JPPIPA 12(6) (2026)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Advancing SDG 4 Quality Education: Needs Analysis and Automated PjBL Essay Assessment Design for *Projek IPAS* in Vocational High Schools

Sayid Juni Riauan[1,2*], Darmansyah[2], Zelhendri Zen[2], Fetri Yeni J[2]

[1] SMK Negeri 1 Batam, Batam, Indonesia.
[2] Educational Technology, Postgraduate, Universitas Negeri Padang, Padang, Indonesia.

Received: April 29, 2026
Revised: May 23, 2026
Accepted: June 25, 2026
Published: June 30, 2026

Corresponding Author:
Sayid Juni Riauan
sayidjunior@smkn1batam.sch.id

DOI: 10.29303/jppipa.v12i6.15088

Abstract: The implementation of Project-Based Learning (PjBL) in vocational high schools requires authentic descriptive essay assessments, yet manual grading frequently induces severe teacher fatigue and subjectivity. This study aims to analyze teachers' assessment needs and design a conceptual model for automated PjBL essay grading specifically for the *Projek IPAS* subject. Utilizing the initial phases of the ADDIE research and development framework, this descriptive study employed preliminary teacher interviews, field observations, and literature reviews. The needs analysis revealed a critical demand for an efficient evaluation system to mitigate educators' cognitive load. Consequently, a conceptual architecture integrating Word Count, Similarity Index, and Keyword Matching was designed strictly as an administrative and technical pre-screening mechanism, while the Timestamp parameter was relegated solely to task management. Ultimately, this automated conceptual model is projected to serve as a cognitive offloading mechanism to substantially reduce administrative burdens and potentially mitigate grading bias, empowering science educators to reallocate their energy toward authentic substantive assessment.

Keywords: Automated assessment; Needs analysis; Project-based learning; *Projek IPAS*; Teacher fatigue

Introduction

The implementation of the Merdeka Curriculum in Indonesian vocational high schools (SMK) has fundamentally transformed the pedagogical paradigm, shifting from rigid summative testing to authentic, project-based assessments (Sari et al., 2025; Yunita & Widodo, 2023). This transformation is highly evident in the Projek IPAS (Applied Natural and Social Sciences Project) subject (BSKAP Kemendikdasmen, 2025). As a core component of vocational education, Projek IPAS demands students to integrate scientific literacy, empirical observation, and social perspectives to solve real-world problems (Alfiah et al., 2025; Arthur et al., 2021; Nadir et al., 2022; Pratama et al., 2025). To accurately measure students' conceptual understanding, critical thinking, and scientific reasoning, educators rely heavily on authentic assessments, primarily through descriptive essays and reflective project narratives (Rustam & Priyanto, 2022). In science education, these written reflections are paramount as they capture the nuances of a student's scientific inquiry process elements often entirely missed by standardized multiple-choice questions. Effective assessment instruments within the PjBL framework are crucial not only for evaluation but also for continuously stimulating students' critical and creative thinking skills during scientific inquiries (Asiah & Festiyed, 2024, 2024; Rosidin et al., 2023).

In the context of science education, descriptive essays serve as a window into a student's cognitive structure (Irdalisa et al., 2024; Nisah et al., 2024; Sidiki et al., 2024; Yıldırır & Kızmaz, 2025; Zakiyah et al., 2024). Unlike multiple-choice questions that often promote rote memorization, essays require students to synthesize

How to Cite:
Riauan, S. J., Darmansyah, Zen, Z., & J, F. Y. (2026). Advancing SDG 4 Quality Education: Needs Analysis and Automated PjBL Essay Assessment Design for Projek IPAS in Vocational High Schools. *Jurnal Penelitian Pendidikan IPA*, *12*(6), 291–300. https://doi.org/10.29303/jppipa.v12i6.15088

scientific concepts, demonstrate causal reasoning, and apply theoretical knowledge to practical project phenomena (Maryani et al., 2021). For instance, in Projek IPAS, when a student describes the process of energy transformation in a project, they are not just reporting data; they are manifesting their level of scientific literacy and conceptual mastery.

The pedagogical value of descriptive essays lies in their ability to demand generative cognitive processing (S. Kim et al., 2021; Liu et al., 2024). While seminal research by Mueller and Oppenheimer (Mueller & Oppenheimer, 2014) suggested that longhand writing results in superior conceptual understanding due to deeper information reframing, recent replications have provided a more nuanced perspective. Findings by Morehead et al. (2019) and Urry et al. (2021) indicate that the performance gap between pen and keyboard may be less significant than previously assumed, provided that the writing task encourages generative synthesis rather than passive transcription (J. Kim et al., 2026). This suggests that digital assessment environments can maintain cognitive integrity if they are specifically designed to evaluate conceptual depth. Furthermore, since motor-sensory engagement remains fundamentally linked to literacy and conceptual representations (Ibaibarriaga et al., 2025; Wiley & Rapp, 2021), the development of automated systems like SAUM must bridge the gap between digital efficiency and the preservation of high-level cognitive engagement.

Despite its pedagogical advantages, the implementation of descriptive essay assessments in Projek IPAS has introduced a severe logistical and administrative bottleneck (Daud et al., 2023; Rasyid et al., 2024). Based on preliminary surveys and structured interviews with vocational science teachers, educators are confronted with an overwhelming evaluation workload. A single educator typically handles four to six classes, amounting to approximately 140 to 210 student essays per week. When evaluated manually, requiring 5 to 7 minutes per essay to provide meaningful formative feedback, the grading process consumes up to 24.5 hours weekly. This labor-intensive reality inevitably triggers severe teacher fatigue (human error). Cognitive exhaustion during manual grading significantly elevates the risk of subjectivity, inconsistency, and the halo effect (Chen & Pan, 2022; Doyle et al., 2024; Gray et al., 2025; Mahshanian & Shahnazari, 2020; Palermo, 2022), where teachers may inadvertently award uniform grades or abandon detailed feedback altogether due to sheer mental exhaustion. Consequently, the formative goal of the assessment fails, and the authenticity of students' original reflections is often compromised by the educator's fatigue.

To mitigate this systemic inefficiency, the integration of educational technology through Automated Essay Scoring (AES) systems offers a promising solution (Banihashem et al., 2025; Nie, 2025; Plasencia-Calaña, 2025). Furthermore, the integration of digital technology in science evaluation is no longer optional; it is a necessity to accustom students to technology-enhanced environments (Fajri et al., 2024; Mellyzar et al., 2024) while utilizing efficient systems to pre-screen their foundational concept mastery before educators perform substantive evaluation (Hasnawati et al., 2022; Yoo et al., 2024). While various AES models have been developed globally, the majority are heavily optimized for language arts or generic humanities subjects, utilizing complex Natural Language Processing (NLP) that is computationally expensive and difficult to deploy in resource-constrained vocational schools (Ayaan & Ng, 2025; Gao et al., 2024; Rumaisa et al., 2021; Shaik et al., 2022; Sun & Wang, 2024). While advanced Large Language Models (LLMs) and semantic NLP models offer superior capabilities for assessing critical thinking, their implementation is currently unfeasible in resource-constrained vocational schools due to high computational costs (Yan et al., 2024; Yigci et al., 2025). Therefore, the SAUM model deliberately employs foundational rule-based algorithms (such as Keyword Matching and Word Thresholds). These are utilized not because they are more accurate than LLMs, but purely due to infrastructural limitations, serving strictly as an initial eligibility pre-screening mechanism. Furthermore, there is a notable research gap concerning automated assessment models specifically tailored for vocational science projects (Siregar et al., 2025), which require multi-parametric evaluation criteria rather than mere grammatical correctness. An ideal automated support system for Projek IPAS must clearly distinguish between the essay's cognitive pre-screening components (such as comprehensiveness and concept relevance) and its administrative management features (such as originality checking and student punctuality) (Fitriani et al., 2022; Hardison et al., 2025; Yolanda & Yohandri, 2025).

Addressing this critical gap, this study aims to conduct an in-depth needs analysis and formulate a conceptual design for a localized, highly efficient Automated PjBL Essay Assessment Model (locally termed as Sistem Analisis Uraian Murid or SAUM). By acting as a cognitive offloading mechanism for science educators, this model seeks to automate the most exhausting aspects of narrative evaluation (Yang et al., 2025; Yusuf et al., 2025). This research details the theoretical architecture of the system, providing a robust technological framework to restore assessment objectivity, reduce teacher workload, and ultimately

enhance the quality of science education in vocational high schools.

The novelty of this research lies in its specific adaptation of automated assessment technology to the unique constraints of vocational science projects, acting as a technical pre-screening tool rather than a generic language evaluator. This research is critically important as it offers a low-computational, practical solution to a severe logistical bottleneck, ensuring the sustainability of authentic assessment in the Merdeka Curriculum without causing mass educator burnout.

Method

This research represents the foundational stages of a comprehensive Research and Development (R&D) initiative, employing the highly structured ADDIE (Analysis, Design, Development, Implementation, and Evaluation) instructional design framework (Chotimah et al., 2024; Darmansyah, 2023; Khairunnisa & Darmansyah, 2022; Putri et al., 2025; Rahmadayani & Darmansyah, 2023).

Figure 1. Diagram Model ADDIE (Crawford: 2004 – Adopted by Darmansyah)

This study utilizes the initial phases (Analysis and Design) of the ADDIE research and development framework, employing a qualitative approach supported by quantitative descriptive data. Given the limited sample size of three Science Project teachers at SMK Negeri 1 Batam, the research refrains from broad quantitative generalizations and instead relies on in-depth qualitative interviews to explore teacher workload and assessment needs, supported by descriptive arithmetic calculations of their grading duration.

During the Analysis phase, data were gathered from 3 *Projek IPAS* educators at SMK Negeri 1 Batam through structured interviews and workload observation logs. The observation specifically tracked the duration required to grade project-based essays for 420 students across various vocational majors. Subsequently, in the Design phase, these data were translated into a conceptual system architecture (SAUM). These four computational parameters (Word Count, Similarity Index, Timestamp, and Keyword Matching) were explicitly designed as an administrative and technical pre-screening mechanism to filter technical aspects of the essays before teachers conduct the actual, substantive authentic assessments based on the 2025 IPAS Project Guidelines Rubric. Data analysis was performed using descriptive statistics for quantitative workload data and thematic analysis for qualitative teacher interviews to ensure the system design directly addresses field-level bottlenecks (Machado et al., 2025).

The research was contextually situated at SMK Negeri 1 Batam, a prominent vocational high school in Indonesia, focusing on educators who actively teach the *Projek IPAS* subject. The data collection strategy during the Analysis phase was multifaceted, integrating three primary instruments into a cohesive methodology. Initially, direct field observations and workload tracking were conducted to map the conventional manual grading workflow. This process involved quantifying the average time required by teachers to evaluate a single scientific narrative and calculating their total weekly cognitive load. Subsequently, semi-structured interviews were held with vocational science teachers to identify specific pedagogical pain points. These discussions focused on the challenges of maintaining grading objectivity, the frequency of teacher fatigue, and the specific parameters educators deem essential for evaluating science project reflections. Finally, a comprehensive systematic review of existing literature on Automated Essay Scoring (AES) and text-mining algorithms was conducted to select the most computationally efficient and pedagogically appropriate parameters for the proposed model.

Data acquired from the analysis phase underwent a rigorous thematic and descriptive analysis. Qualitative insights from interviews were coded to define the system's functional requirements, while workload data were descriptively analyzed to establish the urgency of the automation. Subsequently, during the Design phase, these synthesized needs were translated into a technical blueprint. A conceptual matrix and a procedural system architecture were developed, integrating four primary algorithmic parameters (Word Count, Similarity Index, Timestamp, and Keyword Matching) to address the specific criteria of *Projek IPAS* assessments.

Result and Discussion

Analysis Phase: Mapping the Assessment Crisis

The Halo Effect and Cognitive Fatigue are silent disruptors in science assessment. When a teacher evaluates the 100th project report, their mental grading rubric unconsciously shifts, leading to grading drift. By implementing SAUM as a cognitive offloading tool, we eliminate this physiological bias, ensuring that the first

student and the last student in a cohort are administratively pre-screened with the same level of algorithmic precision and objectivity (Kumar, 2025).

The SAUM model aligns with the Deep Learning framework (*Pembelajaran Mendalam*) mandated by the Indonesian Ministry of Basic and Secondary Education (Kementerian Pendidikan Dasar dan Menengah Republik Indonesia, 2025). By automating the technical pre-screening of reflective essays, SAUM provides educators the crucial pedagogical time required to facilitate the substantive Reflecting (*Merefleksi*) phase of the student experience, ensuring that students do not just know scientific facts, but can internalize and connect them to real-world project contexts through disciplined narrative expression.

The initial phase of the ADDIE framework revealed a significant discrepancy between the pedagogical ideal of the Merdeka Curriculum and the practical reality of assessment at SMK Negeri 1 Batam. Based on the workload tracking, it was found that the authenticity of PjBL projects in *Projek IPAS* has paradoxically created a logistical nightmare for educators (Marentek, 2025; Napitupulu et al., 2023; Arafah et al., 2026). The quantitative data regarding this administrative burden is detailed in Table 1.

Table 1. Workload and Time Allocation Analysis for Manual PjBL Essay Grading

Assessment Variables	Quantitative Data/ Value
Average classes handled by one teacher	4 - 6 Classes
Average students per class	35 Students
Total essay assignments per week	140 - 210 Essays
Average manual grading time per essay	5 - 7 Minutes
Total grading time per week	11.60 - 24.50 Hours

The fluctuation in the grading duration, specifically the lower limit of 11.60 hours, reflects the condition of educators assigned to fewer classes or smaller student cohorts. However, even this minimum duration represents a substantial administrative burden, while the upper limit of up to 24.5 hours unequivocally corroborates the severe fatigue crisis highlighted initially.

The data in Table 1 suggests that a teacher may spend nearly 60% of their weekly working hours solely on grading if they adhere to the minimum feedback standards. Interviews with educators identified that this excessive workload leads to a plateau of subjectivity, where after grading the 50th essay, the teacher's ability to distinguish between Good and Excellent scientific reasoning diminishes significantly due to cognitive exhaustion. This fatigue-induced bias directly undermines the objectivity of the PjBL model, confirming that the current manual system is no longer sustainable for large-scale vocational classes.

Design Phase: The SAUM Conceptual Framework

In response to the identified needs, this study proposes the *Sistem Analisis Uraian Murid (SAUM)*, a conceptual model designed to automate the evaluation of project narratives. Unlike general-purpose AES systems, SAUM is context-aware, focusing on the specific requirements of science-based projects. The system's architecture, as illustrated in Figure 1, operates by decomposing a complex manual rubric into four measurable algorithmic parameters.

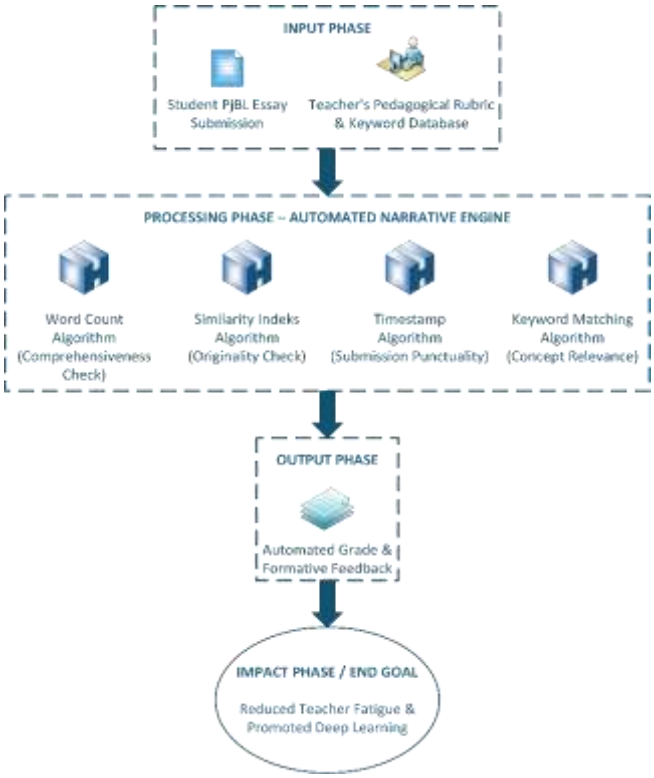

Figure 2. Conceptual architecture of the automated PjBL essay assessment model

Technical Description of the SAUM Architecture

The procedural architecture of the *Sistem Analisis Uraian Murid (SAUM)*, as depicted in Figure 1, represents a systematic filtering-and-scoring workflow designed to digitize the teacher's cognitive process in evaluation. The system operates through a sequential analysis engine that processes raw student narratives into structured performance data.

The workflow begins with the Input Phase, where students submit their project narratives via a digital interface. Once submitted, the SAUM Analysis Engine initiates a four-layered technical pre-screening process. First, the Temporal Filter (Timestamp) captures the submission's temporal data purely as an administrative

task management feature to log assignment punctuality. Following this, the Elaboration Filter (Volume/Word Count) screens the narrative for its quantitative volume. Rather than serving as an absolute measure of cognitive depth, a narrative that falls below a predetermined word threshold is flagged to indicate a potential lack of structural elaboration in the student's report. Next, the Conceptual Presence Filter (Keyword Matching) scans for pre-defined scientific anchor words related to the specific *Projek IPAS* theme. This acts strictly as an early technical indicator of foundational concept presence before educators conduct substantive assessments. Finally, the Authenticity Verification Filter (Similarity Index) employs a detection algorithm to safeguard academic integrity. Given the collaborative nature of PjBL, this administrative filter ensures that the reflective essay remains an individual expression of thought, free from excessive unoriginal content.

The output of these four filters is synthesized into a comprehensive Analysis Result and Grading Recommendation. Crucially, the SAUM architecture maintains a Human in the Loop approach. The system does not unilaterally decide the final grade; instead, it presents the analyzed data on the Teacher Dashboard. This allows the educator to perform a final verification, exercising professional judgment based on the algorithmic insights provided. This synergy between algorithmic precision and human pedagogical wisdom effectively mitigates the risk of grading bias while significantly reducing the educator's cognitive load (Werang et al., 2025).

Algorithmic Synchronization and Parameterization

The core innovation of the SAUM model lies in its synchronization between traditional pedagogical workflows and computational parameters. Rather than serving as direct proxies for scientific quality, these four integrated algorithms function as a technical pre-screening mechanism. First, Volume (Word Count) acts as an indicator of structural elaboration. The system uses a minimum word-threshold algorithm to flag overly brief answers (e.g., responses with less than 20 words), serving as an initial metric for text quantity rather than an absolute measure of cognitive depth (Gong, 2022). Second, the Similarity Index is employed to verify authenticity. To preserve the integrity of PjBL, SAUM incorporates a similarity detection algorithm to ensure that reflections are not merely copied from external sources or peers (Gabatin et al., 2023). Third, the Timestamp parameter tracks project discipline. By capturing submission latency (Wang et al., 2022), it operates strictly as an administrative and behavioral metric that is evaluated entirely separate from the essay's scientific cognitive score. Finally, Keyword Matching is utilized as a foundational text-search tool to detect the initial presence of scientific concepts. Acknowledging its technological limitations, this algorithm acts as an early indicator of thematic relevance rather than a comprehensive measure of critical thinking or deep conceptual accuracy (Novianti et al., 2023). While the anchor words are derived from the 2025 IPAS Guidelines to keep narratives grounded, the system effectively delegates the deeper evaluation of scientific reasoning, critical thinking, and communication competencies back to the human educators (Musahrain et al., 2024).

Restoring the Teacher's Role as a Facilitator

The findings of this study suggest that the SAUM model is projected to provide a vital cognitive offloading mechanism (Peng & Yeh, 2025). By automating the repetitive, data driven aspects of grading (word counts, plagiarism checks, and deadline tracking), the system restores the teacher's capacity to focus on high-level pedagogical tasks (Gnanaprakasam & Lourdusamy, 2024). In the context of vocational science education, this technological intervention is not intended to replace the teacher and theoretically aims to cleanse the assessment process from human error factors such as fatigue and boredom. This alignment between technology and pedagogy is essential for the successful long-term implementation of the Merdeka Curriculum in vocational science education.

This theoretical projection is strongly aligned with recent empirical evidence demonstrating that grading under a high cognitive load significantly exacerbates grading subjectivity and triggers subconscious teacher bias (Doyle et al., 2024). By offloading foundational technical screening to an automated framework like SAUM, educators can effectively mitigate the cognitive fatigue that compromises assessment objectivity. Furthermore, utilizing computational tools to handle repetitive administrative evaluations creates the necessary pedagogical space, allowing teachers to shift their focus from time-consuming technical filtering toward deeper, substantive formative feedback and highly contextualized student interactions (Hopfenbeck et al., 2023).

Conclusion

The implementation of authentic descriptive essay assessments for the *Projek IPAS* subject in vocational high schools currently induces severe teacher fatigue, demanding between 11.60 and 24.50 hours per week from a single educator. To address this crisis, this study designed the conceptual architecture of the SAUM model by synchronizing pedagogical assessment workflows with four measurable pre-screening parameters: Word Count for structural elaboration,

Similarity Index for authenticity, Timestamp for compliance, and Keyword Matching for foundational concept presence. By automating these technical evaluations, this conceptual model is projected to mitigate human-error factors such as fatigue and bias. Furthermore, the development of SAUM transcends mere administrative efficiency; by automating technical screening, it allows educators the cognitive space required to implement the Deep Learning framework established in KepMen 126/P/2025 and conduct substantive assessments. For future research, it is highly recommended to advance this conceptual model into the practical development and implementation phases to empirically test the system's efficacy, as well as to explore the integration of more advanced semantic NLP approaches to enhance cognitive evaluation.

Acknowledgments
The authors would like to express their deepest gratitude to the principal, the management team, and the *Projek IPAS* (Science) educators at SMK Negeri 1 Batam for their invaluable facilitation and active participation during the field observations and data collection phases. The authors also extend their appreciation to the Educational Technology Master's Program, Universitas Negeri Padang, for the continuous academic support and guidance throughout this research.

Author Contributions
Conceptualization, S.J.R. and D.; methodology, S.J.R. and D.; validation, D., Z.Z., and F.Y.J.; formal analysis, S.J.R.; investigation (field observation), S.J.R.; writing—original draft preparation, S.J.R.; writing—review and editing, D., Z.Z., and F.Y.J.; supervision, D., Z.Z., and F.Y.J. All authors have read and agreed to the published version of the manuscript.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

References

Alfiah, M. H., Bramastia, & Sukarmin. (2025). Literature Review: The Impact of Project Based Learning on Scientific Literacy in Secondary Education. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 1–9. https://doi.org/10.29303/jppipa.v11i1.8616

Arafah, K., & Palloan, P. (2026). Development of Project-Based Physics Teaching Modules to Improve Students' Science Process Skills. *Jurnal Penelitian Pendidikan IPA*, *12*(1), 723–734. https://doi.org/10.29303/jppipa.v12i1.14069

Arthur, R., Daryati, Maulana, A., Febiansyah, R., & Kidung, K. (2021). Scientific literacy of vocational school students in building construction. *Journal of Physics: Conference Series*, *1833*(1), 012036. https://doi.org/10.1088/1742-6596/1833/1/012036

Asiah, S., & Festiyed. (2024). Development of Diagnostic, Formative and Summative Assessment Instruments in the PjBL Model to Stimulate Students' Critical and Creative Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, *10*(10), 7486–7492. https://doi.org/10.29303/jppipa.v10i10.8757

Ayaan, A., & Ng, K.-W. (2025). Automated grading using natural language processing and semantic analysis. *MethodsX*, *14*, 103395. https://doi.org/10.1016/j.mex.2025.103395

Banihashem, S. K., Gašević, D., Noroozi, O., Jarodzka, H., Joosten-ten Brinke, D., & Drachsler, H. (2025). Optimizing Formative Assessment with Learning Analytics. *Review of Educational Research*, 00346543251370753. https://doi.org/10.3102/00346543251370753

BSKAP Kemendikdasmen. (2025). *Panduan Mata Pelajaran Projek Ilmu Pengetahuan Alam Dan Sosial (PROJEK IPAS)*. Kemendikdasmen RI. Retrieved from https://kurikulum.kemendikdasmen.go.id/file/panduan/dokumen/01.%20Final%20Panduan%20Mata%20Pelajaran%20Projek%20IPAS_SMK_30_09_2025.pdf

Chen, H., & Pan, J. (2022). Computer or human: A comparative study of automated evaluation scoring and instructors' feedback on Chinese college students' English writing. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 34. https://doi.org/10.1186/s40862-022-00171-4

Chotimah, A. N. I., Sariyatun, S., & Ardianto, D. T. (2024). Development of Virtual Tour Media of Sambisari Temple as a History Learning Media in High Schools. *JTP - Jurnal Teknologi Pendidikan*, *26*(3), 1008–1023. https://doi.org/10.21009/jtp.v26i3.50837

Daud, A., Azhar, F., Isjoni, I., & Chowdhury, R. (2023). Complexities of Authentic Assessment Implementation in English Learning at Rural Areas-Based High Schools. *International Journal of Language Education*. https://doi.org/10.26858/ijole.v7i3.41345

Doyle, L., Harris, P. R., & Easterbrook, M. J. (2024). Quality and quantity: How contexts influence the emergence of teacher bias. *Social Psychology of Education*, *27*(4), 1855–1882. https://doi.org/10.1007/s11218-023-09882-z

Fajri, N., Sriyati, S., & Rochintaniawati, D. (2024). Global Research Trends of Digital Learning Media in Science Education: A Bibliometric Analysis. *Jurnal*

Penelitian Pendidikan IPA, *10*(1), 1–11. https://doi.org/10.29303/jppipa.v10i1.6248

Fitriani, H., Samsuri, T., Rachmadiarti, F., & Raharjo, R. (2022). Characteristics of Evaluation-Process Biology Learning Tools Based on Conceptual Problem-Based Learning Models to Train Critical Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, *8*(1), 269–276. https://doi.org/10.29303/jppipa.v8i1.1168

Gabatin, L. A., Loquias, C. M. M., Redondo, M. M., & Fajardo, D. F. (2023). "Ctrl + C, Ctrl + V": Academic Integrity in Online Language Learning. *Dialektika: Jurnal Bahasa, Sastra Dan Budaya*, *10*(2), 179–204. (NA). https://doi.org/10.33541/dia.v10i2.5392

Gao, R., Merzdorf, H. E., Anwar, S., Hipwell, M. C., & Srinivasa, A. R. (2024). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers and Education: Artificial Intelligence*, *6*, 100206. https://doi.org/10.1016/j.caeai.2024.100206

Gnanaprakasam, J., & Lourdusamy, R. (2024). *The Role of AI in Automating Grading: Enhancing Feedback and Efficiency*. IntechOpen. https://doi.org/10.5772/intechopen.1005025

Gong, W. (2022). An Empirical Study on the Correlation between Critical Thinking Ability and L2 Argumentative Writing Competence. *Frontiers in Educational Research*, *5*(2). https://doi.org/10.25236/FER.2022.050217

Gray, A., Rahat, A., Lindsay, S., Pearson, J., & Crick, T. (2025). *Rendering Transparency to Ranking in Educational Assessment via Bayesian Comparative Judgement*. arXiv. https://doi.org/10.48550/arXiv.2503.15549

Hardison, H., Mukhaiyar, R., Muskhir, M., Waskito, W., & Luthfi, A. (2025). Designing a Web-Based Internship Information System for Vocational High Schools. *Jurnal Penelitian Pendidikan IPA*, *11*(2), 642–653. https://doi.org/10.29303/jppipa.v11i2.10227

Hasnawati, H., Syazali, M., & Widodo, A. (2022). Development of live worksheets assisted diagnostic assessment instruments to measure understanding of science concepts prospective elementary school teachers. *Jurnal Pijar MIPA*, *17*(6), 743–747. https://doi.org/10.29303/jpm.v17i6.4217

Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: A perspective article. *Frontiers in Education*, *8*, 1270700. https://doi.org/10.3389/feduc.2023.1270700

Ibaibarriaga, G., Acha, J., & Perea, M. (2025). The impact of handwriting and typing practice in children's letter and word learning: Implications for literacy development. *Journal of Experimental Child Psychology*, *253*, 106195. https://doi.org/10.1016/j.jecp.2025.106195

Irdalisa, I., Akbar, B., Fuadi, T. M., Maesaroh, M., & Kartikawati, E. (2024). Ricosre Model with Question Formulation Technique (QFT): Enhancing Students' Higher Order Thinking Skills (HOTS) and Science Literacy. *Jurnal Penelitian Pendidikan IPA*, *10*(3), 1175–1178. https://doi.org/10.29303/jppipa.v10i3.6764

Kementerian Pendidikan Dasar dan Menengah Republik Indonesia. (2025). *Keputusan Menteri Pendidikan Dasar Dan Menengah Republik Indonesia Nomor 126/P/2025 Tentang Pedoman Implementasi Pembelajaran Mendalam Pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, Dan Jenjang Pendidikan Menengah*. Retrieved from https://kurikulum.kemendikdasmen.go.id/file/1758017373_manage_file.pdf

Khairunnisa, K., & Darmansyah, D. (2022). Pengembangan Media Video Animasi pada Metode Inkuiri untuk Pembelajaran IPS Kelas IX SMP. *FONDATIA, 6*(4), 972-986. Retrieved from https://ejournal.stitpn.ac.id/index.php/fondatia/article/view/2332

Kim, J., Lee, S.-S., Detrick, R., Wang, J., & Li, N. (2026). Students-Generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*, *38*(1), 504–525. https://doi.org/10.1007/s12528-025-09444-6

Kim, S., Yang, J. W., Lim, J., Lee, S., Ihm, J., & Park, J. (2021). The impact of writing on academic performance for medical students. *BMC Medical Education*, *21*(1), 61. https://doi.org/10.1186/s12909-021-02485-2

Kumar, R. (2025). Transforming student evaluation and feedback through AI-driven automated assessment. *IJFMR - International Journal For Multidisciplinary Research*, *7*(2). https://doi.org/10.36948/ijfmr.2025.v07i02.42212

Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, *211*, 104977. https://doi.org/10.1016/j.compedu.2023.104977

Machado, A., Tenório, K., Santos, M. M., Barros, A. P., Rodrigues, L., Mello, R. F., Paiva, R., & Dermeval, D. (2025). Workload perception in educational resource recommendation supported by artificial intelligence: A controlled experiment with teachers. *Smart Learning Environments*, *12*(1), 20. https://doi.org/10.1186/s40561-025-00373-6

Mahshanian, A., & Shahnazari, M. (2020). The effect of raters fatigue on scoring EFL writing tasks.

Indonesian Journal of Applied Linguistics, *10*(1), 1–13. https://doi.org/10.17509/ijal.v10i1.24956

Marentek, L. K. M. (2025). Teachers' Challenges in Implementing Project-Based Learning in Teaching Science in Inclusive Schools at Tomohon City North Sulawesi. *Jurnal Penelitian Pendidikan IPA*, *11*(8), 850–854. https://doi.org/10.29303/jppipa.v11i8.12377

Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitrianawati, M. (2021). HOTs Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-order Thinking Skills of Prospective Teachers. *Turkish Journal of Science Education*, 4. https://doi.org/10.36681/tused.2021.97

Mellyzar, M., Andriani, R., Lukman, I. R., Muttakin, M., Wati, E., Pasaribu, A. I., & Fadli, M. R. (2024). Digital Assessment Tools: As a Media for Assesing High School Science Learning. *Jurnal Penelitian Pendidikan IPA*, *10*(3), 1362–1374. https://doi.org/10.29303/jppipa.v10i3.6416

Morehead, K., Dunlosky, J., & Rawson, K. (2019). How Much Mightier Is the Pen than the Keyboard for Note-Taking? A Replication and Extension of Mueller and Oppenheimer (2014). *Educational Psychology Review*, *31*. https://doi.org/10.1007/s10648-019-09468-2

Mueller, P. A., & Oppenheimer, D. M. (2014). The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking. *Psychological Science*, *25*(6), 1159–1168. https://doi.org/10.1177/0956797614524581

Musahrain, M., Ainurrahmi, A., Ferniawan, F., & Sabrina, A. (2024). Analisis Kemampuan Berpikir Kritis pada Mata Pelajaran IPA SMP Kelas IX di Kabupaten Sumbawa. *Jurnal Pendidikan Ilmu Pengetahuan Alam (JP-IPA)*, *5*(2), 152–159. https://doi.org/10.56842/jp-ipa.v5i2.439

Nadir, M., Arthur, R., & Daryati, D. (2022). Literature Review: The Role Of E-Modules In Improving Vocational Students' Scientific Literacy Skills. *Jurnal PenSil*, *11*(3), 197–205. https://doi.org/10.21009/jpensil.v11i3.28673

Napitupulu, N. D., Ratu, B., Walanda, D. K., & Napitupulu, M. (2023). Challenges Faced by Science Teachers in Implementing Differentiated Learning in Junior High School Palu Central Sulawesi Indonesia. *Jurnal Penelitian Pendidikan IPA*, *9*(SpecialIssue), 960–966. https://doi.org/10.29303/jppipa.v9iSpecialIssue.6383

Nie, Y. (2025). Automated essay scoring with SBERT embeddings and LSTM-Attention networks. *PeerJ Computer Science*, *11*, e2634. https://doi.org/10.7717/peerj-cs.2634

Nisah, K., Saminan, Syukri, M., Elisa, & Markisni. (2024). Optimizing of Physics Learning through PjBL-STEM Model to Improve Critical Thinking Skills and Students Responsibility Attitudes. *Jurnal Penelitian Pendidikan IPA*, *10*(4), 1770–1778. https://doi.org/10.29303/jppipa.v10i4.6795

Novianti, R., Aisyah, W. N., & Sukmawati, W. (2023). Analysis of Student's Answer Error on Understanding of Energy Concept in Conceptual Change Text (CCT)-Based Learning. *Jurnal Penelitian Pendidikan IPA*, *9*(2), 505–508. https://doi.org/10.29303/jppipa.v9i2.2049

Palermo, C. (2022). Rater characteristics, response content, and scoring contexts: Decomposing the determinates of scoring accuracy. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.937097

Peng, J.-L., & Yeh, S.-L. (2025). Cognitive Offloading in Short-Term Memory Tasks: Trust Toward Tools as a Moderator. *International Journal of Human–Computer Interaction*, *41*(21), 13382–13393. https://doi.org/10.1080/10447318.2025.2474449

Plasencia-Calaña, Y. (2025). *Operationalizing Automated Essay Scoring: A Human-Aware Approach*. arXiv. https://doi.org/10.48550/ARXIV.2506.21603

Pratama, H., Matsun, Puspitasari, Y. D., & Maduretno, T. W. (2025). Science Literacy through STEM-Based Project Based Learning Model. *Jurnal Penelitian Pendidikan IPA*, *11*(7), 320–330. https://doi.org/10.29303/jppipa.v11i7.11306

Putri, D. P., Febianti, Y. N., & Herawan, E. (2025). Augmented Reality Implementation in Financial Regulatory Education for Indonesian High School Students. *JTP - Jurnal Teknologi Pendidikan*, *27*(2), 598–618. https://doi.org/10.21009/jtp.v27i2.57830

Rahmadayani, S., & Darmansyah, D. (2023). Pengembangan Media Video Animasi Pembelajaran Seni Budaya Dalam Metode Demonstrasi Kelas VIII Di MTS. *Jurnal Pendidikan Indonesia*, *4*(02), 167–184. https://doi.org/10.59141/japendi.v4i02.1611

Rasyid, M. N. A., Alfian, A., Djamereng, J., & Nurwahida, N. (2024). English language assessment in Indonesian senior high schools: Perceived challenges and prospective solutions. *Journal on English as a Foreign Language*, *14*(1), 1–25. https://doi.org/10.23971/jefl.v14i1.6963

Rosidin, U., Herliani, D., & Viyanti. (2023). Development of Assessment Instruments in Project-Based Learning to Measure Students Scientific Literacy and Creative Thinking Skills on Work and Energy Materials. *Jurnal Penelitian Pendidikan IPA*, *9*(6), 4484–4494. https://doi.org/10.29303/jppipa.v9i6.2421

Rumaisa, F., Puspitarani, Y., Rosita, A., Zakiah, A., & Violina, S. (2021). Penerapan Natural Language Processing (NLP) di bidang pendidikan. *Jurnal Inovasi Masyarakat*, *1*(3), 232–235. https://doi.org/10.33197/jim.vol1.iss3.2021.799

Rustam, R., & Priyanto, P. (2022). Critical thinking assessment in the teaching of writing Indonesian scientific texts in high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *26*(1), 12–25. https://doi.org/10.21831/pep.v26i1.36241

Sari, D. N., Wagiran, W., & Luriawati, D. (2025). Transformasi Penilaian Di Era Kurikulum Merdeka: Strategi Inovatif Asesmen Sumatif Di Smk Swasta. *Jurnal Selaksa Makna*, *1*(4), 151–156. https://doi.org/10.23960/selaksamakna.v1i4.1225

Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis. *IEEE Access*, *10*, 56720–56739. https://doi.org/10.1109/ACCESS.2022.3177752

Sidiki, S. A. A., Abdjul, T., Lamangantjo, C. J., Buhungo, T. J., Yusuf, F. M., & Hermanto, I. M. (2024). The Implementation of PBL Model on Student Learning Outcomes in the Concept of Life Organizational Systems. *Jurnal Penelitian Pendidikan IPA*, *10*(8), 4486–4497. https://doi.org/10.29303/jppipa.v10i8.8100

Siregar, T. M., Armanto, D., Frisnoiry, S., & Ruslan, D. (2025). Effective Solutions for Implementing Android-Based Computer-Based Testing (CBT) In Vocational High Schools as A Digital Assessment Innovation. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 623–633. https://doi.org/10.29303/jppipa.v11i1.10155

Sun, K., & Wang, R. (2024). *Automatic Essay Multi-dimensional Scoring with Fine-tuning and Multiple Regression*. arXiv. https://doi.org/10.48550/arXiv.2406.01198

Urry, H., Crittle, C., Floerke, V., Leonard, M., Perry, C., Akdilek, N., Albert, E., Block, A., Bollinger, C., Bowers, E., Brody, R., Burk, K., Burnstein, A., Chan, A., Chan, P., Chang, L., Chen, E., Chiarawongse, P., Chin, G., & Zarrow, J. (2021). Don't Ditch the Laptop Just Yet: A Direct Replication of Mueller and Oppenheimer's (2014) Study 1 Plus Mini Meta-Analyses Across Similar Studies. *Psychological Science*, *32*. https://doi.org/10.1177/0956797620965541

Wang, Y., Luo, N., & Zhou, J. (2022). *Mining Assignment Submission Time to Detect At-Risk Students with Peer Information*. https://doi.org/10.5281/ZENODO.6853175

Werang, B. R., Sri, A. A. P., Prawira, A. A. G. O. A., Yani, N. W. M. S. A., & Leba, S. M. R. (2025). Emotional Exhaustion and Its Impact on Teachers' Job Satisfaction and Organizational Commitment in Indonesian Elementary Schools. *European Journal of Science, Innovation and Technology*, *5*(2), 202–214. Retrieved from https://ejsit-journal.com/index.php/ejsit/article/view/641

Wiley, R. W., & Rapp, B. (2021). The Effects of Handwriting Experience on Literacy Learning. *Psychological Science*, *32*(7), 1086–1103. https://doi.org/10.1177/0956797621993111

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Yang, J., Latif, E., He, Y., & Zhai, X. (2025). *Fine-tuning ChatGPT for Automatic Scoring of Written Scientific Explanations in Chinese*. arXiv. https://doi.org/10.48550/arXiv.2501.06704

Yigci, D., Eryilmaz, M., Yetisen, A. K., Tasoglu, S., & Ozcan, A. (2025). Large Language Model-Based Chatbots in Higher Education. *Advanced Intelligent Systems*, *7*(3), 2400429. https://doi.org/10.1002/aisy.202400429

Yıldırır, H. E., & Kızmaz, N. (2025). Descriptive Content Analysis of Research on Cognitive Structures in Science Education in Turkey. *Educational Academic Research*, 153–168. https://doi.org/10.33418/education.1538016

Yolanda, S. A., & Yohandri. (2025). Pengembangan E-Modul IPA Berbasis PjBL (Project Based Learning) pada Kelas VIII Untuk Meningkatkan Keterampilan Berpikir Kritis Peserta Didik. *Jurnal Penelitian Pendidikan IPA*, *11*(10), 873–881. https://doi.org/10.29303/jppipa.v11i10.12283

Yoo, J., Park, J., Ha, M., & Mae Lagmay Darang, C. (2024). Exploring Pre-Service Teachers' Cognitive Processes and Calibration with an Unsupervised Learning-Based Automated Evaluation System. *Sage Open*, *14*(3). https://doi.org/10.1177/21582440241262864

Yunita, L., & Widodo, H. (2023). The Implementation Of Merdeka Curriculum In Islamic Education Learning at SMK Muhammadiyah Lumajang. *Ta Dib Jurnal Pendidikan Islam*, *12*(1), 103–112. https://doi.org/10.29313/tjpi.v12i1.11287

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Towards reducing teacher burden in Performance-Based assessments using aivaluate: An emotionally intelligent LLM-Augmented pedagogical AI conversational agent. *Education and*

Information Technologies, *30*(17), 24649–24693. https://doi.org/10.1007/s10639-025-13755-7

Zakiyah, S., Suwarma, I. R., & Rusdiana, D. (2024). Constructing Assessment for Evaluating Critical Thinking and Creative Problem-Solving Skills in the Lesson on Soundwave. *Jurnal Penelitian Pendidikan IPA*, *10*(8), 4614–4622. https://doi.org/10.29303/jppipa.v10i8.8332