

EVALUASI PROGRAM PELATIHAN PERANCANGAN ASESMEN BERBASIS KECERDASAN ARTIFISIAL BERDASARKAN KUALITAS PRODUK ASESMEN GURU

Rusijono¹, Hirnanda Dimas Pradana², Arqoma Nurveda Carreza³

^{1,2,3}Universitas Negeri Surabaya

¹rusijono@unesa.ac.id, ²hirnandapradana@unesa.ac.id,

³arqomacarreza@unesa.ac.id

ABSTRACT

This study aimed to evaluate a training program on designing artificial intelligence (AI)-based assessments, focusing on the quality of the assessment products created by teachers. Although generative AI offers new opportunities for assessment design, teachers' ability to utilize it pedagogically requires strengthening through structured training. An evaluative research method with a descriptive quantitative approach was applied to 25 teachers participating in a five-stage training program: content delivery, AI demonstration, a design workshop, independent product development, and a presentation and feedback session. The products were assessed using a five-aspect rubric—alignment with learning objectives, instrument quality, authenticity, AI integration, and rubric quality—on a scale of 1 to 4. The results showed an overall average of 3.44 (categorized as "very good"), with the highest score in alignment with learning objectives (M=3.64) and the lowest in AI integration (M=3.08). 56% of participants achieved the "very good" category, while 44% fell into the "good" category. The findings indicate that product-based evaluation provides an authentic picture of teachers' competency attainment while identifying areas requiring improvement, particularly regarding the critical and reflective use of AI.

Keywords: Program Evaluation, Learning Assessment, Artificial Intelligence

ABSTRAK

Penelitian ini bertujuan mengevaluasi program pelatihan perancangan asesmen berbasis kecerdasan artifisial (AI) berdasarkan kualitas produk asesmen yang dihasilkan guru. Meskipun AI generatif membuka peluang baru dalam perancangan asesmen, kemampuan guru memanfaatkannya secara pedagogis masih memerlukan penguatan melalui pelatihan terstruktur. Metode penelitian evaluatif dengan pendekatan kuantitatif deskriptif diterapkan pada 25 guru yang mengikuti pelatihan lima tahap: penyampaian materi, demonstrasi AI, workshop perancangan, penyusunan produk mandiri, serta presentasi dan umpan balik. Produk dinilai menggunakan rubrik lima aspek: kesesuaian dengan tujuan pembelajaran, kualitas instrumen, autentisitas, integrasi AI, dan kualitas rubrik pada skala 1–4. Hasil menunjukkan rata-rata keseluruhan 3,44 (kategori sangat baik) dengan skor

tertinggi pada kesesuaian tujuan pembelajaran ($M=3,64$) dan terendah pada integrasi AI ($M=3,08$). Sebanyak 56% peserta mencapai kategori sangat baik dan 44% kategori baik. Temuan mengindikasikan bahwa evaluasi berbasis produk mampu memberikan gambaran autentik mengenai capaian kompetensi guru, sekaligus mengidentifikasi area yang memerlukan penguatan, khususnya pada aspek pemanfaatan AI secara kritis dan reflektif.

Kata Kunci: Evaluasi Program, Asesmen Pembelajaran, Kecerdasan Artifisial

A. Pendahuluan

Perkembangan kecerdasan artifisial generatif telah menghadirkan perubahan mendasar dalam ekosistem pendidikan, tidak terkecuali pada praktik asesmen pembelajaran. Berbagai platform seperti ChatGPT, Claude, dan model bahasa besar lainnya menawarkan kemampuan untuk menghasilkan butir soal, rubrik penilaian, hingga analisis hasil belajar dalam waktu yang jauh lebih singkat dibandingkan metode konvensional (Celik, Dindar, Muukkonen, & Järvelä, 2022; Kasneci et al., 2023). Sejumlah studi empiris bahkan menunjukkan bahwa model bahasa besar mampu menghasilkan butir soal pilihan ganda dengan indeks kesukaran dan daya beda yang layak digunakan dalam ujian sesungguhnya (Daungsupawong & Wiwanitkit, 2024), meskipun kualitas keluarannya tetap sangat bergantung pada ketepatan rekayasa prompt yang digunakan penggunaannya (Park & Choo, 2025;

Wang, Song, Guo, & Yang, 2025). Kendati menjanjikan efisiensi, integrasi AI ke dalam praktik asesmen tidak dapat dilakukan secara serampangan. Guru perlu memiliki literasi AI yang memadai — mencakup kemampuan memahami, menggunakan, dan mengevaluasi kecerdasan artifisial secara kritis (Ng, Leung, Chu, & Qiao, 2021) — agar mampu memverifikasi luaran AI, menyesuaikannya dengan konteks pembelajaran, dan memastikan bahwa instrumen yang dihasilkan tetap memenuhi kaidah validitas serta reliabilitas (Ng, Leung, Su, Ng, & Chu, 2023; Su, Ng, & Chu, 2023).

Dalam konteks pendidikan kejuruan, tuntutan terhadap kualitas asesmen semakin tinggi karena penilaian harus mencerminkan kesiapan peserta didik menghadapi dunia kerja. Asesmen tidak cukup hanya mengukur penguasaan pengetahuan deklaratif, melainkan harus bersifat autentik, mampu

mengukur keterampilan berpikir tingkat tinggi (HOTS), serta relevan dengan permasalahan di lingkungan profesional (Darling-Hammond, Schachner, Wojcikiewicz, & Flook, 2024). Namun demikian, kajian terhadap ratusan butir soal ujian pendidikan kejuruan menemukan bahwa institusi pendidikan vokasi masih cenderung mengandalkan soal tertutup yang berorientasi hafalan, sehingga prinsip autentisitas belum sepenuhnya tercapai (Villarroel, Melipillán, Santana, & Aguirre, 2024), sementara upaya mengintegrasikan asesmen autentik dengan keterampilan abad ke-21 juga masih menghadapi kendala penerapan di berbagai disiplin ilmu (Vlachopoulos & Makri, 2024). Sayangnya, studi-studi terdahulu menunjukkan bahwa banyak guru masih mengalami kesulitan dalam merancang asesmen yang memenuhi kriteria tersebut, terlebih ketika harus mengintegrasikan teknologi AI ke dalam proses penyusunannya (Chiu, 2024; Zhang & Aslan, 2021).

Sejumlah lembaga telah merespons kesenjangan ini dengan menyelenggarakan pelatihan pemanfaatan AI bagi guru, sejalan dengan seruan untuk merancang

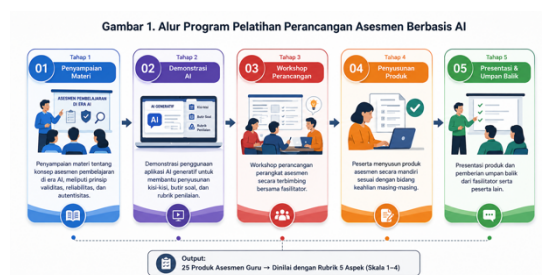
pendidikan AI yang holistik bagi tenaga pendidik (Chiu, 2021; Shi, Gu, & Li, 2025; Zhao, Wu, & Luo, 2022). Akan tetapi, evaluasi terhadap program-program tersebut sebagian besar masih mengandalkan angket kepuasan peserta atau tes pengetahuan deklaratif (pre-test/post-test), yang belum tentu berkorelasi kuat dengan kemampuan aktual guru dalam menghasilkan produk asesmen yang berkualitas (Guskey, 2025; Sancar, Atal, & Deryakulu, 2021). Mengukur dampak pelatihan semata-mata melalui persepsi atau skor tes berpotensi memberikan gambaran yang kurang utuh, karena kompetensi merancang asesmen pada hakikatnya merupakan keterampilan aplikatif yang lebih tepat dinilai melalui artefak atau produk kerja nyata (Aumann, Schnebel, & Weitzel, 2023; Darling-Hammond et al., 2024).

Berdasarkan kesenjangan tersebut, penelitian ini mengevaluasi program pelatihan perancangan asesmen berbasis kecerdasan artifisial menggunakan pendekatan berbasis produk (product-oriented evaluation). Kualitas produk asesmen yang dihasilkan guru dianalisis pada lima aspek: kesesuaian dengan tujuan pembelajaran, kualitas instrumen,

otentisitas, integrasi AI, dan kualitas rubrik penilaian. Pendekatan ini memungkinkan identifikasi aspek-aspek yang telah dikuasai maupun yang masih memerlukan penguatan, sehingga hasil evaluasi dapat menjadi dasar bagi penyempurnaan program pelatihan serupa di masa mendatang.

B. Metode Penelitian

Penelitian ini termasuk dalam kategori penelitian evaluatif (evaluation research) dengan pendekatan kuantitatif deskriptif. Fokus utama adalah mengevaluasi keberhasilan program pelatihan berdasarkan kualitas produk asesmen yang dihasilkan peserta setelah mengikuti seluruh rangkaian kegiatan. Subjek penelitian adalah 25 guru dari berbagai bidang keahlian yang berpartisipasi secara penuh dalam pelatihan yang diselenggarakan di Yayasan Barunawati pada Mei 2026. Program pelatihan terdiri atas lima tahapan sebagaimana disajikan pada Gambar 1.



Gambar 1. Alur Program Pelatihan Perancangan Asesmen Berbasis AI

Tahap pertama berupa penyampaian materi mengenai konsep asesmen pembelajaran di era AI, mencakup prinsip validitas, reliabilitas, dan autentisitas. Tahap kedua merupakan demonstrasi penggunaan aplikasi AI generatif untuk membantu penyusunan kisi-kisi, butir soal, dan rubrik penilaian. Pada tahap ketiga, peserta mengikuti workshop perancangan perangkat asesmen secara terbimbing. Tahap keempat memberi kesempatan kepada peserta untuk menyusun produk asesmen secara mandiri sesuai bidang keahlian masing-masing. Tahap kelima berupa presentasi produk dan pemberian umpan balik dari fasilitator serta peserta lain.

Instrumen pengumpulan data berupa rubrik penilaian produk asesmen yang dikembangkan berdasarkan kajian literatur mengenai asesmen pembelajaran dan integrasi AI dalam pendidikan. Rubrik mencakup lima aspek dengan indikator spesifik sebagaimana dirangkum pada Tabel 1. Setiap aspek dinilai menggunakan skala empat

tingkat: 1 (kurang), 2 (cukup), 3 (baik), dan 4 (sangat baik). Skor seluruh aspek dijumlahkan kemudian dirata-rata untuk memperoleh skor total kualitas produk setiap peserta.

Tabel 1 Aspek dan Indikator Penilaian Produk Asesmen

Aspek Penilaian	Indikator
Kesesuaian dengan tujuan Pembelajaran	Keselaran asesmen dengan capaian pembelajaran, tujuan, dan indikator yang ditetapkan
Kualitas instrumen asesmen	Kejelasan konstruksi soal, penggunaan bahasa, dan ketepatan bentuk asesmen
Autentisitas asesmen	Kemampuan asesmen mengukur HOTS, pemecahan masalah, dan penerapan pengetahuan
Integrasi kecerdasan	Ketepatan pemanfaatan AI serta kemampuan memvalidasi dan memodifikasi hasil AI
Rubrik penilaian	Kelengkapan indikator, kejelasan kriteria, dan kemudahan penggunaan rubrik

Data dianalisis menggunakan statistik deskriptif meliputi rata-rata (mean), simpangan baku (SD), nilai minimum, nilai maksimum, dan persentase capaian. Skor rata-rata selanjutnya dikonversi ke dalam kategori kualitas mengacu pada kriteria: sangat baik (3,26–4,00), baik (2,51–3,25), cukup (1,76–2,50), dan kurang (1,00–1,75). Analisis dilakukan

pada tingkat keseluruhan dan pada setiap aspek penilaian untuk mengidentifikasi kekuatan serta area yang memerlukan pengembangan lebih lanjut.

C.Hasil Penelitian dan Pembahasan

Program pelatihan diikuti oleh 25 guru dari berbagai latar belakang bidang keahlian. Seluruh peserta menyelesaikan rangkaian kegiatan secara penuh dan menghasilkan produk asesmen yang dinilai menggunakan rubrik lima aspek. Gambaran umum hasil penilaian menunjukkan bahwa produk asesmen yang dihasilkan peserta berada pada kategori sangat baik dengan rata-rata skor keseluruhan 3,44 dari skala maksimum 4,00 (SD=0,42).

Tabel 2 Hasil Penilaian Produk Asesmen Guru per Aspek

Aspek Penilaian	Mean	SD	Kategori
Kesesuaian dengan tujuan pembelajaran	3,64	0,38	Sangat Baik
Kualitas instrumen asesmen	3,52	0,41	Sangat Baik
Autentisitas asesmen	3,40	0,43	Sangat Baik
Integrasi kecerdasan artifisial	3,08	0,48	Baik
Rubrik penilaian	3,56	0,37	Sangat Baik
Rata-rata keseluruhan	3,44	0,42	Sangat Baik

Berdasarkan Tabel 2, aspek kesesuaian dengan tujuan pembelajaran mencatat skor tertinggi ($M=3,64$; $SD=0,38$) yang menandakan bahwa peserta telah mampu menyusun asesmen yang selaras dengan capaian dan indikator pembelajaran. Aspek rubrik penilaian menempati posisi kedua ($M=3,56$; $SD=0,37$), diikuti kualitas instrumen ($M=3,52$; $SD=0,41$) dan autentisitas asesmen ($M=3,40$; $SD=0,43$). Keempat aspek tersebut berada pada kategori sangat baik. Sementara itu, aspek integrasi kecerdasan artifisial memperoleh skor terendah ($M=3,08$; $SD=0,48$) dan berada pada kategori baik. Simpangan baku pada aspek ini juga paling besar, mengindikasikan variasi kemampuan peserta yang lebih lebar dalam memanfaatkan AI secara pedagogis.

Tabel 3. Distribusi Kategori Kualitas Produk Asesmen

Kategori	Jumlah Guru	Persentase
Sangat Baik	14	56,0%
Baik	11	44,0%
Cukup	0	0%
Kurang	0	0%
Total	25	100%

Tabel 3 memperlihatkan bahwa lebih dari separuh peserta (56%) mencapai kategori sangat baik dan sisanya (44%) berada pada kategori

baik. Tidak satu pun produk yang masuk ke dalam kategori cukup atau kurang. Temuan ini mengonfirmasi bahwa program pelatihan secara umum berhasil memfasilitasi guru dalam menghasilkan produk asesmen yang memenuhi standar minimum yang ditetapkan. Meskipun demikian, masih terdapat ruang perbaikan, terutama pada aspek integrasi AI yang menjadi satu-satunya aspek dengan rerata di bawah 3,26.

Analisis lebih mendalam terhadap produk peserta mengungkap beberapa temuan kualitatif yang melengkapi data kuantitatif. Pada aspek kesesuaian, sebagian besar peserta telah menyusun indikator yang selaras dengan tujuan, meskipun beberapa indikator masih dirumuskan secara umum. Pada aspek instrumen, mayoritas peserta berhasil mengembangkan butir soal kontekstual, namun variasi bentuk asesmen — seperti portofolio, proyek, atau studi kasus — masih terbatas (Vlachopoulos & Makri, 2024). Temuan paling kontras terdapat pada integrasi AI: peserta cenderung memanfaatkan AI sebatas untuk menghasilkan draf awal tanpa melakukan verifikasi dan modifikasi yang memadai, sebuah pola yang

juga banyak dilaporkan pada penelitian mengenai penggunaan ChatGPT untuk menyusun rubrik maupun butir soal penilaian (Huang, Coleman, Gachago, & Van Belle, 2024; Kiyak & Emekli, 2024). Beberapa produk bahkan masih memuat luaran AI yang belum disesuaikan dengan konteks mata pelajaran, mengindikasikan bahwa keterampilan evaluatif terhadap hasil AI belum berkembang secara optimal.

Hasil penelitian menunjukkan bahwa program pelatihan mampu menghasilkan produk asesmen guru dengan kualitas rata-rata pada kategori sangat baik ($M=3,44$). Capaian ini mengindikasikan bahwa pelatihan tidak semata-mata meningkatkan pengetahuan deklaratif peserta, melainkan juga memfasilitasi pengembangan keterampilan aplikatif dalam menyusun perangkat asesmen. Temuan tersebut sejalan dengan pandangan Guskey (Guskey, 2025) yang menekankan bahwa efektivitas pengembangan profesional guru idealnya diukur melalui perubahan praktik dan produk kerja nyata, bukan hanya melalui angket kepuasan atau skor tes tertulis, sebuah argumen yang juga tercermin dalam kerangka

pengembangan profesional guru kontemporer (Sancar et al., 2021).

Dominasi skor pada aspek kesesuaian dengan tujuan pembelajaran ($M=3,64$) memberi indikasi bahwa prinsip constructive alignment telah terinternalisasi dengan baik oleh mayoritas peserta. Peserta mampu menerjemahkan tujuan pembelajaran ke dalam indikator asesmen yang relevan, sebuah kompetensi fundamental yang menjadi fondasi bagi keseluruhan proses asesmen (Hailikari, Virtanen, Vesalainen, & Postareff, 2022). Capaian ini juga selaras dengan temuan Darling-Hammond et al. (2024) yang menyatakan bahwa pelatihan yang mengintegrasikan teori dengan praktik langsung cenderung lebih efektif dalam membangun kompetensi guru dibandingkan pelatihan berbasis ceramah semata.

Skor terendah pada aspek integrasi AI ($M=3,08$) merupakan temuan yang perlu mendapat perhatian serius. Kecenderungan peserta menggunakan AI sebatas alat generating tanpa diiringi proses validasi dan adaptasi mengonfirmasi kekhawatiran yang diungkap oleh Celik et al. (2022) serta Kasneci et al. (2023) mengenai kesenjangan antara

kemudahan akses terhadap AI dan kesiapan guru dalam memanfaatkannya secara pedagogis. Kecenderungan serupa juga tampak pada penelitian mengenai penyusunan butir soal berbantuan model bahasa besar, yang menunjukkan bahwa kualitas keluaran AI sangat ditentukan oleh ketepatan rekayasa prompt yang digunakan penggunaanya (Lee et al., 2024; Wang et al., 2025). Temuan ini mengisyaratkan bahwa porsi pelatihan pada aspek literasi AI kritis — mencakup verifikasi fakta, deteksi bias, adaptasi kontekstual, dan dokumentasi proses pemanfaatan AI — perlu diperkuat secara signifikan pada program pelatihan berikutnya (Chan, 2023; Ng et al., 2023). Kebutuhan ini semakin relevan pada konteks pendidikan kejuruan, mengingat kerangka literasi AI bagi guru vokasi menuntut kompetensi tambahan berupa kemampuan menyelaraskan luaran AI dengan tuntutan dunia industri (Shi et al., 2025).

Simpangan baku tertinggi pada aspek integrasi AI ($SD=0,48$) menandakan adanya disparitas antarpeserta yang cukup lebar, yang dapat diatribusikan pada perbedaan

tingkat literasi digital awal — sebuah faktor penentu keberhasilan pelatihan berbasis teknologi (Chiu, 2024; Zhang & Aslan, 2021). Disparitas semacam ini juga banyak ditemukan pada kajian mengenai instrumen pengukuran literasi digital guru, yang menunjukkan bahwa kesenjangan kompetensi awal kerap menjadi sumber variasi hasil pelatihan berbasis teknologi (Nguyen & Habók, 2023; Zhao et al., 2022). Dari perspektif metodologis, penelitian ini juga memperlihatkan bahwa evaluasi berbasis produk mampu menyajikan gambaran yang lebih kaya dibandingkan pendekatan konvensional. Dengan menelaah artefak peserta, peneliti tidak hanya memperoleh data capaian, tetapi juga dapat mengidentifikasi pola spesifik yang memerlukan intervensi, seperti kecenderungan menggunakan prompt generik atau mengabaikan verifikasi hasil AI — informasi yang sulit diperoleh dari instrumen self-report semata (Aumann et al., 2023; Park & Choo, 2025).

Beberapa keterbatasan perlu diakui. Pertama, jumlah peserta yang terbatas ($n=25$) dan berasal dari satu lokasi membuat temuan belum dapat digeneralisasikan. Kedua, evaluasi dilakukan terhadap produk yang

dihasilkan selama pelatihan, sehingga belum mencerminkan implementasi asesmen dalam praktik pembelajaran di kelas. Ketiga, reliabilitas antar-penilai belum diukur secara statistik, sebagaimana ditekankan pentingnya pada berbagai kajian tentang konsistensi penilai dalam asesmen berbasis rubrik (Lyness, 2024; Permana & Widodo, 2022). Studi lanjutan disarankan untuk melibatkan sampel yang lebih besar dan beragam, mengukur dampak pelatihan terhadap praktik asesmen di kelas dan hasil belajar peserta didik (Ventista & Brown, 2023), serta menguji pengaruhnya terhadap kemampuan berpikir tingkat tinggi peserta didik menggunakan desain kuasi-eksperimental (Agussuryani, Sudarmin, Sumarni, Cahyono, & Ellianawati, 2022; Zhou, Gan, Chen, Wijaya, & Li, 2023).

D. Kesimpulan

Penelitian ini berhasil mengevaluasi program pelatihan perancangan asesmen berbasis kecerdasan artifisial menggunakan pendekatan berbasis produk. Hasil menunjukkan bahwa produk asesmen yang dihasilkan guru berada pada kategori sangat baik ($M=3,44$),

dengan 56% peserta mencapai kategori sangat baik dan 44% pada kategori baik. Aspek kesesuaian dengan tujuan pembelajaran menjadi kekuatan utama, sedangkan integrasi AI merupakan area yang memerlukan penguatan paling signifikan. Temuan ini menegaskan bahwa evaluasi berbasis produk mampu memberikan gambaran autentik mengenai capaian kompetensi guru pascapelatihan sekaligus mengidentifikasi aspek-aspek spesifik yang perlu menjadi fokus penyempurnaan program.

Berdasarkan hasil penelitian, direkomendasikan agar program pelatihan serupa di masa mendatang: (1) memperkuat modul literasi AI kritis yang mencakup strategi validasi, adaptasi kontekstual, dan dokumentasi pemanfaatan AI; (2) menerapkan asesmen awal literasi digital untuk memungkinkan diferensiasi pembelajaran; serta (3) memperluas evaluasi hingga tahap implementasi asesmen di kelas. Studi lanjutan dengan sampel yang lebih luas dan desain eksperimental diperlukan untuk mengonfirmasi efektivitas pendekatan pelatihan ini pada konteks yang berbeda.

DAFTAR PUSTAKA

- Agussuryani, Q., Sudarmin, S., Sumarni, W., Cahyono, E., & Ellianawati, E. (2022). STEM literacy in growing vocational school student HOTS in science learning: A meta-analysis. *International Journal of Evaluation and Research in Education (IJERE)*, 11(1), 51–60. <https://doi.org/10.11591/IJERE.V11I1.21647>
- Aumann, A., Schnebel, S., & Weitzel, H. (2023). The EnTPACK rubric: development, validation, and reliability of an instrument for measuring pre-service science teachers' enacted TPACK. *Frontiers in Education*, 8, 1190152. <https://doi.org/10.3389/FEDUC.2023.1190152/TEXT>
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The Promises and Challenges of Artificial Intelligence for Teachers: a Systematic Review of Research. *TechTrends*, 66(4), 616–630. <https://doi.org/10.1007/S11528-022-00715-Y/FIGURES/6>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education* 20:1, 20(1), 38-. <https://doi.org/10.1186/S41239-023-00408-3>
- Chiu, T. K. F. (2021). A Holistic Approach to the Design of Artificial Intelligence (AI) Education for K-12 Schools. *TechTrends*, 65(5), 796–807. <https://doi.org/10.1007/S11528-021-00637-1>
- Chiu, T. K. F. (2024). The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187–6203. <https://doi.org/10.1080/10494820.2023.2253861>
- Darling-Hammond, L., Schachner, A. C. W., Wojcikiewicz, S. K., & Flook, L. (2024). Educating teachers to enact the science of learning and development. *Applied Developmental Science*, 28(1), 1–21. <https://doi.org/10.1080/10888691.2022.2130506>
- Daungsupawong, H., & Wiwanitkit, V. (2024). ChatGPT for generating multiple-choice questions: comment. *European Journal of Clinical Pharmacology* 2024 80:8, 80(8), 1253–1254. <https://doi.org/10.1007/S00228-024-03670-0>
- Guskey, T. R. (2025). Explaining the puzzling results in professional learning research. *Phi Delta Kappan*, 106(7–8), 55–60. <https://doi.org/10.1177/00317217>

- 251342372;PAGE:STRING:ARTICLE/CHAPTER
- Hailikari, T., Virtanen, V., Vesalainen, M., & Postareff, L. (2022). Student perspectives on how different elements of constructive alignment support active learning. *Active Learning in Higher Education*, 23(3), 217–231. <https://doi.org/10.1177/1469787421989160>;REQUESTEDJOURNAL:JOURNAL:ALHA;PAGE:STRING:ARTICLE/CHAPTER
- Huang, C. W., Coleman, M., Gachago, D., & Van Belle, J. P. (2024). Using ChatGPT to Encourage Critical AI Literacy Skills and for Assessment in Higher Education. *Communications in Computer and Information Science*, 1862, 105–118. https://doi.org/10.1007/978-3-031-48536-7_8
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/J.LINDIF.2023.102274>
- Kiyak, Y. S., & Emekli, E. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: a literature review. *Postgraduate Medical Journal*, 100(1189), 858–865. <https://doi.org/10.1093/POSTMJ/QGAE065>
- Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english education. *Education and Information Technologies*, 29(9), 11483–11515. <https://doi.org/10.1007/S10639-023-12249-8>
- Lyness, S. A. (2024). More Evidence of Low Inter-Rater Reliability of a High-Stakes Performance Assessment of Teacher Candidates. *Education Sciences* 2024, Vol. 14, Page 300, 14(3), 300. <https://doi.org/10.3390/EDUCSCI14030300>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/J.CAEAI.2021.100041>
- Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development* 2023 71:1, 71(1), 137–161.

- | | |
|--|---|
| <p>https://doi.org/10.1007/S11423-023-10203-6</p> <p>Nguyen, L. A. T., & Habók, A. (2023). Tools for assessing teacher digital literacy: a review. <i>Journal of Computers in Education</i> 2023 11:1, 11(1), 305–346. https://doi.org/10.1007/S40692-022-00257-5</p> <p>Park, J., & Choo, S. (2025). Generative AI Prompt Engineering for Educators: Practical Strategies. <i>Journal of Special Education Technology</i>, 40(3), 411–417. https://doi.org/10.1177/01626434241298954;SUBPAGE:STRING: FULL</p> <p>Permana, R. A. H. A., & Widodo, A. (2022). Validity and Inter-rater Reliability of the Scoring Rubrics for the Science Teacher TPACK Test Instrument. <i>Jurnal Penelitian Dan Pembelajaran IPA</i>, 8(1), 27–44. https://doi.org/10.30870/JPPi.V8I1.11164</p> <p>Sancar, R., Atal, D., & Deryakulu, D. (2021). A new framework for teachers' professional development. <i>Teaching and Teacher Education</i>, 101, 103305. https://doi.org/10.1016/J.TATE.2021.103305</p> <p>Shi, Y., Gu, M., & Li, M. (2025). AI Literacy in Vocational Education: A Framework for Teacher Professional Development. <i>Journal of Educational Theory and Practice</i>, 2(4), 3005–9305.</p> | <p>https://doi.org/10.62177/JETP.V2I4.863</p> <p>Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. <i>Computers and Education: Artificial Intelligence</i>, 4, 100124. https://doi.org/10.1016/J.CAEAI.2023.100124</p> <p>Ventista, O. M., & Brown, C. (2023). Teachers' professional learning and its impact on students' learning outcomes: Findings from a systematic review. <i>Social Sciences & Humanities Open</i>, 8(1), 100565. https://doi.org/10.1016/J.SSAHO.2023.100565</p> <p>Villarroel, V., Melipillán, R., Santana, J., & Aguirre, D. (2024). How authentic are assessments in vocational education? An analysis from Chilean teachers, students, and examinations. <i>Frontiers in Education</i>, 9, 1308688. https://doi.org/10.3389/FEDUC.2024.1308688/TEXT</p> <p>Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. <i>Studies in Educational Evaluation</i>, 83, 101425. https://doi.org/10.1016/J.STUEDUC.2024.101425</p> |
|--|---|

Wang, L., Song, R., Guo, W., & Yang, H. (2025). Exploring prompt pattern for generative artificial intelligence in automatic question generation. *Interactive Learning Environments*, 33(3), 2559–2584. <https://doi.org/10.1080/10494820.2024.2412082>

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/J.CAEAI.2021.100025>

Zhao, L., Wu, X., & Luo, H. (2022). Developing AI Literacy for Primary and Middle School Teachers in China: Based on a Structural Equation Modeling Analysis. *Sustainability 2022, Vol. 14, Page 14549, 14(21)*, 14549. <https://doi.org/10.3390/SU142114549>

Zhou, Y., Gan, L., Chen, J., Wijaya, T. T., & Li, Y. (2023). Development and validation of a higher-order thinking skills assessment scale for preservice teachers. *Thinking Skills and Creativity*, 48. <https://doi.org/10.1016/J.TSC.2023.101272>