faster and potentially more accurate and objective assessments at a lower cost compared to human evaluators (Naidu & Sevnarayan, 2023). AI systems reduce inconsistency (the reliability problem) stemming from fatigue, distraction, or subjective judgments in human assessment by consistently applying scoring criteria. However, the claim that AI increases "objectivity" is a contradictory and incomplete argument in the context of algorithmic bias research in the relevant literature. This is because if the historical data used to train AI systems contains societal and cultural biases (Thalpage et al., 2025), the AI system learns these biases and systematically reproduces them (Whittlestone et al., 2019). Consequently, even if the assessment process is purged of human subjectivity, the algorithm itself can produce unfair and discriminatory outcomes against certain demographic groups (e.g., based on race, gender, or socio-economic status) (Wang & Redelmeier, 2025). This situation suggests that while AI enhances technical reliability, it can significantly compromise its ethical validity and the principle of fairness (Temur, 2025c). Therefore, while emphasizing cost-effectiveness and

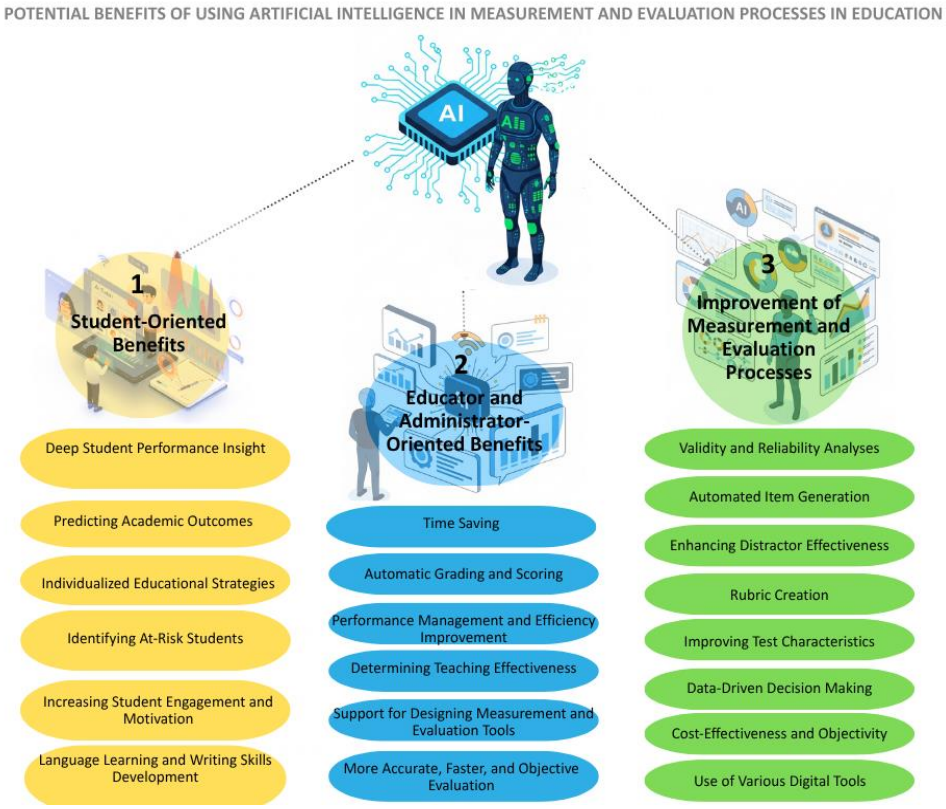
potential objectivity, it must be stated that continuous oversight and corrections to reduce algorithmic bias are vital for AI systems to be fair, inclusive, and ethically valid.

- Student Engagement and Motivation: Çevrimiçi değerlendirme metodolojilerine oyunlaştırmanın entegrasyonu, öğrenci katılımını ve motivasyonunu önemli ölçüde artırabilir. Daha sürükleyici ve keyifli bir değerlendirme deneyimi yaratarak, oyunlaştırma genel öğrenme çıktılarını iyileştirme potansiyeline sahiptir (Kapp, 2012).
- Language Learning and Writing Skills: AI-powered NLP tools provide substantial support for students in language acquisition and enhancing writing skills. These tools facilitate learning by providing targeted feedback on crucial aspects such as grammar, spelling, and punctuation (Owan et al., 2023).
- Determining Instructional Effectiveness: AI-based assessment and evaluation tools can provide educators with valuable insights into the effectiveness of their teaching methodologies. By analyzing assessment data, these tools help teachers identify specific areas where adjustments to their teaching strategies may be necessary (Huang et al., 2023b). This capability supports a data-driven approach to instructional improvement.
- Identification of At-Risk Students: AI-powered tools offer a significant advantage in identifying students who may need additional support. By analyzing student data, including test scores and attendance records, these tools can discern patterns indicative of a need for intervention. This capability helps educators proactively identify students at risk of falling behind, as highlighted by Sato, Shyyan, Chauhan, and Christensen (2024). Consequently, this allows teachers to provide targeted support precisely to those students who need it most, as emphasized by Delgado et al. (2020).
- Performance Management: In performance management processes, AI applications can be used to objectively evaluate and improve teacher performance (Buck & Marrow, 2018).
- Increased Efficiency: In a study on the use of generative AI in assessment and evaluation processes, Terwiesch (2023) reported that the approximately 20 hours of work typically required to create an exam were reduced to 10 hours, resulting in a 50% efficiency increase thanks to generative AI.

As in many disciplines, the increasing use of AI technologies in educational sciences brings significant transformations, especially in measurement and evaluation processes. The time- and labor-intensive nature of traditional measurement and evaluation approaches, along with their limited ability to respond to individual differences, has highlighted AI's potential in this field. AI-based systems have the potential to enhance students' learning processes by making them more dynamic, interactive, and personalized, while also reducing the workload of educators

and administrators and optimizing decision-making processes (Tanberkan et al., 2024). From this perspective, the comprehensive potential benefits that AI can provide to measurement and evaluation processes are detailed in Figure 2, which clarifies the conceptual framework and presents a graphical summary to facilitate understanding. This visualization helps to understand the benefits detailed in the text from a holistic perspective and concretizes AI's transformative power in this field.

Figure 2. Potential Benefits of Using AI in Measurement and Evaluation Processes in Education



In addition to the potential benefits offered by AI technologies in education, particularly in assessment and evaluation processes, significant risks and ethical concerns also arise from their use. The high data processing and computational capacity of AI systems in decision-making processes bring to the forefront issues of moral and social responsibility. This section provides a detailed discussion of the potential risks associated with the use of AI in educational assessment and evaluation processes, in light of international legal and ethical regulations, as well as academic and teacher perspectives.

Risks Associated with the Use of AI in Measurement and Evaluation Processes in Education

While the conveniences offered by AI technology to human life cannot be overlooked, the development of this technology also brings increasing ethical concerns to the forefront. The fact that AI systems have more data and computational power than humans in decision-making processes reinforces the importance of these systems acting in accordance with moral values (Temur, 2024b). Southgate (2019) identified AI bias and the lack of transparency in AI systems as the most urgent ethical issues in AI applications in education. Dignum (2018) drew attention to concerns in the ethical dimension of AI, including the moral, social, and legal consequences of AI's decision-making process, the issue of responsibility, and the problem of control. Furthermore, the concept of AI singularity has also been recently discussed in relation to the ethical implications of AI. This concept entails the prediction that AI systems will develop autonomously and enter a process that will surpass human control, ultimately leading to technological singularity (Kurzweil, 1990). In this context, since the birth of AI, the concern that this technology could develop uncontrollably and pose a threat to humanity has become widespread (Köroğlu, 2017). Therefore, questions such as "Can a robot perform good or bad actions?" and "Are robots a threat to humanity?" are frequently debated in the public sphere. Ersoy (2020) emphasizes that one of the most fundamental questions regarding AI is who will be responsible for the errors caused by AI. Moreover, rapid advancements in technology, along with the development of AI hardware and software, raise new ethical issues regarding the possibility of these systems exceeding human control and making their own independent decisions in the future (Öztürk Dilek, 2019). This situation raises uncertainties about whether AI poses a threat or an opportunity for humanity. Therefore, more comprehensive research in the field of AI is crucial to answering these questions.

As a direct consequence of these advancements, while AI-powered innovations undoubtedly present promising opportunities for developing more robust, efficient, and personalized assessments, they simultaneously introduce serious concerns. These concerns primarily revolve around the critical psychometric properties of validity and reliability, as well as ethical considerations such as fairness, transparency, equity, and overall test security (Hao et al., 2024). These algorithms can inadvertently perpetuate or even exacerbate existing biases present in their training data. Within an educational context, such algorithmic bias has the potential to exacerbate existing inequalities and disadvantage specific student groups by unfairly influencing assessment outcomes, such as test scores or grades (Bulut et al., 2024). The existence of algorithmic bias in AI-based evaluation systems has been demonstrated with concrete examples in the literature. Indeed, in the e-evaluators examined by Bridgeman et al.,

African-American male students were found to systematically receive slightly lower scores from automated evaluation systems compared to human evaluators (Litman et al., 2021). This situation illustrates the general finding presented by Wang and Redelmeier (2025) that AI can produce biases related to race and gender.

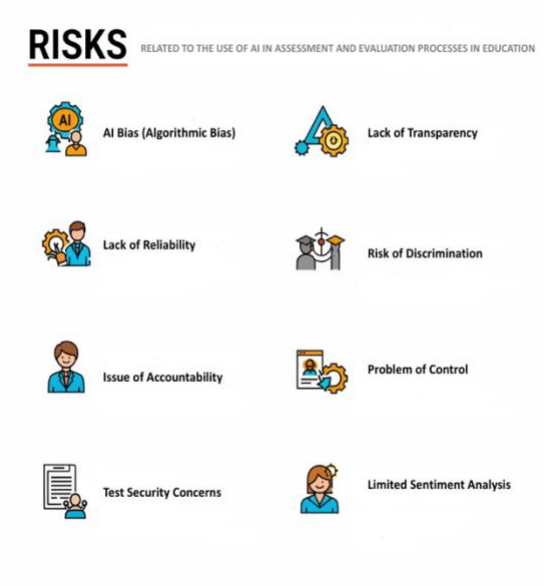
Furthermore, an AI application used to evaluate job applications at St. George's Medical School in London disregarded the academic achievements of female candidates and those from certain countries, such as Pakistan, discriminating solely based on gender and nationality, and automatically rejected many applications (Öztürk Dilek, 2019). These concrete cases confirm the potential of AI systems to create unfair and artificial barriers against minority and disadvantaged groups in education and similar fields. Naidu and Sevnarayan (2023) emphasize that, despite the numerous possibilities offered by the use of AI in online assessment, it is critically important to approach this approach cautiously and consider the ethical implications for academic integrity in online assessment. Furthermore, various academic and teacher groups also draw attention to the risks of using AI. In a study by Güler et al. (2025), academics highlighted potential dangers, including a lack of reliability and the risk of discrimination, in measurement and evaluation processes. Sanca (2022) also reported that science teachers expressed limitations of AI, such as its limited ability to perform sentiment analysis.

To mitigate the inherent risks associated with integrating AI in educational assessment, various international initiatives have been launched. A notable example is the "Duolingo English Test's Responsible AI Standards" (Johnson, 2024), which guides the collaborative and judicious application of AI and human expertise. This framework aims to ensure the development of reliable, safe, and practical assessments for all stakeholders. Similarly, the "National Council of Measurement in Education" established the "Special Interest Group on AI in Measurement and Education" (AIEM) in 2025. AIEM's primary objective is to advance both theoretical and applied research concerning the use of AI in educational measurement, with a key focus on establishing comprehensive guidelines for its ethical implementation.

In conclusion, these risks indicate that a cautious approach must be adopted when integrating AI into educational measurement and assessment processes. Establishing ethical principles, minimizing algorithmic biases, and maintaining human oversight are essential to fully leverage AI's potential in this field and prevent possible negative consequences.

Within this framework, the risks associated with integrating AI into education assessment and evaluation processes are presented in detail in Figure 3, incorporating both visual and textual elements.

Figure 3. Risks Related to the Use of Software in Measurement and Evaluation Processes in Education



The potential benefits and accompanying risks that AI technologies bring to educational measurement and assessment processes offer an important roadmap for future work in this area. The practical and ethical use of AI in education requires technological advancements to progress in harmony with pedagogical principles. Accordingly, it is essential to consider the recommendations detailed below to maximize the potential of AI in measurement and assessment while minimizing potential adverse outcomes. These recommendations are presented under four main headings: technical and pedagogical development, ethical and fair use, implementation and process management, and legal and institutional framework.

Suggestions for Future Work

Given the potential benefits and possible risks mentioned above, it is crucial for future studies to focus on the following recommendations for the practical and ethical use of AI in measurement and evaluation processes:

Technical and Pedagogical Development-Oriented Recommendations

- Adoption of Hybrid Approaches: The development of hybrid models should be encouraged, where AI-supported automated assessment systems are combined with human expertise, especially in evaluating complex or open-ended responses.
- Development of Adaptive Assessment Systems: The focus should be on the widespread adoption of adaptive AI-based assessment systems that can adjust to students' learning paces and styles, dynamically modify difficulty levels, and provide immediate feedback.

- Utilization of AI in Enhancing Teacher-Created Tests: User-friendly AI-based platforms should be developed to enable teachers to more easily and accurately analyze the psychometric properties (validity, reliability, and item analysis) of their self-prepared tests with the aid of AI tools.

Ethical and Fair Use-Oriented Recommendations

- Research on Reducing Algorithmic Biases: In-depth research should be conducted to detect and minimize potential biases in AI algorithms. This research is critical for developing fairer and more equitable assessment systems, as stated by Sato et al. (2024).
- Focus on Transparency and Explainability: Efforts should be made to make the decision-making processes of AI-based assessment systems more transparent and explainable.
- Increasing Student Inclusivity and Accessibility: AI tools should be designed to enable more effective participation of students with learning difficulties or special needs in assessment processes. For example, features such as speech recognition, text-to-speech conversion, and alternative presentation formats can be integrated. This is a prerequisite for inclusive measurement and evaluation (Temur, 2025b).

Implementation and Process Management-Oriented Recommendations

- Establishment of Periodic Evaluation Mechanisms: Regular and systematic evaluation mechanisms should be established to monitor the impact of AI on teaching and learning (Tovar & Gutiérrez Ocegueda, 2025).
- Continuous Analysis of Teacher and Student Feedback: Feedback data should be regularly collected and analyzed from both teachers and students to adjust and improve the implementation of learning strategies continually.
- Development of Teachers' AI Competencies: To ensure the effective and ethical integration of AI-powered assessment systems into teaching and learning processes, teachers need to acquire specific competencies through systematic and structured professional development programs. These programs, based on international standards such as the UNESCO ICT Competency Framework (United Nations Educational, Scientific and Cultural Organization, 2018), should focus on two main competency areas: AI Literacy and Ethical Awareness (ethical issues such as fundamental AI concepts, algorithmic operating principles, data privacy, algorithmic bias, and the preservation of academic integrity) and Competence in Using and Analyzing AI-Powered Assessment Tools (selection and application of tools such as automated scoring systems and adaptive tests, interpretation of large datasets, and personalization

of teaching strategies based on these analytical outputs). The design of professional development programs should include at least 40 hours of theoretical and practical workshops, with an emphasis on practical project development activities that utilize teachers' real classroom materials, aiming to transform competencies into lasting learning. Finally, the success of the program and the level of teacher competency should be measured multidimensionally, not only through knowledge tests but also through portfolio-based practical projects developed and self-assessment forms reflecting adherence to ethical principles, and supported by monitoring in-class application performance.

Legal and Institutional Framework-Oriented Recommendations

- Development of Ethical and Legal Frameworks: Comprehensive ethical and legal frameworks for the use of AI in measurement and evaluation processes in education should continue to be developed at national and international levels.
- Promoting Interdisciplinary Collaboration: Interdisciplinary collaboration should be encouraged among AI experts, educators, ethicists, legal professionals, and policymakers to foster a more comprehensive understanding of the field. Araştırmada verilerin hangi yöntem ve araçlarla elde edildiği belirtilir.

DISCUSSION and CONCLUSION

With the rapid integration of digital technologies, existing educational paradigms are undergoing a radical transformation. In this context, the integration of AI technologies into assessment and evaluation processes has the potential to support student learning and increase educator efficiency (Bolender et al., 2024). However, it is necessary to distinguish between the benefits that AI has been documented to provide in the literature and those that have been practically proven, and those that are still at a theoretical level or in a potential stage. When examining the potential benefits within the scope of the study, these advantages have been classified into three main groups: student-oriented, educator-oriented, and those aimed at improving assessment and evaluation processes. Moreover, the benefits of AI need to be distinguished between those proven and documented in practice in the literature and those that are still theoretical or potential. Indeed, proven benefits include practical applications such as measurable time savings and educator efficiency provided through automated scoring and rapid data analysis. In contrast, theoretical benefits represent ambitious goals that have not yet reached their full potential and require further research and ethical maturity, such as accurately predicting students' risk of academic failure, developing deeply individualized educational strategies, or automatically improving the psychometric properties of tests.

A. Proven and Documented Benefits in Practice (Current Efficiency)

These benefits are areas where AI applications are already providing efficiency gains, and their results are widely documented in academic literature:

1. Educator and Administrator-Focused Benefits:

- Time Saving and Automated Scoring: Automatic grading and scoring of tests, especially multiple-choice, true-false, and even short-answer questions. This provides educators with measurable time savings.
- Data Analysis and Feedback: Faster and more accurate analysis of student performance data using AI tools and provision of initial feedback based on this analysis.

2. Benefits for Assessment and Evaluation Processes:

- Objectivity and Consistency: Automated scoring systems increase the objectivity and consistency of the evaluation process by minimizing subjectivity caused by human factors.
- Use of Various Digital Tools: Support for assessment tool design (e.g., rubric drafting) and facilitation of the use of different digital tools in processes.

B. Theoretical or Potential Benefits (Future Development Areas)

These benefits are based on the advanced analytical capabilities offered by AI, but require further application, ethical regulation, and academic evidence to reach their full potential:

3. Student-Focused Benefits:

- Identifying At-Risk Students and Academic Predictions: Predicting student academic outcomes and early identification of student failures (at-risk students) are key targets for developing learning analytics tools.
- Individualized Strategies: Developing fully individualized educational strategies based on in-depth performance information.
- Deep Performance Analysis: More complex and qualitative outcomes, such as increasing student engagement and motivation, and improving language learning and writing skills.

4. Benefits for Assessment and Evaluation Processes:

- Psychometric Improvement: Automation of validity and reliability analyses of tests and improvement of the psychometric properties of tests based on these analyses.
- Automated Item Generation and Distractor Effectiveness: Fully automated item (question) generation with AI and increased effectiveness of distractors.

Consequently, applications such as automated grading, data analysis, and feedback provision demonstrate that AI is a proven facilitator in measurement and evaluation. Meanwhile,

ambitious outcomes such as risk identification, academic outcome prediction, and individualized strategies represent the transformative potential that AI has yet to achieve fully. Academic and teacher opinions in the relevant literature reflect a broad consensus on the effectiveness of AI in data analysis and automated scoring.

However, the use of AI in educational measurement and evaluation processes also brings certain ethical and practical risks. Algorithmic bias, lack of transparency, data security and privacy, misleading results, technical failures, and over-reliance are significant risks that need to be meticulously addressed for AI systems to be used fairly, reliably, and effectively. Being aware of these risks and adhering to internationally developed ethical guidelines and legal regulations is vital to minimizing the adverse effects of AI. Initiatives such as the "Duolingo English Test's Responsible AI Standards" and the "National Council on Measurement in Education's AIEM Special Interest Group" are important steps taken towards developing ethical and reliable practices in this field. Academician and teacher opinions also highlight the benefits of AI, such as its robust and sensitive data analysis capabilities, qualified data analysis, and process-oriented evaluation, as well as its limitations, including limited emotion analysis and issues with level appropriateness in test creation processes.

Considering all these potential benefits and possible risks, a careful and balanced approach is necessary for the use of AI in educational assessment and evaluation processes. The automation, personalization, and in-depth analysis capabilities provided by AI undoubtedly have the potential to improve the quality of education. However, it is essential that these technologies are developed in line with ethical principles and that algorithmic biases are continuously monitored and minimized. Most importantly, as stated by Dillenbourg (2016), the transition to digital education and the rise of AI do not mean that teachers' traditional roles will be eliminated. AI should not be seen as a tool that replaces the teacher, but rather as a supporter that strengthens the teacher's role, making them more efficient and effective. Teachers will continue to play a central role in interpreting the data provided by AI, adapting it to student needs, and making pedagogical decisions that require a human touch. In this context, it is crucial to develop teachers' AI literacy skills and provide them with the necessary training to utilize AI tools and manage potential risks effectively. Teachers who strategically utilize the opportunities offered by AI to enrich assessment and evaluation processes will ultimately contribute to students having a higher-quality learning experience and to the overall improvement of the education system.

REFERENCES

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), ep429. https://doi.org/10.30935/cedtech/13152

AI in Measurement and Education. (2025). *About*. https://www.ncme-aime.org/about.html

Alers, H., Malinowska, A., Meghoe, G., & Apfel, E. (2024). Using ChatGPT-4 to grade open question exams. In K. Arai (Ed.), *Advances in information and communication* (pp. 1–9). Springer Nature. https://doi.org/10.1007/978-3-031-53960-2_1

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., & Albekairy, A. M. (2023). The emergent role of Artificial Intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy, 19*(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Ayala-Pazmiño, M. F. (2023). Inteligencia artificial en la educación: Explorando los beneficios y riesgos potenciales [Artificial Intelligence in education: Exploring the potential benefits and risks]. *Digital Publisher CEIT, 8*(3), 892–899. https://doi.org/10.33386/593dp.2023.3.1827

Ayva Yörü, F. G., & Büyükkıdık, S. (2024, October 4–6). *Eğitimde ölçme ve değerlendirme alanında Yapay Zekâ: Sistematik bir derleme* [Artificial Intelligence in educational measurement and evaluation: A systematic review] [Oral presentation]. International Measurement, Selection and Placement Symposium (p. 55). Ankara, Türkiye.

Baker, R. S., Martin, T., & Rossi, L. M. (2016). Educational data mining and learning analytics. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 379–396). Wiley. https://doi.org/10.1002/9781118956588.ch16

Baltacı, A. (2019). Nitel araştırma süreci: Nitel bir araştırma nasıl yapılır? [Qualitative research process: How to conduct qualitative research]. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 5*(2), 368–388. https://doi.org/10.31592/aeusbed.598299

Beiting-Parrish, M., & Whitmer, J. (2023). Lessons learned about evaluating fairness from a data challenge to automatically score NAEP reading items. *Chinese/English Journal of Educational Measurement and Evaluation, 4*(3), 1–11. https://doi.org/10.59863/NKCJ9608

Bittencourt, I. I., Cukurova, M., Muldner, K., Luckin, R., & Millán, E. (2020, July 6–10). *Artificial Intelligence in education* [Conference session]. 21st International Conference on Artificial Intelligencein Education (AIED 2020), Ifrane, Morocco.

Boduroğlu, E., & Yiğiter, M. S. (2024). Öğretmen yapımı testler için yapay zekâ destekli geribildirim [AI-powered feedback for teacher-created tests]. *Journal of Applied Measurement and Assessment, 2024*(1), 50–58.

Bolender, B., Vispoel, S., Converse, G., Koprowicz, N., Song, D., & Osaro, S. (2024). Generative AI in K12: Analytics from early adoption. *Journal of Measurement and Evaluation in Education and Psychology, 15*(Special Issue), 361–377. https://doi.org/10.21031/epod.1539710

Buck, B., & Marrow, J. (2018). AI, performance management and engagement: Keeping your best their best. *Strategic HR Review, 17*(5), 261–262. https://doi.org/10.1108/SHR-10-2018-145

Bulut, G., & Akyıldız, M. (2024). Yapay zekâ ile üretilen soruların ve madde parametrelerinin MST test koşullarında karşılaştırılması [Comparison of AI-generated questions and item parameters under MST testing conditions]. *Dijital Teknolojiler ve Eğitim Dergisi, 3*(1), 1–12. https://doi.org/10.5281/zenodo.12637347

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., Ormerod, C., Fabiyi, D. G., Ivan, R., Walsh, C., Rios, O., Wilson, J., Yildirim-Erbasli, S. N., Wongvorachan, T., Liu, J. X., Tan, B., & Morilova, P. (2024). The rise of Artificial Intelligence in educational measurement: Opportunities and ethical challenges. *Chinese/English Journal of Educational Measurement and Evaluation, 5*(3), 1–32. https://doi.org/10.59863/MIQL7785

Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). *Artificial Intelligence and the future of teaching and learning: Insights and recommendations*. U.S. Department of Education, Office of Educational Technology. https://tech.ed.gov/

Ceylan, M., & Gökçe, S. (2024, October 4–6). *Test geliştirmede paradigma değişimi: Yapay zekâ destekli çeldiricilerin etkililiği* [Paradigm shift in test development: The effectiveness of Artificial Intelligence-assisted distractors] [Oral presentation]. International Measurement, Selection and Placement Symposium (pp. 228–229). Ankara, Türkiye.

Cui, W., Xue, Z., & Thai, K. P. (2019). Performance comparison of an AI-based adaptive learning system in China. *2018 Chinese Automation Congress* (pp. 3170–3175). https://doi.org/10.1109/CAC.2018.8623327

Çam, M. B., Çelik, N. C., Turan Güntepe, E., & Durukan, Ü. G. (2021). Öğretmen adaylarının yapay zekâ teknolojileri ile ilgili farkındalıklarının belirlenmesi [Determining prospective teachers' awareness of Artificial Intelligence technologies]. *Hatay Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 18*(48), 263–285.

Çangal, Ö., Çelik, M. E., & Başar, U. (2025). Yabancılara Türkçe öğretiminde yapay zekâ kullanımına yönelik öğretici görüşleri [Instructor opinions on the use of Artificial Intelligence in teaching Turkish to foreigners]. *Aydın Tömer Dil Dergisi, 10*(1), 57–97.

Çavuş, M. N. (2024). Eğitimde yapay zekâ tabanlı ölçme ve değerlendirme üzerine bir derleme [A review on AI-based assessment and evaluation in education]. *International Journal of English for Specific Purposes, 2*(1), 39–54.

Çavuş, M., & Kuzilek, J. (2024). An effect analysis of the balancing techniques on the counterfactual explanations of student success prediction models. *Journal of Measurement and Evaluation in Education and Psychology, 15*(Special Issue), 302–317. https://doi.org/10.21031/epod.1526704

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

Çüm, S. (2024). Eğitimde ölçme ve değerlendirmede insan-yapay zekâ iş birliği: Büyük dil modellerinin kullanımı [Human-Artificial Intelligence collaboration in measurement and evaluation in education: The use of big language models]. *Journal of Applied Measurement and Assessment, 1*(2), 29–39.

Dalton, G. (2024). *Artificial Intelligence: Background, risks, and policies*. Nova Science Publishers.

Danış, F. (2025). *Bilim ve sanat merkezlerinde görev yapan bilişim teknolojileri öğretmenlerinin eğitimde yapay zekâ kullanımına ilişkin görüşleri* [The opinions of information technologies teachers working in science and art centers on the use of Artificial Intelligencein education] [Master's thesis, Fırat University]. YÖK Ulusal Tez Merkezi [HEC National Thesis Center].

Delgado, H. O., de Azevedo Fay, A., Sebastiany, M. J., & Silva, A. D. (2020). Artificial Intelligence adaptive learning tools. *BELT-Brazilian English Language Teaching Journal, 11*(2), 1–19. https://doi.org/10.15448/2178-3640.2020.2.38749

Dignum, V. (2018). Ethics in Artificial Intelligence: Introduction to the special issue. *Ethics and Information Technology, 20*, 1–3. https://doi.org/10.1007/s10676-018-9450-z

Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education, 26*(2), 544–560. https://doi.org/10.1007/s40593-016-0106-z

Dumlu, B. Ö., Gezer, E., & Yıldız, B. (2024). Eşitsizlik konusunda ChatGPT ile hazırlanan ders planlarının incelenmesi [Examination of lesson plans prepared with ChatGPT on inequality]. *Türk Eğitim Bilimleri Dergisi, 22*(1), 337–358. https://doi.org/10.37217/tebd.1338959

Eaton, S. (2023). *Academic integrity and Artificial Intelligence: Implications for plagiarism and academic writing* [Conference session]. PUPP. https://www.youtube.com/watch?v=9QNNPVSC24w

Effendi, T., & Mayuni, I. (2022). Examining teacher-made English test in a language school. *LADU: Journal of Languages and Education, 2*(2), 67–76. https://doi.org/10.56724/ladu.v2i2.109

Ersoy, Ç. (2020). *Robotlar, yapay zekâ ve hukuk* [Robots, Artificial Intelligence and law] (5th ed.). On İki Levha Yayıncılık.

Gao, X., & Grisham-Brown, J. (2011). The use of authentic assessment to report accountability data on young children's language, literacy, and pre-math competency. *International Education Studies, 4*(2), 41–53. https://doi.org/10.5539/ies.v4n2p41

Güler, E., Uğur, S., & Güler, C. (2025). Eğitim fakültesi öğretim elemanlarının üretken yapay zekâya yönelik farkındalıklarının belirlenmesi ve öğretmen yetiştirmede kullanımına yönelik öneriler [Determination of faculty members' awareness of productive Artificial Intelligence and suggestions for its use in teacher training]. *Mehmet Akif Ersoy University Journal of Education Faculty, 74*, 618–647. https://doi.org/10.21764/maeuefd.1627851

Güven, G., & Yumugan, S. (2024, October 4–6). *Okul öncesi eğitimde yapay zekâ: Yapay zekâ ile kişiselleştirilmiş ölçme ve değerlendirme* [Artificial Intelligencein preschool education: Personalized assessment and evaluation with Artificial Intelligence] [Oral presentation]. International Measurement, Selection and Placement Symposium (pp. 86–87). Ankara, Türkiye.

Hao, J., von Davier, A. A., Yaneva, V., Lottridge, S., von Davier, M., & Harris, D. J. (2024). Transforming assessment: The impacts and implications of large language models and generative AI. *Educational Measurement: Issues and Practice, 43*(2), 16−29. https://doi.org/10.1111/emip.12602

Hrastinski, S. O., Arkenback, C., Ekström, S., Ericsson, E., Fransson, G., Jaldemark, J., Ryberg, T., Öberg, L. M., Fuentes, A., Gustafsson, U., Humble, N., Mozelius, P., Sundgren, M., & Utterberg, M. (2019). Critical imaginaries and reflections on Artificial Intelligence and robots in post-digital K-12 education. *Post-Digital Science and Education, 1*(2), 427–445. https://doi.org/10.1007/s42438-019-00046-x

Huang, A. Y., Lu, O. H., & Yang, S. J. (2023a). Effects of Artificial Intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education, 194*, 104684. https://doi.org/10.1016/j.compedu.2022.104684

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023b). Trends, research issues, and applications of Artificial Intelligence in language education. *Educational Technology & Society, 26*(1), 112–131.

Ilham, R., Giatman, M., & Maksun, H. (2024). Artificial Intelligence research in education: A bibliometric analysis. *Journal on Education, 6*(2), 13467–13479.

İnceçay, V., Aktekin, N. Ç., & Kocabaş, C. (2024, October 4–6). *Üniversite eğitimlerini ingilizce alan öğrencilerin yabancı dil yazma becerilerinin yapay zekâ tabanlı puanlama ve insan tarafından değerlendirilmesinin karşılaştırmalı bir analizi* [A comparative analysis of Artificial Intelligence-based scoring and human assessment of foreign language writing skills of university students majoring in English] [Oral presentation]. International Measurement, Selection and Placement Symposium (pp. 148–149). Ankara, Türkiye.

Johnson, M. S. (2024). How do we demonstrate AI responsibility: The devil is in the details. *Journal of Educational and Behavioral Statistics, 49*(5), 723–729. https://doi.org/10.3102/10769986241257963

Jukiewicz, M. (2024). The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process. *Thinking Skills and Creativity, 52*, 101522. https://doi.org/10.1016/j.tsc.2024.101522

Karakaya, İ., Öç, U., & Gümüş, G. (2024). Üst düzey zihinsel becerilerin ölçülmesinde açık uçlu maddelerin yapay zekâ destekli uygulamalarda hazırlanması [Preparation of open-ended items in Artificial Intelligence-supported applications in the measurement of high-level cognitive skills]. *Journal of Applied Measurement and Assessment, 1*(2), 59–70.

Karasar, N. (2023). *Bilimsel araştırma yöntemi: Kavramlar, ilkeler, teknikler* [Scientific research method: Concepts, principles, techniques] (38th ed.). Nobel Akademik Yayıncılık.

Kaya, İ. (2024). *Eğitim yönetiminde yapay zekâ kullanımına ilişkin okul yöneticileri ve öğretmenlerin görüşleri* [Opinions of school administrators and teachers on the use of Artificial Intelligence in educational administration] [Master's thesis, İstanbul Aydın University]. YÖK Ulusal Tez Merkezi [HEC National Thesis Center].

Keleş, O. (2024). *Yapay zekâ destekli müzik dersi aktivitelerinin öğrencilerin ders motivasyonlarına etkisi* [The effect of Artificial Intelligence-supported music lesson activities on students' course motivation] [Master's thesis, Atatürk University]. YÖK Ulusal Tez Merkezi [HEC National Thesis Center].

Khan, W., Topham, L. K., Atherton, P., Al-Shabandar, R., Kolivand, H., Khan, I., & Hussain, A. (2024). Auto-assessment of assessment: A conceptual framework towards fulfilling

the policy gaps in academic assessment practices. *Computers and Society*, 1–20. https://doi.org/10.48550/arXiv.2411.08892

Korkmaz, H., Çiloğlu, R., & Mut, A. İ. (2025). In-service mathematics teachers' considerations of technological developments and the use of Artificial Intelligence in mathematics teaching. *Pedagogical Perspective, 4*(1), 114–127. https://doi.org/10.29329/pedper.2025.95

Köroğlu, Y. (2017). *Yapay Zekâ'nın teorik ve pratik sınırları* [Theoretical and practical limits of Artificial Intelligence]. Boğaziçi Üniversitesi Yayınevi.

Kucam, E., & Malakcıoğlu, İ. (2024, October 4–6). *Psikomotor becerilerin ölçülmesinde yapay zekâ tespit araçlarından elde edilen sonuçların doğruluğunun incelenmesi* [Examining the accuracy of the results obtained from Artificial Intelligence detection tools in measuring psychomotor skills] [Oral presentation]. International Measurement, Selection and Placement Symposium (pp. 70–71). Ankara, Türkiye.

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2019). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education, 30*, 121–204. https://doi.org/10.1007/s40593-019-00186-y

Kurzweil, R. (1990). *The age of intelligent machines*. MIT Press.

Kuzembayeva, G., Sultangubiyeva, A., & Tusupbekova, N. (2025). University instructors' perspectives on pros and cons of Artificial Intelligence in education. *Bulletin of the Khalel Dosmukhamedov Atyrau University, 76*(1), 89–96. https://doi.org/10.47649/vau.25.v76.i1.08

Küçükkara, M. F., Ünal, M., & Sezer, T. (2024). Okul öncesi eğitimi öğretmenlerinin yapay zekâya ilişkin görüşleri [Preschool education teachers' views on Artificial Intelligence]. *Temel Eğitim Araştırmaları Dergisi, 4*(1), 17–28. https://doi.org/10.55008/te-ad.1431142

Litman, D., Zhang, H., Correnti, R., Matsumura, L. C., & Wang, E. (2021). A fairness evaluation of automated methods for scoring text evidence usage in writing. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Artificial Intelligence in education: AIED 2021* (pp. 255–267). Springer, Cham. https://doi.org/10.1007/978-3-030-78292-4_21

McCarthy, J. (2007). *What is artificial?* http://jmc.stanford.edu/articles/whatisai.html

Naidu, K., & Sevnarayan, K. (2023). ChatGPT: An ever-increasing encroachment of Artificial Intelligence in online assessment in distance education. *Online Journal of Communication and Media Technologies, 13*(3), e202336. https://doi.org/10.30935/ojcmt/13291

Nayıroğlu, B., & Tutak, T. (2024). Matematik öğretiminde yapay zekanın rolü: Eğitimde kullanılan araçların incelenmesi [The role of Artificial Intelligence in mathematics teaching: An examination of tools used in education]. *Turkish Journal of Educational Studies, 11*(1), 65–78. https://doi.org/10.33907/turkjes.1415591

Ndukwe, I. G., Daniel, B. K., & Amadi, C. E. (2019). A machine learning grading system using chatbots. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luck (Eds.), *Artificial Intelligence in education: AIED 2019* (Vol. 11626, pp. 365–368). Springer. https://doi.org/10.1007/978-3-030-23207-8_67

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, 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), em2309. https://doi.org/10.29333/ejmste/13428

Ölçer, S., Cesur, B., Kurşun-Şen, Z., Çetinkale, E., & Gök, B. (2025). The opinions of academics on the usage of Artificial Intelligence tools in primary schools. *Educational Academic Research, 57*, 67–80. https://doi.org/10.33418/education.1515939

Önder, A., Öztaş, G. S., & Akçapınar, G. (2023). Öğrenme analitiği sürecine yönelik modellere genel bir bakış: Kavramsal bir çerçeve önerisi [An overview of models for the learning analytics process: A conceptual framework proposal]. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi, 9*(1), 92–117. https://doi.org/10.51948/auad.1163540

Özdemir, Ö. F. (2024). *Güzel sanatlar eğitiminde yapay zekanın resim üzerine katkısına yönelik bir inceleme* [An examination on the contribution of Artificial Intelligence to painting in fine arts education] [Master's thesis, İnönü University]. YÖK Ulusal Tez Merkezi [HEC National Thesis Center].

Öztürk Dilek, G. (2019). Yapay zekânın etik gerçekliği [Ethical reality of Artificial Intelligence]. *Ankara Uluslararası Sosyal Bilimler Dergisi, 2*(4), 47–59.

Peras, C. M. B., Aviluna, J. M., Barbadillo, N. R. A., Canoy, A. T., Eslet, M. E. R., Española, J. B. O., Miras, L. J. R., & Nepangue, J. (2023). Artificial Intelligence as a tool for increasing academic performance. *International Journal of Advanced Multidisciplinary Research and Studies, 3*(6), 1151–1155.

Poole, F. J., & Coss, M. D. (2024). Can ChatGPT reliably and accurately apply a rubric to L2 writing assessments? The devil is in the prompt(s). *Journal of Technology & Chinese Language Teaching, 15*(1), 1–24.

Popenici, S. A., & Kerr, S. (2017). Exploring the impact of Artificial Intelligence on teaching and learning in higher education. *Research and Practice in Technology-Enhanced Learning, 12*(2), 1–13. https://doi.org/10.1186/s41039-017-0062-8

Qin, F., Li, K., & Yan, J. (2020). Understanding user trust in Artificial Intelligence-based educational systems: Evidence from China. *British Journal of Educational Technology, 51*(5), 1693–1710. https://doi.org/10.1111/bjet.12994

Sanca, M. (2022). *Fen öğretiminde yapay zekâ destekli teknolojilerin kullanımına ilişkin fen bilimleri öğretmenlerinin görüşlerinin incelenmesi* [The analyzing of the points of science teachers concerning the usage of Artificial Intelligence technologies in science teaching] [Master's thesis, Van Yüzüncü Yıl University]. YÖK Ulusal Tez Merkezi [HEC National Thesis Center].

Sato, E., Shyyan, V., Chauhan, S., & Christensen, L. (2024). Putting AI in fair: A framework for equity in AI-driven learner models and inclusive assessments. *Journal of Measurement and Evaluation in Education and Psychology, 15*(Special Issue), 263–281. https://doi.org/10.21031/epod.1526527

Schuster, E. (1986). The role of native grammars in correcting errors in second language learning. *Computational Intelligence, 2*(1), 93–98. https://doi.org/10.1111/j.1467-8640.1986.tb00074.x

Senkivska, L. (2022). The role of digital technologies in education. *Journal of Education, Health and Sport, 12*(1), 419–423. http://dx.doi.org/10.12775/JEHS.2022.12.01.036

Seyrek, M., Yıldız, S., Emeksiz, H., Şahin, A., & Türkmen, M. T. (2024). Öğretmenlerin eğitimde yapay zekâ kullanımına yönelik algıları [Teachers' perceptions on the use of Artificial Intelligence in education]. *International Journal of Social and Humanities Sciences Research (JSHSR), 11*(106), 845–856. https://doi.org/10.5281/zenodo.11113077

Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.4378735

Southgate, E. (2019). *Artificial Intelligencein schools: An ethical storm is brewing*. EduResearch Matters, Australia Association for Research in Education. https://blog.aare.edu.au/artificial-intelligence-in-schools-an-ethical-storm-is-brewing/

Stiggins, R. (2014). Improve assessment literacy outside of schools too. *Phi Delta Kappan, 96*(2), 67–72.

Şahin, A., Thompson, N., & Ercikan, K. (2024). Opportunities and challenges of AI in educational assessment. *Journal of Measurement and Evaluation in Education and Psychology, 15*(Special Issue), 260–262. https://doi.org/10.21031/epod.1607441

Şenel, M. (2024). Examining teachers' views on the use of Artificial Intelligence (AI) in education. *i-manager's Journal of Educational Technology, 20*(4), 13–30. https://doi.org/10.26634/jet.20.4.20237

Tan, B., Armoush, N., Mazzullo, E., Bulut, O., & Gierl, M. J. (2025). A review of automatic item generation techniques leveraging large language models. *International Journal of Assessment Tools in Education, 12*(2), 317–340. https://doi.org/10.21449/ijate.1602294

Tanberkan, H., Özer, M., & Gelbal, S. (2024). Impact of Artificial Intelligence on assessment and evaluation approaches in education. *International Journal of Educational Studies and Policy, 5*(2), 139–152. https://doi.org/10.5281/zenodo.14016103

Temur, S. (2024a). Yapay zekânın eğitim sistemine entegrasyonunun potansiyel faydaları [Potential benefits of integrating Artificial Intelligence into the education system]. *Batı Anadolu Eğitim Bilimleri Dergisi, 15*(3), 2621–2656. https://doi.org/10.51460/baebd.1541524

Temur, S. (2024b). Yapay zekânın etik boyutu ve hukuki düzenlemeler: Güncel durum ve gelecek perspektifleri [Ethical dimension of Artificial Intelligence and legal regulations: Current status and future perspectives]. In F. Ayaz & H. Ayaz (Eds.), *Yapay zekâ denkleminde "sorgulananlar" (medya, iletişim ve eğitim çalışmaları)* ["Questioned" in the Artificial Intelligence equation (media, communication and education studies)] (pp. 145–172). Eğitim Yayınevi.

Temur, S. (2025a). 2000-2024 yılları arasında eğitim alanında yapılan yapay zekâ konulu lisansüstü çalışmaların incelenmesi [An investigation of Artificial Intelligence postgraduate studies in the field of education between the years 2000-2024]. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi, 63*, 1181–1218. https://doi.org/10.53444/deubefd.1600146

Temur, S. (2025b). Türkiye yüzyılı maarif modeli öğretim programları ortak metninin kapsayıcı ölçme ve değerlendirme ilkeleri açısından incelenmesi [Examination of the Türkiye century education model curriculum common text in terms of inclusive measurement and evaluation principles]. *Eğitim ve Bilim, 50*(ek sayı 1), 19–41. https://doi.org/10.15390/eb.2025.13988

Temur, S. (2025c). Eğitimde yapay zekâ kullanımı: Etik sorunlar ve çözümler [Using Artificial Intelligencein education: Ethical issues and solutions]. *Mehmet Akif Ersoy University Journal of Education Faculty, 74*, 568–595. https://doi.org/10.21764/maeuefd.1516576

Terwiesch, C. (2023). *Would ChatGPT3 get a Wharton MBA? A prediction based on its performance in the operations management course*. Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania. https://tinyurl.com/3fw4xm6t

Thalpage, N. S., Epa, D., & Jayawardhana, S. (2025). Ethical challenges in explainable AI: A review on cultural and social bias. *JDAH, 6*(1), 30–39. https://doi.org/10.33847/2712-8148.6.1_3

Tovar, I. Z., & Gutiérrez Ocegueda, G. J. R. (2025). Attitudes of university professors towards the use of Artificial Intelligence in teaching and learning. *International Journal of Multidisciplinary Research and Analysis, 8*(1), 364–387. https://doi.org/10.47191/ijmra/v8-i01-46

Türker, O. (2024). *Güzel sanatlar ve eğitim alanlarında yapay zekâ araştırmaları* [Artificial Intelligence research in fine arts and education] (C. Ş. Külük, Ed.). Platanus Publishing.

United Nations Educational, Scientific, and Cultural Organization. (2018). *UNESCO's ICT competency framework for teachers*. https://www.unesco.org/en/digital-competencies-skills/ict-cft

Wang, J., & Redelmeier, D. A. (2025). Forewarning Artificial Intelligence about cognitive biases. *Medical Decision Making, 45*(7), 913–916. https://doi.org/10.1177/0272989X251346788

Whitmer, J., Deng, E. Y., Blankenship, C., Beiting-Parrish, M., Zhang, T., & Bailey, P. (2023). *Results of NAEP reading item automated scoring data challenge (Fall 2021)*. OSF. https://doi.org/10.35542/osf.io/2hevq

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. In AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 195–200). https://doi.org/10.1145/3306618.3314289

Yadav, B., & Anubha. (2024). A bibliometric review of the role of Artificial Intelligencein education. *ShodhKosh: Journal of Visual and Performing Arts, 5*(6), 2593–2603. https://doi.org/10.29121/shodhkosh.v5.i6.2024.3206

Yeşilyurt, Ş. (2025). Measuring c-level Turkish reading skills with chatgpt: Advantages and challenges. *The Journal of International Education Science, 12*(42), 165–179. http://dx.doi.org/10.29228/INESJOURNAL.79971

Yiğiter, M. S., & Boduroğlu, E. (2025). Examining the performance of Artificial Intelligence in scoring students' handwritten responses to open-ended items. *Education and Science, 50*(Supplement 1), 1