

DAFTAR PUSTAKA

- [1] B. M. Dahl, F. Vasset, and M. Frilund, “Students’ approaches to scientific essay writing as an educational method in higher education A mixed methods study,” *Soc. Sci. Humanit. Open*, vol. 7, no. 1, 2023, doi: <https://doi.org/10.1016/j.ssaho.2022.100389>.
- [2] D. Ramesh and S. K. Sanampudi, *An automated essay scoring systems: a systematic literature review*, vol. 55, no. 3. Springer Netherlands, 2022. doi: 10.1007/s10462-021-10068-2.
- [3] A. Tack, “Automated Essay Scoring,” *Comput. Linguist.* 2025, vol. 51, no. 3, pp. 1005–1008, 2025, doi: doi.org/10.1162/coli_r_00513.
- [4] E. B. Page, “The Imminence of... Grading Essays by Computer,” *Phi Delta Kappan*, vol. 47, no. 5, pp. 238–243, 1966, doi: <https://www.jstor.org/stable/20371545>.
- [5] E. B. Page, “Statistical and linguistic strategies in the computer grading of essays,” in *COLING ’67: Proceedings of the 1967 conference on Computational linguistics*, 1967, pp. 1–13. doi: doi.org/10.3115/991566.991598.
- [6] Y. Attali and J. Burstein, “Automated Essay Scoring With e-rater® V.2,” *J. Technol. Learn. Assess.*, vol. 4, no. 3, 2006, [Online]. Available: <https://ejournals.bc.edu/index.php/jtla/article/view/1650>
- [7] F. E. T. Nkoyo, C. F. Ijezue, M. Amjad, A. I. Amjad, S. Butt, and G. Castañeda-garza, “Advances in Auto-Grading with Large Language Models: A Cross-Disciplinary Survey,” in *20th Workshop on Innovative Use of NLP for Building Educational Applications*, Association for Computational Linguistics, 2025, pp. 477–498. [Online]. Available: <https://aclanthology.org/2025.bea-1.35.pdf>
- [8] P. Y. Liew and I. K. T. Tan, “On Automated Essay Grading using Large Language Models,” *CSAI 2024 - Proc. 2024 8th Int. Conf. Comput. Sci. Artif. Intell.*, pp. 204–211, 2025, doi: 10.1145/3709026.3709030.
- [9] W. Xu *et al.*, “Explainable AI for education: Enhancing essay scoring via rubric-aligned chain-of-thought prompting,” *Int. J. Mod. Phys. C*, vol. 2542013, pp. 1–26, 2025, doi: 10.1142/S0129183125420136.
- [10] M. Yamamoto, N. Umemura, and H. Kawano, “Automated Essay Scoring System Based on Rubric,” 2018, doi: 10.1007/978-3-319-64051-8.
- [11] C. Xiao *et al.*, “Human-AI Collaborative Essay Scoring: A Dual-Process Framework with LLMs,” *LAK 2025*, pp. 293–305, 2025, doi: 10.1145/3706468.3706507.
- [12] S. Y. Chu, J. W. Kim, B. Wong, and M. Y. Yi, “Rationale Behind Essay Scores : Enhancing S-LLM ’ s Multi-Trait Essay Scoring with Rationale Generated by LLMs,” no. 2024, pp. 5811–5829, 2025.
- [13] R. G. Sibarani, I. Hidayah, and A. E. Permanasari, “Rubric-based Automated Essay Scoring and Providing Insightful Feedback by Utilizing LLM for Indonesian Language Academic Text,” *2025 Int. Conf. Inf. Technol. Comput.*, no. August, pp. 356–361, 2025, doi: 10.1109/ICITCOM66635.2025.11264872.
- [14] J. Su *et al.*, “EssayJudge : A Multi-Granular Benchmark for Assessing Automated Essay Scoring Capabilities of Multimodal Large Language Models,” *Find. Assoc. Comput. Linguist. ACL 2025*, pp. 6363–6389, 2025, doi: doi.org/10.18653/v1/2025.findings-acl.329.
- [15] M. R. Parvez, “Chain of Evidences and Evidence to Generate : Prompting for

- Context Grounded and Retrieval Augmented Reasoning,” *Proc. 4th Int. Work. Knowledge-Augmented Methods Nat. Lang. Process. (KnowledgeNLP’25)*, pp. 230–245, 2025, doi: 10.18653/v1/2025.knowledgenlp-1.21.
- [16] S. C. Weigle, “English language learners and automated scoring of essays: Critical considerations,” *Assess. Writ.*, vol. 18, no. 1, pp. 85–99, 2013, doi: 10.1016/j.asw.2012.10.006.
- [17] Mark J Gierl, S. Latifi, H. Lai, A.-P. Boulais, and A. De Champlain, “Automated essay scoring and the future of educational assessment in medical education,” *Med. Educ.*, vol. 48, no. 10, pp. 950–962, 2014, doi: <https://doi.org/10.1111/medu.12517>.
- [18] U. Hasanah, A. E. Permanasari, and S. S. Kusumawardani, “A scoring rubric for automatic short answer grading system,” *TELKOMNIKA*, vol. 17, no. 2, pp. 763–770, 2019, doi: 10.12928/TELKOMNIKA.v17i2.11785.
- [19] Gunawansyah, R. Rahayu, Nurwathi, B. Sugiarto, and Gunawan, “Automated Essay Scoring Using Natural Language Processing And Text Mining Method,” pp. 1–4, 2020, doi: 10.1109/TSSA51342.2020.9310845.
- [20] M. M. Saafan, “Automatic Exam Correction Framework (AECF) for the MCQs , Essays , and Equations Matching,” vol. 9, no. 1, 2021, doi: 10.1109/ACCESS.2021.3060940.
- [21] N. Fadilah and S. Priyanta, “Automatic Essay Scoring Using Data Augmentation in Bahasa Indonesia,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 16, no. 4, pp. 401–410, 2022, doi: 10.22146/ijccs.76396.
- [22] M. Idhom, I. G. Putu, A. Buditjahjanto, and M. Samani, “Performance Evaluation of Automated Essay Scoring Online System for Competency Assessment of Community Academy,” *2022 Fifth Int. Conf. Vocat. Educ. Electr. Eng.*, pp. 206–210, 2022, doi: 10.1109/ICVEE57061.2022.9930387.
- [23] A. Muhammad, A. E. Permanasari, and I. Hidayah, “Enhancing GAN-LCS Performance Using an Abbreviations Checker in Automatic Short Answer Scoring,” 2022.
- [24] L. Saskia, I. Hidayah, and S. S. Kusumawardani, “Improvement of GAN-LCS Performance with Synonym Recognition,” *2022 6th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng.*, pp. 40–45, 2022, doi: 10.1109/ICITISEE57756.2022.10057648.
- [25] A. H. Nugroho, I. Hidayah, and S. S. Kusumawardani, “Transformer Model Fine-Tuning for Indonesian Automated Essay Scoring with Semantic Textual Similarity,” *2022 5th Int. Semin. Res. Inf. Technol. Intell. Syst.*, pp. 118–124, 2022, doi: 10.1109/ISRITI56927.2022.10052865.
- [26] S. Prabhu, K. Akhila, and S. Sanriya, “A Hybrid Approach Towards Automated Essay Evaluation based on Bert and Feature Engineering,” *2022 IEEE 7th Int. Conf. Conver. Technol.*, pp. 1–4, 2022, doi: 10.1109/I2CT54291.2022.9824999.
- [27] N. Chamidah, E. Yulianti, and I. Budi, “Evaluating the impact of sentence tokenization on Indonesian automated essay scoring using pretrained sentence embeddings,” *Rev. d’Intelligence Artif.*, vol. 37, no. 5, pp. 1101–1108, 2023, doi: <https://doi.org/10.18280/ria.370502>.
- [28] R. H. Chassab and L. Q. Zakaria, “An Optimized LSTM-Based Augmented Language Model (FLSTM-ALM) Using Fox Algorithm for Automatic Essay Scoring Prediction,” *IEEE Access*, vol. 12, no. March, pp. 48713–48724, 2024, doi: 10.1109/ACCESS.2024.3381619.
- [29] X. He *et al.*, “Beyond Final Products : Multi-Dimensional Essay Scoring Using Keystroke Logs and Deep Learning,” in *LAK ’25: Proceedings of the 15th*

- International Learning Analytics and Knowledge Conference*, ACM Digital Library, 2025. doi: 10.1145/3706468.3706548.
- [30] J. Gao, Q. Yang, Y. Zhang, L. Zhang, and S. Wang, “A Bi-modal Automated Essay Scoring System for Handwritten Essays,” *2021 Int. Jt. Conf. Neural Networks*, pp. 1–8, 2021, doi: 10.1109/IJCNN52387.2021.9533505.
- [31] A. Sharma, R. Katlaa, G. Kaur, and D. B. Jayagopi, “Full-page handwriting recognition and automated essay scoring for in-the-wild essays,” pp. 35253–35276, 2023.
- [32] D. Stribling, Y. Xia, M. K. Amer, K. S. Graim, C. J. Mulligan, and R. Renne, “The model student : GPT - 4 performance on graduate biomedical science exams,” *Sci. Rep.*, pp. 1–11, 2024, doi: 10.1038/s41598-024-55568-7.
- [33] A. Mizumoto and M. Eguchi, “Exploring the potential of using an AI language model for automated essay scoring,” *Res. Methods Appl. Linguist.*, vol. 2, no. 2, p. 100050, 2023, doi: 10.1016/j.rmal.2023.100050.
- [34] A. Gandolfi, *GPT - 4 in Education : Evaluating Aptness , Reliability , and Loss of Coherence in Solving Calculus Problems and Grading Submissions*, no. 0123456789. Springer New York, 2024. doi: 10.1007/s40593-024-00403-3.
- [35] Y. Huang and J. Wilson, “Evaluating LLM-Based Automated Essay Scoring : Accuracy , Fairness , and Validity,” in *Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con)*, National Council on Measurement in Education (NCME), 2025, pp. 71–83. [Online]. Available: <https://aclanthology.org/2025.aimecon-wip.9/>
- [36] X. Wang and X. Wang, “English Composition Image Automatic Scoring Based on Multi-modal Large Language Models,” no. November 2024, 2026, doi: 10.1145/3708394.3708437.
- [37] S. C. Weigle, *Assessing Writing*. Cambridge University Press, 2010. doi: <https://doi.org/10.1017/CBO9780511732997>.
- [38] K. Hyland, *Disciplinary Discourses: Social Interactions in Academic Writing*, Michigan C. University of Michigan Press ELT, 2004. doi: <https://doi.org/10.3998/mpub.6719>.
- [39] H. Cho and Y. Kim, “Comparing the characteristics of EFL students ’ multimodal composing and traditional monomodal writing : The case of a reading-to-write task,” *Lang. Teach. Res.*, vol. 28, no. 6, pp. 2487–2512, 2024, doi: 10.1177/13621688211046740.
- [40] R. E. Ferdig, K. E. Pytash, L. Johnston, K. Wolfe, E. O. Dea, and J. Mull, *Exploring Multimodal Composition and Digital Writing*. IGI Global Scientific Publishing, 2014. doi: <https://doi.org/10.4018/978-1-4666-4345-1>.
- [41] Y. Kim and D. Belcher, “Multimodal Composing and Traditional Essays : Linguistic Performance and Learner Perceptions,” *RELC J.*, vol. 51, no. 1, pp. 86–100, 2020, doi: 10.1177/0033688220906943.
- [42] D. Tranfield, D. Denyer, and P. Smart, “Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review,” *Br. J. Manag.*, vol. 14, no. 3, pp. 207–222, 2003, doi: <https://doi.org/10.1111/1467-8551.00375>.
- [43] B. Kitchenham and S. M. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering, version 2.3,” 2007. [Online]. Available: https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf

- [44] A. Booth, A. Sutton, and D. Papaioannou, *Systematic Approaches to a Successful Literature Review, 2nd Ed.* London: Sage Publications Inc, 2016.
- [45] M. J. Page *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n71.
- [46] B. Kitchenham and P. Brereton, “A systematic review of systematic review process research in software engineering,” *Inf. Softw. Technol.*, vol. 55, no. 12, pp. 2049–2075, 2013, doi: 10.1016/j.infsof.2013.07.010.
- [47] S. M. Brookhart, *How to Create and Use Rubrics: for Formative Assessment and Grading.* ASCD, 2013. [Online]. Available: [https://www.geocities.ws/bdktraining/pdfkur/How to Create and Use Rubrics for Formative Assessment and Grading \(PDFDrive \).pdf](https://www.geocities.ws/bdktraining/pdfkur/How to Create and Use Rubrics for Formative Assessment and Grading (PDFDrive).pdf)
- [48] A. Jönsson, A. Balan, E. Hartell, and A. Jönsson, “Analytic or holistic ? A study about how to increase the agreement in teachers ’ grading agreement in teachers ’ grading,” *Assess. Educ. Princ. Policy Pract.*, vol. 28, no. 3, pp. 212–227, 2021, doi: 10.1080/0969594X.2021.1884041.
- [49] T. Eckes, *Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments, 2nd Ed.* Peter Lang Edition, 2015. doi: 10.3726/978-3-653-04844-5.
- [50] M. D. Shermis and J. Burstein, *Handbook of Automated Essay Evaluation: Current Applications and New Directions.* Routledge/Taylor&Francis Group, 2013. [Online]. Available: <https://www.routledge.com/Handbook-of-Automated-Essay-Evaluation-Current-Applications-and-New-Directions/Shermis-Burstein/p/book/9780415810968>
- [51] R. Bommasani *et al.*, “On the Opportunities and Risks of Foundation Models,” 2022. doi: <https://doi.org/10.48550/arXiv.2108.07258>.
- [52] Z. Ji *et al.*, “Survey of Hallucination in Natural Language Generation,” vol. 55, no. 12, 2026, doi: 10.1145/3571730.
- [53] S. Yin *et al.*, “A survey on multimodal large language models,” *Natl. Sci. Rev.*, no. October, 2024, doi: 10.1093/nsr/nwae403.
- [54] S. Wang, “Capabilities of GPT-5 on Multimodal Medical Reasoning,” 2025, [Online]. Available: <https://arxiv.org/pdf/2508.08224>
- [55] A. Vaswani, “Attention Is All You Need,” no. Nips, 2017.
- [56] G. Yenduri, M. Ramalingam, G. C. Selvi, Y. Supriya, and G. Srivastava, “GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies , Potential Applications , Emerging Challenges , and Future Directions,” *IEEE Access*, vol. 12, no. April, pp. 54608–54649, 2024, doi: 10.1109/ACCESS.2024.3389497.
- [57] OpenAI, “GPT-4o System Card,” pp. 1–33, 2024, doi: <https://doi.org/10.48550/arXiv.2410.21276>.
- [58] OpenAI, “GPT-5 System Card,” 2025. [Online]. Available: <https://cdn.openai.com/gpt-5-system-card.pdf>
- [59] P. Liu and U. States, “Pre-train , Prompt , and Predict : A Systematic Survey of Prompting Methods in Natural Language Processing Pre-train , Prompt , and Predict : A Systematic Survey,” *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–35, 2023, doi: 10.1145/3560815.
- [60] V. Geroimenko, “Key Principles of Good Prompt Design. In: The Essential Guide to Prompt Engineering,” in *SpringerBriefs in Computer Science*, Springer, Cham, 2025. doi: https://doi.org/10.1007/978-3-031-86206-9_2.
- [61] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, and P. Dhariwal,

- “Language models are few-shot learners,” in *NIPS’20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, ACM Digital Library, 2020, pp. 1877–1901. doi: <https://dl.acm.org/doi/abs/10.5555/3495724.3495883>.
- [62] Y. Qian, “Prompt Engineering in Education : A Systematic Review of Approaches and Educational Applications,” *J. Educ. Comput. Res.*, vol. 63, no. 7–8, pp. 1782–1818, 2025, doi: <https://doi.org/10.1177/07356331251365189>.
- [63] C. Zhang and X. Luo, “Scaling Equitable Reflection Assessment in Education via Large Language Models and Role-Based Feedback Agents,” *Proc. IEEE*, 2025, doi: <https://doi.org/10.48550/arXiv.2511.11772>.
- [64] J. Wei, D. Schuurmans, E. H. Chi, Q. V Le, and C. L. Jan, “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models,” in *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, 2022. doi: <https://doi.org/10.48550/arXiv.2201.11903> Focus to learn more.
- [65] S. Mori, C. Y. Suen, and K. Yamamoto, “Historical Review of OCR Research and Development,” *Proc. IEEE*, vol. 80, no. 7, pp. 1029–1058, 1992, doi: 10.1109/5.156468.
- [66] J. R. Landis and G. G. Koch, “The Measurement of Observer Agreement for Categorical Data Published by : International Biometric Society Stable URL : <http://www.jstor.org/stable/2529310>,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
- [67] OpenAI, “File Uploads FAQ.” Accessed: Dec. 19, 2025. [Online]. Available: <https://help.openai.com/es-es/articles/8555545-file-uploads-faq>
- [68] IEEE, “Manuscript Templates for Conference Proceedings.” [Online]. Available: <http://ieee.org/conferences/publishing/templates>