
IMPLEMENTASI KECERDASAN BUATAN DALAM PENYUSUNAN INSTRUMEN TES KEMAMPUAN AKADEMIK (TKA) BAGI GURU MATEMATIKA SMK

Tabah Heri Setiawan^{1*}, Aden², Khoeri Abdul Azis³, Anneke Lutfia Putri⁴, Irawan⁵

^{1,2,3,4,5}Departmen Program Studi Matematika Fakultas MIPA, Universitas Pamulang

*E-mail: tabah.ibnubara@gmail.com

ABSTRAK

Implementasi kecerdasan buatan atau *artificial intelligence* dalam penyusunan instrumen Tes Kemampuan Akademik (TKA) Matematika penting dilakukan sebagai upaya penguatan kompetensi profesional bagi guru-guru pengampu mata pelajaran matematika di Sekolah Menengah Kejuruan (SMK). Sebagai upaya mendorong dan mempercepat transformasi pendidikan berbasis kecerdasan buatan, salah satunya melalui program pengabdian masyarakat berupa pelatihan atau workshop, sehingga guru matematika dapat mengembangkan instrumen asesmen yang relevan dengan dunia kerja vokasional melalui pemanfaatan teknologi kecerdasan buatan. Untuk mengukur keberhasilan kegiatan pelatihan digunakan pendekatan *One-Group Pretest-Posttest Design* pada 76 guru matematika yang terlibat dalam pelatihan. Hasil pretest menunjukkan kemampuan penggunaan kecerdasan buatan dalam penyusunan instrumen asesmen berada pada kategori rendah dan pemahaman terhadap AI juga masih terbatas. Setelah pelatihan, terjadi peningkatan rata-rata kompetensi sebesar 55,48% dan dapat menyusun variasi soal asesmen sebagai bahan latihan atau *try out*. Kegiatan ini membuktikan efektivitas integrasi kecerdasan buatan dalam penyusunan instrumen asesmen, sekaligus menekankan pentingnya pendampingan berkelanjutan dan pendekatan diferensial untuk mengoptimalkan literasi kecerdasan buatan bagi guru dalam transformasi pendidikan di era modern.

Kata kunci: kecerdasan buatan, matematika, tes kemampuan akademik.

ABSTRACT

The implementation of artificial intelligence (AI) in the development of Academic Proficiency Test (APT) for mathematics is important as an effort to strengthen the professional competence of mathematics teachers in vocational high schools. As an effort to encourage and accelerate AI-based education transformation, one of the ways is through community service programs in the form of training or workshops, so that mathematics teachers can develop assessment instruments that are relevant to the vocational world through the use of artificial intelligence technology. To measure the success of the training activities, a One-Group Pretest-Posttest Design approach was used on 76 mathematics teachers who participated in the training. The pretest results showed that the ability to use artificial intelligence in developing assessment instruments was low, and understanding of AI was also limited. After the training, there was an average increase in competence of 55.48%, and participants were able to develop a variety of assessment questions as practice materials or tryouts. This activity proves the effectiveness of integrating artificial intelligence in the development of assessment instruments, while emphasizing the importance of continuous assistance and a differential approach to optimize artificial intelligence literacy for teachers in the transformation of education in the modern era.

Keywords: *academic proficiency test; artificial intelligence; mathematics.*

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*) dalam satu dekade terakhir telah mengubah pendekatan pembelajaran dan asesmen di berbagai negara (Parmar dkk., 2025). Salah satu kemajuan penting adalah pemanfaatan AI untuk *automatic item generation* (AIG), yaitu proses menghasilkan butir soal secara otomatis dengan dukungan model bahasa dan teknik komputasi canggih (Song dkk., 2025). Beberapa penelitian menegaskan bahwa penerapan AI dalam pengukuran pendidikan mampu meningkatkan efisiensi penyusunan instrumen, memperluas cakupan konten, dan mendukung asesmen adaptif yang lebih akurat (Minn, 2022; S. C. dkk., 2024; Salendab dkk., 2025).

Temuan serupa disampaikan oleh May dkk., (2025) yang menyoroti bahwa model AI modern dapat digunakan sebagai alat bantu pengembangan butir dengan tetap mempertahankan peran sentral ahli dalam memastikan validitas isi dan kesesuaian tingkat kognitif.

Dalam konteks pendidikan Indonesia, peningkatan kualitas asesmen menjadi agenda strategis nasional. Rapor Pendidikan Indonesia 2025 dalam tajuk “Capaian dan Agenda Perbaikan” yang diterbitkan oleh Pusat Standar dan Kebijakan Pendidikan Kemendikdasmen menunjukkan bahwa literasi numerasi siswa Indonesia masih perlu diperkuat (Pratiwi dkk., 2025). Data PISA 2022 juga mengonfirmasi bahwa kemampuan numerasi siswa Indonesia masih berada pada kategori menengah-rendah di bawah rata-rata global, serta membutuhkan dukungan intervensi sistematis pada berbagai aspek (OECD, 2023). Kondisi ini menuntut guru untuk mampu mengembangkan instrumen yang tidak hanya mengukur pengetahuan prosedural, tetapi juga kemampuan berpikir tingkat tinggi (*higher-order thinking skills*). Dalam praktiknya, penyusunan instrumen Tes Kemampuan Akademik (TKA) Matematika di sekolah masih menghadapi sejumlah kendala (Nurlaily dkk., 2019). Guru sering mengalami keterbatasan waktu, sumber daya, serta akses terhadap pelatihan penyusunan soal berbasis standar psikometrik (Mardhiyana, 2023). Akibatnya, bank soal yang digunakan cenderung kurang bervariasi dan belum sepenuhnya mencerminkan kompetensi numerasi yang relevan bagi siswa vokasional (Setiani dkk., 2022).

Tantangan ini menjadi semakin signifikan ketika dikaitkan dengan tuntutan pendidikan vokasi, dimana asesmen matematika tidak hanya mengukur penguasaan materi, tetapi juga kemampuan menerapkan konsep dalam situasi dunia kerja. Untuk menjembatani kesenjangan tersebut, integrasi kecerdasan buatan menawarkan solusi melalui otomatisasi pembuatan butir soal, penyediaan variasi tingkat kesukaran soal, serta penyusunan alternatif soal yang selaras dengan indikator kompetensi. Dengan catatan penggunaan kecerdasan buatan tetap harus mengikuti prinsip *human-in-the-loop* untuk menjaga validitas, reliabilitas, serta integritas data asesmen (Wilchek dkk., 2023).

Pada sekolah menengah kejuruan di wilayah Jakarta Timur 2, kebutuhan untuk mengembangkan instrumen asesmen yang bervariasi dan berkualitas menjadi semakin penting mengingat karakteristik siswa yang heterogen serta tuntutan kurikulum yang menekankan relevansi terhadap dunia kerja. Guru matematika membutuhkan dukungan dalam bentuk pelatihan dan pendampingan yang dapat meningkatkan kapasitas mereka dalam menyusun instrumen TKA. Pemanfaatan kecerdasan buatan tidak hanya bertujuan meningkatkan efisiensi, tetapi juga memperluas peluang untuk menghasilkan instrumen asesmen berkualitas yang mampu memotret kemampuan matematika siswa secara lebih akurat. Melalui program pengabdian kepada masyarakat ini bukan hanya kegiatan transfer pengetahuan, tetapi juga kontribusi nyata terhadap peningkatan mutu pendidikan vokasional melalui penguatan kapasitas guru dalam menghadapi tantangan era digital.

METODE

Pengabdian kepada masyarakat (PkM) ini berupa pelatihan/workshop dan pendampingan terarah mengenai pemanfaatan kecerdasan buatan (artificial intelligence/AI) dalam penyusunan instrumen Tes Kemampuan Akademik (TKA) Matematika bagi guru SMK. Kegiatan dilaksanakan pada Kamis, 16 Oktober 2025 di SMKN 68 Jakarta dengan mitra MGMP (Musyawarah Guru Mata Pelajaran) Matematika SMK Wilayah Jakarta Timur 2.

Evaluasi keberhasilan pelatihan menggunakan pendekatan evaluatif dengan desain *One-Group Pretest-Posttest*, yaitu pengukuran kemampuan peserta sebelum dan sesudah pelatihan pada kelompok yang sama (Bonate, 2020). Populasi sasaran adalah guru matematika SMK yang mengikuti program peningkatan kompetensi penyusunan instrumen TKA, dengan sampel sebanyak 76 guru matematika yang terlibat sebagai peserta dan mengikuti rangkaian kegiatan secara penuh. Teknik sampling yang digunakan adalah sampel jenuh (total sampling), karena seluruh peserta yang memenuhi kriteria (guru pengampu matematika, bersedia mengikuti *pretest-posttest*, dan mengikuti sesi praktik penyusunan butir hingga menghasilkan produk) dijadikan responden.



Gambar 1. Pelaksanaan Pelatihan

Pengumpulan data dilakukan melalui tes pretest dan posttest sebanyak 25 butir soal pilihan ganda untuk mengukur peningkatan kompetensi peserta dalam memahami AI dan keterampilan memanfaatkannya untuk menyusun instrumen asesmen, serta penilaian produk berupa butir soal TKA yang disusun peserta selama pelatihan. Selain itu, dokumentasi kegiatan (daftar hadir, catatan fasilitator, dan foto kegiatan) digunakan sebagai data pendukung keterlaksanaan program.

Pelaksanaan pelatihan mencakup tahap persiapan (koordinasi, penyusunan modul, contoh kisi-kisi, template prompt, serta instrumen *pretest-posttest*), tahap implementasi (pengenalan AI dalam asesmen dan prinsip *human-in-the-loop*, penyusunan kisi-kisi/indikator dan pemetaan level kognitif, praktik penyusunan butir berbantuan AI untuk menghasilkan variasi soal dan distraktor, serta penyuntingan dan verifikasi butir oleh guru), dan tahap evaluasi (*posttest*, rekap luaran butir soal/paket latihan, serta refleksi). Alat dan bahan yang digunakan meliputi laptop/komputer peserta dengan akses

internet dan browser terbaru, layanan AI generatif berbasis model bahasa sebagai alat bantu penyusunan draft butir, perangkat presentasi (LCD proyektor dan audio), serta modul pelatihan dan lembar kerja praktik.

Analisis data dilakukan secara deskriptif dan inferensial sederhana meliputi:

1. Statistik deskriptif

Statistik deskriptif meliputi nilai rata-rata, persentase, dan simpangan baku skor pretest-posttest, serta rekap kualitas produk butir.

2. Persentase peningkatan kompetensi dihitung dengan rumus (Sugiyono, 2013):

$$\text{Peningkatan}(\%) = \frac{\bar{X}_{\text{post}} - \bar{X}_{\text{pre}}}{\bar{X}_{\text{pre}}} \times 100\% \quad (1)$$

dengan $\bar{X}_{\text{pre}}$ = rata-rata pretest dan $\bar{X}_{\text{post}}$ = rata-rata posttest.

3. N-gain untuk melihat efektivitas peningkatan relatif (Hake, 1998):

$$g = \frac{\bar{X}_{\text{post}} - \bar{X}_{\text{pre}}}{X_{\text{maks}} - \bar{X}_{\text{pre}}} \quad (2)$$

dengan X_{maks} = skor maksimum (100). Kategori rendah ($g < 0,30$), sedang ($0,30 \leq g \leq 0,70$), tinggi ($g > 0,70$).

HASIL

Kegiatan PkM diikuti oleh 76 guru matematika SMK yang mengikuti pengukuran pretest dan posttest (skor 0-100; 25 butir, benar bernilai 4). Hasil pretest dan posttest ditunjukkan pada Tabel 1.

Tabel 1. Statistik deskriptif skor pretest dan posttest (n = 76)

Variabel	Minimum	Maksimum	Mean	Standar Deviasi
Pretest	16	60	44,68	8,12
Posttest	40	88	69,47	8,56

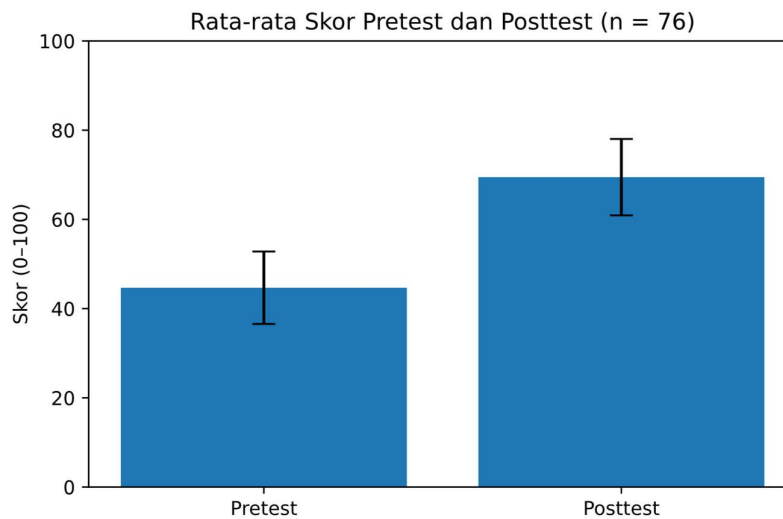
Berdasarkan rata-rata skor, peningkatan dari pretest ke posttest sebesar 24,79 poin. Persentase peningkatan kompetensi dihitung dengan rumus (1) menghasilkan kenaikan 55,48%. Efektivitas peningkatan juga dihitung menggunakan normalized gain (N-gain), dengan rata-rata 0,46 (SD = 0,08) yang termasuk kategori sedang.

Tabel 2. Peningkatan skor dan N-gain (n = 76)

Indikator	Nilai
Selisih mean (Post–Pre)	24,79

Persentase peningkatan	55,48%
Rata-rata N-gain (SD)	0,46 (0,08)
N-gain rendah ($g < 0,30$)	1 (1,32%)
N-gain sedang ($0,30 \leq g \leq 0,70$)	75 (98,68%)
N-gain tinggi ($g > 0,70$)	0 (0,00%)

Berikut grafik batang rata-rata skor *pretest* dan *posttest* (dengan *error bar* = Standar Deviasi)



Gambar 2. Rata-rata Skor *Pretest* dan *Posttest*

PEMBAHASAN

Hasil menunjukkan adanya peningkatan kemampuan guru dalam memanfaatkan kecerdasan buatan (AI) untuk penyusunan instrumen TKA Matematika, tercermin dari kenaikan rerata skor pretest (44,68) menjadi posttest (69,47), peningkatan relatif 55,48%, serta N-gain 0,46 (kategori sedang). Pola ini mengindikasikan bahwa pelatihan berbasis praktik (penyusunan kisi-kisi, perumusan *prompt*, pembuatan variasi butir, dan telaah/revisi) efektif memperkuat literasi operasional guru dalam mengintegrasikan AI secara terarah pada tugas asesmen.

Secara teoretik, temuan tersebut konsisten dengan literatur yang menegaskan bahwa pemanfaatan model bahasa/AI untuk automatic item generation dapat meningkatkan efisiensi produksi butir dan memperkaya variasi soal, tetapi kualitas butir sangat bergantung pada kontrol indikator, batasan kognitif, dan proses peninjauan manusia. Tinjauan riset tentang AIG berbasis LLM menunjukkan peluang besar untuk mempercepat pengembangan butir, namun juga menyoroti keterbatasan seperti

potensi *hallucination*, ketidaktepatan konsep, dan inkonsistensi penalaran jika keluaran AI tidak diverifikasi. (Tan dkk., 2025)

Dari sisi kebijakan, hasil ini relevan dengan kebutuhan penguatan numerasi dan pengambilan keputusan berbasis bukti. Rapor Pendidikan Indonesia 2025 menempatkan capaian literasi numerasi sebagai indikator penting yang perlu terus diperkuat melalui intervensi terarah dan penguatan kapasitas satuan pendidikan. Selain itu, hasil PISA 2022 (catatan negara Indonesia) menunjukkan capaian matematika Indonesia masih di bawah rata-rata OECD dan proporsi siswa yang mencapai level minimum masih relatif rendah, sehingga upaya peningkatan kualitas asesmen (termasuk kualitas butir dan variasi latihan/*try out*) menjadi semakin strategis. Program ini juga sejalan dengan arah penguatan kompetensi digital nasional, termasuk dorongan integrasi keterampilan koding dan AI di ekosistem pendidikan.

Praktik baik kedepannya, agar dampak pelatihan tidak berhenti pada peningkatan skor sesaat, diperlukan (1) pendampingan berkelanjutan melalui komunitas guru (*review* silang butir, klinik *prompt*, dan bank *prompt*), (2) pendekatan diferensial (modul dasar-menengah-lanjutan sesuai literasi AI guru), dan (3) mekanisme *quality assurance* yang lebih kuat melalui checklist telaah butir, uji coba terbatas, serta analisis butir (kesukaran, daya pembeda, dan fungsi distraktor) sebelum butir soal disebarluaskan secara luas. Kerangka ini selaras dengan rekomendasi riset mutakhir yang menempatkan pelatihan guru, tata kelola, dan etika sebagai prasyarat adopsi AI yang bertanggung jawab pada asesmen

SIMPULAN

Kegiatan pengabdian kepada masyarakat berupa pelatihan pemanfaatan kecerdasan buatan (AI) untuk penyusunan instrumen TKA Matematika bagi 76 guru matematika se wilayah Jakarta Timur 2 yang berlangsung di SMKN 68 Jakarta menunjukkan hasil positif, ditandai dengan peningkatan rerata skor dari 44,68 (*pretest*) menjadi 69,47 (*posttest*) atau naik 55,48% dengan N-gain 0,46 (kategori sedang). Temuan ini menegaskan bahwa integrasi AI efektif meningkatkan kompetensi guru dalam merancang butir soal yang lebih bervariasi dan terarah, terutama ketika diterapkan dengan prinsip *human-in-the-loop* untuk menjaga validitas dan kualitas asesmen. Program lanjutan untuk kedepannya, dampak program perlu diperkuat melalui pendampingan berkelanjutan, pendampingan komunitas, serta mekanisme penjaminan mutu (telaah butir dan analisis butir) agar pemanfaatan AI selaras dengan agenda penguatan numerasi dan kebijakan transformasi digital pendidikan.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada LPPM Universitas Pamulang atas dukungan pendanaan dan memfasilitasi pelaksanaan kegiatan pengabdian kepada masyarakat berdasarkan Kontrak Nomor 0001/D5/SPKPM/LPPM/UNPAM/IV/2025. Ucapan terima kasih juga disampaikan

kepada Musyawarah Guru Mata Pelajaran (MGMP) Matematika SMK Wilayah Jakarta Timur 2 atas kerja sama, koordinasi, serta dukungan partisipasi guru dalam pelaksanaan kegiatan sehingga program dapat berjalan dengan baik dan mencapai luaran yang diharapkan.

DAFTAR PUSTAKA

- Bonate, P. L. (2020). *Analysis of Pretest-Posttest Designs*. CRC PRESS.
- Hake, R. R. (1998). Interactive-engagement Versus Traditional Methods: A Six-Thousand-student Survey of Mechanics Test Data for Introductory Physics Courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Mardhiyana, D. (2023). Analisis Kesulitan Menyusun Instrumen Tes Pada Penilaian Pembelajaran Matematika. *Jurnal Pemikiran dan Penelitian Pendidikan Matematika (JP3M)*, 6(2), 68–80. <https://doi.org/10.36765/jp3m.v6i2.677>
- May, T. A., Provinzano, K., Koskey, K. L. K., Sondergeld, C. J., Stone, G. E., Archer, J. N., & Rimkunas, N. (2025). Expanding the Team: Integrating Generative Artificial Intelligence into the Assessment Development Process. *Applied Sciences*, 15(18), 9976. <https://doi.org/10.3390/app15189976>
- Minn, S. (2022). AI-Assisted Knowledge Assessment Techniques for Adaptive Learning Environments. *Computers and Education: Artificial Intelligence*, 3, 100050. <https://doi.org/10.1016/j.caeai.2022.100050>
- Nurlaily, V. A., Soegiyanto, H., & Usodo, B. (2019). Elementary School Teacher's Obstacles in The Implementation of Problem-Based Learning Model in Mathematics Learning. *Journal on Mathematics Education*, 10(2), 229–238. <https://doi.org/10.22342/jme.10.2.5386.229-238>
- OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Parmar, K. J., Palaniappan, D., Premavathi, T., & Jain, R. (2025). Revolutionizing Educational Assessments With AI Technology. Dalam C. C. Chinedu, N. Mafarja, M. H. Hussein, & A. Saleem (Ed.), *Advances in Computational Intelligence and Robotics* (hlm. 227–252). IGI Global. <https://doi.org/10.4018/979-8-3693-9062-7.ch011>
- Pratiwi, I., Samosir, I., Ekowati, D. R., Wicaksono, M. E., Krismahardika, B. G., Rihardika, A., Krisna, F. N., & Martak, Y. F. (2025). Rapor Pendidikan Indonesia 2025: Capaian dan Agenda Perbaikan. Pusat Standar dan Kebijakan Pendidikan Kemendikdasmen.
- S. C., V., Vidhyapriya, P., & V. P., A. (2024). The Role of AI in Transforming Assessment Practices in Education: Dalam K. L. Keeley (Ed.), *Advances in Educational Marketing, Administration, and Leadership* (hlm. 43–70). IGI Global. <https://doi.org/10.4018/979-8-3693-5443-8.ch003>
- Salendab, F. A., Cogo, D. A., Catipay, A. Z., & Millendez, G. J. T. (2025). The Role of Artificial Intelligence (AI) in Personalized Learning and Adaptive Assessments: Dalam J. Moore, S. Gupta, M. Sharma, A. Garg, & H. Josephine V. L. (Ed.), *Advances in Computational Intelligence and Robotics* (hlm. 93–122). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-2185-1.ch004>
- Setiani, N. W., Asikin, M., & Dewi, N. R. (2022). Numerical Literacy Skills of Vocational High School Students in Solving HOTS Problems. *AlphaMath : Journal of Mathematics Education*, 8(2), 121. <https://doi.org/10.30595/alphamath.v8i2.14161>
- Song, Y., Du, J., & Zheng, Q. (2025). Automatic Item Generation for Educational Assessments: A Systematic Literature Review. *Interactive Learning Environments*, 33(9), 5386–5405. <https://doi.org/10.1080/10494820.2025.2482588>
- Sugiyono. (2013). *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D* (Cet. 6). Alfabeta.
- Tan, B., Armoush, N., Mazzullo, E., Bulut, O., & Gierl, M. (2025). A Review of Automatic Item Generation Techniques Leveraging Large Language Models. *International Journal of Assessment Tools in Education*, 12(2), 317–340. <https://doi.org/10.21449/ijate.1602294>

Wilchek, M., Hanley, W., Lim, J., Luther, K., & Batarseh, F. A. (2023). Human-in-the-loop for Computer Vision Assurance: A Survey. *Engineering Applications of Artificial Intelligence*, 123, 106376. <https://doi.org/10.1016/j.engappai.2023.106376>