

Pendampingan Guru Matematika SMP dalam Penyusunan Soal Tes Kemampuan Akademik Berbasis Konteks Berbantuan *Artificial Intelligence* di Kabupaten Jombang

Yurizka Melia Sari^{1*}, Endah Budi Rahaju¹, Nonik Indrawatiningsih¹,
Riska Wahyu Romadhonia², Danang Ariyanto³

¹Program Studi Pendidikan Matematika, Universitas Negeri Surabaya, Surabaya, Indonesia

²Program Studi Matematika, Universitas Negeri Surabaya, Surabaya, Indonesia

³Program Studi Sains Aktuaria, Universitas Negeri Surabaya, Surabaya, Indonesia

*Corresponding Author: yurizkasari@unesa.ac.id

Dikirim: 11-08-2026; Direvisi: 30-08-2026; Diterima: 01-09-2026

Abstrak: Tes Kemampuan Akademik (TKA) menuntut guru mampu menyusun butir soal berbasis konteks yang mengukur tiga level kognitif, yaitu pengetahuan dan pemahaman, aplikasi, serta penalaran. Hasil koordinasi dengan MGMP Matematika SMP Kabupaten Jombang menunjukkan bahwa guru masih kesulitan merumuskan konteks yang relevan dan belum terampil memanfaatkan Artificial Intelligence sebagai alat bantu penyusunan soal. Kegiatan pengabdian ini bertujuan meningkatkan kompetensi guru matematika SMP di Kabupaten Jombang dalam menyusun soal TKA berbasis konteks berbantuan AI. Metode yang digunakan adalah pendekatan pelatihan dan pendampingan berbasis prinsip pembelajaran orang dewasa yang dilaksanakan melalui lima tahap, yaitu analisis situasi dan koordinasi awal, penyusunan materi dan instrumen, pelaksanaan pelatihan, pendampingan pasca pelatihan, serta evaluasi. Pelatihan dilaksanakan secara luring pada 25 Juli 2026 bersama guru anggota MGMP Matematika SMP Kabupaten Jombang. Evaluasi dilakukan melalui pretest, posttest, dan angket respon peserta. Hasil analisis terhadap 27 peserta yang mengikuti kedua tes menunjukkan peningkatan rerata penguasaan konsep dari 69,8 menjadi 87,3 dengan N-gain 0,46 berkategori sedang dan perbedaan yang signifikan secara statistik. Kelengkapan komponen prompt yang disusun peserta meningkat dari rerata 2,66 menjadi 4,13 dari tujuh komponen, sedangkan proporsi peserta yang mampu menyusun prompt lengkap naik dari 25,0 persen menjadi 54,8 persen. Angket respon menunjukkan 89,5 persen jawaban peserta berada pada kategori setuju dan sangat setuju. Disimpulkan bahwa pelatihan yang memadukan penguatan kerangka asesmen dengan praktik prompting efektif meningkatkan kompetensi guru, meskipun keterampilan teknis penyusunan prompt masih memerlukan pendampingan lanjutan. Pengabdian selanjutnya direkomendasikan menyelenggarakan pendampingan berkelanjutan berbasis sesi praktik iteratif dan instrumen telaah butir mandiri agar keterampilan prompting guru berkembang sampai tataran teknis.

Kata Kunci: pendampingan guru; Tes Kemampuan Akademik; soal berbasis konteks; Artificial Intelligence; MGMP Matematika.

Abstract: The Academic Ability Test requires teachers to construct context-based items measuring three cognitive levels, namely knowledge and comprehension, application, and reasoning. Coordination with the Junior High School Mathematics Teacher Working Group of Jombang Regency revealed that teachers still struggle to formulate relevant contexts and lack the skills to use Artificial Intelligence as a tool for item construction. This community service program aimed to strengthen the competence of junior high school mathematics teachers in Jombang Regency in developing context-based test items with AI assistance. The program applied a training and mentoring approach grounded in adult learning principles, carried out in five stages, namely situation analysis and initial coordination, material and instrument development, training implementation, post-training mentoring, and evaluation.

The offline training was held on 25 July 2026 for mathematics teachers of the Jombang working group. Evaluation employed a pretest, a posttest, and a participant response questionnaire. Analysis of 27 participants who completed both tests showed an increase in mean conceptual mastery from 69,8 to 87,3, with a normalised gain of 0,46 in the moderate category and a statistically significant difference. The completeness of participant-written prompts rose from a mean of 2,66 to 4,13 out of seven components, while the proportion of participants producing complete prompts increased from 25,0 percent to 54,8 percent. The questionnaire indicated that 89,5 percent of the responses fell into the agree and strongly agree categories. Training that combines strengthening of the assessment framework with prompting practice effectively improves teacher competence, although the technical skill of writing prompts still requires further mentoring.

Keywords: teacher mentoring; academic ability test; context-based items; artificial intelligence; mathematics teacher working group.

PENDAHULUAN

Pemberlakuan Tes Kemampuan Akademik (TKA) sejak tahun 2025 membawa konsekuensi langsung pada praktik asesmen di satuan pendidikan. TKA merupakan asesmen terstandar berbasis komputer yang dirancang untuk mengukur capaian akademik murid secara objektif pada sejumlah mata pelajaran, termasuk matematika. Karakteristik butir TKA yang menyajikan stimulus kontekstual dan menjangkau tiga level kognitif menuntut pola penyusunan soal yang berbeda dari soal bertipe prosedural yang selama ini banyak digunakan di kelas. Konsekuensinya, guru dituntut tidak hanya memahami karakteristik butir soal TKA, tetapi juga mampu memproduksi butir soal setara sebagai bahan latihan dan asesmen internal di sekolah masing-masing. Tuntutan tersebut bersifat berkelanjutan karena kebutuhan butir soal muncul pada setiap periode asesmen, sehingga kemampuan menyusun soal secara mandiri menjadi kompetensi yang tidak dapat dihindari oleh guru matematika.

Soal TKA yang berkualitas tidak berhenti pada pengukuran pengetahuan faktual, melainkan menuntut kemampuan analisis, evaluasi, dan kreativitas siswa dalam mengaplikasikan pengetahuan untuk menyelesaikan masalah kontekstual (Zulkardi et al., 2025). Karakteristik ini sejalan dengan taksonomi Bloom yang telah direvisi, yang menempatkan proses kognitif dalam jenjang berpikir dari mengingat hingga mencipta (Mulyani et al., 2025). Dalam kerangka TKA Matematika SMP, jenjang tersebut diterjemahkan ke dalam tiga level kognitif, yaitu pengetahuan dan pemahaman, aplikasi, serta penalaran, yang dikenakan pada empat elemen materi: bilangan, aljabar, geometri dan pengukuran, serta data dan peluang. Penyusunan butir yang mampu membedakan ketiga level tersebut memerlukan pemahaman mendalam tentang taksonomi kognitif, pemilihan konteks yang relevan, dan keterampilan teknis merumuskan stimulus serta pengecoh (Sari et al., 2018).

Di sisi lain, perkembangan teknologi kecerdasan buatan generatif membuka peluang baru dalam proses penyusunan soal. Kajian sistematis terhadap 95 artikel penelitian yang terbit antara tahun 2015 dan 2024 menunjukkan bahwa pemanfaatan Artificial Intelligence (AI) dalam pendidikan berkembang pesat, tetapi hanya 35% studi yang menyoroti peran AI dalam pengembangan profesional guru (Tan et al., 2025a). AI dapat membantu guru menghasilkan ide soal, menyusun konteks yang menarik, memvariasikan tingkat kesulitan, serta memberi umpan balik terhadap butir yang telah disusun (Chen et al., 2020). Namun pemanfaatan tersebut menuntut literasi digital, pemahaman tentang cara berinteraksi secara efektif dengan AI melalui prompt



(Nazari & Saadi, 2024), serta kemampuan mengevaluasi dan menyempurnakan keluaran AI agar sesuai standar pendidikan (Picciano, 2024). Dimensi etika juga tidak dapat diabaikan, mengingat penggunaan AI di ruang kelas menyentuh persoalan tanggung jawab profesional dan akuntabilitas guru (Du et al., 2024). Kebutuhan verifikasi tersebut ditegaskan oleh analisis kesalahan GPT-4 pada butir ujian, yang menemukan bahwa kekeliruan model justru paling banyak terjadi pada level mengingat dan memahami, sehingga keluaran AI dapat tampak meyakinkan tetapi keliru secara konsep (Herrmann-Werner et al., 2024).

Berdasarkan hasil observasi dan diskusi dengan MGMP Matematika SMP Kabupaten Jombang, ditemukan empat tantangan utama dalam penyusunan soal TKA. Pertama, guru masih kesulitan menyusun soal berbasis konteks yang relevan dengan kehidupan sehari-hari siswa, khususnya konteks lokal Kabupaten Jombang. Kedua, pemahaman tentang karakteristik soal yang mengukur kemampuan berpikir tingkat tinggi masih minim. Ketiga, keterbatasan waktu dan referensi menghambat pengembangan bank soal yang berkualitas. Keempat, pemanfaatan teknologi AI sebagai alat bantu penyusunan soal belum optimal.

Kondisi tersebut dikonfirmasi oleh hasil pretest yang diberikan kepada 32 guru sebelum pelatihan. Kemampuan membedakan level kognitif merupakan titik terlemah: hanya 53% peserta yang tepat mengidentifikasi ciri proses berpikir pada level aplikasi dan 56% pada level penalaran. Sebanyak 34% peserta juga belum mengenali bentuk soal Multiple Choice Multiple Answer (MCMA) yang lazim digunakan pada TKA, dan 31% belum memahami cara kerja Large Language Model. Pada aspek keterampilan, prompt yang disusun peserta rata-rata hanya memuat 2,66 dari 7 komponen yang diperlukan, dengan panjang rerata 22 kata, sehingga instruksi yang diberikan kepada AI cenderung terlalu umum untuk menghasilkan butir soal sesuai spesifikasi.

Bertolak dari kondisi tersebut, MGMP Matematika SMP Kabupaten Jombang menghadapi dua permasalahan pokok. Pertama, guru mengalami kesulitan dalam mengidentifikasi dan merumuskan konteks yang tepat untuk soal TKA, sehingga soal yang disusun cenderung bersifat prosedural dan kurang mengembangkan kemampuan analisis serta pemecahan masalah siswa. Kedua, guru belum memiliki keterampilan memadai dalam memanfaatkan AI untuk membantu penyusunan soal, mencakup teknik prompting, evaluasi kualitas keluaran AI, dan penyempurnaan hasil agar memenuhi standar butir TKA. Kedua permasalahan tersebut saling berkaitan, sebab keterbatasan pemahaman terhadap kerangka asesmen membuat guru sulit menilai apakah butir yang dihasilkan AI benar-benar sesuai dengan level kognitif yang dituju.

METODE PELAKSANAAN KEGIATAN

Kegiatan ini menggunakan pendekatan pelatihan dan pendampingan (*training and mentoring approach*) yang mengintegrasikan prinsip pembelajaran orang dewasa dan praktik langsung (Hemmler & Ifenthaler, 2024). Pendekatan tersebut menekankan pembelajaran aktif, kolaboratif, dan reflektif, dengan pemanfaatan teknologi AI yang diintegrasikan secara bertahap sesuai kebutuhan dan kesiapan guru peserta.

Kegiatan pengabdian kepada masyarakat ini bertujuan meningkatkan kompetensi guru matematika SMP di Kabupaten Jombang dalam menyusun soal TKA berbasis konteks berbantuan AI, sekaligus menumbuhkan pemahaman tentang pemanfaatan AI secara bertanggung jawab dalam pendidikan. Melalui pelatihan dan



pendampingan, diharapkan terbentuk bank soal TKA hasil karya guru yang relevan dengan konteks lokal dan siap digunakan di sekolah masing-masing.

Mitra kegiatan adalah MGMP Matematika SMP Kabupaten Jombang, Jawa Timur. Pelatihan dilaksanakan secara luring pada Sabtu, 25 Juli 2026, pukul 08.00 sampai 12.00 WIB. Peserta merupakan guru matematika SMP anggota MGMP yang berasal dari sekolah negeri maupun swasta di seluruh wilayah Kabupaten Jombang. Sebanyak 32 peserta mengisi pretest dan 31 peserta mengisi posttest, dengan 27 di antaranya terpasangkan pada kedua tes dan digunakan sebagai dasar analisis peningkatan.

Prosedur kerja kegiatan terdiri atas lima tahap sebagaimana disajikan pada Gambar 1. Tahap pertama adalah analisis situasi dan koordinasi awal, meliputi identifikasi kebutuhan mitra, koordinasi dengan Dinas Pendidikan Kabupaten Jombang dan pengurus MGMP, serta penentuan jadwal pelaksanaan. Tahap kedua adalah penyusunan materi dan instrumen pelatihan. Tahap ketiga adalah pelaksanaan pelatihan yang mencakup pemahaman soal TKA berbasis konteks, pengenalan teknologi AI dan teknik prompting, praktik penyusunan soal, serta evaluasi kualitas soal. Tahap keempat adalah pendampingan pasca pelatihan secara daring maupun luring untuk membantu guru menyempurnakan soal dan membangun bank soal. Tahap kelima adalah evaluasi dan penyusunan laporan beserta luaran.



Gambar 1. Alur Pelaksanaan Kegiatan Pengabdian kepada Masyarakat

Materi pelatihan disusun dalam dua bagian. Materi pertama, Framework Tes Kemampuan Akademik Matematika SMP, membahas konsep dan karakteristik soal TKA berbasis konteks, analisis capaian pembelajaran dan indikator soal, penyusunan stimulus kontekstual, serta pengembangan butir sesuai level kognitif yang diharapkan. Materi kedua, Penyusunan Soal TKA Berbantuan AI, mencakup pengenalan AI dalam pendidikan, teknik prompting, praktik penyusunan soal berbantuan AI, serta evaluasi dan penyempurnaan kualitas soal yang dihasilkan AI. Teknik prompting diperkenalkan melalui akronim KORAN LAPANGAN, yaitu Konteks, Peran, Larangan, Panjang, dan Tambahan, sebagai kerangka praktis agar guru dapat menuliskan instruksi yang spesifik dan terukur.

Instrumen yang digunakan meliputi instrumen pretest dan posttest berbasis Google Form untuk mengukur pemahaman awal dan akhir peserta (Lihat Tabel 1), lembar tugas penyusunan soal TKA sebagai panduan praktik, serta angket respon

peserta yang terdiri atas sepuluh pernyataan dengan skala Likert 1 sampai 5 dan tiga pertanyaan terbuka (Lihat Tabel 2). Seluruh instrumen divalidasi oleh anggota tim sebelum digunakan.

Tabel 1. Struktur Instrumen Pretest dan Posttest

No	Aspek yang diukur	Indikator butir	Bentuk butir	Kunci
1	Karakteristik TKA sebagai asesmen terstandar	Peserta menentukan pernyataan yang paling tepat menggambarkan karakteristik TKA	Pilihan ganda	Asesmen terstandar berbasis komputer untuk mengukur capaian akademik secara objektif
2	Cakupan elemen TKA Matematika SMP	Peserta memilih seluruh elemen materi yang termasuk cakupan TKA Matematika SMP	Isian centang (jawaban ganda)	Bilangan; Aljabar; Geometri dan Pengukuran; Data dan Peluang
3	Identifikasi level pengetahuan dan pemahaman	Peserta memasang ciri melakukan perhitungan atau prosedur rutin dan memahami informasi dari tabel atau grafik dengan level kognitifnya	Menjodohkan	Pengetahuan dan Pemahaman
4	Identifikasi level aplikasi	Peserta memasang ciri memodelkan masalah kontekstual dan menerapkan strategi pada masalah rutin dengan level kognitifnya	Menjodohkan	Aplikasi
5	Identifikasi level penalaran	Peserta memasang ciri mengaitkan beberapa konsep untuk menyelesaikan masalah non-rutin serta mengevaluasi strategi dengan level kognitifnya	Menjodohkan	Penalaran
6	Pengenalan bentuk soal MCMA	Peserta menentukan istilah bentuk soal yang menuntut siswa memilih semua pilihan benar	Pilihan ganda	MCMA (Multiple Choice Multiple Answer)
7	Pemahaman cara kerja Large Language Model	Peserta menentukan pernyataan yang tepat mengenai prinsip kerja LLM	Pilihan ganda	Memprediksi kata atau token berikutnya berdasarkan pola bahasa dari data pelatihan

Tabel 2. Struktur Angket Respon Siswa

No	Dimensi	Indikator	Pernyataan angket
1	Kualitas penyelenggaraan	Kejelasan dan keterpahaman penyampaian materi	Materi disampaikan dengan jelas dan mudah dipahami
2	Kualitas penyelenggaraan	Kesesuaian materi dengan kebutuhan penyusunan soal TKA	Materi sesuai kebutuhan penyusunan soal TKA
3	Penguasaan kerangka TKA	Pemahaman karakteristik dan kerangka soal TKA	Memahami karakteristik dan kerangka soal TKA
4	Penguasaan kerangka TKA	Pemahaman prosedur penyusunan kisi-kisi soal TKA	Memahami cara menyusun kisi-kisi soal TKA
5	Keterampilan pemanfaatan AI	Pemahaman teknik penyusunan prompt yang efektif	Memahami cara menyusun prompt yang efektif
6	Keterampilan pemanfaatan AI	Keterikutan dan keterpraktikan contoh penggunaan AI	Contoh penggunaan AI mudah diikuti dan dipraktikkan



No	Dimensi	Indikator	Pernyataan angket
7	Kemampuan validasi	Pemahaman prosedur validasi dan penyempurnaan keluaran AI	Memahami cara validasi dan penyempurnaan soal hasil AI
8	Dampak terhadap produk	Kontribusi kegiatan terhadap mutu soal yang dihasilkan	Kegiatan membantu menghasilkan soal yang lebih berkualitas
9	Dampak afektif	Peningkatan kepercayaan diri dalam memanfaatkan AI	Meningkatkan kepercayaan diri memanfaatkan AI
10	Keterterapan	Kebermanfaatan dan keterterapan materi di sekolah	Materi bermanfaat dan dapat diterapkan di sekolah

Analisis peningkatan pemahaman dilakukan terhadap tujuh butir objektif yang identik pada pretest dan posttest agar perbandingan bersifat setara. Skor dikonversi ke skala 100, kemudian dihitung N-gain ternormalisasi dengan kriteria rendah, sedang, dan tinggi, serta diuji dengan uji-t berpasangan. Keterampilan prompting dianalisis melalui kelengkapan tujuh komponen prompt, yaitu jenjang atau kelas, materi, bentuk soal, level kognitif, konteks, kunci dan pembahasan, serta alasan distraktor. Respon peserta dianalisis secara deskriptif berdasarkan rerata skor dan proporsi jawaban positif.

Kegiatan dinyatakan berhasil apabila minimal 70% peserta memberikan respon positif terhadap pelaksanaan kegiatan, minimal 70% peserta mampu menyusun soal TKA berbasis konteks berbantuan AI dengan kategori baik, dan terbentuk bank soal TKA yang terdiri atas minimal 20 butir soal berkualitas.

IMPLEMENTASI KEGIATAN DAN PEMBAHASAN

Koordinasi Awal dan Analisis Kebutuhan Mitra

Tim pelaksana melakukan koordinasi dengan Dinas Pendidikan Kabupaten Jombang pada 13 Mei 2026. Koordinasi tersebut menghasilkan kesepakatan mengenai sasaran peserta, dukungan fasilitas, dan penetapan pelaksanaan kegiatan pada Sabtu, 25 Juli 2026. Melalui diskusi dengan pengurus MGMP, teridentifikasi bahwa kebutuhan paling mendesak guru bukan pada penguasaan materi matematika, melainkan pada penerjemahan capaian pembelajaran menjadi butir soal yang benar-benar membedakan level kognitif, serta pada keterampilan teknis memanfaatkan AI. Temuan ini menjadi dasar penyusunan komposisi materi, yakni satu sesi untuk kerangka TKA dan satu sesi untuk praktik prompting.



Gambar 2. Koordinasi Awal dengan Dinas Pendidikan Kab. Jombang

Hasil pretest memperkuat arah tersebut. Sebagaimana disajikan pada Tabel 3, penguasaan peserta relatif baik pada butir yang bersifat pengenalan, seperti cakupan elemen TKA Matematika SMP yang dijawab benar oleh 88% peserta, tetapi menurun tajam pada butir yang menuntut pembedaan level kognitif. Pola ini menegaskan bahwa kesulitan guru terletak pada tataran operasional penyusunan butir, bukan pada pengenalan kebijakan TKA secara umum.

Penyusunan Instrumen dan Materi Pelatihan

Sebelum pelaksanaan, tim menyusun instrumen pretest dan posttest, materi kerangka TKA, materi penyusunan soal berbantuan AI, serta lembar tugas praktik. Lembar tugas dirancang agar setiap peserta menyusun satu paket soal TKA Matematika SMP yang terdiri atas tiga butir dengan konten sama pada tiga level kognitif berbeda, disertai dokumentasi prompt yang digunakan, keluaran mentah AI, catatan hasil validasi, serta soal final beserta kunci dan pembahasan. Rancangan ini secara sengaja memisahkan keluaran AI dari soal final agar peserta terbiasa memperlakukan hasil AI sebagai draf yang wajib divalidasi, bukan sebagai produk jadi. Seluruh instrumen dan materi disusun dengan mempertimbangkan karakteristik guru peserta dan konteks lokal Kabupaten Jombang agar relevan dan aplikatif.

C. TUGAS PRAKTIK (KHUSUS POSTTEST)

PENYUSUNAN PAKET SOAL TKA MATEMATIKA SMP

BERBANTUAN AI BERDASARKAN LEVEL KOGNITIF

Setelah mengikuti workshop, Bapak/Ibu diminta menyusun satu paket soal TKA Matematika SMP dengan bantuan AI. Paket terdiri atas tiga butir soal yang mengukur satu konten/topik yang sama — yaitu topik yang menurut Bapak/Ibu paling sulit dipahami siswa — pada ketiga level kognitif TKA.

KETENTUAN

- Pilih satu konten/topik Matematika SMP dari elemen Bilangan, Aljabar, Geometri & Pengukuran, atau Data & Peluang yang menurut Bapak/Ibu paling sulit dikuasai siswa. Tuliskan alasan/justifikasi kesulitannya.
- Susun 3 butir soal dengan konten yang sama untuk tiga level kognitif TKA:
- Gunakan bantuan AI dalam proses penyusunan dan dokumentasikan prompt yang digunakan untuk setiap butir.
- Lakukan validasi terhadap hasil AI: kebenaran matematis, kesesuaian level kognitif, kesesuaian dengan jenjang SMP, serta kaidah penulisan soal.

LEVEL KOGNITIF TKA

- Butir 1:** Level Pengetahuan dan Pemahaman
- Butir 2:** Level Aplikasi
- Butir 3:** Level Penalaran

FORMAT PENGUMPULAN

1. Identitas - Nama lengkap - Asal sekolah	2. Topik/Konten dan Justifikasi - Topik/konten yang dipilih - Alasan/justifikasi kesulitannya bagi siswa	3. Untuk Tiap Butir Soal (1, 2, dan 3) <table style="width: 100%;"> <tr> <td style="width: 20%;">(a) Prompt yang digunakan</td> <td style="width: 20%;">(b) Output mentah dari AI</td> <td style="width: 20%;">(c) Catatan hasil validasi</td> <td style="width: 40%;">(d) Soal Final + Kunci + Pembahasan</td> </tr> </table> Pastikan soal sesuai level kognitif, benar secara matematis, sesuai jenjang SMP, dan mengikuti kaidah penulisan soal TKA.				(a) Prompt yang digunakan	(b) Output mentah dari AI	(c) Catatan hasil validasi	(d) Soal Final + Kunci + Pembahasan
(a) Prompt yang digunakan	(b) Output mentah dari AI	(c) Catatan hasil validasi	(d) Soal Final + Kunci + Pembahasan						
4. Refleksi Singkat (3–5 kalimat) Tuliskan kendala yang dihadapi dan manfaat penggunaan AI dalam menyusun soal TKA.									

LINK PENGUMPULAN TUGAS

<http://bit.ly/TKAdenganAI>

DEADLINE

31 Juli 2026,
Pukul 15.00 WIB

Tugas Sebagai Syarat untuk Mendapatkan Sertifikat

★ Kerjakan dengan sungguh-sungguh dan manfaatkan AI secara bijak untuk hasil terbaik!

Gambar 3. Lembar Tugas Penyusunan Soal TKA Berbantuan AI

Pelaksanaan Pelatihan

Kegiatan diawali registrasi peserta, dilanjutkan pembukaan dan sambutan, kemudian pengerjaan pretest untuk mengukur pemahaman awal. Kegiatan inti terdiri atas dua sesi materi. Sesi pertama membahas Framework Tes Kemampuan Akademik Matematika SMP, meliputi konsep, karakteristik, dan kriteria penyusunan soal TKA berbasis konteks. Sesi kedua membahas penyusunan soal TKA berbantuan AI, yang membekali peserta dengan teknik prompting serta praktik langsung pemanfaatan AI. Kedua sesi dilanjutkan dengan tanya jawab bersama narasumber, pengerjaan posttest, penjelasan tugas praktik dan mekanisme pengumpulan, serta penutupan.

Pada sesi tanya jawab, isu yang paling banyak diangkat peserta adalah bagaimana memastikan soal hasil AI benar-benar sesuai level kognitif yang dituju. Tim menegaskan bahwa AI bekerja dengan memprediksi kata atau token berikutnya berdasarkan pola bahasa dari data pelatihan, sehingga keluarannya dapat tampak meyakinkan tetapi belum tentu akurat secara matematis. Penegasan ini penting karena, sebagaimana ditunjukkan pada Tabel 1, hampir sepertiga peserta pada pretest masih memahami AI sebagai mesin pencari jawaban dari basis data soal atau sebagai mesin komputasi simbolik yang selalu akurat. Miskonsepsi semacam ini berpotensi membuat guru menerima keluaran AI tanpa verifikasi, sehingga pembahasan mengenai validasi dan etika penggunaan AI ditempatkan sebagai bagian yang tidak terpisahkan dari sesi praktik (Papagiannidis et al., 2025).



Gambar 4. Pelaksanaan Pelatihan Bersama Guru MGMP Matematika SMP Kabupaten Jombang

Peningkatan Pemahaman Peserta

Perbandingan capaian peserta pada tujuh butir objektif yang identik antara pretest dan posttest disajikan pada Tabel 3.

Tabel 3. Persentase Jawaban Benar per Butir pada Pretest dan Posttest

Aspek yang diukur	Pretest (%)	Posttest (%)	Selisih
Karakteristik TKA sebagai asesmen terstandar	72	94	+22
Cakupan elemen TKA Matematika SMP	88	97	+9
Identifikasi level pengetahuan dan pemahaman	66	81	+15
Identifikasi level aplikasi	53	74	+21
Identifikasi level penalaran	56	74	+18
Pengenalan bentuk soal MCMA	66	97	+31

Aspek yang diukur	Pretest (%)	Posttest (%)	Selisih
Pemahaman cara kerja Large Language Model	69	74	+5

Keterangan: pretest n = 32; posttest n = 31.

Peningkatan terbesar terjadi pada pengenalan bentuk soal MCMA sebesar 31 poin dan pemahaman karakteristik TKA sebesar 22 poin, yaitu dua aspek yang bersifat konseptual dan dapat langsung dikoreksi melalui pemaparan materi. Kemampuan mengidentifikasi level aplikasi dan penalaran, yang semula menjadi titik terlemah, juga meningkat masing-masing 21 dan 18 poin. Sebaliknya, pemahaman tentang cara kerja Large Language Model hanya naik 5 poin dan tetap menjadi butir dengan capaian akhir terendah bersama level kognitif. Temuan ini mengindikasikan bahwa pemahaman konseptual tentang mekanisme kerja AI memerlukan pembahasan yang lebih intensif dibanding keterampilan prosedural, sejalan dengan pandangan bahwa integrasi AI dalam pendidikan menuntut pemahaman kritis terhadap cara kerja teknologi tersebut, bukan sekadar kemampuan mengoperasikannya (Tan et al., 2025). Pola capaian pada level kognitif juga sejalan dengan temuan Akdeniz et al. (2025), yang menunjukkan bahwa butir soal hasil AI selaras dengan tugas kognitif tingkat rendah, sedangkan pada tugas tingkat tinggi butir buatan guru tetap dinilai lebih selaras dengan taksonomi kognitif. Ringkasan capaian 27 peserta yang mengikuti kedua tes disajikan pada Tabel 4.

Tabel 4. Ringkasan Peningkatan Pemahaman Peserta (n = 27)

Aspek	Hasil
Rerata skor pretest	69,8
Rerata skor posttest	87,3
Selisih rerata	17,5
N-gain ternormalisasi	0,46 (sedang)
Uji-t berpasangan	t = 3,65; p = 0,001
Peserta dengan skor meningkat	17 orang (63,0%)
Peserta dengan skor tetap	6 orang (22,2%)
Peserta dengan skor menurun	4 orang (14,8%)

Rerata capaian peserta meningkat dari 69,8 menjadi 87,3 pada skala 100, dengan N-gain 0,46 yang termasuk kategori sedang. Uji-t berpasangan menunjukkan perbedaan yang signifikan secara statistik, sehingga peningkatan tersebut tidak dapat dijelaskan sebagai variasi acak semata. Sebanyak 63,0% peserta mengalami kenaikan skor, sementara 14,8% mengalami penurunan. Penurunan pada sebagian kecil peserta patut dicermati dan diduga berkaitan dengan kondisi pengerjaan posttest yang dilaksanakan di akhir rangkaian kegiatan menjelang tengah hari, ketika konsentrasi peserta telah menurun. Temuan ini menjadi catatan bagi penjadwalan evaluasi pada kegiatan berikutnya.

Perubahan Kualitas Prompt yang Disusun Peserta

Selain penguasaan konsep, indikator yang lebih dekat dengan keterampilan praktis adalah kualitas prompt yang disusun peserta. Hasil analisis kelengkapan komponen prompt disajikan pada Tabel 5.

Tabel 5. Kualitas Prompt yang Disusun Peserta

Indikator	Pretest	Posttest	Selisih
Rerata komponen prompt (dari 7)	2,66	4,13	+1,47
Prompt memuat 5 komponen atau lebih	25,0%	54,8%	+29,8



Indikator	Pretest	Posttest	Selisih
Rerata panjang prompt (kata)	22	51	+29

Rerata kelengkapan prompt meningkat dari 2,66 menjadi 4,13 dari tujuh komponen, dan proporsi peserta yang mampu menyusun prompt dengan minimal lima komponen naik lebih dari dua kali lipat, dari 25,0% menjadi 54,8%. Perubahan tersebut juga tampak secara kualitatif. Pada pretest, prompt yang dituliskan peserta umumnya berupa instruksi singkat seperti permintaan membuat soal materi peluang tanpa menyebutkan kelas, level kognitif, maupun komponen keluaran yang diharapkan. Pada posttest, sebagian besar peserta telah menuliskan prompt berstruktur yang secara eksplisit memuat peran yang diminta dijalankan AI, sub kompetensi, elemen, level kognitif, bentuk soal, konteks, batasan panjang, larangan, serta permintaan kunci jawaban, pembahasan, dan alasan distraktor. Struktur ini mencerminkan penerapan kerangka KORAN LAPANGAN yang diperkenalkan pada sesi kedua dan sejalan dengan prinsip bahwa prompt yang efektif bersifat spesifik, kontekstual, dan menyertakan format keluaran yang diharapkan (Lee & Palmer, 2025).

Meskipun demikian, capaian 54,8% pada indikator prompt lengkap masih berada di bawah target 70%. Analisis terhadap pertanyaan terbuka menunjukkan bahwa kendala utama bukan pada pemahaman konsep prompting, melainkan pada keterbatasan waktu praktik. Sejumlah peserta menyatakan bahwa sesi praktik terasa singkat sehingga belum sempat menguji dan merevisi prompt secara berulang. Hal ini konsisten dengan karakteristik keterampilan prompting yang bersifat iteratif dan menuntut siklus coba dan revisi berulang, sehingga penguasaannya sulit dicapai dalam satu sesi tatap muka. Kesenjangan antara ketersediaan pelatihan dan kebutuhan guru dalam mengintegrasikan AI memang menjadi catatan utama kajian mutakhir mengenai pengembangan profesional guru (Li et al., 2025).

Di Pasar Tradisional, toko Kain Batik memberi promo setiap 12 hari sekali dan toko Keramik memberi promo setiap 18 hari sekali. Hari ini kedua toko sama-sama sedang promo.

Pernyataan berikut yang **BENAR** adalah ...
(pilih semua jawaban yang benar)

- A Promo kedua toko akan bersamaan lagi dalam 6 hari.
- B Promo kedua toko akan bersamaan lagi dalam 36 hari.
- C Dalam 30 hari ke depan, promo bersamaan terjadi 1 kali.
- D Dalam 60 hari ke depan, promo bersamaan terjadi 3 kali.
- E FPB dari 12 dan 18 adalah 6.

KUNCI JAWABAN
B, D, E

PEMBAHASAN
KPK dari 12 dan 18 menentukan kapan promo bersamaan terjadi lagi.
Faktorisasi: $12 = 2^2 \times 3$, $18 = 2 \times 3^2$
 $KPK = 2^2 \times 3^2 = 4 \times 9 = 36$ hari $\rightarrow$ jadi pernyataan B benar.
• Promo bersamaan setiap 36 hari.
• Dalam 30 hari ke depan (tidak sampai 36), terjadi 0 kali $\rightarrow$ C salah.
• Dalam 60 hari terjadi pada hari ke-36 dan ke-72 (di luar 60), jadi hanya 1 kali? Tunggu: hitung dari hari ini.
Mulai hari ini (hari ke-0), dalam 60 hari ke depan terjadi pada hari ke-36 saja $\rightarrow$ 1 kali? Tetapi pernyataan D mengatakan 3 kali. Perhatikan: hari ke-0, 36, 72, ... Dalam rentang 60 hari (hari ke-0 sampai hari ke-60) terjadi di hari ke-0 dan ke-36 $\rightarrow$ 2 kali. Namun soal biasanya menghitung setelah hari ini. Jika setelah hari ini, maka hanya 1 kali. Tetapi pilihan D menyatakan 3 kali $\rightarrow$ salah.
• (A) 6 hari adalah FPB, bukan KPK $\rightarrow$ salah.
• FPB(12,18) = 6 $\rightarrow$ E benar.

ALASAN DISTRAKSI

- A Menggunakan FPB (6) sebagai waktu bersamaan, padahal yang tepat adalah KPK.
- C Menghitung kurang tepat: mengira dalam 30 hari pasti terjadi 1 kali.
- D Kekeliruan menghitung banyak kejadian dalam rentang waktu.

Gambar 5. Hasil Tugas Peserta workshop

Pada aspek validasi keluaran AI, respons peserta menunjukkan kemajuan pada kualitas argumentasi. Alasan yang dituliskan peserta pada pretest cenderung singkat dan umum, sedangkan pada posttest sebagian peserta telah mampu menunjukkan letak kesalahan secara spesifik, termasuk mengidentifikasi ketidaksesuaian antara perintah soal, kunci jawaban, dan pembahasan yang dihasilkan AI, serta menamainya sebagai

bentuk halusinasi AI yang tampak meyakinkan tetapi keliru secara konsep. Kemampuan mengenali kekeliruan semacam ini merupakan inti dari peran guru sebagai penjamin mutu butir soal, sebagaimana ditegaskan bahwa teknologi AI diposisikan sebagai alat bantu dan bukan pengganti pertimbangan profesional pendidik (Nyaaba & Zhai, 2024).

Respon Peserta terhadap Pelaksanaan Kegiatan

Angket respon diisi oleh 19 peserta dengan hasil sebagaimana disajikan pada Tabel 6.

Tabel 6. Respon Peserta terhadap Pelaksanaan Kegiatan (n = 19)

Pernyataan	Rerata	Respon positif
Materi disampaikan dengan jelas dan mudah dipahami	4,68	95%
Materi sesuai kebutuhan penyusunan soal TKA	4,58	89%
Memahami karakteristik dan kerangka soal TKA	4,32	89%
Memahami cara menyusun kisi-kisi soal TKA	4,26	84%
Memahami cara menyusun prompt yang efektif	4,47	89%
Contoh penggunaan AI mudah diikuti dan dipraktikkan	4,53	89%
Memahami cara validasi dan penyempurnaan soal hasil AI	4,47	89%
Kegiatan membantu menghasilkan soal yang lebih berkualitas	4,53	89%
Meningkatkan kepercayaan diri memanfaatkan AI	4,47	89%
Materi bermanfaat dan dapat diterapkan di sekolah	4,63	89%
Rerata keseluruhan	4,49	89,5%

Keterangan: skala 1 sampai 5; respon positif adalah jawaban setuju dan sangat setuju.

Rerata respon keseluruhan sebesar 4,49 dengan 89,5% jawaban berada pada kategori setuju dan sangat setuju menunjukkan bahwa indikator keberhasilan pertama, yaitu minimal 70% peserta memberikan respon positif, telah terlampaui. Skor tertinggi diperoleh pada kejelasan penyampaian materi dan kebermanfaatan materi untuk diterapkan di sekolah. Skor terendah, meskipun tetap tinggi, terdapat pada pemahaman cara menyusun kisi-kisi soal TKA. Pola ini konsisten dengan hasil posttest yang menunjukkan bahwa aspek teknis penyusunan butir memerlukan latihan lebih panjang dibanding aspek pengenalan konsep.

Pada pertanyaan terbuka mengenai materi yang paling bermanfaat, jawaban peserta terpusat pada teknik penyusunan prompt, khususnya kerangka KORAN LAPANGAN, serta pada kegiatan verifikasi butir soal hasil AI. Adapun bagian yang dinilai masih perlu diperjelas berkaitan dengan penyusunan prompt yang lebih detail dan perbedaan antara soal level aplikasi dan penalaran, yang memang merupakan dua butir dengan capaian posttest terendah. Konvergensi antara data kuantitatif dan jawaban terbuka ini memberi arah yang jelas bagi penajaman materi pada kegiatan berikutnya.

Evaluasi Ketercapaian dan Kendala Pelaksanaan

Ditinjau dari tiga indikator keberhasilan, indikator pertama telah tercapai dengan respon positif 89,5%. Indikator kedua tercapai pada aspek penguasaan konsep, ditunjukkan oleh rerata posttest 87,3 yang melampaui ambang 70, namun baru tercapai sebagian pada aspek keterampilan menyusun prompt lengkap sebesar 54,8%. Indikator ketiga, yaitu terbentuknya bank soal minimal 20 butir, masih dalam proses penyelesaian melalui tugas praktik dan pendampingan pasca pelatihan yang dijadwalkan secara daring.



Kendala yang teridentifikasi selama pelaksanaan meliputi tiga hal. Pertama, alokasi waktu praktik dinilai terlalu singkat oleh sebagian peserta sehingga proses uji coba dan revisi prompt belum tuntas. Kedua, keterbatasan koneksi internet di lokasi kegiatan sempat menghambat kelancaran praktik pemanfaatan AI secara serentak. Ketiga, terdapat catatan peserta mengenai ketepatan waktu pelaksanaan dan kejelasan alur registrasi. Perbaikan yang direncanakan mencakup penambahan durasi sesi praktik, pemilihan lokasi dengan dukungan jaringan yang memadai, penyediaan panitia pendamping pada meja registrasi, serta penyelenggaraan sesi lanjutan secara daring agar peserta memperoleh kesempatan iterasi yang lebih panjang.

Secara keseluruhan, kegiatan ini menunjukkan bahwa pendekatan pelatihan yang memadukan penguatan kerangka asesmen dengan praktik langsung pemanfaatan AI efektif meningkatkan kompetensi guru dalam menyusun soal TKA berbasis konteks. Pemosisian AI sebagai alat bantu yang keluarannya wajib divalidasi menjadi kunci agar peningkatan efisiensi tidak mengorbankan mutu butir soal.

KESIMPULAN

Kegiatan pendampingan ini mencapai tujuan yang ditetapkan, yaitu meningkatnya kompetensi guru matematika SMP di Kabupaten Jombang dalam menyusun soal TKA berbasis konteks berbantuan AI. Peningkatan tersebut berlangsung pada dua tataran. Pada tataran konseptual, guru menjadi lebih mampu mengenali karakteristik TKA sebagai asesmen terstandar, membedakan ketiga level kognitif, serta mengidentifikasi bentuk soal yang digunakan di dalamnya, dengan capaian peningkatan berkategori sedang. Pada tataran operasional, guru mampu menyusun prompt yang lebih spesifik dan terstruktur serta memperlakukan keluaran AI sebagai draf yang wajib divalidasi, bukan sebagai produk yang siap pakai. Tujuan kedua, yaitu menumbuhkan pemahaman tentang pemanfaatan AI secara bertanggung jawab, turut tercapai dan ditandai oleh kemampuan sebagian peserta mengenali kekeliruan keluaran AI serta menunjukkan letak kesalahannya secara spesifik. Respon peserta yang sangat positif menegaskan bahwa program dinilai relevan dengan kebutuhan guru dan dapat diterapkan di sekolah masing-masing.

Temuan pokok kegiatan ini adalah bahwa penguatan kompetensi asesmen berbantuan AI tidak berlangsung merata pada semua aspek. Pemahaman konseptual dapat dibangun secara efektif melalui pemaparan materi dalam satu sesi, sedangkan keterampilan menyusun prompt dan kepekaan membedakan soal level aplikasi dari level penalaran menuntut latihan berulang yang belum terwadahi oleh pelatihan tatap muka sekali jalan. Temuan tersebut menegaskan bahwa pemosisian AI sebagai alat bantu hanya bermakna apabila guru memiliki kemampuan menilai mutu butir yang dihasilkannya. Oleh karena itu, kegiatan lanjutan disarankan memperpanjang porsi praktik atau menyelenggarakan sesi praktik terpisah secara daring, menajamkan materi pada perbedaan level kognitif dan pemahaman cara kerja AI, serta mengembangkan instrumen telaah butir yang dapat digunakan guru secara mandiri untuk memvalidasi keluaran AI. Perluasan pemanfaatan platform AI lain juga perlu dipertimbangkan agar guru memiliki alternatif sesuai ketersediaan fasilitas di sekolah.



UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Negeri Surabaya atas dukungan pendanaan melalui skema Pengabdian kepada Masyarakat Kebijakan Fakultas MIPA, kepada Dinas Pendidikan Kabupaten Jombang atas fasilitasi koordinasi kegiatan, serta kepada pengurus dan seluruh anggota MGMP Matematika SMP Kabupaten Jombang atas partisipasi aktifnya selama pelaksanaan kegiatan.

DAFTAR PUSTAKA

- Akdeniz, H., Clark, T., & Roberts, J. L. (2025). Can AI Generate Questions Aligned with Bloom's Taxonomy? A Framework for Gifted Education to Support Teachers. *Journal of Advanced Academics*, 36(4 Special Issue on Artificial Intelligence in Advanced Academics), 671–694. <https://doi.org/10.1177/1932202X251349917>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Du, H., Sun, Y., Jiang, H., Islam, A. Y. M. A., & Gu, X. (2024). Exploring the effects of AI literacy in teacher learning: an empirical study. *Humanities and Social Sciences Communications*, 11(1), 559-. <https://doi.org/10.1057/s41599-024-03101-6>
- Hemmler, Y. M., & Ifenthaler, D. (2024). Self-regulated learning strategies in continuing education: A systematic review and meta-analysis. *Educational Research Review*, 45, 100629. <https://doi.org/10.1016/J.EDUREV.2024.100629>
- Herrmann-Werner, A., Festl-Wietek, T., Holderried, F., Herschbach, L., Griewatz, J., Masters, K., Zipfel, S., & Mahling, M. (2024). Assessing ChatGPT's Mastery of Bloom's Taxonomy Using Psychosomatic Medicine Exam Questions: Mixed-Methods Study. *Journal of Medical Internet Research*, 26(1). <https://doi.org/10.2196/52113>
- Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: a systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, 22(1), 7-. <https://doi.org/10.1186/S41239-025-00503-7>
- Li, K., Wang, P., & Chen, G. (2025). How can AI be integrated into teacher professional development programs? A systematic review based on an adapted technology-based learning model. *Teaching and Teacher Education*, 168, 105219. <https://doi.org/10.1016/J.TATE.2025.105219>
- Mulyani, A., Masriyah, & Sari, Y. M. (2025). Karakteristik Butir Soal Setara AKM Numerasi Berbasis Jotform untuk Peserta Didik SMP. *MATHEdunesa*, 14(3), 892–901. <https://doi.org/10.26740/MATHEDUNESA.V14N3.P892-901>
- Nazari, M., & Saadi, G. (2024). Developing effective prompts to improve communication with ChatGPT: a formula for higher education stakeholders. *Discover Education*, 3(1), 45-. <https://doi.org/10.1007/S44217-024-00122-W>



- Nyaaba, M., & Zhai, X. (2024). Generative AI Professional Development Needs for Teacher Educators. *Journal of AI*, 8(1), 1–13. <https://doi.org/10.61969/JAI.1385915>
- Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, 34(2), 101885. <https://doi.org/10.1016/J.JSIS.2024.101885>
- Picciano, A. G. (2024). Graduate Teacher Education Students Use and Evaluate ChatGPT as an Essay-Writing Tool. *Online Learning*, 28(2), 1–20. <https://doi.org/10.24059/olj.v28i2.4373>
- Sari, Y. M., Kartowagiran, B., Retnawati, H., & Fiangga, S. (2018). The Characteristics of Digital Assessment Bloom for Indonesian Junior High School. *Proceeding of International Conference on Research, Implementation and Education of Mathematics and Sciences*. <https://www.atlantispress.com/proceedings/miseic-18/25905000>
- Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, 8, 100355. <https://doi.org/10.1016/J.CAEAI.2024.100355>
- Zulkardi, Z., Putri, R. I. I., Sari, N., & Utari, R. S. (2025). Pelatihan dan Pendampingan Guru dalam Pendesainan Soal PISA dan TKA Menggunakan Konteks Kearifan Lokal. *Jurnal Inovasi Penelitian Dan Pengabdian Masyarakat*, 5(2), 401–415. <https://doi.org/10.53621/JIPPMAS.V5I2.661>

