

Transforming Authentic Assessment in Chemistry Learning Through the Collaboration of Merdeka Curriculum and AI Technology

Salma Mardhiyah Nuraini^{1*}, Nahadi²

^{1,2} Department of Chemistry Education, Faculty of Mathematics and Natural Sciences Education, UPI, Indonesia

*Corresponding author. Email: mardhsalma02@upi.edu

ABSTRACT

The study investigated the transformation of authentic assessment in education, which represented a shift from product-based evaluation toward contextual, reflective, and competency-oriented learning. Within the *Merdeka Curriculum* framework, authentic assessment functioned as both an evaluative and developmental tool to cultivate 21st century skills (critical thinking, creativity, collaboration, and communication) while supporting the *Profil Pelajar Pancasila*. However, limited research had examined how Artificial Intelligence (AI) could operationalize authentic assessment principles within this curriculum. An integrative literature review was conducted to analyze thirteen peer-reviewed journal articles published between 2023 and 2026. The inclusion criteria covered three aspects: a focus on authentic assessment in education, the integration of AI or digital technologies, and relevance to 21st-century learning competencies. The synthesis showed three major findings. First, AI enhanced experiential and reflective learning through generative tools and personalized feedback that deepened students' conceptual understanding and inquiry processes. Second, ethical and pedagogical tensions appeared between authenticity and academic integrity due to the increased use of AI in completing authentic tasks. Third, AI fostered innovative assessment models that redefined teacher roles and learner agency, enabling data-informed reflection and adaptive evaluation. The study concluded by proposing a conceptual model for integrating AI-based authentic assessment aligned with the *Merdeka Curriculum*, emphasizing ethical implementation, human-AI collaboration, and scientific literacy as essential components of transformative learning.

Keywords: *artificial intelligence, authentic assessment, Merdeka Curriculum, Profil Pelajar Pancasila.*

1. INTRODUCTION

The rapid advancement of the digital era has prompted a fundamental shift in educational assessment practices, moving away from purely summative evaluation toward more authentic and process-oriented approaches that are capable of capturing students' real competencies. In chemistry education, this shift is particularly crucial due to the dual emphasis on conceptual understanding and experimental skills, which necessitates comprehensive assessment across cognitive, affective, and psychomotor domains (Nahadi & Firman, 2019; Supardi, 2015). Authentic assessment highlights learners' ability to apply scientific knowledge in meaningful contexts that reflect real-world scientific practices, thereby providing a more holistic representation of learning outcomes (Darling-Hammond & Snyder, 2000; Wiggins, 1993).

Within the Indonesian educational context, the Merdeka Curriculum reinforces the importance of authentic assessment as a strategic instrument for nurturing the Profil Pelajar Pancasila, particularly in fostering critical thinking, creativity, collaboration, and communication skills (Sugiri & Priatmoko, 2020). In chemistry learning, these principles are commonly operationalized through project-based laboratory activities, portfolio assessments, and inquiry-oriented evaluation models that encourage active student engagement and reflective learning processes (Pahriah et al., 2025; Setyawarno & Kurniawati, 2018).

Despite its pedagogical potential, the implementation of authentic assessment in chemistry classrooms remains challenging. Previous studies report that limited teacher readiness, time constraints in designing assessment instruments, and varying levels of digital literacy among educators hinder its effective application (Devi & Fatonah, 2025; Hanifah & Irambona, 2019). Alongside these challenges, recent technological developments have introduced artificial intelligence as a promising tool to support assessment practices, particularly through automated scoring systems, adaptive feedback mechanisms, and data-driven learning analytics (Bulut & Beiting-Parrish, 2024; J, 2024). Nevertheless, the integration of AI in educational assessment has also raised concerns related to academic integrity, student originality, and the ethical boundaries of automated evaluation (Chan & Lee, 2025).

Although research on AI in education has grown substantially, a critical gap persists in understanding how AI can be meaningfully aligned with authentic assessment principles within the framework of the Merdeka Curriculum, particularly in the context of chemistry education and scientific literacy development. Existing studies tend to emphasize general AI functionalities or technological affordances (Kofinas et al., 2025; Thanh et al., 2023), while paying limited attention to the human-centered and ethical dimensions required to support authentic and reflective assessment practices (Picasso et al., 2024; Wahidin et al., 2025). Therefore, this integrative literature review synthesizes findings from thirteen international studies published between 2023 and 2026 to examine how the collaboration between AI and the Merdeka Curriculum can transform chemistry assessment into a more adaptive, ethical, and student-centered system.

2. METHODOLOGY

2.1. Research Design

This study employed an Integrative Literature Review (ILR) design to synthesize and critically evaluate research on the transformation of authentic assessment through the integration of Artificial Intelligence (AI) in education. The ILR approach allows the integration of theoretical, empirical, and conceptual studies to produce a comprehensive understanding of emerging educational phenomena (Snyder, 2019; Whitemore &

Knafel, 2005). This design was selected because it facilitates an in-depth examination of how AI-based tools influence authentic assessment design, implementation, and evaluation within 21st-century learning frameworks.

2.2. Sources of Data

The data of this study consisted of thirteen peer-reviewed journal articles published between 2023 and 2026. The selection process followed a systematic screening procedure to ensure transparency, as visualized in the PRISMA flow diagram (see Figure 1) (Page et al., 2021). Articles were purposively selected from reputable databases using the search terms “artificial intelligence,” “authentic assessment,” “learning,” and “education.” The inclusion criteria were as follows:

1. Articles focusing on authentic assessment in educational or learning contexts;
2. Studies integrating AI-based or digital assessment technologies for feedback, evaluation, or learning enhancement; and
3. Research demonstrating relevance to 21st-century competencies, critical thinking, and creative problem-solving.

The final corpus included the following thirteen studies:

(Chan & Lee, 2025; Cole, 2023; Fatima et al., 2024; Kirsanov et al., 2026; Kofinas et al., 2025; Lawrie, 2023; Mokganya & Zitha, 2025; Moreira & Hadjipieris, 2024; Picasso et al., 2024; Salinas-Navarro et al., 2024; Thanh et al., 2023; Wahidin et al., 2025; Yung et al., 2024)

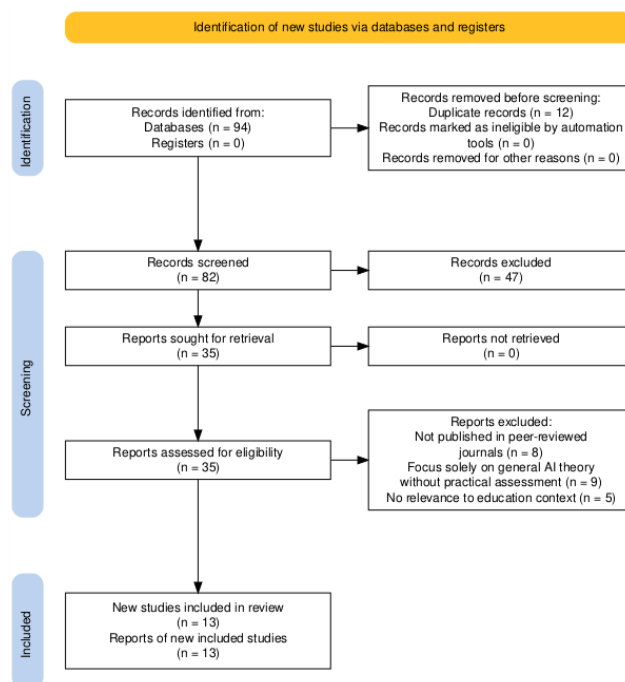


Figure 1. PRISMA Flow Diagram of the Literature Selection Process

2.3. Data Analysis Techniques

Data were analyzed through four systematic stages adapted from integrative review procedures:

1. Identification: reviewing the objectives, designs, and outcomes of each study;
2. Screening and Selection: applying inclusion and exclusion criteria to ensure focus on AI-integrated authentic assessment;
3. Categorization: organizing studies into emerging thematic clusters such as AI-enhanced learning, academic integrity, pedagogical authenticity, and ethical implications; and
4. Synthesis and Interpretation: integrating findings across studies to construct a conceptual framework describing how AI supports authentic assessment in education.

3. RESULT AND DISCUSSION

The synthesis of thirteen peer-reviewed journal articles published between 2023 and 2026 revealed three central findings regarding the transformation of authentic assessment through Artificial Intelligence (AI) integration in educational and scientific contexts.

Table 1. Synthesis of AI-Integrated Authentic Assessment Transformation (2023–2026)

Theme	Transformation & Key Findings	Supporting Sources	Relevance to Merdeka Curriculum & Chemistry
Experiential & Reflective Learning	AI enhances inquiry skills, conceptual understanding, and provides dialogic feedback for deeper engagement.	Salinas-Navarro et al. (2024); Lawrie (2023); Cole (2023)	Supports project-based laboratory activities and reflective scientific inquiry.
Ethical & Integrity Tension	Tension between AI's capability to mimic authentic responses and the need for academic honesty.	Thanh et al. (2023); Kofinas et al. (2025); Kirsanov et al. (2026)	Focuses on developing character and integrity in the Profil Pelajar Pancasila.
Pedagogical Innovation & Agency	Redefining teacher roles through data justice, adaptive feedback, and fostering student self-agency.	Picasso et al. (2024); Moreira & Hadjipieris (2024); Fatima et al. (2024); Chan & Lee (2025); Mokganya & Zitha (2025)	Promotes learner-centeredness and inclusivity in diverse chemistry classrooms.
Future Model: AI-Reflective	A shift toward augmented human reflection and professional identity cultivation.	Wahidin et al. (2025); Yung et al. (2024)	Prepares students for AI-mediated scientific environments.

First, AI acted as a catalyst for experiential and reflective learning. Studies by Salinas-Navarro et al. (2024) and Lawrie (2023) demonstrated that generative AI tools enhanced students' conceptual understanding, inquiry skills, and reflective experimentation. Similarly, Cole (2023) reported that professional conversation-based assessments supported by AI facilitated dialogic feedback and deeper engagement with authentic tasks.

Second, a recurring challenge involved the tension between authenticity, academic integrity, and technological dependence. Thanh et al. (2023) and Kofinas et al. (2025) found that generative AI could reproduce authentic responses, potentially compromising originality and integrity. Kirsanov et al. (2026) identified students' deliberate concealment of AI-generated content in online assessments.

Third, AI fostered pedagogical innovation by redefining teachers roles and promoting learner agency. Picasso et al. (2024) and Moreira and Hadjipieris (2024) showed that data-driven AI analytics supported adaptive feedback and inclusivity through a data justice approach. Fatima et al. (2024) and Chan and Lee (2025) highlighted the potential of AI-based assessments to facilitate collaboration and self-evaluation, while Mokganya and Zitha (2025) emphasized the importance of maintaining scientific interpretation and epistemological integrity in AI-mediated tasks.

Collectively, the reviewed studies demonstrated a shift from AI-assisted assessment toward AI-integrated and AI-reflective models (Wahidin et al., 2025; Yung et al., 2024), where authenticity was redefined through the balance between technological mediation, human judgment, and ethical awareness.

The findings further indicated that AI integration reshaped the epistemological and ethical dimensions of learning evaluation. These transformations aligned with the Merdeka Curriculum's emphasis on critical thinking, creativity, collaboration, and communication (4C) within the Profil Pelajar Pancasila framework.

The first theme suggested that AI-supported tools enabled experiential authenticity, allowing learners to engage in simulated scientific practices mirroring real-world inquiry (Lawrie, 2023; Salinas-Navarro et al., 2024). Such practices enhanced reflective learning cycles by linking experimentation with personalized feedback, consistent with constructivist principles in chemistry education.

The second theme revealed critical ethical and pedagogical challenges. The capacity of generative AI to perform complex tasks blurred the boundaries of originality and authorship (Kirsanov et al., 2026; Thanh et al., 2023). Kofinas et al. (2025) recommended redesigning assessment tasks that foreground process transparency and contextual adaptation. These insights highlighted the need for teacher assessment literacy to ensure fair and human-centered AI utilization.

The third theme underscored a paradigm shift toward human-AI collaboration in assessment design. Picasso et al. (2024) advocated for ethical governance of AI through

fairness and inclusivity principles, while Fatima et al. (2024) and Moreira and Hadjipieris (2024) demonstrated how AI-enhanced analytics could improve formative feedback loops and learner self-agency. In chemistry education, these models were relevant for assessing experimental reasoning and problem-solving skills while maintaining the interpretive nature of scientific inquiry (Lawrie, 2023).

Furthermore, emerging AI-reflective assessment models (Mokganya & Zitha, 2025; Wahidin et al., 2025; Yung et al., 2024) indicated that the future of assessment lies not in automation but in augmented human reflection. Within the Merdeka Curriculum context, this transformation fosters ethically grounded, data-literate, and self-regulated learners capable of navigating AI-mediated learning environments.

Overall, the integrative synthesis confirmed that AI-based authentic assessment can enhance educational equity and innovation while preserving humanistic values. Sustainable transformation depends on balancing technological efficiency with ethical awareness and pedagogical authenticity, ensuring that assessment continues to serve as a process of learning rather than a mere act of measurement.

4. CONCLUSION

This integrative literature review concluded that the integration of Artificial Intelligence (AI) into authentic assessment practices has redefined the nature, purpose, and ethics of educational evaluation, particularly in science and chemistry learning. The synthesis of thirteen international studies published between 2023 and 2026 revealed that AI functions as both a catalyst and a disruptor within the assessment landscape.

First, AI-enabled authentic assessments enhance experiential and reflective learning by simulating real-world inquiry, providing automated feedback, and personalizing student learning trajectories. Second, the growing tension between authenticity and academic integrity underscores the necessity for ethical literacy and assessment transparency among educators and learners alike. Third, AI promotes pedagogical innovation and teacher–learner collaboration, emphasizing assessment as an interactive, data-informed, and human-centered process.

In the context of the Merdeka Curriculum, these insights suggest that AI-driven authentic assessments can support the development of Profil Pelajar Pancasila by fostering critical thinking, creativity, collaboration, and scientific literacy. However, sustainable implementation requires educators to adopt a balanced approach—using AI as a co-assessor rather than a replacement for human judgment.

Suggestions arising from this study emphasize the importance of teacher professional development in AI literacy, ethical assessment design, and digital pedagogical competence. Curriculum developers are encouraged to integrate AI-based tools within national frameworks to enhance feedback quality and equity in chemistry education. Future research should investigate context-specific models of AI-supported

authentic assessment in Indonesian schools, particularly those aligned with the reflective and inquiry-oriented principles of science learning.

ACKNOWLEDGMENTS

The author wishes to express sincere gratitude to the Curriculum Development Study Program, UPI, as the organizer of the 6th International Conference on Educational Research and Development (ICERD) 2025, for providing an inspiring platform to share and discuss innovative research in education. Deep appreciation is extended to Prof. Dr. Nahadi, M.Si., M.Pd., for his invaluable supervision, encouragement, and insightful feedback throughout the preparation of this literature review. The author also gratefully acknowledges the support of the Department of Chemistry Education, UPI, for providing academic guidance and resources that supported the completion of this work.

REFERENCES

- Bulut, O., & Beiting-Parrish, M. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *Chinese/English Journal of Educational Measurement and Evaluation*, 5(3). <https://doi.org/10.59863/MIQL7785>
- Chan, C. K. Y., & Lee, K. K. W. (2025). The balancing act between AI and authenticity in assessment: A case study of secondary school students' use of GenAI in reflective writing. *Computers and Education*, 238(September 2024), 105399. <https://doi.org/10.1016/j.compedu.2025.105399>
- Cole, D. (2023). Authentic Assessment Through Professional Conversations: An AI friendly assessment method? *Journal of Perspectives in Applied Academic Practice*, 11(3), 10–16. <https://doi.org/10.56433/jpaap.v11i3.586>
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5–6), 523–545. [https://doi.org/10.1016/S0742-051X\(00\)00015-9](https://doi.org/10.1016/S0742-051X(00)00015-9)
- Devi, D. R., & Fatolah, S. F. (2025). Obstacles and Challenges in Implementing Authentic Assessment in Elementary Madrasah. *International Journal of Basic Educational Research*, 2(1), 21–25. <https://doi.org/10.14421/ijber.v2i1.8800>
- Fatima, S. S., Sheikh, N. A., & Osama, A. (2024). Authentic assessment in medical education: exploring AI integration and student-as-partners collaboration. *Postgraduate Medical Journal*, 100(1190), 959–967. <https://doi.org/10.1093/postmj/qgae088>
- Hanifah, M., & Irambona, A. (2019). Authentic assessment: Evaluation and its application in science learning. *Psychology, Evaluation, and Technology in Educational Research*, 1(2), 81. <https://doi.org/10.33292/petier.v1i2.4>
- J, K. S. (2024). the Role of Artificial Intelligence in Educational Process of University. *Eurasian Experiment Journal of Scientific and Applied Research*, 5(2), 44–49.
- Kirsanov, O., Kushwah, L., & Selvaretnam, G. (2026). Beyond Detection: How Students Use—and Hide—AI in Online Assessments and What Authentic Tasks Can Do About It. *Journal of Academic Ethics*, 24(1), 1–24. <https://doi.org/10.1007/s10805-025-09691-3>
- Kofinas, A. K., Tsay, C. H. H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56(6), 2522–2549. <https://doi.org/10.1111/bjet.13585>
- Lawrie, G. (2023). Establishing a delicate balance in the relationship between artificial intelligence and authentic assessment in student learning. *Chemistry Education Research and Practice*, 24(2), 392–393. <https://doi.org/10.1039/d3rp90003g>
- Mokganya, M. G., & Zitha, I. (2025). Epistemologies of Authentic Assessments Amidst the AI Advancement: An Evaluation of the Divergence Between Traditional and AI-Generated Scientific Illustrations. *Journal of Education and Learning Technology*, November, 996–1009. <https://doi.org/10.38159/jelt.20256105>

- Moreira, D., & Hadjipieris, P. (2024). Leveraging AI for Innovative Authentic Assessment and Enhanced Learning in Political Science. *Transformative Dialogues: Teaching & Learning Journal*, 17(2), 196–212.
<https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=eue&AN=182130558&site=eds-live&scope=site>
- Nahadi, & Firman, H. (2019). *Asesmen: Pembelajaran Kimia*. UPI Press.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, n71.
<https://doi.org/10.1136/bmj.n71>
- Pahriah, P., Ratnaya, I. G., Parwata, I. G. L. A., & Budhyani, I. D. A. M. (2025). Authentic Assessment in Chemistry Education at Senior High Schools and Higher Education: A Systematic Literature Review (2015–2025). *Path of Science*, 11(7), 1013. <https://doi.org/10.22178/pos.120-10>
- Picasso, F., Atenas, J., Havemann, L., & Serbati, A. (2024). Advancing Critical Data and AI Literacies Through Authentic and Real-World Assessment Design Using a Data Justice Approach. *Open Praxis*, 16(3), 291–310. <https://doi.org/10.55982/openpraxis.16.3.667>
- Salinas-Navarro, D. E., Vilalta-Perdomo, E., Michel-Villarreal, R., & Montesinos, L. (2024). Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment. *Education Sciences*, 14(1). <https://doi.org/10.3390/educsci14010083>
- Setyawarno, D., & Kurniawati, A. (2018). Implementation of Authentic Assessment in Science Learning at Indonesian Schools. *Journal of Science Education Research*, 2(2), 47–55.
<https://doi.org/10.21831/jser.v2i2.22468>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sugiri, W. A., & Priatmoko, S. (2020). Perspektif asesmen autentik sebagai alat evaluasi dalam merdeka belajar. *At-Thullab: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, 4(1), 53.
<https://doi.org/10.30736/atl.v4i1.119>
- Supardi. (2015). *Penilaian Autentik: Pembelajaran Afektif, Kognitif, dan Psikomotor: Konsep dan Aplikasi*. Rajawali Pers.
- Thanh, N. B., Diem, T. H. V., Minh, N. N., Thi, T. T. P., Hieu, T. T., & Son, H. X. (2023). Race with the machines: Assessing the capability of generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*, 202(5), 39.
- Wahidin, Soelestiono, S., & Tanjung, R. (2025). Education Beyond AI: Building Integrity Through Authentic Assessment. *ENDLESS: International Journal of Futures Studies*, 8(2), 1–11. <http://endless-journal.com/index.php/endless>
- Whittemore, R., & Knaf, K. (2005). The integrative review: updated methodology. *Journal of Advanced Nursing*, 52(5), 546–553. <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. Jossey-Bass.
- Yung, R., Moyle, B., & Yang, E. C. L. (2024). Mitigating the Impact of ChatGPT: Cultivating Professional Identity Through Interactive Oral Authentic Assessments. *Journal of Hospitality and Tourism Education*, 00(00), 1–7. <https://doi.org/10.1080/10963758.2024.2331516>