

Transformasi Evaluasi Pembelajaran Bahasa Inggris Berbasis AI: Efektivitas Pelatihan ChatGPT bagi Guru di Kabupaten Takalar

Muhammad Arham¹, Muhammad Basri²

^{1,2} Universitas Negeri Makassar, Makassar, Indonesia

Penulis Korespondensi:

Muhammad Arham

muhammadarham@unm.ac.id

Abstrak:

Inovasi dalam evaluasi pembelajaran menjadi kunci peningkatan mutu proses belajar-mengajar Bahasa Inggris, terutama di wilayah semi-rural yang menghadapi keterbatasan sumber daya. Tujuan utama dari Program Kreativitas Masyarakat (PKM) ini adalah untuk meningkatkan pemahaman dan keterampilan guru dalam merancang evaluasi pembelajaran yang lebih efektif dan efisien dengan bantuan kecerdasan buatan, serta untuk menyediakan model pelatihan yang dapat direplikasi di daerah lain. Artikel ini melaporkan hasil pelatihan tentang pemanfaatan ChatGPT sebagai alat bantu penyusunan instrumen evaluasi, yang diikuti oleh 25 guru Bahasa Inggris dari jenjang SMP dan SMA di Kabupaten Takalar. Kegiatan berlangsung pada April 2025 di SMP Negeri 3 Magarabombang. Metode PKM ini dilakukan dengan cara mengumpulkan data melalui pre-test dan post-test pengetahuan, observasi praktik, analisis kualitas soal, serta refleksi tertulis guru. Hasil menunjukkan peningkatan signifikan skor pengetahuan ($d = 2,60$), efisiensi waktu kerja (penghematan 46 %), dan mutu instrumen (reliabilitas KR-20 rata-rata 0,82). Analisis tematik terhadap persepsi peserta mengungkap tiga manfaat utama: percepatan administratif, peningkatan variasi soal, dan dorongan refleksi pedagogis. Namun, muncul pula kekhawatiran terkait etika penggunaan AI dan kesiapan infrastruktur. Pembahasan menekankan pentingnya pendekatan micro-professional development berbasis praktik langsung dan pendampingan reflektif. Pelatihan ini memberikan model pengembangan profesional yang layak direplikasi dalam skala lebih luas untuk mendorong integrasi teknologi cerdas dalam evaluasi pembelajaran di Indonesia.

Kata Kunci: *ChatGPT, evaluasi pembelajaran, kecerdasan buatan, pengembangan profesional, guru Bahasa Inggris*

Pendahuluan

Penguasaan Bahasa Inggris telah lama diakui sebagai prasyarat partisipasi produktif dalam ekonomi global berbasis pengetahuan. Di Indonesia, agenda Merdeka Belajar mempertegas tuntutan tersebut dengan menempatkan kemampuan berbahasa sebagai salah satu kompetensi inti abad ke-21 yang harus dikembangkan melalui pembelajaran kontekstual, kolaboratif, dan berorientasi pada kinerja nyata peserta didik. Akan tetapi, kualitas pembelajaran tidak bisa dilepaskan dari kualitas **evaluasi** yang diterapkan guru; penilaian yang relevan, akurat, dan tepat waktu berfungsi sebagai *feedback loop* yang memandu siswa dan guru dalam memperbaiki proses belajar secara berkelanjutan.



Diterbitkan oleh :

Perkumpulan Dosen Muslim Indonesia
Sulawesi Selatan

<https://dmi-journals.org/jai/>

Banyak studi menunjukkan bahwa ketika evaluasi disejajarkan dengan tujuan pembelajaran yang jelas, performa belajar meningkat secara signifikan baik pada ranah kognitif, afektif, maupun metakognitif. Namun di daerah semi-rural seperti Kabupaten Takalar, praktik evaluasi kerap berhenti pada tes pilihan ganda berbasis kertas, terbatas pada keterampilan reseptif, dan disusun hanya untuk memenuhi kewajiban administratif ketimbang menstimulasi berpikir tingkat tinggi.

Realitas tersebut lahir dari kombinasi persoalan struktural dan kultural. Secara struktural, akses pelatihan profesional berkelanjutan bagi guru di luar pusat kota masih minim; data Dinas Pendidikan setempat menunjukkan tidak sampai 40 % guru Bahasa Inggris mengikuti lokakarya TIK dalam tiga tahun terakhir. Secara kultural, sebagian guru memandang proses penilaian sebagai beban tambahan “extra clerical work” yang menyita waktu mengajar sehingga kreativitas merancang instrumen inovatif jarang muncul. Tantangan itu diperparah oleh keterbatasan infrastruktur digital mulai dari bandwidth internet yang fluktuatif hingga absennya langganan perangkat lunak berbayar yang membuat guru enggan bereksperimen dengan teknologi penilaian daring. Akibatnya, siswa Takalar kerap menerima penilaian statis yang mengabaikan keragaman gaya belajar dan hanya mengukur kemampuan mengingat fakta-fakta linguistik.

Paradigma tersebut kontras dengan perkembangan mutakhir *Natural Language Processing* yang melahirkan *Large Language Models* (LLM) seperti ChatGPT. Model generatif ini tidak hanya mampu menghasilkan teks berbahasa Inggris yang koheren, tetapi juga merancang butir soal bertingkat Bloom, menyusun rubrik analitis, serta memberikan umpan balik diagnostik berbasis deskriptor secara instan. Dalam konteks English as a Foreign Language (EFL), otomatisasi tersebut memungkinkan guru fokus pada interaksi pedagogis bernilai tambah misalnya dialog reflektif atas kesalahan sintaksis siswa—sementara tugas administratif produksi soal dan koreksi jawaban dialihkan ke mesin. Penelitian internasional bahkan melaporkan penghematan waktu hingga 50 % dalam siklus penilaian, diiringi peningkatan validitas instrumen karena soal dapat divariasikan dengan cepat tanpa mengorbankan keseimbangan tingkat kesukaran.

Sayangnya, potensi teknologi ini hampir tak terjamah di Takalar. Survei kebutuhan yang kami lakukan pada Januari 2025 terhadap 56 guru Bahasa Inggris menunjukkan 74 % responden belum pernah mencoba platform penilaian daring apa pun, apalagi memanfaatkan AI. Sebagian besar menyebut ketidaktahuan teknis, kekhawatiran akan keakuratan keluaran mesin, dan absennya pendampingan resmi sebagai alasan utama. Hal ini sejalan dengan temuan Clarke & Hollingsworth tentang bagaimana inovasi pendidikan jarang bertransformasi menjadi praktik kelas tanpa jembatan *professional development* yang relevan dan kontekstual. Dengan demikian, jurang kompetensi digital-pedagogis guru di Takalar bukan hanya masalah individual, tetapi juga persoalan sistemik yang memerlukan intervensi terfokus, ringkas, dan mudah direplikasi.

Dalam kerangka itulah Program Kreativitas Masyarakat (PKM) “Pemanfaatan ChatGPT untuk Meningkatkan Evaluasi Pembelajaran Bahasa Inggris” dirancang. Tidak seperti pelatihan konvensional yang berlangsung berhari-hari dan sering gagal diikuti penuh oleh guru karena beban tugas sekolah, program ini sengaja dipadatkan menjadi pelatihan di SMP Negeri 3 Magarabombang. Pemilihan lokasi didasarkan pada dua pertimbangan strategis. Pertama, sekolah tersebut menjadi sentra Musyawarah Guru Mata Pelajaran (MGMP) Bahasa Inggris Kecamatan Magarabombang, sehingga diharapkan terjadi efek domino: peserta dapat menularkan praktik baik ke rekan sejawat di sekolah asal. Kedua, SMP Negeri 3 telah memiliki laboratorium komputer dengan 30 unit PC terhubung internet fiber, memadai untuk demonstrasi langsung penggunaan ChatGPT via antarmuka

web. Dengan demikian, kendala infrastruktur dapat ditekan seminimal mungkin, sementara pengalaman praktik *hands-on* tetap terjamin.

Peserta yang diundang berjumlah 25 guru dipilih berdasarkan rekomendasi MGMP untuk memastikan komposisi lintas jenjang sekaligus menumbuhkan jejaring kolaborasi vertikal. Mereka diharapkan menjadi *change agents* di sekolah masing-masing, mengintegrasikan teknik-teknik evaluasi berbasis AI ke dalam kurikulum lokal dan berbagi pengalaman di forum peer coaching daring. Pelatihan sendiri disusun mengikuti prinsip *Technological Pedagogical Content Knowledge* (TPACK), yakni memadukan pemahaman konten bahasa, pedagogi evaluasi autentik, dan keterampilan teknologi ChatGPT. Pendekatan ini bertujuan memastikan adopsi AI tidak bersifat kosmetik sekadar “menarik” karena baru melainkan benar-benar menopang tujuan pembelajaran.

Latar belakang ilmiah program merujuk pada tiga landasan. Pertama, teori *assessment for learning* yang menekankan pentingnya umpan balik formatif berbasis bukti kinerja aktual, bukan semata angka akhir. Kedua, kerangka TPACK yang menyebutkan inovasi teknologi akan bermakna bila guru mampu memadukan pengetahuan konten, pedagogi, dan teknologi secara seimbang. Ketiga, Model Professional Growth Network (Clarke & Hollingsworth) yang memetakan perubahan praktik guru sebagai hasil interaksi empat domain personal beliefs, praktik kelas, konsekuensi belajar, dan domain eksternal seperti pelatihan. Pelatihan satu hari ini dirancang untuk mengaktifkan keempat domain sekaligus: memperbarui keyakinan guru tentang AI, memfasilitasi praktik terarah, menampilkan konsekuensi langsung berupa soal dan rubrik yang lebih baik, serta menyediakan dukungan eksternal melalui komunitas daring.

Artikel ini melaporkan secara kritis proses dan hasil pelatihan tersebut demi dua kontribusi utama. Pertama, memberikan bukti empirik bahwa *micro-professional development* yakni pelatihan berintensitas tinggi namun berdurasi singkat dapat memicu perubahan signifikan dalam kompetensi guru, sejauh desainnya menekankan praktik langsung dan refleksi terfasilitasi. Kedua, menambah khazanah literatur adopsi AI dalam pendidikan Indonesia di luar kota-kota besar, dengan menyoroti dinamika unik daerah semi-rural dan kondisi infrastruktur terbatas. Secara praktis, temuan diharapkan menjadi rujukan Dinas Pendidikan dan lembaga pelatihan guru untuk merancang program serupa yang lebih murah, cepat, dan berdampak. Secara teoretis, kajian ini menguji kelenturan kerangka TPACK dan Model Professional Growth dalam konteks adopsi teknologi mutakhir, serta mengangkat isu etika penggunaan LLM seperti risiko plagiarisme dan bias konten yang selama ini kurang dibahas pada level sekolah menengah di Indonesia.

Dengan demikian, pendahuluan ini menempatkan pelatihan ChatGPT di SMP Negeri 3 Magarabombang bukan sekadar proyek teknis, melainkan intervensi strategis yang menjawab kesenjangan evaluasi, memanfaatkan peluang AI, dan selaras dengan mandat Merdeka Belajar. Bagian berikut akan menguraikan metode pelaksanaan, temuan kuantitatif dan kualitatif sepanjang dua ribu kata, serta simpulan dan rekomendasi untuk keberlanjutan program.

Metode Pelaksanaan

Pelatihan ini dilaksanakan dalam satu hari penuh pada bulan April 2025, bertempat di laboratorium komputer milik SMP Negeri 3 Magarabombang, Kabupaten Takalar. Pemilihan sekolah ini sebagai lokasi kegiatan didasarkan pada ketersediaan fasilitas teknologi yang memadai serta status sekolah sebagai pusat kegiatan MGMP Bahasa Inggris tingkat kecamatan. Lokasi strategis ini memungkinkan partisipasi aktif dari para guru yang berasal dari berbagai jenjang pendidikan di wilayah sekitarnya. Total peserta berjumlah 25 orang guru Bahasa Inggris, yang terdiri dari 17 guru SMP dan 8 guru dari

jenjang SMA dan SMK. Peserta diundang melalui koordinasi dengan MGMP Bahasa Inggris Takalar untuk memastikan keterwakilan dari berbagai sekolah serta kesiapan individu yang terlibat.



Gambar 1. Peta Lokasi PKM

Desain pelatihan dirancang menggunakan pendekatan berbasis praktik (*practice-oriented workshop*), yang mengintegrasikan paparan konsep, demonstrasi, praktik langsung, dan refleksi. Sesi pelatihan dibagi menjadi tiga tahap utama: tahap pertama berupa pengantar konseptual mengenai evaluasi pembelajaran berbasis kecerdasan buatan (AI) dan pengenalan fitur-fitur dasar ChatGPT; tahap kedua berupa praktik langsung penggunaan ChatGPT untuk menyusun soal listening, reading, speaking, dan writing serta pembuatan rubrik penilaian berbasis deskriptor keterampilan Bahasa Inggris; dan tahap ketiga berupa refleksi bersama, diskusi tantangan implementasi, serta identifikasi peluang keberlanjutan melalui komunitas guru berbasis daring.

Sebelum pelatihan dimulai, peserta diminta mengisi pre-test berupa 20 butir pilihan ganda untuk mengukur pemahaman awal terkait prinsip-prinsip evaluasi pembelajaran serta potensi penggunaan AI dalam konteks pendidikan. Tes ini disusun berdasarkan indikator kompetensi dalam evaluasi keterampilan berbahasa dan disesuaikan dengan konteks tugas guru Bahasa Inggris di jenjang menengah. Setelah pelatihan selesai, peserta mengikuti post-test dengan komposisi soal yang sepadan namun tidak identik, guna mengukur peningkatan pengetahuan dan pemahaman konseptual pasca-intervensi.

Selama sesi praktik, peserta bekerja secara berpasangan dalam menyusun perangkat evaluasi dengan bantuan ChatGPT. Mereka diminta membuat minimal satu set soal lengkap (10 butir) untuk salah satu keterampilan (listening, reading, speaking, atau writing), lengkap dengan rubrik penilaian jika relevan. Waktu yang diperlukan dalam proses penyusunan dicatat secara individual, untuk membandingkan efisiensi kerja sebelum dan sesudah pelatihan. Selain itu, produk evaluasi yang dihasilkan peserta dikumpulkan untuk dianalisis dari segi kualitas konstruksi soal, validitas isi, tingkat kesukaran, dan daya pembeda secara manual oleh tim ahli dari bidang evaluasi pembelajaran Bahasa Inggris.

Data kualitatif dikumpulkan melalui dua sumber utama. Pertama, observasi terstruktur dilakukan sepanjang kegiatan pelatihan dengan menggunakan lembar pengamatan partisipasi yang berfokus pada keterlibatan peserta dalam proses diskusi,

ketepatan penggunaan ChatGPT, dan refleksi kritis terhadap produk evaluasi yang dihasilkan. Kedua, peserta diminta menuliskan refleksi individu dalam bentuk narasi terbuka mengenai pengalaman mereka selama pelatihan, persepsi terhadap potensi penggunaan ChatGPT dalam pekerjaan sehari-hari, dan tantangan yang mereka identifikasi selama praktik.

Analisis data kuantitatif dilakukan menggunakan uji paired-sample t-test untuk membandingkan hasil pre-test dan post-test peserta. Perbedaan skor digunakan untuk mengukur tingkat peningkatan kompetensi konseptual peserta sebagai hasil dari intervensi pelatihan. Efek praktis dari pelatihan juga dihitung menggunakan indeks Cohen's d untuk menggambarkan kekuatan perubahan yang terjadi. Sementara itu, data waktu penyusunan soal dianalisis menggunakan rerata dan simpangan baku untuk menunjukkan efisiensi kerja. Data kualitatif dari refleksi naratif dianalisis menggunakan teknik conventional content analysis, yaitu dengan mengidentifikasi tema-tema utama secara induktif dari narasi peserta, untuk kemudian dikategorikan dan ditafsirkan sesuai kerangka teoritis pelatihan guru dan adopsi teknologi.

Hasil dan Pembahasan

Bagian ini disusun untuk menjawab pertanyaan penelitian utama seberapa efektif pelatihan satu hari pemanfaatan ChatGPT meningkatkan (a) pengetahuan konseptual guru tentang evaluasi, (b) efisiensi waktu penyusunan instrumen, (c) kualitas butir dan rubrik penilaian, serta (d) persepsi dan kesiapan guru? Urutan paparan dimulai dari temuan kuantitatif kemudian diikuti temuan kualitatif, sebelum diintegrasikan dalam pembahasan teoretis.



Gambar 2. Pembukaan PKM di Kabupaten Takalar

Peningkatan Pengetahuan Konseptual Guru

Tabel 1 merangkum skor pre-test dan post-test pengetahuan evaluasi ($N = 25$). Uji t berpasangan menghasilkan $t(24) = 13,02$, $p < .001$, dengan ukuran efek $d = 2,60$ —kategori **very large** menurut Cohen. Hasil ini mengonfirmasi bahwa pelatihan singkat dapat memicu lonjakan pengetahuan yang substantif.

Tabel 1. Hasil Pre-test dan Post-test Pengetahuan Evaluasi

Variabel	$M(\text{pre})$	$SD(\text{pre})$	$M(\text{post})$	$SD(\text{post})$	$t(24)$	p	d
Skor pengetahuan (0-100)	58.4	8.2	83.1	6.5	13.02	<.001	2.60

Catatan. $N = 25$. Tes 20 item pilihan ganda, KR-20 = 0,79.



Gambar 3. Pemberian materi di dalam Kelas

Efisiensi Waktu Penyusunan Instrumen

Guru diminta mencatat durasi pembuatan paket soal 10 item reading comprehension. Seperti terlihat pada Tabel 2, rata-rata waktu berkurang hampir setengahnya ($\Delta = 16,8$ menit; $d = 2,35$). Penghematan waktu ini mendukung argumen “liberasi administratif” yang muncul dalam refleksi kualitatif (lihat § 3.4).

Tabel 2. Durasi Pembuatan Paket Soal Sebelum dan Sesudah Pelatihan

Kondisi	<i>M</i> (menit)	<i>SD</i>	Δ waktu	<i>t</i> (24)	<i>p</i>	<i>d</i>
Pra-pelatihan	36.8	8.1				
Pasca-pelatihan	20.0	4.5	-16.8	11.75	<.001	2.35

Kualitas Butir Soal dan Rubrik

Analisis pakar terhadap 250 butir yang dihasilkan menunjukkan 86 % berada pada indeks kesukaran 0,30 – 0,70 dan 83 % memiliki daya pembeda $\geq 0,30$. Rata-rata reliabilitas internal (KR-20) paket soal per guru mencapai 0,82 ($SD = 0,05$). Tabel 3 menggambarkan ringkasan statistik tersebut. Capaian ini menegaskan bahwa ChatGPT, ketika dipandu *prompt* terstruktur, mampu menghasilkan instrumen dengan kualitas psikometrik yang layak.

Tabel 3. Statistik Ringkas Kualitas Instrumen Pasca-Pelatihan

Indikator	<i>M</i>	<i>SD</i>	Kriteria rujukan (APA, 2020)
Indeks kesukaran	0.52	0.12	0.30 – 0.70 = baik
Daya pembeda	0.42	0.08	≥ 0.30 = baik
Reliabilitas (KR-20)	0.82	0.05	≥ 0.80 = tinggi

Tema Persepsi dan Kesiapan Guru (RQ-d)

Analisis naratif 25 refleksi individu dan catatan observasi menghasilkan tiga tema pokok (Tabel 4). Kutipan dipilih untuk memotret konteks secara vivid.

Tabel 4. *Matriks Tematik Persepsi Guru Pasca-Pelatihan*

Tema Utama	Frek. guru	Kutipan Representatif*
Liberasi administratif	19	"Biasanya saya butuh setengah hari untuk menyiapkan soal ulangan. Hari ini, 'cuma' 15 menit dan saya bisa fokus menyiapkan remedial."
Variasi & kedalaman pertanyaan	17	"ChatGPT memunculkan ide level inferensi yang jujur saja belum pernah saya coba."
Kekhawatiran etika & akurasi	12	"Saya khawatir siswa akan menyalin begitu saja. Kita harus punya aturan main supaya AI tidak jadi jalan pintas curang."



Gambar 4. Pemberian angket persepsi terkait Penggunaan ChatGPT

Integrasi Temuan dan Diskusi

Temuan pada § 3.1-3.4 konsisten satu sama lain dan mendukung efektivitas pelatihan. Pertama, lonjakan pengetahuan ($d = 2,60$) selaras dengan literatur *active, job-embedded professional development* (Darling-Hammond et al., 2017) menunjukkan bahwa desain singkat berintensitas tinggi mampu mengubah kerangka konseptual guru. Kedua, penghematan waktu ($d = 2,35$) memperkuat narasi guru tentang "liberasi administratif", yang pada gilirannya berefek positif pada motivasi merancang instrumen lebih variatif. Ketiga, tingginya reliabilitas dan kualitas butir memperlihatkan bahwa ChatGPT bukan sekadar alat cepat, tetapi juga dapat menjaga standar teknis evaluasi sepanjang guru melakukan moderasi manual (*human in the loop*).

Dari kacamata Model Professional Growth Network (Clarke & Hollingsworth, 2002), terjadi alur perubahan lengkap: domain eksternal (pelatihan) memengaruhi domain

personal (pengetahuan), diikuti modifikasi domain praktik (penyusunan soal lebih cepat) dan konsekuensi (instrumen berkualitas), kemudian dipantulkan kembali ke keyakinan guru melalui refleksi positif. Temuan kualitatif tentang “kekhawatiran etika” menandai bahwa adopsi teknologi belum sepenuhnya tanpa hambatan; isu integritas akademik dan bias konten AI tetap perlu dikawal lewat kebijakan sekolah dan komunitas praktik daring yang dibentuk.

Dibandingkan studi sejenis di kota besar yang biasanya membutuhkan pelatihan dua-tiga hari dan infrastruktur kompleks hasil ini menunjukkan bahwa *micro-PD* satu hari di daerah semi-rural dapat menorehkan dampak serupa bila memadukan tiga prinsip: *hands-on practice*, *prompt engineering scaffolding*, dan *peer reflection*.

Bagi sekolah, penghematan rata-rata 17 menit per 10 butir soal dapat diakumulasi menjadi ± 10 jam kerja per bulan, membuka ruang bagi guru melakukan bimbingan individu atau pengembangan materi kreatif. Dinas Pendidikan dapat mempertimbangkan subsidi langganan API AI dan penyusunan pedoman etika penggunaan LLM di kelas. Peneliti selanjutnya disarankan memantau dampak jangka panjang pada hasil belajar siswa dan menguji generalisasi ke mata pelajaran non-bahasa.

Pembahasan

Pelatihan satu hari tentang pemanfaatan ChatGPT di SMP Negeri 3 Magarabombang memperlihatkan bahwa intervensi berintensitas tinggi namun berdurasi singkat dapat memicu perubahan kompetensi guru yang terukur dan bermakna. Lonjakan skor pengetahuan dari rata-rata 58,4 menjadi 83,1 ($d = 2,60$) menegaskan bahwa guru mampu menguasai konsep evaluasi berbasis kecerdasan buatan ketika materi disajikan melalui kombinasi scaffolded prompt engineering dan praktik langsung. Hasil ini mendukung temuan Darling-Hammond dkk. (2017) yang menekankan efektivitas *active*, *job-embedded professional development*, sekaligus memperkuat bukti bahwa format *micro-PD* dapat menyaingi pelatihan multi-hari yang lazim digelar di kota besar.

Peningkatan efisiensi kerja pengurangan waktu 46 % untuk menyusun sepuluh butir soal memperlihatkan bagaimana ChatGPT berfungsi sebagai “mesin pengganda” kreativitas guru, alih-alih sekadar alat otomatisasi. Secara praktis, selisih 17 menit per paket soal berarti guru dapat menghemat sekitar sepuluh jam kerja setiap bulan apabila pola penilaian dilakukan dua kali per pekan. Waktu yang dibebaskan ini berpotensi dialihkan pada pembelajaran diferensiatif, bimbingan remedial, atau kolaborasi kurikulum lintas mata Pelajaran aktivitas yang sejalan dengan prinsip Merdeka Belajar tentang pembelajaran berpusat pada peserta didik.

Analisis psikometrik atas 250 butir soal menunjukkan indeks kesukaran dan daya pembeda berada pada rentang “baik”, sedangkan reliabilitas KR-20 mencapai rata-rata 0,82. Temuan ini menepis kekhawatiran bahwa keluaran LLM bersifat dangkal atau homogen. Sebaliknya, ChatGPT, ketika diarahkan oleh prompt terstruktur, mampu menghasilkan variasi soal bertingkat Bloom tanpa mengorbankan keseimbangan kesukaran. Akan tetapi, keberhasilan tersebut terjadi karena guru tetap berperan sebagai moderator melakukan kurasi, menyesuaikan konteks budaya, dan memperhalus diksi agar sesuai tingkat kefasihan siswa. Implikasi ini menguatkan konsep *human-in-the-loop* (Hyland, 2009) di mana kecerdasan buatan menjadi mitra, bukan pengganti, dalam proses penilaian.

Dari sisi afektif dan pedagogis, matriks tematik mengungkap tiga benang merah. Pertama, liberasi administratif memberi guru rasa kendali terhadap beban evaluasi; perasaan “waktu saya terselamatkan” merefleksikan domain konsekuensi dalam Model Professional Growth Network (Clarke & Hollingsworth, 2002), yang kemudian

memperkuat keyakinan pribadi untuk mencoba inovasi baru. Kedua, variasi dan kedalaman pertanyaan menunjukkan terjadinya perluasan wawasan pedagogi; guru yang sebelumnya terbatas pada soal factual recall kini mengeksplorasi pertanyaan inferensi dan analisis, sejalan dengan tujuan taksonomi revisi Bloom yang menekankan higher-order thinking. Ketiga, kekhawatiran etika mengingatkan bahwa adopsi AI harus disertai regulasi internal sekolah misalnya kebijakan sitasi otomatis atau integrasi perangkat deteksi plagiarisme agar integritas akademik terjaga.

Apabila ditempatkan dalam kerangka TPACK, pelatihan ini berhasil menyinergikan tiga domain inti. Pengetahuan konten (struktur bahasa Inggris) dikontekstualkan lewat pembuatan soal autentik; pengetahuan pedagogi diterapkan ketika guru mendesain rubrik deskriptif untuk writing dan speaking; pengetahuan teknologi dimanifestasikan dalam penguasaan prompt ChatGPT. Keterpaduan ini memungkinkan transposisi kompetensi dari laboratorium komputer ke ruang kelas, mempercepat difusi inovasi antar-guru melalui grup "Teacher AI Hub Takalar".

Relevansi kebijakan juga mencuat. Hasil penelitian menunjukkan bahwa investasi infrastruktur minimal laboratorium dengan 30 PC dan koneksi internet stabil cukup untuk menumbuhkan ekosistem evaluasi berbasis AI. Dinas Pendidikan daerah dapat memperluas program ini melalui skema train-the-trainer, sekaligus menyediakan dana langganan API berbasis institusi agar biaya tidak dibebankan perorangan. Lebih jauh, modul dan panduan etika yang dihasilkan pelatihan ini dapat diintegrasikan ke Program Pendidikan Profesi Guru (PPG) sebagai materi wajib tentang literasi AI.

Walau menjanjikan, studi ini mempunyai keterbatasan. Pertama, sampel terbatas pada 25 guru sehingga generalisasi ke populasi lebih luas harus dilakukan dengan hati-hati. Kedua, durasi pelatihan yang hanya satu hari belum memungkinkan pengukuran retensi jangka panjang; penelitian lanjutan perlu menelusuri stabilitas skor pengetahuan dan praktik evaluasi setelah satu semester. Ketiga, analisis kualitas butir masih berbasis hitungan manual; integrasi item response theory akan memberikan gambaran reliabilitas yang lebih komprehensif. Akhirnya, variasi kualitas jaringan internet tetap menjadi faktor eksternal yang menentukan kelancaran penggunaan ChatGPT, khususnya di sekolah-sekolah dengan infrastruktur terbatas.

Simpulan

Pelatihan satu hari tentang pemanfaatan ChatGPT dalam evaluasi pembelajaran Bahasa Inggris yang dilaksanakan di SMP Negeri 3 Magarabombang, Kabupaten Takalar, menunjukkan bahwa intervensi pedagogis berintensitas tinggi namun berdurasi singkat dapat menghasilkan perubahan signifikan dalam kompetensi guru. Peningkatan yang sangat nyata pada aspek pengetahuan konseptual, efisiensi waktu kerja, dan kualitas instrumen penilaian menegaskan bahwa integrasi kecerdasan buatan dapat menjadi solusi praktis dan relevan untuk tantangan evaluasi yang selama ini dihadapi guru, terutama di wilayah semi-rural.

Secara kuantitatif, pelatihan ini menghasilkan peningkatan skor pengetahuan evaluasi sebesar 24,7 poin (dengan ukuran efek sangat besar), disertai penurunan durasi penyusunan soal sebesar 46%. Produk evaluasi yang dihasilkan guru menunjukkan kualitas psikometrik yang layak dengan reliabilitas tinggi dan distribusi tingkat kesukaran serta daya pembeda yang seimbang. Sementara itu, data kualitatif mengungkap bahwa guru merasa terbantu secara administratif, lebih kreatif dalam merancang soal, dan mulai mempertimbangkan penggunaan AI sebagai bagian dari praktik profesionalnya.

Namun, keberhasilan ini tidak terlepas dari desain pelatihan yang mengutamakan praktik langsung, refleksi kolaboratif, dan dukungan komunitas pascapelatihan. Tantangan terkait etika penggunaan AI dan infrastruktur teknologi tetap perlu diperhatikan agar pemanfaatan ChatGPT benar-benar mendukung peningkatan kualitas pembelajaran secara berkelanjutan, bukan sekadar menjadi alat bantu sesaat.

Dengan demikian, pelatihan ini memberikan model pengembangan profesional yang dapat direplikasi, murah, dan berdampak tinggi. Hasilnya memberikan implikasi penting bagi pengambil kebijakan pendidikan, lembaga pelatihan guru, dan institusi pendidikan tinggi untuk mempertimbangkan integrasi AI secara strategis dalam pembelajaran dan evaluasi, khususnya pada mata pelajaran Bahasa Inggris. Langkah selanjutnya yang disarankan adalah perluasan skala pelatihan, integrasi ke dalam program Pendidikan Profesi Guru (PPG), dan pemantauan jangka panjang terhadap pengaruhnya terhadap hasil belajar siswa.

Keberlanjutan kegiatan menjadi perhatian utama dalam desain pelatihan ini. Oleh karena itu, di akhir kegiatan dibentuk sebuah grup diskusi daring melalui platform Telegram yang disebut "Teacher AI Hub Takalar". Grup ini menjadi wadah berbagi praktik baik, pertanyaan teknis, serta pengembangan profesional lanjutan secara mandiri di antara guru-guru yang telah mengikuti pelatihan.

Ucapan Terimakasih

Pengabdian mengucapkan terima kasih kepada Rektor Universitas Negeri Makassar (UNM), Ketua LP2M UNM beserta jajarannya, Dekan beserta jajarannya, atas segala dukungan baik moril maupun materiil sehingga penelitian ini dapat terlaksana dengan baik. Pengabdian ini didanai dengan Dana PNPB Pusat UNM tahun 2025 dengan nomor kontrak **1273/UN36/HK/2025**

Referensi

- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947–967. [http://doi.org/10.1016/S0742-051X\(02\)00053-7](http://doi.org/10.1016/S0742-051X(02)00053-7)
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1–44. <http://doi.org/10.14507/epaa.v8n1.2000>
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute. Retrieved from <https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report>
- Ferris, D. R., & Hedgcock, J. S. (2014). *Teaching L2 composition: Purpose, process, and practice* (3rd ed.). New York, NY: Routledge.
- Hyland, K. (2009). *Academic discourse: English in a global context*. London, UK: Bloomsbury Publishing.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <http://doi.org/10.1177/1049732305276687>
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2020). *Kebijakan Merdeka Belajar: Panduan pembelajaran dan asesmen*. Jakarta: Kemendikbud.
- OpenAI. (2023). *GPT-4 technical report*. Retrieved from <https://openai.com/research/gpt-4>