Nusantara 1

Artikel Galleys Scientica-Vol2.4-1986.pdf

Check - No Repository 19

New Test

Konsorsium PTS Batch 2

Document Details

Submission ID

trn:oid:::1:3355379636

Submission Date

Sep 29, 2025, 2:49 PM GMT+7

Download Date

Sep 29, 2025, 2:51 PM GMT+7

File Name

Artikel_Galleys_Scientica-Vol2.4-1986.pdf

File Size

619.0 KB

15 Pages

6,631 Words

40,247 Characters

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

38 Not Cited or Quoted 6%
Matches with neither in-text citation nor quotation marks

13 Missing Quotations 2%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

6% Internet sources
6% Publications
1% Submitted works (Student Papers)

Match Groups

- **38** Not Cited or Quoted 6%
  Matches with neither in-text citation nor quotation marks
- **13** Missing Quotations 2%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

6% Internet sources
6% Publications
1% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
slejournal.springeropen.com		<1%
2	Publication	
"Reimagining Transformative Educational Spaces", Springer Science and Business...		<1%
3	Publication	
Eileen du Plooy, Daleen Casteleijn, Denise Franzsen. "Personalized adaptive learni...		<1%
4	Internet	
pubmed.ncbi.nlm.nih.gov		<1%
5	Publication	
Samritin, Reni Astuti, Nita Sri Budiarti, Desna Sari, Selviana Selviana, Mikrahma A...		<1%
6	Publication	
Trey Martindale, Tonya B. Amankwatia, Lauren Cifuentes, Anthony A. Piña. "Hand...		<1%
7	Internet	
api.aesop-planning.eu		<1%
8	Internet	
www.nature.com		<1%
9	Internet	
www.researchsquare.com		<1%
10	Student papers	
De LaSalle University - College of Saint Benilde		<1%

11	Internet	
blog.talentgarden.com		<1%
12	Internet	
www.scribd.com		<1%
13	Student papers	
Adtalem Global Education		<1%
14	Student papers	
Mapua Institute of Technology		<1%
15	Publication	
Yoshihiko Shoji, Shuji Munejiri, Eiko Kaga. "Validity of Force Concept Inventory ev...		<1%
16	Internet	
ejournal.staialhikmahpariangan.ac.id		<1%
17	Internet	
intapi.sciendo.com		<1%
18	Internet	
isarpublisher.com		<1%
19	Internet	
www.compadre.org		<1%
20	Publication	
Karla Menezes, Conceição Aleixo, Susana Silva, Francisco Carreira, Silvio Farias M...		<1%
21	Internet	
eldorado.tu-dortmund.de		<1%
22	Internet	
www.andhraloyolacollege.ac.in		<1%
23	Internet	
www.frontiersin.org		<1%
24	Publication	
Woo Sub Kim, Hong-Geun Jung. "Supramalleolar Osteotomy for Treatment of the ...		<1%

25	Internet	
ijip.in		<1%
26	Internet	
kuey.net		<1%
27	Internet	
pure.rug.nl		<1%
28	Internet	
www.judijournal.com		<1%
29	Publication	
"Computer-Human Interaction Research and Applications", Springer Science and ...		<1%
30	Publication	
Beri Nopriansyah, Titi Fitri, Bilqis Arifah, Miftahul Huda, Umi Machamudah. "Goo...		<1%
31	Publication	
Yuli Rahmawati, Peter Charles Taylor. "Empowering Science and Mathematics for ...		<1%
32	Internet	
ejournal.uinmadura.ac.id		<1%
33	Internet	
epub.uni-regensburg.de		<1%
34	Internet	
japendi.publikasiindonesia.id		<1%
35	Internet	
jmedcasereportsimages.org		<1%
36	Internet	
journal-center.litpam.com		<1%
37	Internet	
rgsa.emnuvens.com.br		<1%
38	Internet	
sprcopen.org		<1%

39 Internet

www.drugsandalcohol.ie <1%

40 Publication

Maharani Kartikasari, Triyanto, Laila Fitriana, Farida Nurhasanah. "Using Interac... <1%

41 Publication

Stephen J. Farenga, Salvatore G. Garofalo, Daniel Ness. "International Handbook ... <1%

42 Publication

Jozsef Katona, Klara Ida Katonane Gyonyoru. "Integrating AI-Based Adaptive Lear... <1%

Journals Scientica Education Journal
E-ISSN : 3046-8523
https://nawalaeducation.com/index.php/SEJ/index
Vol.2.No.4 September 2025

Applying AI-Based Adaptive Assessment to Map and Enhance Science Literacy Skills in Generation Z: A Systematic Literature Review

Gde Wikan Pradnya Dana
Computer Engineering Study Program, Warmadewa University, Denpasar, Bali, Indonesia
Email: wikanpdana8044@warmadewa.ac.id

Inputted : August 18, 2025
Accepted : August 29, 2025
Revised : September 08, 2025
Published : September 29, 2025

ABSTRACT

Science literacy is a critical 21st-century competency, yet student achievement in this area presents significant challenges globally. This gap is exacerbated by conventional assessment methods that are misaligned with the learning characteristics of the digital-native Generation Z, who expect personalized, interactive, and instantaneous feedback. This study aims to analyze and synthesize the current research landscape on the application of AI-based adaptive assessment to map and enhance the science literacy skills of Generation Z. Employing a Systematic Literature Review (SLR) guided by the PRISMA framework, this study identifies existing theoretical models, platforms, evidence of effectiveness, and implementation challenges. The findings indicate that AI-based adaptive assessment platforms, such as Inq-ITS and ALEKS, effectively measure various science sub-skills and improve learning outcomes. Key features including personalization, interactivity, and immediate feedback closely align with the preferences of Generation Z, thereby enhancing student motivation and engagement. Nevertheless, implementation faces significant challenges related to infrastructure, teachers' pedagogical readiness, and crucial ethical considerations, including data privacy and algorithmic bias. This study concludes that AI-adaptive assessment holds transformative potential, yet its effective and equitable adoption requires addressing existing challenges and future research gaps.

Keywords: *adaptive assessment, artificial intelligence, science literacy, generation Z, systematic literature Review*

INTRODUCTION

Global civilization is currently in the midst of a fundamental digital transformation, an era characterized by rapid technological advancements and the increasing complexity of global challenges. This development necessitates a redefinition of various aspects of life, not least the education sector, which plays a crucial role as the primary foundation for developing a skilled workforce capable of driving national progress (Firdaus et al., 2025). Amidst this dynamic landscape, science literacy has transformed from merely an academic domain into an essential competency for every individual. The capacity for science literacy is no longer limited to the ability to recall scientific facts but has evolved into a holistic set of skills. This skill set encompasses the ability to think critically, analyze and evaluate evidence objectively, understand the essence of the scientific inquiry process, and apply scientific concepts to formulate informed and responsible decisions in daily life (Patigu et al., 2024). This competency has become a prerequisite life skill in the 21st century, preparing individuals to face contemporary challenges and compete effectively on the global stage.

Despite its critical importance, education systems in many countries, including Indonesia, still face significant challenges in equipping students with adequate levels of science literacy. Reports from international assessment programs such as the Programme for International Student Assessment (PISA) consistently indicate that the science literacy level of Indonesian students remains below the international average. According to the 2022 PISA data, Indonesia ranks 6th among participating ASEAN countries and 69th globally.(OECD, 2023)

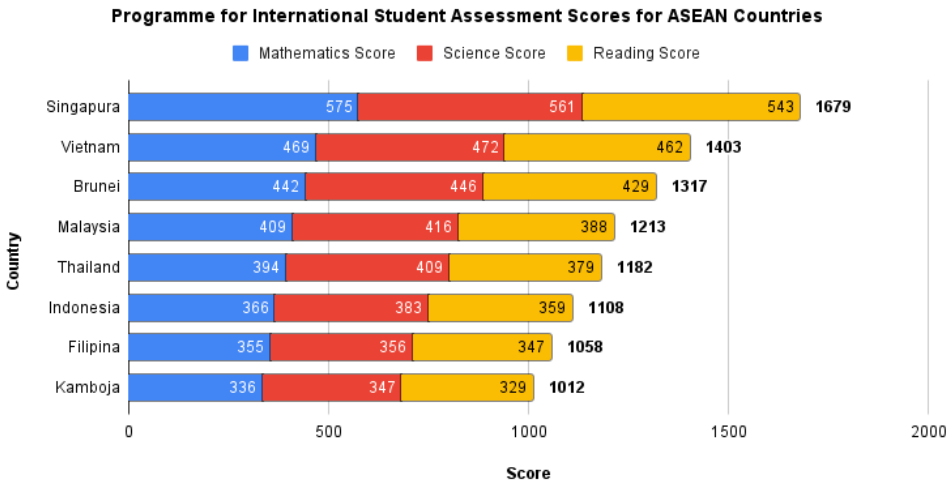

Figure 1. Programme for International Student Assessment (PISA) Scores for ASEAN Countries in 2022. Source: (OECD, 2023)

This low achievement indicates that conventional teaching and assessment approaches are not fully effective. Traditional assessment methods, often designed to be static, uniform, and focused on final outcomes, have inherent limitations. These methods frequently fail to effectively and accurately measure the broad and complex spectrum of science literacy skills, which includes higher-order cognitive processes such as analysis, evaluation, and synthesis. Consequently, the resulting depiction of student abilities is often incomplete and incapable of providing constructive feedback to improve the learning process (Ananda et al., 2023).

This challenge becomes increasingly urgent when contextualized for the unique characteristics of Generation Z, a cohort that has grown up in an environment fully saturated with digital technology. As true digital natives, their expectations, preferences, and learning styles differ fundamentally from previous generations. (Bafadal & Rosyid, 2024). Generation Z expects learning experiences that mirror the digital world they inhabit daily; they anticipate personalization, a high degree of interactivity, collaboration, and rapid, continuous feedback mechanisms. These traits are often unfulfilled by conventional educational models, which tend to be linear, passive, and employ a "one-size-fits-all" approach. The failure of the education system to adapt its pedagogical and assessment methods to this generation's characteristics risks creating an engagement gap, reducing learning motivation, and ultimately hindering a deep understanding of scientific concepts.

It is within this context that advancements in Artificial Intelligence (AI) offer extraordinary transformative potential. One of the most relevant applications of AI to address assessment challenges is the development of adaptive assessment systems. Fundamentally different from conventional linear tests, an adaptive assessment system dynamically adjusts the difficulty, sequence, and even content of questions based on a student's real-time performance and responses. By leveraging intelligent algorithms, this approach enables a far more precise, efficient, and personalized measurement of

student abilities. More than a summative evaluation tool, adaptive assessment functions as a powerful formative diagnostic instrument, capable of mapping the specific strengths and weaknesses of each student and providing actionable feedback to support a continuous learning cycle.

Although the potential of AI-based adaptive assessment to revolutionize science education is immense, its specific application to map and enhance the science literacy skills of Generation Z remains a relatively new and evolving field of research. There is a pressing academic need to comprehensively understand the landscape of existing theoretical models and technological platforms, to systematically evaluate the evidence of their effectiveness, and to navigate the accompanying technical, pedagogical, and ethical challenges. Without a holistic, evidence-based understanding, the adoption of this advanced technology risks being ineffective, inequitable, or even counterproductive.

Therefore, this systematic literature review is designed to synthesize and analyze the existing body of research to build a solid foundation of knowledge on this topic. The study aims to provide a clear overview of the current status, potential, and challenges within the domain of AI adaptive assessment for science education. Specifically, this review will be guided by the following research questions:

- **RQ1:** How are current AI-based adaptive assessment models and platforms used in the literature to identify, measure, and profile the various sub-skills of science literacy in students?
- **RQ2:** Based on existing literature, what empirical evidence demonstrates the effectiveness of AI-based adaptive assessment in enhancing the conceptual understanding and science process skills of Generation Z students?
- **RQ3:** What are the most significant challenges, limitations, and ethical considerations identified in the literature regarding the implementation of AI-based adaptive assessment, and what research gaps are recommended for future studies?

By answering these questions through a systematic and rigorous literature review methodology, this report aims to provide a comprehensive, evidence-based understanding. It is hoped that the findings will offer valuable theoretical and practical insights for researchers, educators, educational technology developers, and policymakers in their collective effort to responsibly leverage AI's potential to improve the quality of science education and prepare Generation Z for future challenges.

METHOD

This study adopts a Systematic Literature Review (SLR) methodology to comprehensively answer the formulated research questions (RQs). This approach was selected for its capacity to systematically identify, evaluate, and synthesize relevant research evidence. To ensure methodological rigor, the entire review process was strictly guided by the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework. Adherence to this standard ensures that the resulting analysis is transparent, replicable, and free from the selection bias associated with anecdotal literature reviews.

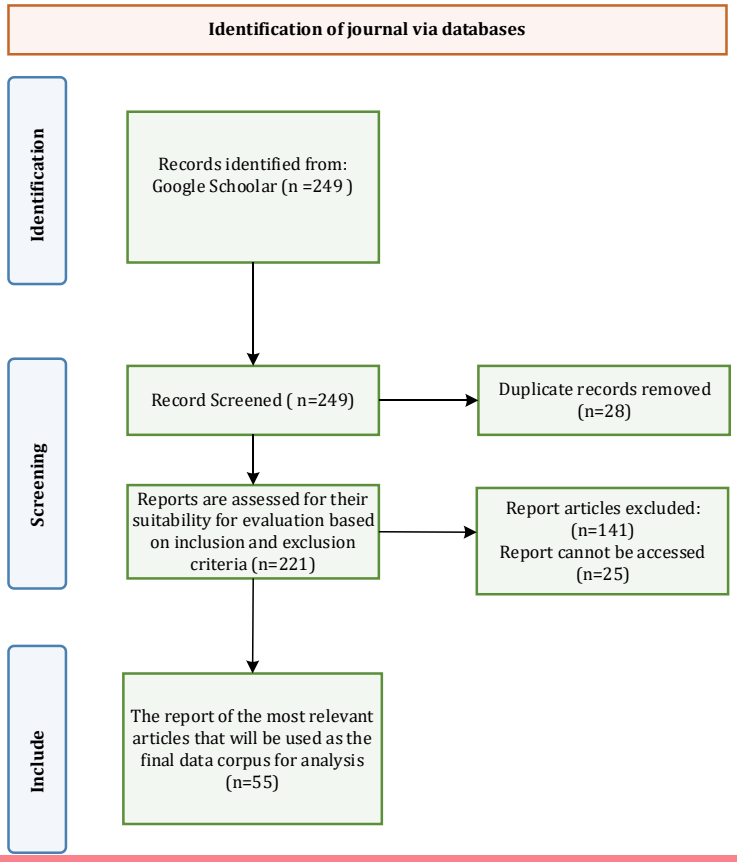

Figure 2. *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart describing the inclusion and exclusion criteria.*

The literature selection process for this study was systematically conducted following the PRISMA framework, beginning with the identification stage. A search was conducted in the Google Scholar database to ensure broad coverage, targeting publications from 2020 to 2025. This initial search yielded 249 articles.

The subsequent stage was screening, during which 28 duplicate articles were removed, resulting in 221 unique articles. These articles were then screened by title and abstract for relevance to the research questions, leading to the exclusion of 98 articles. Of the remaining 123 articles, full-text retrieval was attempted; however, 25 reports could not be accessed.

The final stages involved eligibility assessment and inclusion. The 98 retrieved reports were then assessed for eligibility through a thorough full-text review. At this stage, an additional 43 articles were excluded for not meeting the specific inclusion criteria. This rigorous selection process yielded a final corpus of 55 relevant studies. This final data corpus forms the basis for all the analysis and qualitative synthesis presented in this study to provide in-depth answers to the research questions. Analisis lebih banyak menonjolkan potensi dan keunggulan AI-based adaptive assessment, tetapi kritik terhadap keterbatasan dan risiko masih minim.

RESULTS AND DISCUSSION

RQ1: How are current AI-based adaptive assessment models and platforms used in the literature to identify, measure, and profile the various sub-skills of science literacy in students?

The Evolution of Adaptive Assessment Models and Platforms from Modern Psychometrics to Generative AI

Current AI-based adaptive assessment models and platforms have evolved from the foundations of modern psychometrics toward the utilization of advanced artificial intelligence. This evolution not only changes how assessments are delivered but also fundamentally transforms their purpose from mere measurement into a diagnostic and formative tool integrated into the learning process. The initial step toward adaptive assessment was the shift from static linear test models to Computerized Adaptive Testing (CAT). Conventional assessments present the same set of questions to all students, an approach with inherent weaknesses. Questions that are too easy for high-ability students or too difficult for low-ability students do not provide accurate information, often resulting in floor and ceiling effects (the failure of a test to differentiate proficiency at the extremes of the ability spectrum). CAT addresses this problem by using algorithms to select the next test item based on a student's performance on the previous one, thus delivering assessments that are more precise, efficient, and tailored to each individual's unique ability level (Akhmetova et al., 2025).

The backbone that enables CAT to function effectively is Item Response Theory (IRT). IRT is a statistical framework that links the probability of a student's correct answer to their ability level (theta) and the characteristics of the test item (e.g., difficulty, discrimination). The primary advantage of IRT is its ability to place student ability and item difficulty on the same scale. This allows the CAT algorithm to intelligently select the most informative items typically those where a student has approximately a 50% probability of answering correctly at each step of the test. This approach has become the foundation for large-scale assessments such as PISA to measure science literacy accurately and efficiently.

These theoretical models are realized in various practical platforms. A widely known example of an adaptive assessment system is ALEKS (Assessment and Learning in Knowledge Spaces). ALEKS uses AI algorithms to accurately diagnose student knowledge and then presents the most relevant learning materials and problems to build a personalized learning path (Harati et al., 2021). This goal of personalization aligns with broader trends in educational technology, where platforms like Learning Management Systems (LMS) are also used to provide different learning paths and flexible access to resources, ultimately aiming to support personalized learning (Attard & Holmes, 2022). Despite its noble objectives, the implementation of platforms like ALEKS faces practical challenges. Studies indicate that without careful design, isolated and highly individualized learning environments can decrease student motivation and self-regulated learning skills due to a lack of social interaction and perceived system complexity (Harati et al., 2021)

With advancements in computation, assessment models have evolved from merely measuring a single dimension of ability to more holistic modeling. Approaches such as Bayesian Networks (BN) from the field of Educational Data Mining (EDM) allow for the modeling of complex probabilistic relationships among various variables. Instead of generating a single score for science literacy, BN can predict student success by considering interconnected factors, such as individual variables, family background, and school factors. The flexibility of this model marks a shift from summative assessment toward predictive modeling that can provide systemic insights into the factors influencing science learning.

Recent advancements in AI, particularly Generative AI (GenAI), have opened a new paradigm that surpasses CAT. While CAT selects items from a pre-existing item bank, GenAI-based systems have the potential to create unique questions, problem scenarios, and assessment tasks dynamically and on-the-fly (Gunsaldi et al., 2025). Platforms like ChatGPT are now being explored to generate contextual physics problems or to

compose lesson plans based on the 5E model, enriched with various scenarios and activities (Kotsis, 2025).

This approach enables the assessment of higher-order skills that are difficult to measure with traditional question formats, such as scientific creativity, collaborative problem-solving, and science process skills (Jalaludin et al., 2024; Ješková et al., 2022). By generating contextually relevant tasks, GenAI can make assessment a more authentic and engaging learning experience. Studies show that the use of GenAI tools significantly contributes to conceptual understanding, scientific literacy, and student motivation (Gunsaldi et al., 2025)

The progression from CAT/IRT to BN and then to GenAI reflects a philosophical evolution in the purpose of assessment::

a) Assessment of Learning: Initially, technology (CAT/IRT) was used to make existing methods more efficient and accurate in measuring learning outcomes.
b) Assessment for Learning: Subsequently, technology (EDM/BN) was used to understand the broader systems behind learning and to provide diagnostic feedback.
c) Assessment as Learning: Now, with GenAI, the goal is to seamlessly integrate assessment into the learning process itself, making it a powerful formative tool that encourages reflection and active student engagement.

To provide a concrete illustration of how these theoretical models are implemented, an analysis of several leading platforms is highly relevant. Each platform, from large-scale assessments to classroom-based tutoring systems, adopts a different approach to identify, measure, and profile a diverse range of science literacy sub skills.

1. PISA: CAT for Large-Scale Science Literacy Assessment

21 In contrast to platforms focused on classroom environments, the Programme for International Student Assessment (PISA) utilizes adaptive assessment for the purpose of large-scale educational system evaluation.

a) Objective and Implementation: PISA employs CAT to measure the science literacy of nationally representative samples of 15 year old students across dozens of countries. Its objective is to obtain an accurate and internationally comparable overview of educational system performance.
b) Advantages of CAT: In this context, the primary advantages of CAT are efficiency and accuracy. Because the test adapts to each student's ability level, PISA can obtain reliable estimates using fewer items than a linear test would require. This is crucial for large-scale assessments that cover multiple domains and is effective in providing accurate measurements across the entire ability spectrum (Hopfenbeck et al., 2018)

2. ALEKS: Mapping the Knowledge Space of Science

13 Assessment and Learning in Knowledge Spaces (ALEKS) is a widely used platform in mathematics and science education (such as chemistry), which focuses on mapping conceptual knowledge.

a) Theoretical Foundation: ALEKS is built upon Knowledge Space Theory, which models a knowledge domain as a structured network where mastery of certain topics is a prerequisite for learning others.
b) Adaptive Process: Students begin with an initial assessment to determine their "knowledge state." An AI algorithm then infers which topics have been mastered and identifies the topics a student is "ready to learn." In learning mode, ALEKS presents only these topics, ensuring students are neither faced with overly difficult material nor repeating content they have already mastered.

c) Content Granularity and Implications: The strength of ALEKS lies in its highly granular content mapping, which allows for a very precise diagnosis of a student's knowledge gaps. However, studies show that the highly isolated learning environment of ALEKS can decrease motivation and self-regulated learning skills if not supplemented with social interaction and adequate pedagogical support (Harati et al., 2021)

3. Inq-ITS: Dissecting Science Inquiry Skills

The Inquiry Intelligent Tutoring System (Inq-ITS) is a platform specifically designed to assess and develop science inquiry skills in-depth, in accordance with modern educational standards.

a) Dual Assessment Mechanism: Inq-ITS employs a dual-pronged approach. First, it assesses "doing science" by analyzing patterns of student actions within virtual simulations (microworlds) using data mining to detect inquiry behaviors such as formulating hypotheses or controlling variables. Second, the platform assesses "writing science" by evaluating students' written explanations using Natural Language Processing (NLP) algorithms.

b) Focus on the "Messy Middle": The unique contribution of Inq-ITS is its ability to reveal discrepancies between what students do and what they write. Many students are in the "messy middle" able to conduct experiments but unable to explain them, or vice versa. By assessing both competencies separately, Inq-ITS provides far more nuanced diagnostic insights. The platform also provides scaffolding and direct feedback to help students overcome difficulties as they conduct practical work (Alfakihuddin et al., 2022)

Tabel 1. Comparison of AI-Based Adaptive Assessment Platforms for Science Education

Platform Name	Theoretical Model/Core Approach	Science Sub-Skills Measured	Key Features
Inq-ITS	Educational Data Mining (EDM), Log-file Analysis, Natural Language Processing (NLP)	Science inquiry skills (e.g., formulating hypotheses, collecting and interpreting data), Scientific argumentation (CER)	Virtual simulations (microworlds), Real-time feedback, Analytical dashboard for teachers (Inq-Blotter), Dual assessment ("doing" vs. "writing")
ALEKS	Knowledge Space Theory	Conceptual and procedural knowledge in a specific domain (e.g., Chemistry, Mathematics)	Comprehensive initial assessment, Personalized learning paths, Pie chart visualization of state of knowledge, Focus on “ready to learn”
PISA	Computerized Adaptive Testing (CAT), Item Response Theory (IRT)	General scientific literacy (explaining phenomena, evaluating scientific designs, interpreting data & evidence)	High test efficiency, Better measurement accuracy across the ability spectrum, Used for large-scale assessments

Sumber: Hasil Pengolahan Data 2025

Although these platforms demonstrate significant transformative potential, the implementation of these advanced models is not without challenges. The successful adoption of this technology is contingent upon factors such as digital infrastructure

readiness, AI literacy among teachers, and ethical considerations regarding data privacy (Kotsis, 2025). In many regions, the digital divide, a lack of teacher training, and limited resources serve as primary barriers to the equitable and effective utilization of adaptive assessment platforms (Okoye et al., 2023).

Thus, the literature indicates that AI-based adaptive assessment models and platforms have evolved significantly, from systems that optimize item selection to platforms capable of generating fully dynamic and personalized assessment experiences. However, realizing the full potential of this technology requires a collaborative effort to overcome structural and pedagogical barriers, ensuring that these assessment innovations are accessible and beneficial to all students..

RQ2: Based on existing literature, what empirical evidence demonstrates the effectiveness of AI-based adaptive assessment in enhancing the conceptual understanding and science process skills of Generation Z students?

Following the mapping of the theoretical and technical landscape, the next crucial question is: do these systems actually work? This literature review identifies a substantial body of empirical evidence demonstrating the effectiveness of AI-based adaptive assessment in enhancing various aspects of science learning, from both cognitive and affective standpoints. This evidence points to a dual benefit: a pedagogical advantage through improved student learning outcomes, and a logistical advantage through increased assessment efficiency.

Quantitative Impact on Academic Performance

Experimental and quasi-experimental studies consistently report statistically significant positive impacts from the use of adaptive learning and assessment platforms on students' academic performance.

a. Enhancement of Conceptual Understanding and Skills

1) A study evaluating an Intelligent Tutoring System (ITS) for students in STEM (Science, Technology, Engineering, and Mathematics) found that the experimental group using the system demonstrated a significant improvement in problem-solving precision compared to the control group using traditional methods. Specifically, the experimental group achieved an average precision of 85% in programming and 78% in mathematics. This difference was highly statistically significant, with a p-value of 0.002 (Villegas-Ch et al., 2025)
2) Similarly, a study involving 100 high school students using a real-time adaptive quiz platform found a "marked improvement" in academic performance and knowledge retention in the experimental group. This finding was also supported by strong statistical significance ($p < 0.001$), indicating that the intervention was highly effective. (Ramadhan et al., 2025)
3) Another study that developed a Science Education Adaptive Learning System (ScEd-ASL) found that the system's effectiveness varied depending on students' learning styles. The most prominent result was observed among students with a kinesthetic learning style, where 100% achieved mastery of the material. Although lower, effectiveness was also noted for aural (63%), read/write (55%), and visual (20%) learning styles, highlighting the potential for deeper personalization based on individual cognitive profiles. (Zulfiani et al., 2018)

b. Assessment Efficiency

In addition to enhancing learning outcomes, adaptive systems have also proven to be significantly more efficient. A study that developed a CAT version of the Force Concept Inventory (FCI), a widely used diagnostic test in physics education, found that test length could be drastically reduced. The FCI-CAT was able to reduce the number of test items by 50-63% from the original 30 items, while only experiencing a 5-10% decrease in accuracy and precision. This trade-off between a substantially shorter testing time and a minimal reduction in measurement quality demonstrates a considerable logistical advantage, freeing up valuable classroom time for other learning activities (Yasuda et al., 2021).

Tabel 2. Synthesis of Empirical Studies on the Effectiveness of Adaptive Assessment

Study (Citations and Sources)	Research Design	Sample Size and Population	Science Domain	Key Quantitative Findings	Implications
ITS STEM Studies(Villegas-Ch et al., 2025)	Experimental (Control Group)	450 university students	STEM (Mathematics, Physics, Programming)	Significant increase in precision (p=0.002). The experimental group achieved 78% (Mathematics) & 85% (Programming) precision.	ITS is significantly more effective than traditional methods in improving problem-solving accuracy at the university level.
Adaptive Quiz Study (Ramadhan et al., 2025)	Mixed Methods (Quasi-Experimental)	100 high school students	Not specifically mentioned (General Academic Performance)	Statistically significant (p<0.001) improvement in academic performance in the experimental group.	Real-time adaptive quiz platforms are effective in improving learning outcomes and knowledge retention at the secondary level.
FCI-CAT Study (Yasuda et al., 2021)	Simulation (Monte Carlo & Post Hoc)	Data from physics students	Physics (Concept of Force)	Test length can be reduced by 50-63% with a decrease in accuracy and precision of only 5-10%.	CAT provides significant logistical efficiency gains, saving test time without substantially sacrificing validity.

Sumber: Hasil Pengolahan Data 2025

Contextualization for Generation Z

The effectiveness of adaptive assessment and learning platforms can be understood more deeply when analyzed in the context of Generation Z's unique characteristics. As digital natives raised in an interactive technological environment, this generation has learning preferences and expectations that closely align with the features offered by AI-based systems. Traditional higher education, often rigid and one-

size-fits-all, frequently fails to meet modern students' expectations for engagement and personalization (Kotsis, 2025)

Generation Z is accustomed to digital environments that are interactive, personal, and provide instant feedback. They tend to reject the linear and passive "one-size-fits-all" learning models characteristic of traditional education. AI-based adaptive systems succeed in enhancing engagement and learning outcomes because their core features directly address the psychological needs and learning styles of Generation Z.

1) Personalization and Learner Autonomy: Generation Z is accustomed to digital environments that offer them control and choice. Adaptive platforms meet this need by providing personalized learning paths. The ability to learn at one's own pace (self-paced learning) and focus on material appropriate to one's individual level of understanding fosters a sense of autonomy. This feature stands in stark contrast to the traditional classroom experience, which is often too fast for some students and too slow for others, and personalization has been proven to enhance learners' intrinsic motivation (Halkiopoulos & Gkintoni, 2024).
2) Gamification and Interactivity: To enhance motivation and engagement, many adaptive platforms integrate gamification elements. The use of mechanisms such as points, badges, and leaderboards has been empirically shown to increase student motivation and cognitive achievement (Erlangga et al., 2023; Papp, 2017) (Erlangga et al., 2023; Papp, 2017). Furthermore, platforms that utilize interactive simulations, such as those found in Inq-ITS, enable "learning by doing" (Alfakihuddin et al., 2022). This approach not only makes the learning process more engaging but is also effective in building deep conceptual understanding, which is highly suitable for this demographic.
3) Instant and Continuous Feedback: Having grown up in the era of social media and online gaming, Generation Z is accustomed to receiving instant feedback. Adaptive assessment platforms meet this expectation by providing real-time evaluation and guidance immediately after a student completes a task or answers a question. The system dynamically adjusts the difficulty of questions based on student performance, thus creating a rapid and continuous cycle of learning, feedback, and improvement (Halkiopoulos & Gkintoni, 2024). This direct and relevant feedback is crucial for maintaining learning momentum and preventing demotivation due to uncertainty.

In summary, empirical evidence from various studies consistently shows that the integration of AI in e-learning can enhance student performance, engagement, and motivation. This success can be attributed to the strong alignment between the core features of adaptive technology personalization, interactivity, and instant feedback and the inherent learning styles, needs, and expectations of Generation Z.

RQ3: What are the most significant challenges, limitations, and ethical considerations identified in the literature regarding the implementation of AI-based adaptive assessment, and what research gaps are recommended for future studies?

Although the potential of AI-based adaptive assessment to transform science education is immense, its adoption and scalability are not without significant challenges. This literature review identifies a series of technical, pedagogical, and institutional barriers, as well as a profound set of ethical considerations that must be carefully navigated. Addressing these issues is crucial to ensure that this technology is used responsibly, equitably, and effectively.

Analysis of Technical, Pedagogical, and Institutional Challenges

The implementation of adaptive assessment faces various multidimensional challenges that can hinder its effectiveness if not properly managed.

1) Technical and Financial Challenges: A primary barrier is the high initial cost and technical complexity associated with developing high-quality adaptive platforms. Creating and calibrating the large and diverse item banks required for Computerized Adaptive Testing (CAT) is an expensive and time-consuming endeavor (Halkiopoulos & Gkintoni, 2024). Furthermore, implementing these systems requires adequate technological infrastructure, including reliable hardware, stable internet connectivity, and ongoing technical support. This need becomes a particularly significant challenge in under-resourced schools or in developing countries, risking the widening of the existing digital divide (Okoye et al., 2022).
2) Pedagogical Challenges: Successful implementation demands more than just providing technology; it requires a substantial pedagogical shift. Teachers need extensive training not only on how to use the software but also on how to interpret data from analytics dashboards to inform their teaching practices. The transition from the role of an information provider to a learning facilitator necessitates continuous professional development (Kotsis, 2025). Additionally, there is a risk of over-reliance on technology, which could diminish the human-to-human interaction essential for students' socio-emotional development.

Ethical Framework in AI-Based Assessment

The application of AI in assessments that can influence students' academic trajectories raises a series of complex ethical questions. These issues reflect a fundamental tension between the capabilities of AI systems and core ethical principles.

1) Data Privacy and Security: This is the most frequently cited ethical concern. To effectively personalize learning, AI systems need to collect and analyze vast amounts of granular student data every click, answer, and pause. This massive data collection creates significant risks related to data breaches, misuse of information, or undue surveillance (Kotsis, 2025). The principles of data protection and privacy must be a top priority throughout the entire lifecycle of an AI system.
2) Algorithmic Bias and Fairness: This is perhaps the most socially pernicious ethical risk. AI models learn from data. If the training data reflects existing societal biases and inequalities (e.g., historical achievement gaps between groups), the algorithm will likely replicate or even amplify those biases (Halkiopoulos & Gkintoni, 2024). This can lead to assessment systems that systematically disadvantage students from marginalized groups, thereby threatening the fundamental principle of educational equity.
3) Transparency and Accountability ("The Black Box Problem"): Many powerful AI algorithms, especially deep learning models, operate as "black boxes," where even their developers often cannot fully explain why the system made a particular decision (Kotsis, 2025). This lack of transparency creates serious accountability problems. If a student receives a low score from an AI system, educators and parents have a right to know why. Without the ability to audit the logic behind AI decisions, ensuring fairness and accountability is difficult.
4) Human Oversight and Autonomy: There is a critical need to ensure that AI remains a tool to support, not replace, human decision-making. Final responsibility for high-stakes decisions, such as graduation assessments or

course placements, must remain with human educators (Kotsis, 2025). A balance must be struck to prevent the erosion of the teacher's pedagogical role and to ensure that human qualitative judgment remains an integral part of the evaluation process.

This analysis highlights an inherent tension between how AI systems function and ethical principles. Effective personalization requires massive data, which clashes with the right to privacy. The complexity that makes algorithms powerful also makes them opaque, challenging the need for accountability. Therefore, solutions cannot be purely technical but must involve comprehensive governance frameworks, including an "*ethics by design*" approach, clear regulations, and robust human oversight.

Identification of Research Gaps and Recommendations

The analysis of these challenges and ethical considerations naturally leads to the identification of areas where further research is urgently needed. These gaps in the current evidence base create risks for real-world implementation and should be a priority for the research community.

1) Lack of Longitudinal Studies: The majority of existing research is short-term. There is a pressing need for longitudinal studies that track the long-term impact of adaptive platforms on student learning outcomes, STEM career choices, and the development of 21st-century skills (Akhmetova et al., 2025).
2) Need for More Rigorous Research Designs: The literature reports mixed results regarding the effectiveness of Intelligent Tutoring Systems (ITS), highlighting the need for more rigorous experimental designs. Future research should better isolate variables to determine which specific features (e.g., feedback type, scaffolding strategies) are most effective for different learning contexts and student populations (Erlangga et al., 2023).
3) Underrepresented Contexts: The majority of published research originates from developed nations. A significant gap exists in research on the effectiveness and implementation challenges of AI adaptive assessment in low-resource settings or in developing countries. Solutions effective in one cultural and infrastructural context may not be directly transferable to another (Okoye et al., 2022).
4) Research on Bias Mitigation: While the problem of algorithmic bias is widely acknowledged, research on effective technical and procedural methods for detecting, measuring, and mitigating bias in educational AI systems is still in its early stages (Halkiopoulos & Gkintoni, 2024). More interdisciplinary research is needed to develop reliable bias audit tools.

These research gaps directly create implementation gaps. Without long-term evidence, investments in this technology may not be sustainable. Without research in diverse contexts, there is a risk of applying ineffective or even harmful solutions. Therefore, the future research agenda must directly aim to fill these gaps to inform more ethical, equitable, and effective implementation practices globally.

CONCLUSION

This systematic literature review confirms that AI-based adaptive assessment has transformative potential for mapping and enhancing the science literacy skills of Generation Z. The evolution of assessment models, from those based on modern psychometrics like Computerized Adaptive Testing (CAT) to those utilizing Generative AI, demonstrates a fundamental shift from summative evaluation to a diagnostic and formative tool integrated into the learning process.

Leading platforms such as Inq-ITS and ALEKS effectively implement these models in practice. Inq-ITS successfully dissects science inquiry skills by analyzing the

processes of 'doing' and 'writing' science, while ALEKS granularly maps students' conceptual mastery through Knowledge Space Theory. Existing empirical evidence consistently shows that interventions using these platforms yield significant improvements in learning outcomes, conceptual understanding, and science process skills.

This success is highly relevant for Generation Z, as the core features of adaptive assessment such as personalization, interactivity, gamification, and instant feedback are closely aligned with their preferences and learning styles as digital natives. By providing an autonomous and engaging learning experience, this technology is capable of effectively enhancing student motivation and engagement.

However, the implementation of this technology is not without challenges. The literature clearly identifies significant barriers related to cost, technical infrastructure, and the need for teachers' professional development to pedagogically utilize analytical data. Furthermore, profound ethical considerations, particularly concerning data privacy, algorithmic bias, and 'black box' accountability, are crucial issues that must be addressed to ensure fair and responsible application.

Future research is recommended to focus on filling existing gaps, especially through longitudinal studies to measure long-term impact, the use of more rigorous research designs to isolate effective variables, exploration in the context of developing countries, and the development of methods for algorithmic bias mitigation. By addressing these challenges, AI-based adaptive assessment can be optimized to become a powerful tool in preparing Generation Z with the science literacy needed in the 21st century.

REFERENCES

Akhmetova, A. I., Sovetkanova, D. M., Komekbayeva, L. K., Abdrakhmanov, A. E., Yessenuly, D., & Serikova, O. S. (2025). A systematic review of artificial intelligence in high school STEM education research. *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(4). https://doi.org/10.29333/ejmste/16222

Alfakihuddin, M. L. B., Surahman, E., & Haryani, F. (2022). The Application of Inquiry Intelligent Tutoring System in Biology Practicum. *Journal of Education Technology*, *6*(4), 634–642. https://doi.org/10.23887/jet.v6i4.48466

Ananda, L. J., Simanihuruk, L., Ratno, S., Zati, V. D. A., & Sembiring, M. M. (2023). Caplaire: Inovasi Model Pembelajaran IPA Berbasis Literasi Sains. *Jurnal Sekolah*, *8*(1), 125. https://doi.org/10.24114/js.v8i1.54017

Attard, C., & Holmes, K. (2022). An exploration of teacher and student perceptions of blended learning in four secondary mathematics classrooms. *Mathematics Education Research Journal*, *34*(4), 719–740. https://doi.org/10.1007/s13394-020-00359-2

Bafadal, R., & Rosyid, F. (2024). Memahami Kebutuhan Belajar Generasi Z melalui Asesmen Personal Berbasis Artificial Intelegence. *Journal of Innovation and Teacher Professionalism*, *3*(1), 182–188. https://doi.org/10.17977/um084v3i12025p182-188

Erlangga, Munir, Septem Riza, L., Piantari, E., Junaeti, E., & Seanaldi Permana, I. (2023). Implementation of the Gamification Concept in the Development of a Learning Management System to Improve Students' Cognitive In Basic Programming Subjects Towards a Smart Learning Environment. *ADI Journal on Recent Innovation (AJRI)*, *5*(1), 43–53. https://doi.org/10.34306/ajri.v5i1.902

Firdaus, F. M., Wahidin, W., Ramdani, D., Hernawati, D., & Badriah, L. (2025). Profil Keterampilan Literasi Digital Peserta Didik Dalam Pembelajaran Ipa. *SCIENCE :*

Jurnal Inovasi Pendidikan Matematika Dan IPA, *4*(4), 697–707. https://doi.org/10.51878/science.v4i4.4252

Gunsaldi, M. S., Guner, E. G., Uckan, M., & Bati, K. (2025). The Impact of Generative AI Applications on Student Learning Outcomes in Science Education: A Systematic Review. *Journal of Education in Science, Environment and Health*, *11*(3), 196–208. https://doi.org/10.55549/jeseh.840

Halkiopoulos, C., & Gkintoni, E. (2024). Leveraging AI in E-Learning: Personalized Learning and Adaptive Assessment through Cognitive Neuropsychology—A Systematic Analysis. *Electronics (Switzerland)*, *13*(18). https://doi.org/10.3390/electronics13183762

Harati, H., Sujo-Montes, L., Tu, C. H., Armfield, S. J. W., & Yen, C. J. (2021). Assessment and learning in knowledge spaces (Aleks) adaptive system impact on students' perception and self-regulated learning skills. *Education Sciences*, *11*(10). https://doi.org/10.3390/educsci11100603

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2018). Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*, *62*(3), 333–353. https://doi.org/10.1080/00313831.2016.1258726

Jalaludin, N. A., Salim, M. H. M., Rasul, M. S., Amin, A. F. M., & Saari, M. A. (2024). The Impact of Virtual Collaboration Tools on 21st-Century Skills, Scientific Process Skills and Scientific Creativity in STEM. *International Journal of Advanced Computer Science and Applications*, *15*(8), 1340–1347. https://doi.org/10.14569/IJACSA.2024.01508130

Ješková, Z., Lukáč, S., Šnajder, Ľ., Guniš, J., Klein, D., & Kireš, M. (2022). Active Learning in STEM Education with Regard to the Development of Inquiry Skills. *Education Sciences*, *12*(10). https://doi.org/10.3390/educsci12100686

Kotsis, K. T. (2025). Integrating Artificial Intelligence for Science Teaching in High School. *LatIA*, *3*. https://doi.org/10.62486/latia202589

OECD. (2023). PISA 2022 Results (Volume I) : The State of Learning and Equity in Education. In *Factsheets: Vol. I*. OECD Publishing. https://doi.org/doi.org/10.1787/53f23881-en.

Okoye, K., Hussein, H., Arrona-Palacios, A., Quintero, H. N., Ortega, L. O. P., Sanchez, A. L., Ortiz, E. A., Escamilla, J., & Hosseini, S. (2023). Impact of digital technologies upon teaching and learning in higher education in Latin America: an outlook on the reach, barriers, and bottlenecks. In *Education and Information Technologies* (Vol. 28, Issue 2). Springer US. https://doi.org/10.1007/s10639-022-11214-1

Papp, T. A. (2017). Gamification Effects on Motivation and Learning: Application to Primary and College Students. *International Journal for Cross-Disciplinary Subjects in Education*, *8*(3), 3193–3201. https://doi.org/10.20533/ijcdse.2042.6364.2017.0428

Patigu, Y. P. W., Rahmah, N., & Zulnuraini, Z. (2024). Literasi Sains dan Digital dalam Pembelajaran IPA. *Journal of Education Research*, *5*(3), 3103–3110. https://doi.org/10.37985/jer.v5i3.1404

Ramadhan, S., Atmazaki, A., Ningsih, A. G., Hayati, Y., Henanggil, M. D. F., Nursaid, N., Rahman, F., & Ghaluh, B. M. (2025). Exploring the Impact of Adaptive Real-Time Quiz Platforms with Differentiated Learning Features on Student Engagement and Learning Outcomes: A Mixed-Methods Approach. *International Journal of Information and Education Technology*, *15*(6), 1261–1276. https://doi.org/10.18178/ijiet.2025.15.6.2329

Villegas-Ch, W., Buenano-Fernandez, D., Navarro, A. M., & Mera-Navarrete, A. (2025). Adaptive intelligent tutoring systems for STEM education: analysis of the learning impact and effectiveness of personalized feedback. *Smart Learning Environments*, *12*(1). https://doi.org/10.1186/s40561-025-00389-y

Yasuda, J. I., Mae, N., Hull, M. M., & Taniguchi, M. A. (2021). Optimizing the length of computerized adaptive testing for the Force Concept Inventory. *Physical Review Physics Education Research*, *17*(1), 10115. https://doi.org/10.1103/PhysRevPhysEducRes.17.010115

Zulfiani, Z., Suwarna, I. P., & Miranto, S. (2018). Science education adaptive learning system as a computer-based science learning with learning style variations. *Journal of Baltic Science Education*, *17*(4), 711–727. https://doi.org/10.33225/jbse/18.17.711