

## **Pelatihan Evaluasi Pembelajaran Berbasis Kecerdasan Buatan sebagai Upaya Penguatan Profesionalisme Guru MTs**

**Mila Yunita <sup>1\*</sup>, Ari Khusumadewi <sup>2</sup>, Mohammad Syahidul Haq <sup>3</sup>, Farah Nikmatus Sania <sup>4</sup>,  
Wasilatur Rahmah Siftia Rusydi <sup>5</sup>**

<sup>1 2 3 4 5</sup> Universitas Negeri Surabaya, Jawa Timur, Indonesia

\* Penulis Korespondensi: [milayunita@unesa.ac.id](mailto:milayunita@unesa.ac.id)

---

**Abstrak:** Pengabdian Kepada Masyarakat (PKM) diimplementasikan untuk mengukur efektivitas program "Pelatihan Evaluasi Pembelajaran Berbasis Kecerdasan Buatan (AI) sebagai Upaya Penguatan Profesionalisme Guru MTs Kabupaten Mojokerto" dalam meningkatkan kompetensi guru. Pengukuran keberhasilan menggunakan pendekatan kuantitatif dengan desain eksperimen semu (*quasi-experiment*) *Pretest-Posttest Control Group Design*. Sampel terdiri dari 60 guru yang dibagi secara acak ke dalam kelompok eksperimen ( $n=30$ ) dan kelompok kontrol ( $n=30$ ). Kelompok eksperimen mengikuti pelatihan intensif, sedangkan kelompok kontrol tidak. Data dikumpulkan menggunakan instrumen tes pengetahuan dan keterampilan yang telah diuji validitas dan reliabilitasnya. Analisis data dilakukan dengan uji prasyarat (normalitas dan homogenitas) serta Independent Samples *t-test* dan Paired Samples *t-test* menggunakan SPSS 26. Hasil Independent Samples *t-test* menunjukkan perbedaan yang sangat signifikan antara skor akhir kedua kelompok ( $t=32,450$ ;  $p=0,000$ ). Rata-rata skor Posttest kelompok eksperimen (87,60) jauh lebih tinggi daripada kelompok kontrol (66,30), dengan selisih rata-rata 21,30 poin. Paired Samples *t-test* mengungkap peningkatan (gain) yang sangat dramatis pada kelompok eksperimen sebesar 28,70 poin (dari 58,90 menjadi 87,60;  $t=-49,480$ ;  $p=0,000$ ), sementara peningkatan pada kelompok kontrol hanya 8,50 poin. Program pelatihan evaluasi pembelajaran berbasis AI terbukti efektif secara signifikan dan memberikan dampak yang besar dalam meningkatkan kompetensi profesional guru. Temuan ini menegaskan pentingnya intervensi pelatihan yang terstruktur untuk membekali guru dengan kecakapan *Technological Pedagogical Content Knowledge (TPACK)* spesifik di era digital, sehingga dapat menjadi pendorong peningkatan mutu pembelajaran.

**Kata kunci:** Pembelajaran Berbasis Kecerdasan Buatan, Profesionalisme, Guru, MTs

---

### **Pendahuluan**

Era revolusi industri 4.0 dan pesatnya perkembangan teknologi digital telah membawa transformasi mendasar dalam berbagai aspek kehidupan, tak terkecuali di dunia pendidikan. Kecerdasan Buatan (Artificial Intelligence/AI) telah muncul sebagai salah satu terobosan teknologi yang berpotensi merevolusi praktik pembelajaran dan penilaian hasil belajar. Dalam konteks pembelajaran, AI tidak hanya menawarkan efisiensi administratif, tetapi juga kemampuan untuk menyediakan analisis data pembelajaran yang mendalam, personalisasi materi, serta sistem evaluasi yang lebih adaptif dan obyektif (Chen et al., 2020); Holmes et al., 2022). Kemampuan ini menjadi

krusial untuk mengakomodasi keragaman gaya belajar dan kebutuhan peserta didik di abad ke-21 (Zawacki-richter et al., 2019).

Namun, potensi besar ini belum sepenuhnya dapat dioptimalkan, khususnya di tingkat satuan pendidikan menengah seperti Madrasah Tsanawiyah (MTs). Guru sebagai ujung tombak proses pembelajaran sering kali menghadapi kendala dalam mengadopsi teknologi baru, baik akibat keterbatasan akses, pelatihan, maupun pemahaman tentang integrasi AI yang efektif dan etis dalam evaluasi pembelajaran (Francesc et al., 2019). Padahal, evaluasi yang berkualitas adalah komponen esensial untuk mengukur capaian pembelajaran, memberikan umpan balik yang bermakna, dan mendorong perbaikan proses pembelajaran secara berkelanjutan (Black & Wiliam, 2018). Selain itu, melalui evaluasi yang berkualitas, peserta didik dapat mengetahui tingkat efektivitas usaha belajar yang telah dilakukan, dimana hasil belajar yang baik dapat menjadi motivasi untuk mempertahankan dan meningkatkan prestasi, sedangkan hasil yang belum optimal diharapkan dapat mendorong peserta didik untuk memperbaiki proses belajarnya (Juli & Fitria, 2024).

Di Kabupaten Mojokerto, guru-guru MTs dituntut untuk terus menguatkan profesionalismenya dalam merespons tantangan zaman dan tuntutan Kurikulum Merdeka yang menekankan asesmen untuk pembelajaran (assessment for learning). Tuntutan tersebut menempatkan kemampuan adaptif sebagai kompetensi kunci, yaitu kesiapan guru dalam menyiapkan perubahan; mencakup kesiapan menyesuaikan praktik pembelajaran dan evaluasi secara berkelanjutan (Waita et al., 2025). Tanpa kompetensi yang memadai dalam memanfaatkan alat berbasis AI, terdapat risiko bahwa proses evaluasi tetap berjalan secara konvensional, kurang mampu memberikan gambaran holistik tentang perkembangan peserta didik, dan berpotensi tertinggal dari perkembangan metodologi pendidikan global (Luckin et al., 2022).

Oleh karena itu, upaya strategis untuk memberdayakan guru melalui pelatihan yang komprehensif dan aplikatif menjadi sebuah keniscayaan (Istiningsih & Hasbullah, 2015). Pelatihan yang dimaksud tidak hanya bersifat teknis pengoperasian tool AI, tetapi juga harus mencakup aspek pedagogis, analitis, dan etika dalam pemanfaatan AI untuk evaluasi (Akgun & Greenhow, 2022). Dengan demikian, guru dapat beralih dari peran sebagai pemberi tes konvensional menjadi fasilitator yang mampu mendesain, menganalisis, dan memanfaatkan hasil evaluasi berbasis AI untuk meningkatkan kualitas pembelajaran secara personal dan kolektif.

Berdasarkan latar belakang tersebut, program “Pelatihan Evaluasi Pembelajaran Berbasis Kecerdasan Buatan sebagai Upaya Penguatan Profesionalisme Guru MTs Kabupaten Mojokerto” dirancang. Program ini bertujuan untuk membekali guru dengan pengetahuan, keterampilan, dan mindset yang diperlukan dalam mengintegrasikan teknologi AI ke dalam praktik evaluasi mereka. Diharapkan, melalui penguatan kompetensi ini, profesionalisme guru akan meningkat, proses penilaian menjadi lebih efektif dan informatif, dan pada akhirnya dapat mempercepat peningkatan mutu pendidikan di lingkungan MTs Kabupaten Mojokerto secara signifikan dan berkelanjutan.

## Metode

### Jenis & Desain

Pengabdian Kepada Masyarakat (PKM) ini akan menggunakan pendekatan kuantitatif untuk mengukur dampak dari program pelatihan. Pendekatan ini dipilih guna memperoleh data numerik yang objektif dan terukur, sehingga efektivitas intervensi pelatihan dapat diuji secara statistik. Untuk mencapai tujuan tersebut, desain eksperimen semu akan diterapkan. Metode eksperimen semu (quasi-experiment) dengan desain Pretest-Posttest Control Group Design. Pemilihan desain ini

mempertimbangkan konteks lapangan di mana penugasan subjek secara acak penuh (true random assignment) sulit dilakukan, namun kontrol terhadap variabel pengganggu tetap diperlukan. Dalam desain ini, satu subjek pengabdian akan menerima perlakuan (pelatihan) sebagai kelompok eksperimen, sementara subjek pengabdian lain berfungsi sebagai pembanding (kontrol). Pengukuran awal (pretest) dan pengukuran akhir (posttest) akan dilakukan pada kedua kelompok untuk melihat perbedaan peningkatan yang terjadi.

#### Waktu & Tempat

Pengabdian akan dilaksanakan selama empat bulan, terhitung sejak tahap persiapan instrumen hingga analisis data akhir. Lokasi kegiatan difokuskan pada beberapa Madrasah Tsanawiyah (MTs) negeri dan swasta di wilayah Kabupaten Mojokerto. Pelaksanaan pelatihan inti sebagai bentuk intervensi direncanakan berlangsung secara intensif dalam bentuk workshop. Subjek dalam kegiatan PKM ini mencakup seluruh guru mata pelajaran yang aktif mengajar di MTs se-Kabupaten Mojokerto. Sampel penelitian berjumlah 60 orang guru yang akan dibagi secara acak ke dalam dua kelompok, yaitu 30 guru untuk kelompok eksperimen dan 30 guru untuk kelompok kontrol. Teknik pengambilan sampel dilakukan secara bertahap; pertama dengan purposive sampling untuk memilih guru-guru yang memenuhi kriteria inklusi (seperti memiliki pengalaman mengajar minimal dua tahun dan kesediaan mengikuti seluruh rangkaian penelitian), kemudian dilanjutkan dengan random assignment untuk menempatkan mereka ke dalam salah satu dari kedua kelompok.

#### Instrumen Pengumpulan Data

Instrumen pengumpulan data yang akan digunakan disusun secara khusus. Untuk mengukur peningkatan kemampuan evaluasi pembelajaran berbasis AI digunakan angket pemahaman evaluasi pembelajaran berbasis AI. Seluruh instrumen akan diuji validitas isi oleh ahli dan diuji reliabilitasnya secara statistik sebelum digunakan. Prosedur pengumpulan data diawali dengan pemberian pretest kepada kedua kelompok untuk mengetahui kondisi awal. Selanjutnya, kelompok eksperimen mengikuti program pelatihan penuh, sedangkan kelompok kontrol tidak. Setelah intervensi selesai, kedua kelompok kembali menjalani pengukuran yang sama (posttest) untuk melihat perubahan. Seluruh proses pengumpulan data dilakukan dalam kondisi yang terkontrol untuk menjaga konsistensi.

#### Teknik Analisis Data

Teknik analisis data akan dilakukan secara statistik dengan desain Independent Sample t-test. Analisis deskriptif juga akan disajikan untuk memberikan gambaran umum karakteristik sampel dan kecenderungan data. Analisis data dalam penelitian ini menggunakan teknik statistik inferensial untuk menguji efektivitas program pelatihan. Prosedur analisis dilakukan dengan bantuan software (SPSS) versi 26 melalui dua tahap utama: uji prasyarat analisis dan uji hipotesis. Tahap pertama merupakan uji prasyarat analisis yang meliputi uji normalitas dan homogenitas. Uji normalitas distribusi data dilakukan dengan Uji Shapiro-Wilk pada skor pretest, posttest, dan gain dari kedua kelompok. Hasil pengujian menunjukkan seluruh nilai signifikansi ( $p > 0,05$ ), yang mengindikasikan data berdistribusi normal. Selanjutnya, Uji Homogenitas Varians dengan Uji Levene diterapkan pada skor posttest kedua kelompok. Hasil uji menunjukkan nilai signifikansi 0,269 ( $p > 0,05$ ), yang membuktikan bahwa varians data antara kelompok eksperimen dan kontrol adalah homogen (equal variances assumed). Pemenuhan kedua asumsi parametrik ini menjadi landasan untuk menggunakan uji-t dalam pengujian hipotesis. Tahap kedua merupakan uji hipotesis yang dilakukan dengan dua jenis uji statistik. Pertama, untuk mengetahui perbedaan hasil akhir yang signifikan antara kelompok eksperimen dan kontrol, digunakan Independent Samples t-test.

---

## Hasil & Pembahasan

### Hasil

Kegiatan Pengabdian Kepada Masyarakat (PKM) ini menyajikan temuan empiris mengenai dampak pelatihan evaluasi pembelajaran berbasis kecerdasan buatan terhadap peningkatan kompetensi guru MTs Kabupaten Mojokerto. Paparan hasil mencakup tiga bagian utama: (1) analisis statistik deskriptif untuk memberikan gambaran umum distribusi data, (2) hasil uji *independent samples t-test* yang mengukur perbedaan efektivitas antara kelompok eksperimen dan kontrol, serta (3) hasil uji *paired samples t-test* yang menganalisis peningkatan kompetensi dalam masing-masing kelompok dari tahap *pretest* ke *posttest*.

#### Analisis Statistik Deskriptif

Berdasarkan data yang diperoleh dari 30 responden pada kelompok eksperimen dan 30 responden pada kelompok kontrol, berikut adalah paparan statistik deskriptif untuk skor *posttest* kompetensi guru. Rata-rata skor akhir (*posttest*) kelompok eksperimen yang mengikuti pelatihan evaluasi pembelajaran berbasis kecerdasan buatan adalah 87,60 dengan simpangan baku (*standard deviation*) sebesar 2,80. Nilai simpangan baku yang relatif rendah ini menunjukkan bahwa sebaran skor di sekitar rata-rata cukup homogen, atau dengan kata lain, pencapaian peserta dalam kelompok ini cenderung merata dan tidak memiliki variasi yang besar. Nilai *standard error of mean* sebesar 0,51 mengindikasikan tingkat akurasi estimasi rata-rata populasi berdasarkan sampel ini.

Sementara itu, rata-rata skor akhir kelompok kontrol yang tidak mendapatkan intervensi pelatihan adalah 66,30 dengan simpangan baku 1,90. Simpangan baku yang juga rendah menandakan konsistensi pencapaian di dalam kelompok ini, meskipun pada level kompetensi yang lebih rendah dibandingkan kelompok eksperimen. Nilai *standard error of mean* untuk kelompok kontrol adalah 0,35. Secara komparatif, terdapat selisih rata-rata (*mean difference*) sebesar 21,30 poin antara skor akhir kelompok eksperimen dan kelompok kontrol. Perbedaan yang cukup besar ini memberikan gambaran awal yang kuat mengenai dampak positif dari perlakuan yang diberikan. Deskripsi ini menjadi dasar yang penting sebelum dilakukan uji statistik inferensial untuk membuktikan signifikansi perbedaan tersebut.

**Tabel 1. Group Statistics**

Kelompok	N	Mean	Std. Deviation	Std. Error Mean
Eksperimen	30	87,60	2,80	0,51
Kontrol	30	66,30	1,90	0,35

#### Independent Samples T-Test

Berdasarkan hasil uji prasyarat Levene's Test for Equality of Variances, diperoleh nilai signifikansi (Sig.) sebesar 0,269. Karena nilai ini lebih besar dari tingkat signifikansi 0,05 ( $p > 0,05$ ), dapat disimpulkan bahwa varians antara kelompok eksperimen dan kelompok kontrol adalah homogen. Dengan terpenuhinya asumsi homogenitas ini, analisis uji-t yang digunakan mengacu pada baris Equal Variances Assumed. Hasil uji Independent Samples T-Test menunjukkan nilai t hitung sebesar 32,450 dengan derajat kebebasan (df) 58. Nilai signifikansi dua-arah (Sig. 2-tailed) adalah 0,000, yang jauh lebih kecil dari tingkat kesalahan 0,05 ( $p < 0,05$ ). Hal ini secara statistik membuktikan bahwa terdapat perbedaan yang sangat signifikan antara rata-rata skor post-test kelompok eksperimen dan kelompok kontrol. Selisih rata-rata (Mean Difference) antara

kedua kelompok adalah 21.300 poin, dengan standard error difference sebesar 0.657. Selang kepercayaan 95% untuk perbedaan rata-rata ini berkisar dari 19.985 (batas bawah) hingga 22.615 (batas atas). Karena selang kepercayaan ini tidak mencakupi angka nol (0), dapat disimpulkan bahwa perbedaan tersebut signifikan secara statistik. Temuan ini menguatkan bukti bahwa intervensi pelatihan memberikan dampak yang nyata dalam meningkatkan kompetensi peserta dibandingkan dengan kelompok yang tidak mendapat perlakuan.

**Tabel 2. Independent Samples Test**

Levene's Test for Equality of Variances		t-test for Equality of Means					
F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
Lower							
1,245	0,269	32,450	58	0,000	21,300	0,657	19,985

#### *Paired Samples T-Test*

Berdasarkan hasil uji Paired Samples T-Test, dapat diinterpretasikan bahwa terdapat peningkatan yang signifikan dari skor *pretest* ke *posttest* pada kedua kelompok penelitian, namun dengan besaran (*effect size*) yang sangat berbeda.

#### *Paired Samples Statistics*

Statistik Deskriptif (Tabel 3) menunjukkan bahwa pada kelompok eksperimen, skor rata-rata meningkat dari 58,90 (*pretest*) menjadi 87,60 (*posttest*), yang merepresentasikan peningkatan absolut sebesar 28,70 poin. Simpangan baku (*Std. Deviation*) untuk setiap pengukuran relatif stabil (sekitar 2,75-2,80), menunjukkan konsistensi dalam distribusi data. Pada kelompok kontrol, peningkatan yang terjadi lebih rendah, yaitu dari rata-rata 57,80 menjadi 66,30, dengan selisih 8,50 poin. Perbedaan besar peningkatan ini memberikan gambaran awal tentang dampak intervensi.

**Tabel 3. Paired Samples Statistics**

Kelompok	N	Mean	Std. Deviation	Std. Error Mean
Eksperimen – Pretest	30	58,90	2,75	0,50
Eksperimen – Posttest	30	87,60	2,80	0,51
Kontrol – Pretest	30	57,80	2,60	0,47
Kontrol – Posttest	30	66,30	1,90	0,35

#### *Paired Samples Correlations*

Uji Korelasi Berpasangan (Tabel 4) menunjukkan bahwa terdapat hubungan positif yang kuat dan signifikan antara skor *pretest* dan *posttest* pada kedua kelompok. Nilai korelasi 0,812 (sig. 0,000) untuk kelompok eksperimen dan 0,694 (sig. 0,000) untuk

kelompok kontrol mengindikasikan bahwa peserta yang memiliki skor awal tinggi cenderung juga memiliki skor akhir yang tinggi. Hal ini memperkuat reliabilitas pengukuran yang dilakukan.

**Tabel 4. Paired Samples Correlations**

Pasangan	N	Correlation	Sig.
Pretest & Posttest Eksperimen	30	0,812	0,000
Pretest & Posttest Kontrol	30	0,694	0,000

*Paired Samples Test*

Uji Perbedaan Berpasangan (Tabel 5) memberikan bukti statistik inferensial. Pada kelompok eksperimen, hasil uji menghasilkan nilai  $t = -49,48$  dengan derajat kebebasan ( $df$ ) 29 dan signifikansi (Sig. 2-tailed) 0,000 ( $p < 0,01$ ). Nilai *mean difference* negatif (-28,70) menegaskan bahwa skor *posttest* secara signifikan lebih tinggi daripada skor *pretest*. Selang kepercayaan 95% untuk perbedaan mean adalah -29,89 hingga -27,51, yang tidak mencakupi angka nol, sehingga perbedaan tersebut signifikan secara statistik.

**Tabel 5. Paired Samples Test**

Pasangan	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pretest–Posttest Eksperimen	-28,70	3,20	0,58	-29,89	-27,51	-49,48	29	
Pretest–Posttest Kontrol	-8,50	2,40	0,44	-9,41	-7,59	-19,32	29	

Pada kelompok kontrol, meskipun juga menunjukkan perbedaan yang signifikan dengan nilai  $t = -19,32$  ( $df = 29$ ,  $sig. 0,000$ ), besaran peningkatan (mean difference = -8,50) jauh lebih kecil dibandingkan kelompok eksperimen. Selang kepercayaan 95% nya adalah -9,41 hingga -7,59. Kesimpulan interpretasi adalah bahwa intervensi pelatihan terbukti efektif dalam menghasilkan peningkatan kompetensi yang jauh lebih besar (28,70 poin) dibandingkan dengan peningkatan alami atau faktor lain yang mungkin dialami kelompok kontrol (8,50 poin). Signifikansi statistik pada kedua kelompok ( $p = 0,000$ ) menunjukkan bahwa pengukuran ulang (*pretest-posttest*) memang sensitif mendeteksi perubahan, namun besaran perubahan pada kelompok eksperimen secara substansial lebih meaningful secara praktis. Hasil uji paired samples t-test menunjukkan bahwa pada kelompok eksperimen terdapat peningkatan yang signifikan antara skor pretest dan posttest dengan nilai signifikansi 0,000 ( $p < 0,05$ ). Sementara itu, kelompok kontrol juga mengalami peningkatan, namun dengan rata-rata selisih yang lebih kecil. Temuan ini menegaskan bahwa pelatihan penyusunan evaluasi pembelajaran berbasis kecerdasan buatan (AI) memberikan dampak yang lebih efektif dibandingkan pembelajaran konvensional.

## Pembahasan

Hasil kegiatan pengabdian kepada masyarakat (PKM) ini secara konsisten dan komprehensif membuktikan efektivitas program pelatihan evaluasi pembelajaran berbasis kecerdasan buatan (AI) dalam meningkatkan kompetensi profesional guru Madrasah Tsanawiyah (MTs) Kabupaten Mojokerto. Temuan ini memberikan dukungan empiris terhadap sejumlah teori dan penelitian terdahulu yang menegaskan peran transformatif teknologi AI dalam pendidikan dan pengembangan profesi guru.

Pertama, analisis data deskriptif menunjukkan adanya kesenjangan kompetensi yang sangat jelas antara kelompok eksperimen dan kontrol setelah intervensi. Rata-rata skor akhir (posttest) kelompok eksperimen (87,60) jauh melampaui kelompok kontrol (66,30) dengan selisih rata-rata 21,30 poin. Secara teoritis, hal ini selaras dengan konsep technological pedagogical content knowledge (TPACK) (Koehler, 2006), yang menekankan bahwa integrasi teknologi yang efektif membutuhkan pengetahuan sinergis antara teknologi, pedagogi, dan materi ajar. Sejalan dengan konsep tersebut, penelitian terdahulu oleh Sari & Dwikurnaningsih (2025), menunjukkan bahwa bahwa pengembangan modul pelatihan berbasis TPACK yang terintegrasi dengan kecerdasan buatan mampu meningkatkan kompetensi pedagogis guru secara signifikan. Pelatihan yang dirancang tidak hanya mengajarkan penggunaan tool AI (technological knowledge), tetapi juga bagaimana menerapkannya dalam konteks evaluasi yang sesuai dengan tujuan pembelajaran (pedagogical knowledge) untuk mata pelajaran tertentu (content knowledge), sehingga menghasilkan peningkatan kompetensi yang holistik.

Kedua, hasil uji Independent Samples t-test yang signifikan ( $t=32,450$ ;  $p=0,000$ ) mengkonfirmasi bahwa perbedaan hasil akhir tersebut bukanlah suatu kebetulan, melainkan benar-benar disebabkan oleh intervensi pelatihan. Keberhasilan ini didukung oleh terpenuhinya asumsi homogenitas varians antar kelompok (Levene's test  $sig.=0,269$ ), yang menunjukkan bahwa karakteristik awal kedua kelompok relatif setara. Dengan kata lain, peningkatan yang superior pada kelompok eksperimen dapat diatribusikan secara kausal kepada program pelatihan. Temuan ini mendukung penelitian Chen et al., (2020) yang menyatakan bahwa program pengembangan guru yang terstruktur dalam pemanfaatan AI dapat secara signifikan meningkatkan kapasitas mereka dalam merancang dan melaksanakan asesmen yang lebih personal dan data-driven.

Ketiga, analisis yang lebih mendalam melalui Paired Samples t-test memberikan wawasan penting tentang dinamika peningkatan di dalam masing-masing kelompok. Kelompok eksperimen mengalami peningkatan yang sangat dramatis sebesar 28,70 poin (dari 58,90 menjadi 87,60), dengan nilai t-test yang sangat signifikan ( $-49,480$ ;  $p=0,000$ ). Sementara itu, kelompok kontrol juga menunjukkan peningkatan yang signifikan secara statistik sebesar 8,50 poin, namun dengan besaran (effect size) yang jauh lebih kecil. Peningkatan pada kelompok kontrol ini mungkin disebabkan oleh faktor Hawthorne Effect atau pembelajaran mandiri selama periode penelitian (Mccarney et al., 2007). Namun, perbandingan besaran peningkatan yang sangat timpang (28,70 vs 8,50) secara tegas menegaskan bahwa pelatihan memberikan nilai tambah (value-added) yang sangat besar dibandingkan dengan pembelajaran alami atau konvensional.

Korelasi positif yang kuat dan signifikan antara skor pretest dan posttest pada kedua kelompok (0,812 untuk eksperimen dan 0,694 untuk kontrol) mengindikasikan konsistensi performa individu. Artinya, peserta yang memiliki dasar pemahaman baik sejak awal cenderung mampu mengoptimalkan pelatihan untuk mencapai hasil yang lebih tinggi. Hal ini sejalan dengan prinsip scaffolding dalam pembelajaran, di mana pengetahuan awal menjadi fondasi untuk akuisisi pengetahuan yang lebih kompleks (Vitgosky & Cole, 1978). Pelatihan berhasil berfungsi sebagai scaffold yang efektif, terutama bagi mereka yang telah memiliki modal pengetahuan dasar yang memadai.

Secara implikatif, Kegiatan ini membuktikan bahwa penguatan kompetensi asesmen berbasis AI guru di era digital tidak dapat mengandalkan pembelajaran mandiri atau kebijakan top-down semata. Diperlukan intervensi sistematis dan pelatihan praktis yang membekali guru dengan kompetensi TPACK spesifik, dalam hal ini untuk evaluasi berbasis AI. Sebagaimana diungkapkan oleh (Luckin et al., 2022), guru perlu menjadi "AI-ready", tidak sekadar pengguna pasif, tetapi sebagai desain dan pengambil keputusan yang kritis berdasarkan analisis AI. Program pelatihan dalam penelitian ini telah menunjukkan potensi untuk mencapai tujuan tersebut. Ke depannya, sustainabilitas kompetensi ini perlu dijaga melalui komunitas praktik (community of practice) dan dukungan berkelanjutan dari lembaga pendidikan.

### Simpulan

Berdasarkan hasil analisis, disimpulkan bahwa Pelatihan Evaluasi Pembelajaran Berbasis AI efektif meningkatkan kompetensi profesional guru MTs Kabupaten Mojokerto. Hal ini dibuktikan dengan tiga temuan utama: (1) Terdapat perbedaan signifikan antara kelompok eksperimen dan kontrol. Skor akhir (posttest) kelompok eksperimen (87,60) secara statistik jauh lebih tinggi ( $p=0,000$ ) daripada kelompok kontrol (66,30) dengan selisih rata-rata 21,30 poin. (2) Peningkatan kompetensi pada kelompok eksperimen bersifat transformatif. Analisis paired samples t-test menunjukkan peningkatan (gain) yang sangat besar, yaitu 28,70 poin, dari pretest ke posttest. Angka ini jauh melampaui peningkatan alami pada kelompok kontrol yang hanya 8,50 poin, (3) Keberhasilan pelatihan terletak pada pendekatan komprehensif dan integratif yang selaras dengan kerangka TPACK (Technological Pedagogical Content Knowledge), yang menggabungkan aspek teknis, pedagogis, dan etika dalam evaluasi berbasis AI. Implikasi penelitian ini menegaskan bahwa pemberdayaan guru di era digital memerlukan intervensi pelatihan yang terstruktur dan kontekstual. Program semacam ini merupakan strategi kunci untuk menutup kesenjangan kompetensi digital dan mempersiapkan guru menjadi pendidik yang mampu memanfaatkan inovasi teknologi guna meningkatkan mutu pembelajaran.

### Daftar Pustaka

- Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education : Addressing ethical challenges in K - 12 settings. *AI and Ethics*, 2(3), 431–440. <https://doi.org/10.1007/s43681-021-00096-7>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 1–25. <https://doi.org/10.1080/0969594X.2018.1441807>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education : A Review. 8. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Francesc, P., Miguel, S., Axel, R., & Paula, V. (2019). Artificial Intelligence in Education: Challenges and opportunities for Sustainable Development. UNESCO.
- Holmes, W., Persson, J., Chounta, I. A., Wasson, B., & Dimitrova, V. (2022). Artificial Intelligence and Education - A Critical View Through the Lens of Human Rights, Democracy and the Rule of Law. Council of Europe.
- Istiningsih, S., & Hasbullah. (2015). Blended Learning, Trend Strategi Pembelajaran. *Jurnal Elemen*, 1(1), 49–56.
- Juli, N., & Fitria, N. A. (2024). Langkah-langkah Evaluasi Pembelajaran. 4(3).
- Koehler, M. J. (2006). Technological Pedagogical Content Knowledge : A Framework for Teacher Knowledge. 108(6), 1017–1054.
- Luckin, R., Cukurova, M., & Kent, C. (2022). Computers and Education : Artificial

- Intelligence Empowering educators to be AI-ready. Computers and Education: Artificial Intelligence, 3, 100076. <https://doi.org/10.1016/j.caerai.2022.100076>
- Mccarney, R., Warner, J., Iliffe, S., Haselen, R. Van, Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised , controlled trial. 8, 1–8. <https://doi.org/10.1186/1471-2288-7-30>
- Sari, Y. N., & Dwikurnaningsih, Y. (2025). Development of a TPACK Training Module through AI Integration to Improve the Pedagogical Skills of Arts Teachers. 11(3), 1293–1303.
- Vitgosky, L. S., & Cole, M. (1978). *Mind in Society*. Harvard University Press.
- Waita, B. C., Yiswi, T. A., Kristiahadi, A., Kristen, U., & Wacana, S. (2025). Dampak Artificial Intelligence ( AI ) Terhadap Pendidikan di Indonesia. 6(7), 3112–3121.
- Zawacki-richter, O., Marín, V. I., & Bond, M. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators ?