

# Identifying the Applications of Artificial Intelligence in the Assessment of Medical Students

Ahmad Keykha<sup>1\*</sup>, Hananeh Mohammadi<sup>2</sup>, Fatemeh Darabi<sup>3</sup>, Sajedeh Sadat Hosseini<sup>4</sup>

<sup>1</sup>Postdoctoral Researcher at Sharif University of Technology, Sharif Policy Research Institute, Tehran, Iran

<sup>2</sup>Master's Degree in Educational Psychology, University of Tehran, Faculty of Psychology and Educational Sciences, University of Tehran, Tehran, Iran

<sup>3</sup>PhD Student, Education Psychology, Islamic Azad University Science and Research University, Tehran, Iran

<sup>4</sup>PhD Student, Educational Research, Werklund School of Education, University of Calgary, Calgary, Canada

**Received:** 2025 January 16

**Revised:** 2025 March 12

**Accepted:** 2025 April 20

**Published online:** 2025 April 20

**\*Corresponding author:**

Postdoctoral Researcher at Sharif University of Technology, Sharif Policy Research Institute, Tehran, Iran.  
Email: ahmad.keykha72@sharif.edu

**Citation:**

Keykha A, Mohammadi H, Darabi F, Hosseini SS. Identifying the Applications of Artificial Intelligence in the Assessment of Medical Students. Strides Dev Med Educ. 2025 April; 22(1):e1512. doi:10.22062/sdme.2025.200833.1512

## Abstract

**Background:** AI has rapidly transformed education, research, and community services in medical universities, surpassing earlier expectations about its integration. A key area of this transformation is student assessment, which plays a vital role in shaping learning outcomes, faculty workload, and public trust in medical education.

**Objectives:** This study aims to explore the applications of AI in the assessment of medical students through a content analysis of relevant scholarly literature.

**Methods:** This qualitative study employed a meta-synthesis method following Walsh and Downe's seven-step framework. Using targeted keywords, a comprehensive search was conducted across major databases, including ScienceDirect, Springer, ERIC, Emerald, Sage Journals, Wiley Online Library, PubMed, and Google Scholar, covering publications from 2015 to 2024. A total of 200 articles were initially retrieved; after applying quality appraisal criteria, this number was narrowed down to 24 studies. To ensure the credibility of the findings, Whittemore et al.'s ten indicators for methodological rigor were applied.

**Results:** Six key themes emerged regarding AI applications in medical student assessment: (a) feedback, (b) online exam, (c) instrument design, (d) assessment process, (e) student learning management, and (f) faculty workload management, along with 19 sub-themes. These findings reflect the diverse and evolving impact of AI in assessment practices.

**Conclusion:** This study underscores the multifaceted and transformative impact of AI in medical student assessment across six key domains. These applications serve as a strategic roadmap for seamlessly integrating AI into the assessment of medical students while effectively adapting to evolving educational paradigms.

**Keywords:** Artificial Intelligence; Medical Education; Student Assessment; Meta-Synthesis; Educational Technology

## Background

The rapid and continuous advancement of Artificial Intelligence (AI) within higher education has made predicting its trajectory with certainty increasingly difficult. However, what remains clear is that AI's integration into education is inevitable. Policymakers and stakeholders must remain vigilant in monitoring emerging challenges and opportunities, adapting their strategies accordingly. Efforts to prohibit or delay its adoption are largely ineffective, as the momentum of technological progress continues to accelerate at an unprecedented pace (1).

To date, AI has significantly influenced both education (1) and research (2). One educational domain that has undergone substantial transformation is online student assessment, which has gained even more significance following the onset of the COVID-19 pandemic (3). Assessment, along with the feedback it generates, plays a central role in shaping and advancing student learning. Far beyond serving as a mere tool for measuring achievement, assessment functions as a core component of the learning cycle—guiding students, informing instruction, and fostering reflective thinking. When thoughtfully designed and paired with timely,

constructive feedback, assessment becomes a powerful mechanism for deepening understanding, enhancing motivation, and promoting meaningful academic growth. Accordingly, ensuring the effectiveness of assessment practices remains a critical concern in contemporary educational systems (4).

AI holds the potential to revolutionize student assessment approaches (5). Yet, alongside these opportunities, significant challenges remain—particularly within the field of medical education. Ethical concerns (6), unrealistic or fabricated outcomes (7), and academic dishonesty (8) pose substantial barriers to effective implementation. Despite these challenges, the focus of this study is on exploring the potential applications and benefits of AI in assessing medical students.

In today's rapidly evolving educational landscape, stakeholders no longer have the luxury of deciding whether to engage with AI. Its integration is not a matter of choice, but of necessity. Particularly in medical education, the role of assessment is critical—not only in evaluating academic achievement but also in safeguarding the competence and credibility of future healthcare professionals.

Poorly designed or executed assessment practices can have severe and irreversible consequences. Incompetent graduates may jeopardize patient safety and diminish the quality of healthcare services. Moreover, flawed assessments can lead to educational inequities by allowing unqualified individuals to obtain medical credentials, thereby eroding public trust in the medical profession. Ineffective evaluation practices may also fail to identify students' weaknesses, limiting opportunities for improvement and demotivating learners. As such, a precise, evidence-based, and technologically informed assessment system is essential to maintain the integrity of medical education.

Recent scholarship offers valuable insights into the integration of AI in educational assessment. For instance, Perkins et al. (9), in *A Framework for Ethical Integration of Generative AI in Educational Assessment*, introduced the AI Assessment Scale (AIAS), a practical tool for determining the appropriate use of generative AI (GenAI) based on learning objectives. This tool promotes transparency and fairness in educational policies while emphasizing a balanced approach to adopting AI, shifting away from a focus solely on negative aspects such as facilitating cheating.

Similarly, Mahamuni et al. (10), in *Enhancing Educational Assessment with Artificial Intelligence:*

*Challenges and Opportunities*, examined AI's role in improving educational assessments, highlighting how AI technologies can enhance accuracy, fairness, and efficiency. While traditional assessment methods often face limitations in scalability, adaptability, and the provision of personalized feedback, AI-driven approaches, particularly those using machine learning and natural language processing, offer promising solutions to these challenges. The study also proposes an innovative framework that incorporates algorithms and mathematical models to improve decision-making processes within educational systems. At the same time, it underscores the importance of a balanced integration of AI, particularly concerning ethical issues, data privacy, and digital inequality.

Stanoyevitch (11), in *Online Assessment in the Age of Artificial Intelligence*, examined how AI technologies and supporting platforms have impacted the integrity of online assessments, particularly in relation to rising incidences of academic dishonesty during virtual exams. Drawing on data from an introductory statistics course, the study reported a significant increase in online exam scores following the introduction of tools such as ChatGPT, despite no changes in exam difficulty or grading practices compared to pre-COVID in-person assessments.

In a complementary line of inquiry, Salinas-Navarro et al. (12), in *Designing Experiential Learning Activities with Generative Artificial Intelligence Tools for Authentic Assessment*, explored the potential of GenAI to enhance experiential learning in higher education. Their findings suggest that GenAI can support reflective thinking, hands-on learning, and the development of authentic assessment tasks. The authors emphasize, however, that to fully harness its potential, GenAI must be used responsibly, with thoughtful attention to pedagogical goals.

Similarly, Jayawardena et al. (13), in *Dental Students' Learning Experience: Artificial Intelligence vs Human Feedback on Assignments*, compared the quality of feedback provided by ChatGPT-4 and a human tutor on dental students' assignments. While students rated both sources similarly across most dimensions, they reported feeling more comfortable with human feedback. However, expert evaluations found the AI-generated feedback to be clearer and more constructive, highlighting the potential of AI tools to complement human tutors in educational settings.

This study adopts a meta-synthesis approach combined with content analysis to examine the multifaceted applications of AI in the assessment of medical students. It aims to offer a comprehensive and systematic overview of AI's role in this domain by synthesizing existing empirical findings. In doing so, the study not only highlights the opportunities presented by AI-driven assessment methods but also critically examines the challenges they pose. Ultimately, this research seeks to deepen understanding of the transformative implications of AI in medical education and to inform stakeholders' decision-making as the integration of AI into educational systems becomes increasingly inevitable.

### Objectives

The primary objective of this research is to identify and analyze the applications of AI in the assessment of medical students by conducting a content analysis of existing scholarly literature in the field.

### Methods

This study employed a meta-synthesis methodology, following the seven-step framework proposed by Walsh and Downe (14). This systematic approach involves the following steps: (a) framing the meta-synthesis, (b) identifying relevant studies, (c) establishing inclusion criteria, (d) appraising selected studies, (e) comparing and contrasting findings, (f) conducting reciprocal translation, and (g) synthesizing the translated concepts. This structured process ensures a rigorous and comprehensive synthesis of research findings.

#### Framing the Meta-Synthesis

The first step involved identifying a clear and relevant research focus to guide the inquiry. In this study, the research objective sought to uncover key themes and concepts in the existing literature on the use of AI in medical student assessment. The goal was to develop a comprehensive conceptual framework that can support further academic inquiry in this field. At this stage, the research evidence was mapped, content analysis was conducted, and the selected studies were prepared for in-depth review. This structured approach provided a foundation for identifying prevailing trends, valuable insights, and notable gaps in the current body of research.

#### Locating Relevant Studies, Determining Inclusion, and Appraising Quality

This stage involved a thorough literature search to identify studies relevant to the topic through electronic

database searches and the collection of all potentially eligible sources. Following the systematic review framework proposed by Walsh and Downe (15), a targeted effort was made to locate all pertinent literature on the use of AI in medical student assessment.

A range of reputable academic databases was searched, including ScienceDirect, Springer, ERIC, Emerald, Sage Journals, Wiley Online Library, PubMed, and Google Scholar. The search focused on studies published between 2015 and 2024, a period selected to capture the most recent advancements in AI and its increasing integration into educational and assessment practices. These databases were chosen for their comprehensive coverage of peer-reviewed research in education, medicine, and technology, thereby ensuring access to high-quality, credible sources.

To ensure the breadth and depth of the literature search, a series of specialized keywords was employed. These included:

(Artificial intelligence AND student evaluation) OR (Artificial intelligence AND medical student evaluation) OR (Artificial intelligence AND student assessment) OR (Artificial intelligence AND medical student assessment) OR (Artificial intelligence AND feedback students) OR (Artificial intelligence AND feedback medical students) OR (ChatGPT AND student assessment) OR (ChatGPT AND medical student assessment) OR (ChatGPT AND student evaluation) OR (ChatGPT AND student appraisal) OR (ChatGPT AND medical student appraisal) OR (ChatGPT AND medical student evaluation) OR (ChatGPT AND feedback students) OR (ChatGPT AND feedback medical students).

These search terms were designed to capture a wide range of studies across different contexts, ensuring comprehensive coverage of AI's role in student evaluation and feedback, particularly in medical education.

To determine article eligibility, the review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The screening process was conducted in three stages: 1. title screening, 2. abstract screening, and 3. full-text screening. These stages ensured the inclusion of studies that were both relevant and methodologically sound. The overall screening process is illustrated in [figure 1](#).

The inclusion criteria specify the conditions under which studies were included in this review, which are as follows: (a) Study design: Empirical studies employing quantitative, qualitative, or mixed-methods approaches,

as well as systematic reviews and meta-analyses, were included.

(b) Topic Relevance: The studies' focus must remain on the use of AI or ChatGPT in student evaluation, assessment, or feedback, with specific emphasis on medical student assessment or general student evaluation in educational settings, (c) Publication Period: Only studies published between 2015 and 2024 were included to capture recent trends and developments in AI applications in education, (d) Language: Only studies published in English were included due to limitations in accessing and interpreting non-English texts.

The exclusion criteria outline the conditions under which studies have been excluded from this review, which are as follows: (a) Lack of Relevance: Studies not addressing AI, ChatGPT, or student assessment and feedback in educational or medical contexts were excluded; (b) Methodological Weakness: Studies lacking methodological rigor or failing to present credible findings were excluded. (c) Duplicates: Duplicate publications or studies reporting identical data or findings were excluded. (d) Non-Peer-Reviewed Sources: Unpublished or non-peer-reviewed literature, including dissertations, conference abstracts, and grey literature, were excluded to maintain the quality and reliability of the review. The article screening process is illustrated in the [figure 1](#).

Based on the PRISMA diagram, an initial pool of 200 sources were identified. Out of these, 98 were excluded during the title screening phase. As additional 41 sources were removed after abstract screening, and 22 were excluded following full-text screening, leaving a total of 39 sources eligible for quality assessment.

To enhance the rigor of the meta-synthesis, as researchers used an additional tool to assess the quality of the selected articles, resulting in 24 studies remaining for the final analysis. Each researcher independently reviewed and completed the evaluation checklists based on predefined quality criteria. After this individual assessment, the results were compared and discussed among all team members. Any discrepancies in scoring or interpretation were carefully examined, and consensus was reached through discussions or, if necessary, by consulting a third expert. This collaborative approach would ensure reliability, reduce bias, and enhance the overall validity of the quality assessment process.

In the final appraisal step, studies were evaluated based on sample quality criteria from Atkins et al., (16) and low-quality studies were excluded to enhance the rigor of the meta-synthesis process. Although the focus of the review was initially on qualitative research, the limited availability of relevant studies warranted the inclusion of quantitative and mixed methods designs that met the inclusion and quality criteria. The inclusion criteria prioritized studies with clearly articulated research questions, a well-justified methodological approach, and adequate contextual descriptions, including researcher positioning and sampling strategies. Data collection methods were required to be thoroughly documented and aligned with the research objective. Similarly, data analysis procedures needed to be explicitly outlined, logically connected to the research questions, and supported by sufficient evidence. Based on these criteria, 24 articles were selected for the final meta-synthesis. To ensure alignment with the study's objective, a purposive sampling strategy was employed during the screening process. In [table 1](#), the bibliographic details of the selected studies are provided, including the reference number, names of the researchers, year, and title.

### Comparing and Contrasting Findings

The purpose of this stage was to identify and categorize the various applications of AI in the assessment of medical students. To accomplish this, the included studies were examined through a detailed review of their titles, key concepts, and content. This comparative analysis allowed the researchers to systematically explore the range and nature of AI applications. [Table 2](#) presents a comparative overview of the selected studies, organized by specific application areas of AI in medical student assessment.

### Reciprocal Translation

In the next phase, thematic coding was employed to identify key concepts and recurring patterns across the selected studies. This process involved the development of main themes, subthemes, and codes through a reciprocal translation approach. As defined by Noblit and Hare (41), reciprocal translation refers to the inductive and interpretive emergence of concepts and metaphors through iterative comparison and classification of themes across studies.

Following this approach, relevant themes were synthesized by compiling and interrelating them across the data set. This method was selected for its appropriateness



and practicality in capturing the nuanced similarities and differences among studies. The final thematic categories included: (a) feedback applications, (b) online exam applications, (c) instrument design applications, (d) applications in the assessment process, (e) applications in student learning management, and (f) applications in faculty workload management.

Key excerpts from the selected articles were compiled into a Word document and thoroughly re-read multiple times by the researchers to foster a deep engagement with the data. The sentence was selected as the unit of analysis, and key concepts were extracted based on their semantic density. These concepts were then grouped into subthemes according to their conceptual similarities. Subthemes were further refined and merged into broader, more abstract categories to form the main themes.

### Synthesis of Translations

The final step of the meta-synthesis involved synthesizing the translated and reconsolidated concepts and themes into an overarching framework. This stage aimed to construct a qualitative conceptual model that captured the core applications of AI in the assessment of medical students. The synthesis process involved proposing a general interpretation of the phenomenon that integrated the themes, codes, and categories derived from the prior phases of analysis.

To ensure the credibility and rigor of the research, several quality assurance measures were implemented in line with recognized standards for qualitative research validity. Drawing on Whittemore et al. (42), who synthesized 13 scholarly reports on qualitative validity, ten indicators of research quality were applied across four main criteria and six sub-criteria:

- **Credibility:** The researchers achieved coding consensus through multiple iterative rounds of peer review, ensuring that data interpretation was accurate and supported by all members of the research team. Triangulation of perspectives enhanced the trustworthiness of the findings.
  - **Authenticity:** Efforts were made to accurately reflect the perspectives of the original authors in the selected studies. Direct attention was given to preserving diverse viewpoints and ensuring that interpretations remained true to the intent and context of each source.
  - **Criticality:** A reflective and analytical stance was maintained throughout the study. Researchers actively questioned assumptions considered
- underlying power dynamics, and critically examined the roles of various actors (e.g., students, faculty, assessment systems) to offer a deeper understanding of the phenomena.
  - **Integrity:** The research process was marked by transparency and consistency. An audit trail was maintained, and team members collaboratively engaged in all stages—from article selection to data synthesis—ensuring methodological coherence and ethical alignment.
  - **Ethical Rigor:** Ethical considerations were embedded throughout the research process, from article selection to reporting. The team prioritized respect for original sources, avoided misrepresentation, and remained sensitive to the implications of interpreting others' work.
  - **Methodological Congruence:** The study demonstrated strong alignment between the research objective, inclusion criteria, and analytical strategies. This coherence ensured that the findings were both methodologically sound and directly relevant to the study's purpose.
  - **Creativity:** The development of a novel categorization of AI applications in student assessment reflected the study's innovative contribution. This framework addressed gaps in the literature and provided a fresh lens for future research and practice.
  - **Analytical Depth:** The synthesis process yielded nuanced and detailed insights into the themes and patterns emerging from the selected studies. Through deep engagement with the data, the study moved beyond surface-level coding to offer meaningful interpretations.
  - **Contextualization:** The findings were interpreted with careful attention to the specific social, educational, and institutional contexts of the original studies. This approach ensured relevance across diverse settings and strengthened the applicability of the results.

These rigorous procedures, grounded in Whittemore et al.'s indicators of research quality, contributed to establishing the validity and reliability of the research process, ensuring that it was both methodologically robust and ethically sound.

To further verify the reliability of the findings, a test-retest strategy was employed for assessing coding consistency. In this approach, all extracted key concepts (codes) were re-categorized by the same researcher after

a 20-day interval. This method evaluates the stability of the categorization process by comparing the consistency of code assignments over time. Concepts that were labeled identically across both coding sessions were counted as “agreements,” while discrepancies were labeled as “disagreements.” The frequency of agreement and disagreement was then used to calculate the overall reliability score. In this study, 301 concepts were extracted in total, and 135 showed consistent labeling between the two coding sessions. The resulting test-retest reliability was calculated at 89%, which exceeds the commonly accepted threshold of 60%, thereby confirming the stability and internal consistency of the coding process.

$$\text{Test-retest Reliability Percentage} = \frac{135 * 2}{301} * 100$$

## Results

To address the main research objective, the extracted key concepts were integrated and categorized based on their similarities and differences. This process resulted in the identification of 6 main themes and 19 sub-themes related to the applications of AI in medical student assessment. These themes are presented in the [table 3](#):

As shown in the [table 3](#), the applications of AI in medical student assessment are diverse and span across a range of domains. The findings highlight the transformative potential of AI in reshaping how assessment is conceptualized, implemented, and managed in medical education. An analysis of the frequency of coded concepts under each main theme revealed the following statistics:

- Mean frequency: 24.17 coded concepts per sub-theme
- Minimum frequency: 9 concepts (in the “Instrument Design Applications” theme)

Maximum frequency: 43 concepts (in the “Applications in the Assessment Process” theme)

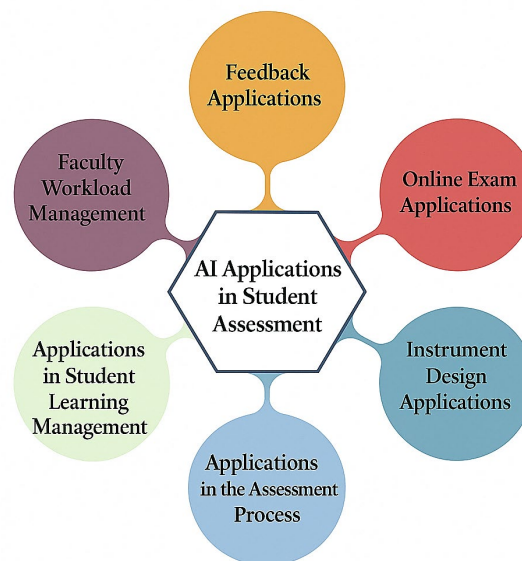
[Figure 2](#) presents a visualization that illustrates the relative frequency of each main theme, providing a visual representation of the emphasis placed on different AI applications.

A schematic model of the research findings is presented below to summarize and conceptualize the thematic structure derived from the synthesis.

## Discussion

This section analyzes the study’s findings through the lens of the six main themes identified, each

representing a key area where AI is transforming student assessment. These six main themes are depicted in [Figure 3](#).



**Figure 3.** Schematic model of research findings

### Feedback Applications

Feedback is broadly conceptualized as information that serves two essential functions: (a) identifying students’ current performance levels and (b) guiding them toward their learning goals by outlining how to progress from point A to point B ([43](#)). Effective feedback is pivotal to student learning as it helps learners recognize their achievements and address performance gaps, ultimately fostering academic motivation and a sense of self-efficacy ([44](#)).

Hattie and Timperley ([45](#)) identified four distinct types of feedback: (a) task, (b) process, (c) self-regulation, and (d) self-feedback. Each serves a different purpose and exerts varying levels of influence on student learning. Consequently, implementing effective feedback strategies requires an understanding of these differences. Providing feedback on students’ performance remains a fundamental component of teaching and learning. Instructors invest considerable time and effort into offering meaningful feedback, which plays a vital role in shaping students’ academic progress by highlighting both their strengths and areas for improvement. When delivered appropriately, feedback enhances students’ sense of responsibility for their learning and motivates them to strive toward their educational goals.

With the rise of AI, feedback practices are being redefined. AI-powered tools facilitate the delivery of

intelligent, automated, and timely feedback that is closely aligned with students' performance. Moreover, these systems can generate comparative and differentiated feedback tailored to various student groups, thereby supporting more personalized learning experiences. The findings of this review reflect these developments and align with existing literature (17, 19, 20, 23, 24, 27-29, 33, 34, 36, 39), as well as other studies (46-48) that examine the role of AI in enhancing feedback within the teaching and learning process.

### Online Exam Applications

Online exams have emerged as a vital tool for assessing student performance, especially in the wake of the COVID-19 pandemic. Compared to traditional assessments, they offer enhanced flexibility and support a range of assessment formats, making them particularly suitable for both diagnostic and formative evaluations (3). Online exams typically incorporate diverse question types, including multiple-choice, true/false, matching, sequencing, fill-in-the-blank, and essay responses. These exams are created using specialized software designed to evaluate students' performance across various learning dimensions. It is important to distinguish online exams from computer-based assessments (CBAs). While both utilize digital platforms, CBAs are typically conducted using specific software installed on standalone devices, independent of internet connectivity (49). In contrast, online exams rely on networked systems and often integrate learning management platforms.

As digitalization continues to reshape educational practices, online examinations are expected to become even more prominent—potentially replacing traditional paper-based tests altogether. The integration of AI in online exam systems further enhances their functionality. AI enables the design of dynamic, personalized assessments tailored to individual students' needs, interests, and ability levels. This level of customization provides more nuanced and realistic data for evaluating students' competencies and identifying both strengths and areas for growth. Moreover, AI-powered grading systems automate the evaluation process, reducing the likelihood of human error and enhancing fairness, accuracy, and efficiency—particularly valuable when assessing large cohorts of students. The findings related to this theme are consistent with previous research (18, 20, 21, 23, 24, 29) and further supported by other studies examining AI applications in assessment contexts (50-52).

### Instrument Design Applications

Assessment instruments are foundational to effective teaching and learning, serving as essential tools for evaluating student performance and guiding instructional decisions. Well-designed instruments enable faculty to accurately measure learning outcomes, identify students' strengths and weaknesses, and ensure alignment between assessment tasks and educational objectives. Beyond individual evaluation, they offer valuable data to inform curriculum development and support the continuous refinement of teaching strategies and course content. In the context of medical education, recent research has highlighted the limitations of many current assessment instruments, emphasizing the urgent need for redesign and modernization to better reflect evolving competencies, clinical demands, and educational goals (53). The findings of the present review echo these concerns in alignment with previous literature (18, 28, 32, 33), reinforcing the call for continuous innovation in assessment design. In this regard, AI plays an increasingly important role in offering tools to support the development of more dynamic, data-driven, and personalized assessment instruments.

### Applications in the Assessment Process

Assessment should be strategically employed to enhance teaching and learning by sending clear signals to students about what knowledge and skills they should prioritize. When implemented effectively, assessment acts as a bridge between instruction and learning, guiding both educators and learners toward meaningful educational outcomes. To accurately evaluate the learning process, it is essential to incorporate both subjective perceptions—those of students and educators—and direct indicators such as exam scores and analytic rubrics (54). This dual perspective ensures that assessment not only reflects academic achievement but also captures the lived experience of learning. Assessment plays a multifaceted role in education. Beyond simply measuring performance, it monitors students' cognitive, affective, and psychomotor development, thereby supporting individual growth and institutional accountability. It also serves as a mechanism to support academic mobility, promote learner autonomy, and improve the overall quality of education by offering continuous feedback on student progress. Through regular assessment, both students and faculty can track developmental trajectories and

make informed adjustments to teaching and learning strategies.

Among the various forms of assessment, authentic assessment stands out as a particularly valuable approach. It evaluates learning through sustained, real-world tasks that mirror professional and academic contexts. Unlike one-time tests, authentic assessment emphasizes ongoing performance and development over time, providing richer, more nuanced insights into students' learning processes and outcomes (55). As such, the evaluation of teaching and learning requires a systematic approach—one that integrates data collection, analysis, and interpretation into a coherent and reflective cycle of improvement. In recent years, advances in AI have further enhanced assessment practices. AI technologies offer powerful tools for tracking student progress longitudinally, predicting future academic outcomes, and improving the precision and reliability of assessments. By minimizing human error and reducing potential bias, AI fosters a more equitable evaluation process. Furthermore, AI systems can analyze emotional and behavioral data, offering educators deeper insights into student engagement and well-being, which are critical yet often overlooked aspects of academic success.

The findings of the present review echo a growing body of literature (17-20, 22, 32, 35, 39, 40) and are consistent with prior research (56, 59) on the transformative role of AI in educational assessment. Several empirical studies have demonstrated how AI can support early identification of learning difficulties and enable timely interventions. For example, Pande (60) used AI algorithms to analyze various determinants of academic success, including age, parental education, occupation, and student health. The Support Vector Machine (SVM) model achieved a precision rate of 84.37%, highlighting AI's capacity to synthesize complex data for predictive purposes. Similarly, Jiao et al. (61) identified class participation, knowledge acquisition, and overall performance as key predictors of academic success, which were effectively modeled using AI techniques. Haron et al. (62) employed the RepTree algorithm to classify students into high- and low-risk categories based on predicted academic outcomes, enabling targeted support. In another study, Wazir et al. (63) explored how time management, study habits, and collaborative learning contributed to academic performance. Their AI-driven models not only detected at-risk students early but also facilitated

interventions that reduced dropout rates by 25%. Together, these studies underscore the growing relevance of AI-enhanced assessment systems in fostering more adaptive, data-informed, and student-centered learning environments.

### Student Learning Management Applications

AI in student assessment helps learners choose more effective and personalized educational pathways by more accurately evaluating their abilities, learning styles, and individual characteristics. AI-driven assessment systems offer a personalized and adaptive approach that promotes learner autonomy, fosters self-directed learning, and enriches the overall educational experience (64). One particularly impactful application of AI in medical education is its role in fostering the development of self-regulated learning (SRL) skills. SRL is widely recognized as a key determinant of academic success, as it enables students to set specific goals, monitor their progress, and implement strategies to achieve desired outcomes (65).

AI-based assessment systems contribute to this process by delivering personalized and timely feedback derived from the analysis of students' behavioral and performance data. This feedback supports learners in setting meaningful goals, tracking their development, and selecting effective strategies to enhance their academic performance.

In addition, AI-driven assessments assist students in gaining a deeper understanding of their learning processes, strengths, and areas for improvement. By recommending tailored educational resources and strategies, these systems help learners identify personalized development paths and support the achievement of individual learning objectives. Furthermore, through continuous monitoring and analysis of student progress, AI systems can guide learners toward ongoing improvement and sustained academic growth. The findings related to this theme are consistent with a growing body of literature emphasizing the role of AI in supporting student learning management (17-20, 23, 28, 29, 31-33, 37, 38). They also align with prior studies (66-69) that underscore the potential of AI applications in managing and enhancing student learning experiences.

### Faculty Workload Management Applications

AI-based assessment systems also offer significant advantages for faculty members, notably in reducing workload and streamlining student evaluation tasks. For full-time faculty, maintaining a balanced workload is



essential—not only to foster innovation within and beyond the classroom but also to prevent burnout and allow time for creative, research-driven contributions (70). By automating the analysis of tests, assignments, and other forms of assessment, AI tools can handle tasks that traditionally demand considerable time and energy. Moreover, these systems can generate immediate, personalized feedback for students, minimizing the need for repetitive, manual grading and enhancing the overall efficiency of the assessment process.

This technological support enables educators to devote more attention to instructional quality and student engagement. Another important function of AI in faculty development is its ability to inform and refine teaching strategies. By analyzing patterns in student performance, learning behaviors, and areas of difficulty or strength, AI can provide actionable insights that help educators adjust their methods to better meet learners' needs. In doing so, teaching becomes more responsive, dynamic, and personalized—contributing to more effective learning outcomes. The findings related to this theme are consistent with those of other studies (17-20, 22-24, 26, 28-30, 33-40) and supported by additional research highlighting the potential of AI to assist faculty in assessment and instruction (71-74). A comparative and analytical summary of traditional versus AI-based assessment methods for students is presented in table 4, reflecting the implications drawn from these findings.

**Limitations:** its strengths, the meta-synthesis approach employed in this study presents several limitations. First, because it relies exclusively on existing literature, there is a risk that some relevant or recently published studies may have been inadvertently excluded, potentially limiting the breadth and comprehensiveness of the findings? Second, the methodological heterogeneity of the included studies posed challenges for synthesis. Variations in research design, data collection techniques, and analytical approaches may have led to divergent interpretations of similar phenomena, complicating the effort to draw cohesive and universally applicable conclusions.

Third, the generalizability of the findings is constrained by the specific educational and cultural contexts in which the included studies were conducted. As a result, the insights derived may not be directly transferable to other settings with different institutional structures or student populations. Finally, the potential for researcher bias must be acknowledged. Although clear

inclusion criteria were applied throughout the review process, subjective judgments during study selection and thematic synthesis may have unintentionally favored literature aligned with the researchers' theoretical orientations, potentially influencing the neutrality and objectivity of the analysis.

**Future Research Directions:** To advance the field, future research is encouraged to adopt diverse methodological approaches that address both the technical and ethical dimensions of AI in medical student assessment. First, experimental and quasi-experimental designs are essential for conducting comparative analyses between traditional and AI-based assessment methods. Such studies would enable researchers to rigorously evaluate the accuracy, efficiency, and educational impact of AI tools relative to conventional practices. Second, qualitative research methods are needed to investigate the ethical implications of using AI in assessment, including concerns about transparency, fairness, accountability, and the potential depersonalization of learner evaluation. These methods can provide rich, contextual insights into the experiences and perceptions of students, educators, and other stakeholders. Finally, a mixed-methods approach is recommended for the development and evaluation of innovative assessment tools that integrate AI into the evaluation of clinical competencies. Combining quantitative performance data with qualitative feedback would ensure that these tools are both pedagogically sound and ethically responsible, ultimately contributing to more effective and equitable medical education practices.

## Conclusion

This study highlighted the transformative impact of AI on medical student assessment, emphasizing its role in shaping education, research, and social services within medical universities. By identifying six key themes through content analysis, the research provided a comprehensive understanding of AI's potential in this domain. These themes included feedback applications, online exam applications, instrument design applications, applications in the assessment process, applications in student learning management, and applications in faculty workload management.

The findings underscored the substantial advantages AI offers in enhancing assessment practices for both students and faculty members. For students, AI promotes the development of self-regulated learning by enabling personalized goal setting, performance

tracking, and strategy adoption. It also supports deeper insight into students' strengths, weaknesses, and learning processes by offering timely and individualized feedback and recommendations. For faculty, AI facilitates assessment tasks, reduces repetitive grading, and provides real-time insights that can inform instructional strategies. It enables educators to tailor their teaching methods and styles based on students' learning patterns and needs, thereby improving the overall quality of education.

However, despite its advantages and various applications, AI can lead to negative consequences and numerous ethical challenges if used irresponsibly. One of the key ethical concerns is algorithmic bias, as AI systems may reflect racial prejudices and gender biases embedded in their training data (75). Another serious concern involves privacy issues, particularly regarding access and security of student data (76). Ensuring the validity and reliability of AI in education demands a high level of accuracy and consistency. Although AI systems are designed to provide objective results, they have not yet reached the adaptability and nuanced understanding of human intelligence. Moreover, the quality of data used in AI systems must be closely monitored and regulated to ensure accurate outcomes (77). Another important challenge lies in data governance, which encompasses the collection, organization, control, usage, storage, archiving, and destruction of data. The implementation of data governance must be guided by structured frameworks, supported by institutional policies and procedures, and communicated effectively through leadership and management (78).

The cost of implementing AI systems also represents a major barrier across various sectors, including education, healthcare, and industry. These costs include the initial investment in developing hardware and software infrastructure, expenses related to processing and storing large volumes of data, and the cost of hiring and training specialized personnel. Additionally, ongoing system maintenance and updates are necessary to ensure accuracy and efficiency, creating further financial pressure on organizations. As a result, many institutions may be unable to adopt AI due to financial constraints, potentially increasing inequality in access to the benefits of this technology. Therefore, developing more cost-effective models, utilizing cloud services, and introducing supportive policies can help reduce these costs and improve institutional access to AI. The

findings of this review emphasize the need for a strategic approach to AI integration in education, one that maximizes its benefits while addressing ethical, technical, and financial challenges. Effective adaptation to AI-driven changes will strengthen medical education and enhance student assessment, ultimately contributing to improved healthcare outcomes.

**Acknowledgements:** We would like to express our sincere gratitude to all those who contributed to the successful completion of this research. Special thanks to the authors and scholars whose work formed the foundation of this meta-synthesis. We are also grateful to our colleagues for their thoughtful feedback and constructive suggestions, which helped refine the direction and clarity of this study.

**Conflict of interests:** There is no conflict of interest.

**Ethical approval:** Not applicable.

**Funding/Support:** This study was conducted independently and did not receive any financial support from public, commercial, or not-for-profit funding agencies at any stage.

## References

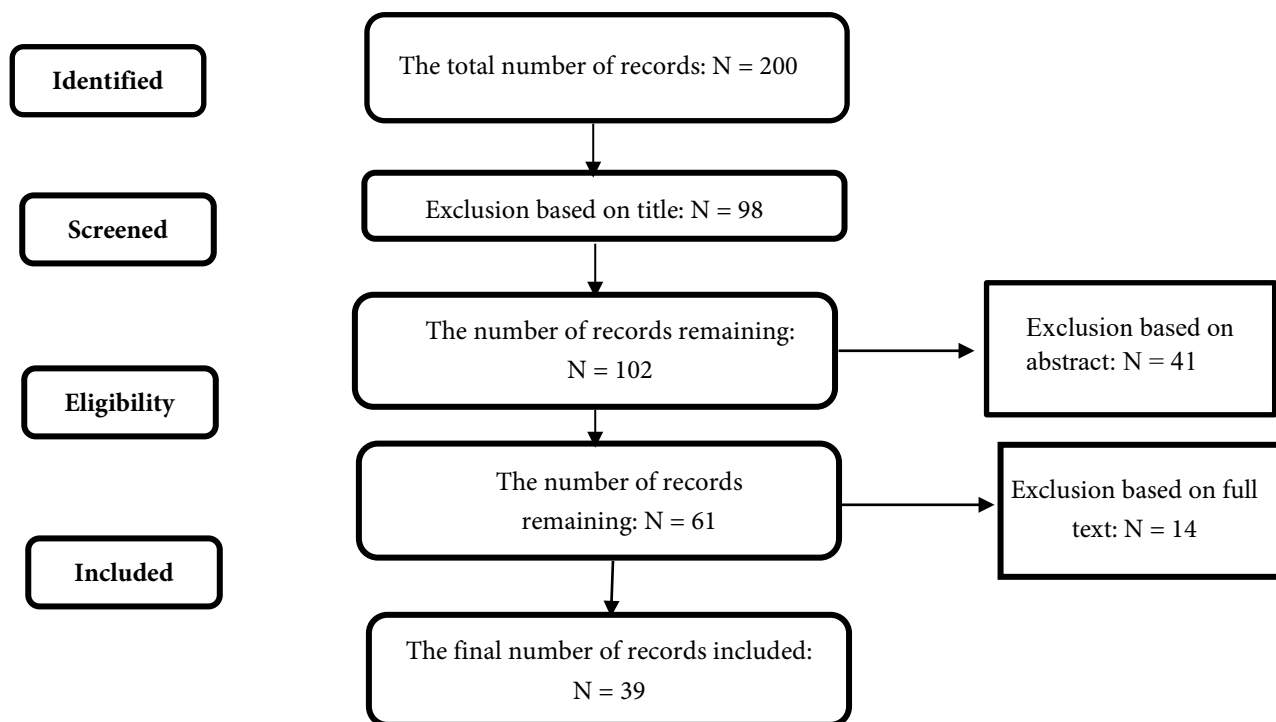
1. Jafari F, Keykha A. Identifying the opportunities and challenges of artificial intelligence in higher education: a qualitative study. *Journal of Applied Research in Higher Education*. 2024 Jul 9;16(4):1228-45. doi:[10.1108/JARHE-09-2023-0426](https://doi.org/10.1108/JARHE-09-2023-0426).
2. Keykha A, Behraves S, Ghaemi F. ChatGPT and medical research: a meta-synthesis of opportunities and challenges. *J Adv Med Educ Prof*. 2024 Jul 1;12(3):135-147. doi:[10.30476/JAMP.2024.101068.1910](https://doi.org/10.30476/JAMP.2024.101068.1910). [PMID: [39175584](https://pubmed.ncbi.nlm.nih.gov/39175584/)] [PMCID: [PMC11336189](https://pubmed.ncbi.nlm.nih.gov/PMC11336189/)]
3. Keykha A, Imanipour M, Shahrokhi J, Amiri M. The advantages and challenges of electronic exams: qualitative research based on Shannon Entropy Technique. *J Adv Med Educ Prof*. 2025 Jan 1;13(1):1-11. doi: [10.30476/jamp.2024.102951.1987](https://doi.org/10.30476/jamp.2024.102951.1987). [PMID: [39906078](https://pubmed.ncbi.nlm.nih.gov/39906078/)] [PMCID: [PMC11788768](https://pubmed.ncbi.nlm.nih.gov/PMC11788768/)]
4. Rutherford S, Pritchard C, Francis N. Assessment IS learning: developing a student-centered approach for assessment in Higher Education. *FEBS Open Bio*. 2025 Jan;15(1):21-34. doi: [10.1002/2211-5463.13921](https://doi.org/10.1002/2211-5463.13921). [PMID: [39487560](https://pubmed.ncbi.nlm.nih.gov/39487560/)] [PMCID: [PMC11705397](https://pubmed.ncbi.nlm.nih.gov/PMC11705397/)]
5. Serhani MA, Bouktif S, Al-Qirim N, El Kassabi HT. Automated system for evaluating higher education programs. *Education and Information Technologies*. 2019; 24:3107-28. doi:[10.1007/s10639-019-09910-6](https://doi.org/10.1007/s10639-019-09910-6).
6. Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*. 2023 Jul 3;25(3):277-304. doi:[10.1080/15228053.2023.2233814](https://doi.org/10.1080/15228053.2023.2233814).
7. Wang N, Wang X, Su YS. Critical analysis of the technological affordances, challenges and future directions of Generative AI in

- education: a systematic review. *Asia Pacific Journal of Education*. 2024 Jan 2;44(1):139-55. doi:[10.1080/02188791.2024.2305156](https://doi.org/10.1080/02188791.2024.2305156).
8. Dawson P. *Defending assessment security in a digital world: preventing e-cheating and supporting academic integrity in higher education*. Abingdon, UK: Routledge; 2021. doi:[10.4324/9780429324178](https://doi.org/10.4324/9780429324178).
  9. Perkins M, Furze L, Roe J, MacVaugh J. The Artificial Intelligence Assessment Scale (AIAS): a framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*. 2024 May;21(6):49-66. doi:[10.53761/q3azde36](https://doi.org/10.53761/q3azde36).
  10. Mahamuni AJ, Tonpe SS. Enhancing educational assessment with artificial intelligence: challenges and opportunities. *Proceedings of the International Conference on Knowledge Engineering and Communication Systems (ICKECS)*; 2024 Aug 7; Chikkaballapur, India. 2024:1-5. doi: [10.1109/ICKECS61492.2024.10616620](https://doi.org/10.1109/ICKECS61492.2024.10616620).
  11. Stanoyevitch A. Online assessment in the age of artificial intelligence. *Discover Education*. 2024 Aug 19;3(1):126. doi:[10.1007/s44217-024-00212-9](https://doi.org/10.1007/s44217-024-00212-9).
  12. Salinas-Navarro DE, Vilalta-Perdomo E, Michel-Villarreal R, Montesinos L. Designing experiential learning activities with generative artificial intelligence tools for authentic assessment. *Interactive Technology and Smart Education*. 2024 Oct 30;21(4):708-34. doi:[10.1108/ITSE-12-2023-0236](https://doi.org/10.1108/ITSE-12-2023-0236).
  13. Jayawardena CK, Gunathilake Y, Ihalagedara D. Dental students' learning experience: artificial intelligence vs human feedback on assignments. *Int Dent J*. 2025 Feb;75(1):100-108. doi: [10.1016/j.identj.2024.12.022](https://doi.org/10.1016/j.identj.2024.12.022). [PMID: [39799065](https://pubmed.ncbi.nlm.nih.gov/39799065/)] [PMCID: [PMC11806320](https://pubmed.ncbi.nlm.nih.gov/PMC11806320/)]
  14. Walsh D, Downe S. Meta-synthesis method for qualitative research: a literature review. *J Adv Nurs*. 2005 Apr;50(2): 204-11. doi: [10.1111/j.1365-2648.2005.03380.x](https://doi.org/10.1111/j.1365-2648.2005.03380.x). [PMID: [15788085](https://pubmed.ncbi.nlm.nih.gov/15788085/)]
  15. Walsh D, Downe S. Appraising the quality of qualitative research. *Midwifery*. 2006 Jun; 22(2): 108-19. doi: [10.1016/j.midw.2005.05.004](https://doi.org/10.1016/j.midw.2005.05.004). [PMID: [16243416](https://pubmed.ncbi.nlm.nih.gov/16243416/)]
  16. Atkins S, Lewin S, Smith H, Engel M, Fretheim A, Volmink J. Conducting a meta-ethnography of qualitative literature: lessons learnt. *BMC Med Res Methodol*. 2008 Apr 16;8:21. doi: [10.1186/1471-2288-8-21](https://doi.org/10.1186/1471-2288-8-21). [PMID: [18416812](https://pubmed.ncbi.nlm.nih.gov/18416812/)] [PMCID: [PMC2374791](https://pubmed.ncbi.nlm.nih.gov/PMC2374791/)]
  17. Lee AV. Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies in Educational Evaluation*. 2023;77:101250. doi:[10.1016/j.stueduc.2023.101250](https://doi.org/10.1016/j.stueduc.2023.101250).
  18. Swiecki Z, Khosravi H, Chen G, Martinez-Maldonado R, Lodge JM, Milligan S, et al. Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*. 2022;3:100075. doi: [10.1016/j.caeai.2022.100075](https://doi.org/10.1016/j.caeai.2022.100075).
  19. Maier U, Klotz C. Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence*. 2022;3:100080. doi:[10.1016/j.caeai.2022.100080](https://doi.org/10.1016/j.caeai.2022.100080).
  20. Yildirim-Erbasli SN, Bulut O. Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment. *Computers and Education: Artificial Intelligence*. 2023;4:100135. doi:[10.1016/j.caeai.2023.100135](https://doi.org/10.1016/j.caeai.2023.100135).
  21. Smerdon D. AI in essay-based assessment: student adoption, usage, and performance. *Computers and Education: Artificial Intelligence*. 2024;7:100288. doi:[10.1016/j.caeai.2024.100288](https://doi.org/10.1016/j.caeai.2024.100288).
  22. Martínez-Comesana M, Rigueira-Díaz X, Larrañaga-Janeiro A, Martínez-Torres J, Ocarranza-Prado I, Kreibel D. Impact of artificial intelligence on assessment methods in primary and secondary education: systematic literature review. *Revista de Psicodidáctica* (English ed.). 2023; 28(2): 93-103. doi:[10.1016/j.psicoe.2023.06.002](https://doi.org/10.1016/j.psicoe.2023.06.002).
  23. Memarian B, Doleck T. A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*. 2024 Jan 1; 2(1): 100040. doi: [10.1016/j.chbah.2023.100040](https://doi.org/10.1016/j.chbah.2023.100040).
  24. Tonbuloglu B. An evaluation of the use of artificial intelligence applications in online education. *Journal of Educational Technology and Online Learning*. 2023; 6(4): 866-84. doi:[10.31681/jetol.1335906](https://doi.org/10.31681/jetol.1335906).
  25. González-Calatayud V, Prendes-Espinosa P, Roig-Vila R. Artificial intelligence for student assessment: a systematic review. *Applied sciences*. 2021 Jun 12; 11(12): 5467. doi:[10.3390/app11125467](https://doi.org/10.3390/app11125467).
  26. Casey D. ChatGPT in public policy teaching and assessment: an examination of opportunities and challenges. *Aust J Publ Admin*. 2024;1-15. doi:[10.1111/1467-8500.12647](https://doi.org/10.1111/1467-8500.12647).
  27. Lu Q, Yao Y, Xiao L, Yuan M, Wang J, Zhu X. Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment & Evaluation in Higher Education*. 2024; 49(5): 616-33. doi:[10.1080/02602938.2024.2301722](https://doi.org/10.1080/02602938.2024.2301722).
  28. Williams P. AI, Analytics and a new assessment model for universities. *Education Sciences*. 2023 Oct 17;13(10):1040. doi:[10.3390/educsci13101040](https://doi.org/10.3390/educsci13101040).
  29. Salinas-Navarro DE, Vilalta-Perdomo E, Michel-Villarreal R, Montesinos L. Using generative artificial intelligence tools to explain and enhance experiential learning for authentic assessment. *Education Sciences*. 2024 Jan 12;14(1):83. doi:[10.3390/educsci14010083](https://doi.org/10.3390/educsci14010083).
  30. Lynam S, Cachia M. Students' perceptions of the role of assessments at higher education. *Assessment & Evaluation in Higher Education*. 2018 Feb 17; 43(2): 223-34. doi:[10.1080/02602938.2017.1329928](https://doi.org/10.1080/02602938.2017.1329928).
  31. Fahmy Y. *Student perception on AI-driven assessment: motivation, engagement and feedback capabilities* (Dissertation). Enschede, Netherlands: University of Twente; 2023.
  32. Koh E, Doroudi S. Learning, teaching, and assessment with generative artificial intelligence: towards a plateau of productivity. *Learning: Research and Practice*. 2023 Jul 3;9(2):109-16. doi:[10.1080/23735082.2023.2264086](https://doi.org/10.1080/23735082.2023.2264086).
  33. Hooda M, Rana C, Dahiya O, Rizwan A, Hossain MS. Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*. 2022;2022(1):5215722. doi:[10.1155/2022/5215722](https://doi.org/10.1155/2022/5215722).
  34. Zirar A. Exploring the impact of language models, such as ChatGPT, on student learning and assessment. *Review of Education*. 2023 Dec;11(3):e3433. doi:[10.1002/rev3.3433](https://doi.org/10.1002/rev3.3433).
  35. Mao J, Chen B, Liu JC. Generative artificial intelligence in education and its implications for assessment. *Tech Trends*. 2024 Jan;68(1):58-66. doi:[10.1007/s11528-023-00911-4](https://doi.org/10.1007/s11528-023-00911-4).
  36. Kooli C, Yusuf N. Transforming educational assessment: insights into the use of ChatGPT and large language models in grading. *International Journal of Human-Computer Interaction*. 2024; 41(5):1-2. doi:[10.1080/10447318.2024.2338330](https://doi.org/10.1080/10447318.2024.2338330).
  37. Chaudhry IS, Sarwary SA, El Refae GA, Chabchoub H. Time to revisit existing student's performance evaluation approach in higher education sector in a new era of ChatGPT—a case study. *Cogent Education*. 2023 Dec 31; 10(1): 2210461. doi:[10.1080/2331186X.2023.2210461](https://doi.org/10.1080/2331186X.2023.2210461).
  38. Ouyang F, Dinh TA, Xu W. A systematic review of AI-driven

- educational assessment in STEM education. *Journal for STEM Education Research*. 2023 Dec; 6(3): 408-26. doi:[10.1007/s41979-023-00112-x](https://doi.org/10.1007/s41979-023-00112-x).
39. Xia Q, Weng X, Ouyang F, Lin TJ, Chiu TK. A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*. 2024 May 24;21(1): 40. doi:[10.1186/s41239-024-00468-z](https://doi.org/10.1186/s41239-024-00468-z).
  40. Fuller KA, Morbitzer KA, Zeeman JM, Persky A M, Savage A C, McLaughlin JE. Exploring the use of ChatGPT to analyze student course evaluation comments. *BMC Med Educ*. 2024 Apr 19;24(1):423. doi: [10.1186/s12909-024-05316-2](https://doi.org/10.1186/s12909-024-05316-2). [PMID: [38641798](https://pubmed.ncbi.nlm.nih.gov/38641798/)] [PMCID: [PMC11031883](https://pubmed.ncbi.nlm.nih.gov/PMC11031883/)]
  41. Noblit GW, Hare RD. *Meta-ethnography: synthesizing qualitative studies*. Newbury Park, CA: Sage Publications; 1988. doi:[10.4135/9781412985000](https://doi.org/10.4135/9781412985000).
  42. Whittemore R, Chase SK, Mandle CL. Validity in qualitative research. *Qual Health Res*. 2001 Jul;11(4):522-37. doi: [10.1177/104973201129119299](https://doi.org/10.1177/104973201129119299). [PMID: [11521609](https://pubmed.ncbi.nlm.nih.gov/11521609/)]
  43. Van der Kleij FM, Lipnevich AA. Student perceptions of assessment feedback: a critical scoping review and call for research. *Educational assessment, evaluation and accountability*. 2021;33:345-73. doi:[10.1007/s11092-020-09331-x](https://doi.org/10.1007/s11092-020-09331-x).
  44. Hill J, West H. Improving the student learning experience through dialogic feed-forward assessment. *Assessment & Evaluation in Higher Education*. 2020; 45(1): 82-97. doi:[10.1080/02602938.2019.1608908](https://doi.org/10.1080/02602938.2019.1608908).
  45. Hattie J, Timperley H. The power of feedback. *Review of Educational Research*. 2007; 77(1): 81-112. doi:[10.3102/003465430298487](https://doi.org/10.3102/003465430298487).
  46. Xu W, Meng J, Raja SK, Priya MP, Kiruthiga Devi M. Artificial intelligence in constructing personalized and accurate feedback systems for students. *International Journal of Modeling, Simulation, and Scientific Computing*. 2023 Feb 22;14(1):2341001. doi:[10.1142/S1793962323410015](https://doi.org/10.1142/S1793962323410015).
  47. Ma X. English teaching in artificial intelligence-based higher vocational education using machine learning techniques for students' feedback analysis and course selection recommendation. *Journal of Universal Computer Science*. 2022 Sep 1;28(9):898. doi:[10.3897/jucs.94160](https://doi.org/10.3897/jucs.94160).
  48. Meyer J, Jansen T, Schiller R, Liebenow LW, Steinbach M, Horbach A, et al. Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*. 2024;6:100199. doi:[10.1016/j.caeai.2023.100199](https://doi.org/10.1016/j.caeai.2023.100199).
  49. Ahmed FR, Ahmed TE, Saeed RA, Alhumyani H, Abdel-Khalek S, Abu-Zinadah H. Analysis and challenges of robust E-exams performance under COVID-19. *Results Phys*. 2021 Apr;23:103987. doi: [10.1016/j.rinp.2021.103987](https://doi.org/10.1016/j.rinp.2021.103987). [PMID: [36338375](https://pubmed.ncbi.nlm.nih.gov/36338375/)] [PMCID: [PMC9616684](https://pubmed.ncbi.nlm.nih.gov/PMC9616684/)]
  50. Susnjak T, McIntosh TR. ChatGPT: The end of online exam integrity? *Education Sciences*. 2024;14(6):656. doi: [10.3390/educsci14060656](https://doi.org/10.3390/educsci14060656).
  51. Newton P, Xiromeriti M. ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review. *Assessment & Evaluation in Higher Education*. 2024; 49(6):781-98. doi:[10.1080/02602938.2023.2299059](https://doi.org/10.1080/02602938.2023.2299059).
  52. Sumbal A, Sumbal R, Amir A. Can ChatGPT-3.5 pass a medical exam? A systematic review of ChatGPT's performance in academic testing. *J Med Educ Curric Dev*. 2024 Mar 13;11:23821205241238641. doi: [10.1177/23821205241238641](https://doi.org/10.1177/23821205241238641). [PMID: [38487300](https://pubmed.ncbi.nlm.nih.gov/38487300/)] [PMCID: [PMC10938614](https://pubmed.ncbi.nlm.nih.gov/PMC10938614/)]
  53. Passeri S, Li LM, Nadruz Jr W, Bicudo AM. Medical students progress in the practice assessment of knowledge, skills, and attitudes. *Creative Education*. 2015; 6(08): 805. doi:[10.4236/ce.2015.68084](https://doi.org/10.4236/ce.2015.68084).
  54. Preston R, Gratani M, Owens K, Roche P, Zimanyi M, Malau-Aduli B. Exploring the impact of assessment on medical students' learning. *Assessment & Evaluation in Higher Education*. 2020;45(1):109-24. doi:[10.1080/02602938.2019.1614145](https://doi.org/10.1080/02602938.2019.1614145).
  55. Maison M, Darmaji D, Astalini A, Perdana R. Supporting assessment in education: e-assessment interest in physics. *Universal Journal of Educational Research*. 2020;8(1):89-97. doi:[10.13189/ujer.2020.080110](https://doi.org/10.13189/ujer.2020.080110).
  56. Rudolph J, Tan S, Tan S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning and teaching*. 2023 Jan 25; 6(1): 342-63. doi:[10.37074/jalt.2023.6.1.9](https://doi.org/10.37074/jalt.2023.6.1.9).
  57. Hopfenbeck TN. The future of educational assessment: self-assessment, grit and ChatGPT?. *Assessment in Education: Principles, Policy & Practice*. 2023 Mar 4;30(2):99-103. doi:[10.1080/0969594X.2023.2212192](https://doi.org/10.1080/0969594X.2023.2212192).
  58. Gamage KA, Dehideniya SC, Xu Z, Tang X. ChatGPT and higher education assessments: more opportunities than concerns? *Journal of Applied Learning and Teaching*. 2023;6(2): 358-69. doi:[10.37074/jalt.2023.6.2.32](https://doi.org/10.37074/jalt.2023.6.2.32).
  59. Kolade O, Owoseni A, Egbetokun A. Is AI changing learning and assessment as we know it? Evidence from a ChatGPT experiment and a conceptual framework. *Heliyon*. 2024 Feb 10;10(4):e25953. doi: [10.1016/j.heliyon.2024.e25953](https://doi.org/10.1016/j.heliyon.2024.e25953). [PMID: [38379960](https://pubmed.ncbi.nlm.nih.gov/38379960/)] [PMCID: [PMC10877295](https://pubmed.ncbi.nlm.nih.gov/PMC10877295/)]
  60. Pande SM. Machine learning models for student performance prediction. In *Proceedings of the 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*; 2023 Mar 14; Coimbatore, India. Piscataway, NJ: IEEE; 2023: 27-32. doi: [10.1109/ICIDCA56705.2023.10099503](https://doi.org/10.1109/ICIDCA56705.2023.10099503).
  61. Jiao P, Ouyang F, Zhang Q, Alavi AH. Artificial intelligence-enabled prediction model of student academic performance in online engineering education. *Artificial Intelligence Review*. 2022 Dec;55(8):6321-44. doi:[10.1007/s10462-022-10155-y](https://doi.org/10.1007/s10462-022-10155-y).
  62. Haron NH, Mahmood R, Amin NM, Ahmad A, Jantan SR. An artificial intelligence approach to monitor and predict student academic performance. *Journal of advanced research in applied sciences and engineering technology*. 2024;44(1):105-19. doi:[10.37934/araset.44.1.105119](https://doi.org/10.37934/araset.44.1.105119).
  63. Wazir S, Mohani SS, Affandi H, Rafique AA, Soomro M. Impact of artificial intelligence and machine learning on predicting student performance and engagement. *Dialogue Social Science Review (DSSR)*. 2025;3(1):1298-311.
  64. Pratama MP, Sampelolo R, Lura H. Revolutionizing education: harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of education, language teaching and science*. 2023; 5(2):350-7. doi:[10.52208/klasikal.v5i2.877](https://doi.org/10.52208/klasikal.v5i2.877).
  65. Jin SH, Im K, Yoo M, Roll I, Seo K. Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*. 2023;20(1):37. doi: [10.1186/s41239-023-00406-5](https://doi.org/10.1186/s41239-023-00406-5).
  66. Fauzi F, Tuhuteru L, Sampe F, Ausat AM, Hatta HR. Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education*. 2023;5(4):14886-91. doi:[10.31004/joe.v5i4.2563](https://doi.org/10.31004/joe.v5i4.2563).



67. Rawas S. ChatGPT: Empowering lifelong learning in the digital age of higher education. *Education and Information Technologies*. 2024;29(6):6895-908. doi:[10.1007/s10639-023-12114-8](https://doi.org/10.1007/s10639-023-12114-8).
68. Monib WK, Qazi A, Mahmud MM. Exploring learners' experiences and perceptions of ChatGPT as a learning tool in higher education. *Education and Information Technologies*. 2025; 30(1): 917-39. doi:[10.1007/s10639-024-13065-4](https://doi.org/10.1007/s10639-024-13065-4).
69. Abas MA, Arumugam SE, Yunus MM, Rafiq KR. ChatGPT and personalized learning: opportunities and challenges in higher education. *International Journal of Academic Research in Business and Social Sciences*. 2023;13(12): 3536-45. doi:[10.6007/IJARBS/v13-i12/20240](https://doi.org/10.6007/IJARBS/v13-i12/20240).
70. Griffith AS, Altinay Z. A framework to assess higher education faculty workload in US universities. *Innovations in Education and Teaching International*. 2020; 57(6): 691-700. doi:[10.1080/14703297.2020.1786432](https://doi.org/10.1080/14703297.2020.1786432).
71. Cambra-Fierro JJ, Blasco MF, López-Pérez ME, Trifu A. ChatGPT adoption and its influence on faculty well-being: an empirical research in higher education. *Education and Information Technologies*. 2024;30:1517-38. doi:[10.1007/s10639-024-12871-0](https://doi.org/10.1007/s10639-024-12871-0).
72. Sun GH, Hoelscher SH. The ChatGPT storm and what faculty can do. *Nurse Educ*. 2023 May-Jun;48(3):119-124. doi:[10.1097/NNE.0000000000001390](https://doi.org/10.1097/NNE.0000000000001390). [PMID: [37043716](https://pubmed.ncbi.nlm.nih.gov/37043716/)]
73. Nikoçeviq-Kurti E, Bërdynaj-Syla L. ChatGPT integration in higher education: impacts on teaching and professional development of university professors. *Educational Process: International Journal*. 2024; 13(3): 22-39. doi:[10.22521/edupij.2024.133.2](https://doi.org/10.22521/edupij.2024.133.2).
74. Cross J, Robinson R, Devaraju S, Vaughans A, Hood R, Kayalackakom T, et al. Transforming medical education: assessing the integration of ChatGPT into faculty workflows at a Caribbean medical school. *Cureus*. 2023 Jul 5;15(7):e41399. doi:[10.7759/cureus.41399](https://doi.org/10.7759/cureus.41399). [PMID: [37426402](https://pubmed.ncbi.nlm.nih.gov/37426402/)] [PMCID: [PMC10328790](https://pubmed.ncbi.nlm.nih.gov/PMC10328790/)]
75. Stahl BC, Wright D. Ethics and privacy in AI and big data: implementing responsible research and innovation. *IEEE Security & Privacy*. 2018;16(3):26-33. doi: [10.1109/MSP.2018.2701164](https://doi.org/10.1109/MSP.2018.2701164).
76. Hesham A, Dempere J, Akre V, Flores P. Artificial intelligence in education (AIED): implications and challenges. In *Proceedings of the HCT International General Education Conference (HCT-IGEC 2023)*; 2023 Nov 24; Abu Dhabi, UAE: Atlantic Press; 2023: 126-40. doi: [10.2991/978-94-6463-286-6\\_1](https://doi.org/10.2991/978-94-6463-286-6_1).
77. Rahayu S. The impact of artificial intelligence on education: opportunities and challenges. *Jurnal Educatio FKIP UNMA*. 2023; 9(4), 2132-40.
78. Owoc ML, Sawicka A, Weichbroth P. Artificial intelligence technologies in education: benefits, challenges and strategies of implementation. In: Ligeza A, Potempa T. *IFIP International Workshop on Artificial Intelligence for Knowledge Management*; 2019 Aug 11; Cham, Switzerland. Springer International Pub; 2019: 37-58. doi: [10.1007/978-3-030-85001-2\\_4](https://doi.org/10.1007/978-3-030-85001-2_4).



**Figure 1.** PRISMA flow diagram for information search

**Table 1.** Demographic characteristics of selected studies

References	Authors	Year	Title
17	Lee	2023	Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation
18	Swiecki et al.	2022	Assessment in the age of artificial intelligence
19	Maier & Klotz	2022	Personalized feedback in digital learning environments: Classification framework and literature review
20	Yildirim-Erbasli & Bulut	2023	Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment
21	Smerdon	2024	AI in essay-based assessment: Student adoption, usage, and performance
22	Martínez-Comesană et al.	2023	Impact of artificial intelligence on assessment methods in primary and secondary education: Systematic literature review
23	Memarian, & Doleck	2024	A review of assessment for learning with artificial intelligence
24	Tonbuloglu	2023	An evaluation of the use of artificial intelligence applications in online education
25	González-Calatayud et al.	2021	Artificial intelligence for student assessment: A systematic review
26	Casey	2024	ChatGPT in public policy teaching and assessment: An examination of opportunities and challenges
27	Lu et al.	2024	Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing?
28	Williams	2023	AI, analytics and a new assessment model for universities
29	Salinas-Navarro et al.	2024	Using generative artificial intelligence tools to explain and enhance experiential learning for authentic assessment
30	Lye & Lim	2024	Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations
31	Fahmy	2024	Student perception on AI-driven assessment: Motivation, engagement and feedback capabilities
32	Koh & Doroudi	2023	Learning, teaching, and assessment with generative artificial intelligence: Towards a plateau of productivity
33	Hooda et al.	2022	Artificial intelligence for assessment and feedback to enhance student success in higher education
34	Zirar	2023	Exploring the impact of language models, such as ChatGPT, on student learning and assessment
35	Mao et al.	2024	Generative artificial intelligence in education and its implications for assessment
36	Kooli & Yusuf	2024	Transforming educational assessment: Insights into the use of ChatGPT and large language models in grading.
37	Chaudhry et al.	2023	Time to revisit existing student's performance evaluation approach in higher education sector in a new era of ChatGPT-a case study
38	Ouyang et al.	2023	A systematic review of AI-driven educational assessment in STEM education
39	Xia et al.	2024	A scoping review on how generative artificial intelligence transforms assessment in higher education
40	A. Fuller et al.	2024	Exploring the use of ChatGPT to analyze student course evaluation comments

**Table 2.** Analytical comparison of selected studies in the main research themes

References	Feedback Applications	Online Exam Applications	Instrument Design Applications	Assessment Process Applications	Student Learning Management Applications	Faculty Workload Management Applications
17	*			*	*	*
18		*	*	*	*	*
19	*			*	*	*
20	*	*		*	*	*
21		*				
22				*		*

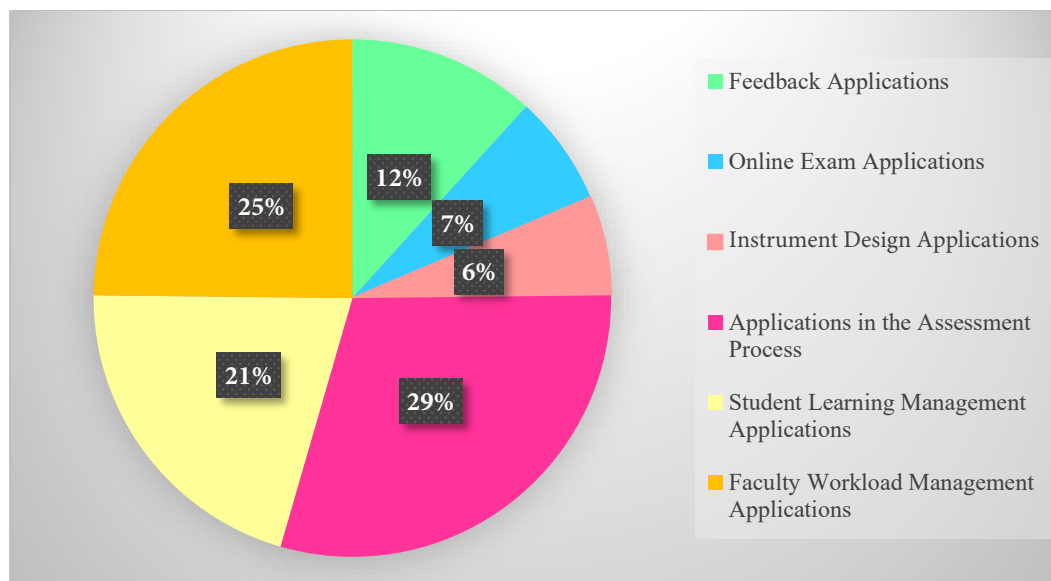
23	*	*		*	*	*
24	*	*		*		*
25				*		
26				*		*
27	*					
28	*		*	*	*	*
29	*	*		*	*	*
30				*		*
31				*	*	
32			*	*	*	
33	*		*	*	*	*
34	*					*
35				*		*
36	*					*
37					*	*
38					*	*
39	*			*		*
40				*		

**Table 3.** Research findings

Main Theme	Sub-theme	Key Concepts
Feedback Applications	Interaction and feedback quality	Constructive and critical feedback; Encouraging and motivating feedback; Improving the quality of feedback; Reducing errors and bias in feedback; Making feedback understandable for students
	Timing and sustainability of feedback	Timely and prompt feedback; Real-time feedback; Feedback timing tailored to each student; Developing sustainable and cyclical feedback mechanisms
	Personalization and flexibility in feedback	Adaptive feedback; Focusing on individual learner goals when providing feedback; Customizing feedback generation rules; Flexibility in providing feedback
	Technology and innovation in feedback	Machine and intelligent feedback; High computational power and analytical framework in feedback delivery; Optimizing learning through feedback; Providing feedback anytime and anywhere
Online Exam Applications	Improving the design and execution process of exams	Improving the quality of exam questions; Multimedia questions; Randomizing exam questions; Fine-tuning the difficulty level of questions; Designing mixed-type exams; Simulating exam questions
	Personalization and optimization of the exam experience	Computerized adaptive testing; Personalizing exam questions; Overcoming student exam anxiety; Assisting in conducting exams in digital environments
Instrument Design Applications	Design and innovation in assessment instrument	Designing creative assessment instruments for students; Assisting in designing standardized instruments; Designing advanced instruments for plagiarism detection; Diversifying grading criteria in instruments
	Improving the accuracy and efficiency of assessment instrument	Enhancing the reliability of measurement instruments; Adjusting scoring weight methods; More diverse and accurate scaling in instruments; Increasing the parameters of instruments; Developing assessment rubrics
Applications in the Assessment Process	Innovation and development of assessment systems	Developing peer assessment mechanisms; Web-based interactive assessment system; Development of electronic assessment platforms; Digital assessment; Creating learner performance dashboards; Integration of online assessment; Innovation in assessment; Development of covert assessment techniques; Assessment management
	Improving assessment processes and standards	Improving current assessment processes; Developing assessment standards; Improving the quality of assessment; Reducing human error in assessment; Reducing human bias in assessment; Greater transparency in the assessment process; Enhanced accountability; Ethics in assessment

	Personalized and learning-aligned assessment	Personalizing assessment; Assessment aligned with learning objectives; Assessment based on previous performances; Objective assessment of outcomes; Collecting multifaceted data in assessment; Collecting behavioral data in assessment; Real-time learning analysis; Analyzing learning behaviors; More accurate prediction of future student grades; Predicting the likelihood of student dropout
	Dynamic assessment strategies and approaches	Diversifying assessment approaches; Game-based assessment; Evidence-based assessment; Developing formative assessment; Group assessment; Stage-based assessments; Developing assessment sequence throughout the semester; Assessment for current and future needs
	Analysis and data-driven assessment	Achieving assessment goals; Learning analysis; Diagnostic analysis of performance causes; Identifying key learning trends; Identifying patterns in behaviors, preferences, and learning achievements of students; More conscious assessment; Assessment for learning; Cognitive analysis-based assessment
Student Learning Management Applications	Supporting personalized and self-regulated learning	Assisting with self-assessment; Helping with self-regulated learning; Developing future learning pathways for students; Helping students identify strengths and weaknesses; Providing recommendations for learners' next steps; Increasing independence in student learning; Enhancing self-directed learning capacity; Helping students choose the most optimal learning path; Assisting students in identifying knowledge gaps; Helping students correct learning behaviors; Helping students understand their own learning
	Developing sustainable learning skills and capabilities	Improving writing skills in exams; Developing students' thinking about their own learning methods; Helping to develop metacognitive skills; Developing metacognitive capabilities; Sustainable and lifelong learning; Problem-based learning; Helping students connect new and prior knowledge; Supporting students in self-development
	Enhancing the quality of learning experience and motivation	Helping improve the quality of class activities; Providing new opportunities to increase student participation; Enhancing learning motivation; Improving satisfaction with education; Helping students better understand learning objectives; Meeting the diverse needs of students; Improving quality learning experiences; Assisting in creating new learning opportunities; Providing students with sufficient opportunity for private learning; Providing individual support to students in identifying areas for improvement and strengthening learning; Helping students organize their activities based on feedback
Faculty Workload Management Applications	Improving efficiency and reducing faculty workload	Eliminating repetitive tasks for faculty members; Reducing the workload of faculty members; Saving time; Facilitating assessment tasks; Grading student assignments; Helping with classroom management; Providing new methods for monitoring learners; Enabling timely and appropriate interventions; Effectiveness in assessment with large classroom scales
	Improving learning and teaching processes	Helping design customized digital tasks for students; Assisting in systematic monitoring of student performance; Designing guidelines for implementing educational corrections; Helping align with teaching strategies; Helping improve the effectiveness of teaching strategies; Designing learning activities based on needs; Designing tasks based on abilities and performance levels; Assisting in adjusting teaching strategies; Designing learning activities based on students' characteristics and abilities; Developing skills and the ability to transfer learning to new situations; Helping review learning objectives; Helping improve educational processes
	Data analysis and supporting student learning	Better understanding of learner behavior in exams; More accurate estimation of students' abilities; Supporting learning-based simulation assessment; Creating performance mapping for students; Helping deliver assignments and learning tasks according to each student's learning level; Identifying areas where students need improvement; Assisting in identifying learning gaps in students; Helping analyze complex data; Assisting in big data analysis; Enabling daily monitoring of student performance; Providing reliable and valid inferences about what students know and can do; Early detection of learning problems; Helping assess student learning outcomes in relation to learning objectives; Determining student competency levels based on assessment results; Helping identify students who need additional support and assistance more quickly





**Figure 2.** Main themes by frequency

**Table 4.** Comparative and analytical summary of traditional and AI-based assessment methods for students

Criteria	Traditional Assessment Methods	AI-Based Assessment Methods
Feedback Applications	<p>Interaction and Feedback Quality: Feedback is often subjective, delayed, and less personalized.</p> <p>Timing: Feedback is given after a delay, typically after grading assignments or exams.</p> <p>Personalization: Feedback is generalized and does not cater to individual learning styles.</p> <p>Technology: Limited technological support for feedback delivery.</p>	<p>Quality: AI can offer constructive, real-time, and personalized feedback tailored to individual student needs, reducing errors and bias.</p> <p>Timing: AI provides real-time feedback, immediately addressing performance and offering ongoing support.</p> <p>Personalization: AI customizes feedback based on the learner's performance and goals, creating a more adaptive learning environment.</p> <p>Innovation: AI enables feedback anytime and anywhere, integrating machine learning and high computational power for more accurate insights.</p>
Online Exam Applications	<p>Design and Execution: Exams are often fixed in format and are not easily adaptable</p> <p>Customization: Exams are standardized, offering little flexibility for individual student needs.</p> <p>Anxiety: Traditional exams may increase student anxiety.</p>	<p>Design: AI allows multimedia, randomized questions, and adaptive testing that changes based on the student's ability level.</p> <p>Personalization: AI personalizes exam questions based on prior student performance and adapts difficulty levels.</p> <p>Optimization: AI can help reduce exam-related stress by providing customized test difficulty and predictive tools.</p>
Instrument Design Applications	<p>Design: Assessment tools are manual and rigid, with fixed grading rubrics and criteria.</p> <p>Efficiency: Instruments are typically slow and require manual intervention.</p>	<p>Design &amp; Innovation: AI allows the creation of dynamic and advanced tools for plagiarism detection, diverse grading, and customized rubrics.</p> <p>Accuracy and Efficiency: AI enhances accuracy by refining scoring methods, increasing diversity in assessment, and improving reliability.</p>
Applications in the Assessment Process	<p>Current Systems: Standardized assessment methods with limited interactivity and personalization.</p> <p>Bias and Error: Human bias and error are inherent in manual assessments.</p>	<p>Innovation: AI enables dynamic and web-based interactive systems, real-time performance tracking, and digital assessment platforms.</p> <p>Bias Reduction: AI reduces human error, providing transparent, reliable, and accurate assessments.</p>

Student Learning Management Applications	<p>Traditional Support: Faculty manually supports student learning, with limited monitoring.</p> <p>Learning Pathways: Limited tools for guiding student learning outside of direct faculty interaction.</p>	<p>Self-Regulated Learning: AI assists students with self-assessment, helps track strengths and weaknesses, and provides recommendations for improvement.</p> <p>Learning Pathways: AI offers customized learning paths based on student performance and helps develop lifelong learning skills.</p>
Faculty Workload Management Applications	<p>Workload: Faculty spends significant time on grading, managing assignments, and overseeing learning.</p> <p>Monitoring: Limited ability for monitoring large numbers of students effectively.</p>	<p>Efficiency: AI automates grading, streamlines task management, and supports timely interventions, reducing faculty workload.</p> <p>Data-Driven Decisions: AI helps faculty monitor and analyze student behavior in real-time, improving teaching effectiveness and student performance.</p>