PENINGKATAN KOMPETENSI GURU DALAM KONSTRUKSI INSTRUMEN PENILAIAN: PENDEKATAN BARU ITEM PILIHAN GANDA

Riyadi[1*], Achmad Ridwan[2], Iva Sarifah[3], Awaluddin Tjalla[4], Wardani Rahayu[5]
Program Studi Doktor Penelitian dan Evaluasi Pendidikan, Universitas Negeri Jakarta, Jakarta, Indonesia
riyadi@unj.ac.id, achmadridwan@unj.ac.id, iva.sarifah@unj.ac.id, awaluddin.tjalla@unj.ac.id, wardani.rahayu@unj.ac.id.

Abstract

This community service activity aims to enhance teachers' competencies in developing Deep Learning-based assessment instruments that support 21st-century learning. Needs analysis at the Al Ashriyyah Nurul Iman Islamic Boarding School Foundation shows that teachers are still dominated by the use of conventional multiple-choice questions and face challenges in developing Higher Order Thinking Skills (HOTS) instruments, designing effective distractors, and creating complex multiple-choice formats, such as Multiple Choice Multiple Answer (MCMA) and Multiple Choice with Category (MCC). The activities were carried out thru participatory training, instrument preparation practice, mentoring, and evaluation of 30 teachers. The results show a significant increase in competence, indicated by the rise in the average score from 5.50 on the pre-test to 9.23 on the post-test with an average N-Gain value of 0.84 (high category). In addition, the participants provided very positive feedback and were able to independently create assessment instruments. This activity contributes to strengthening teachers' capacity to develop more authentic, valid, and higher-order thinking-oriented assessments.

Keywords: *Learning Assessment; Deep Learning; HOTS; Teacher Competence; Multiple Choice*

Abstrak

Kegiatan pengabdian ini bertujuan meningkatkan kompetensi guru dalam mengembangkan instrumen penilaian berbasis Deep Learning yang mendukung pembelajaran abad ke-21. Analisis kebutuhan di Yayasan Al Ashriyyah Nurul Iman Islamic Boarding School menunjukkan bahwa guru masih didominasi penggunaan soal pilihan ganda konvensional serta menghadapi kendala dalam menyusun instrumen Higher Order Thinking Skills (HOTS), merancang distraktor yang efektif, dan mengembangkan format pilihan ganda kompleks, seperti Multiple Choice Multiple Answer (MCMA) dan Multiple Choice with Category (MCC). Kegiatan dilaksanakan melalui pelatihan partisipatif, praktik penyusunan instrumen, pendampingan, dan evaluasi terhadap 30 guru. Hasil menunjukkan peningkatan kompetensi yang signifikan, ditunjukkan oleh kenaikan skor rata-rata dari 5,50 pada pre-test menjadi 9,23 pada post-test dengan nilai rata-rata N-Gain sebesar 0,84 (kategori tinggi). Selain itu, peserta memberikan respons yang sangat positif dan mampu menyusun instrumen penilaian secara mandiri. Kegiatan ini berkontribusi terhadap penguatan kapasitas guru dalam mengembangkan asesmen yang lebih autentik, valid, dan berorientasi pada pengembangan kemampuan berpikir tingkat tinggi.

Kata Kunci: *Asesmen Pembelajaran; Deep Learning; HOTS; Kompetensi Guru; Pilihan Ganda;*

1. PENDAHULUAN

Perkembangan pendidikan abad ke-21 menuntut transformasi tidak hanya pada proses pembelajaran, tetapi juga pada sistem asesmen yang mampu mengukur kompetensi berpikir tingkat tinggi. Paradigma pembelajaran berbasis *Deep Learning* menempatkan peserta didik sebagai pembelajar aktif melalui proses belajar yang bermakna, reflektif, dan kontekstual sehingga asesmen tidak lagi berfungsi hanya sebagai alat pengukur hasil belajar (*assessment of learning*), tetapi juga sebagai sarana untuk mendorong dan memperbaiki proses belajar (*assessment for learning*) (Fullan et al., 2018). Oleh karena itu, guru dituntut memiliki kompetensi dalam merancang instrumen penilaian yang valid, autentik, dan mampu mengukur kemampuan berpikir kritis, pemecahan masalah, serta transfer pengetahuan yang menjadi tuntutan pembelajaran abad ke-21 (Brookhart, 2010; Haladyna & Rodriguez, 2013; Learning, 2019; OECD, 2019).

Berbagai penelitian dan kegiatan Pengabdian kepada Masyarakat (PkM) telah menunjukkan bahwa peningkatan kompetensi guru dalam penyusunan instrumen penilaian berbasis *Higher Order Thinking Skills* (HOTS) masih menjadi kebutuhan yang berkelanjutan. Riyadi et al. (2020) melaporkan bahwa pelatihan penyusunan instrumen HOTS bagi guru produktif SMK meningkatkan pemahaman peserta terhadap konsep HOTS serta kemampuan mengembangkan instrumen penilaian berbasis HOTS. Hasil serupa juga ditunjukkan oleh Riyadi et al. (2021), yang menemukan bahwa pelatihan penyusunan instrumen penilaian kinerja praktik berbasis HOTS mampu meningkatkan kemampuan guru dalam menganalisis kompetensi dasar, menyusun kisi-kisi, mengembangkan instrumen, dan melakukan validasi instrumen. Selanjutnya, Riyadi et al. (2022) menunjukkan bahwa pelatihan penyusunan instrumen penilaian berbasis HOTS meningkatkan pemahaman guru dari 47% menjadi 63% serta memperkuat kemampuan menyusun soal pilihan ganda berbasis HOTS. Temuan tersebut sejalan dengan Otaya et al. (2025) yang melaporkan bahwa kompetensi guru dalam mengembangkan soal HOTS masih berada pada kategori sedang, terutama pada penyusunan stimulus, indikator, dan penyelarasan level kognitif. Secara umum, berbagai penelitian tersebut menunjukkan bahwa program pelatihan efektif meningkatkan kompetensi guru, namun masih berfokus pada pengembangan instrumen HOTS konvensional.

Di sisi lain, perkembangan kajian evaluasi pendidikan menunjukkan bahwa tes pilihan ganda (*Multiple Choice Questions*/MCQ) konvensional memiliki sejumlah keterbatasan, antara lain masih memberikan peluang *guessing*, kurang mampu mengukur kemampuan berpikir tingkat tinggi secara optimal, dan belum dapat merepresentasikan pemahaman parsial peserta didik. Oleh karena itu, berbagai penelitian mulai merekomendasikan penggunaan format soal objektif kompleks, seperti *Multiple Choice Multiple Answer* (MCMA) dan *Multiple Choice with Category* (MCC), yang dipadukan dengan sistem penskoran parsial agar mampu memberikan informasi yang lebih akurat mengenai tingkat penguasaan konsep peserta didik. Kajian Hastuti, Sarifah, dan Wahyuni (2026) menunjukkan bahwa integrasi MCMA dengan *Partial Credit Model* menghasilkan pengukuran yang lebih adil dibandingkan penskoran dikotomis karena mampu mengakomodasi pemahaman parsial peserta didik. Selain itu, kajian internasional juga menegaskan bahwa kualitas asesmen tidak hanya ditentukan oleh isi soal, tetapi juga oleh desain item dan strategi penskorannya sehingga mampu menghasilkan informasi yang lebih valid mengenai kemampuan peserta didik (Haladyna & Rodriguez, 2013).

Sejalan dengan berkembangnya paradigma *Deep Learning*, asesmen juga dituntut tidak hanya mengevaluasi hasil akhir, tetapi mampu menangkap proses berpikir peserta didik secara lebih komprehensif. Padmarani, Luriawati, and Wagiran (2026) menunjukkan bahwa sebagian besar soal pilihan ganda yang digunakan di sekolah masih didominasi level kognitif rendah sehingga perlu direkonstruksi agar selaras dengan karakteristik pembelajaran berbasis *Deep Learning*. Dengan demikian, pengembangan kompetensi guru tidak lagi cukup berfokus pada penyusunan soal HOTS konvensional, tetapi perlu diarahkan pada kemampuan mengembangkan format asesmen objektif yang lebih kompleks dan sistem penskoran yang mampu mengidentifikasi tingkat pemahaman peserta didik secara lebih autentik.

Berdasarkan kajian tersebut, masih terdapat kesenjangan antara pelatihan penyusunan instrumen yang telah banyak dilaksanakan dengan kebutuhan implementasi asesmen pada paradigma *Deep Learning*. Sebagian besar program PkM sebelumnya berfokus pada penyusunan soal HOTS, instrumen kinerja, atau penyempurnaan kualitas butir pilihan ganda konvensional (Riyadi et al. 2020, 2021, 2022), sedangkan pelatihan yang mengintegrasikan penyusunan item objektif kompleks, seperti MCMA dan MCC, dengan strategi penskoran modern, seperti *Modified Plus Minus* (MPM) dan *Partial Credit*, masih sangat terbatas. Oleh karena itu, kebaruan program pengabdian ini terletak pada integrasi paradigma *Deep Learning*, pengembangan format soal objektif kompleks, dan penerapan strategi penskoran modern dalam

satu model pelatihan untuk meningkatkan kompetensi guru dalam mengembangkan asesmen yang lebih autentik, informatif, dan mampu mengidentifikasi pemahaman parsial peserta didik.

Yayasan Al Ashriyyah Nurul Iman Islamic Boarding School di Parung, Bogor, merupakan lembaga pendidikan terpadu yang berkomitmen meningkatkan mutu pembelajaran. Namun, hasil analisis kebutuhan dan diskusi bersama mitra menunjukkan bahwa sebagian besar guru masih mengandalkan tes pilihan ganda konvensional dalam proses evaluasi pembelajaran. Sebanyak 56,67% guru menyatakan belum pernah menyusun instrumen menggunakan format MCMA maupun MCC. Guru juga mengalami kesulitan dalam merancang butir soal HOTS, menyusun distraktor yang efektif, serta menerapkan sistem penskoran yang mampu mengakomodasi pemahaman parsial peserta didik. Kondisi tersebut menyebabkan asesmen yang dilaksanakan belum sepenuhnya mampu menggambarkan profil kemampuan peserta didik secara komprehensif dan kurang selaras dengan karakteristik pembelajaran berbasis *Deep Learning*.

Berdasarkan kondisi tersebut, kegiatan Pengabdian kepada Masyarakat ini bertujuan meningkatkan kompetensi guru dalam mengonstruksi instrumen penilaian berbasis *Deep Learning* melalui pelatihan penyusunan soal MCMA dan MCC beserta penerapan strategi penskoran *Modified Plus Minus* (MPM) dan *Partial Credit*, sehingga guru mampu menghasilkan instrumen penilaian yang lebih valid, autentik, dan mendukung peningkatan kualitas pembelajaran.

2. TINJAUAN LITERATUR

Pemberdayaan masyarakat dalam bidang pendidikan merupakan proses sistematis untuk meningkatkan kapasitas individu maupun organisasi agar mampu mengidentifikasi kebutuhan, mengembangkan kompetensi, dan memecahkan permasalahan secara mandiri sehingga tercipta keberlanjutan program. Dalam konteks sekolah, pemberdayaan guru tidak hanya berorientasi pada transfer pengetahuan, tetapi juga pada penguatan kemampuan profesional melalui proses pembelajaran yang partisipatif, reflektif, dan berkelanjutan. Pendekatan ini menempatkan guru sebagai subjek perubahan yang berperan aktif dalam mengembangkan praktik pembelajaran dan asesmen sesuai kebutuhan peserta didik.

Salah satu strategi yang banyak digunakan dalam pemberdayaan guru adalah *capacity building* atau pengembangan kapasitas. Pengembangan kapasitas guru merupakan upaya sistematis untuk memperkuat pengetahuan, keterampilan, dan praktik profesional guru agar mampu meningkatkan kualitas pembelajaran secara berkelanjutan. Dalam bidang pendidikan, pengembangan kapasitas guru dapat dilakukan melalui pengembangan profesional yang efektif, yaitu kegiatan yang berfokus pada materi pembelajaran, melibatkan praktik aktif, menyediakan kesempatan kolaborasi, memberikan umpan balik dan refleksi, serta dilaksanakan secara berkelanjutan (Darling-Hammond et al., 2017). Stosich (2016) menegaskan pentingnya membangun kapasitas guru dan sekolah melalui dukungan yang terintegrasi dengan praktik pembelajaran, sehingga guru dapat menerjemahkan pengetahuan dan keterampilan yang diperoleh ke dalam pelaksanaan pembelajaran. Oleh karena itu, pelatihan yang hanya berorientasi pada penyampaian materi perlu dilengkapi dengan kesempatan bagi guru untuk menerapkan pengetahuan dan keterampilan dalam praktik, mengembangkan produk atau solusi pembelajaran, serta memperoleh bimbingan dan umpan balik selama proses implementasi. Pendekatan tersebut lebih selaras dengan prinsip pengembangan profesional guru yang efektif dibandingkan pelatihan yang hanya berfokus pada penyampaian informasi (Darling-Hammond et al., 2017; Stosich, 2016).

Pengembangan kompetensi guru dalam bidang asesmen merupakan salah satu aspek penting dari profesionalisme pendidik. Kompetensi asesmen mencakup kemampuan merancang instrumen, melaksanakan penilaian, menginterpretasikan hasil, serta memanfaatkan informasi asesmen untuk memperbaiki proses pembelajaran (Brookhart, 2013). Dalam pembelajaran abad ke-21, asesmen tidak lagi dipandang sebagai kegiatan administratif,

melainkan sebagai bagian integral dari proses pembelajaran yang mendukung pengembangan kemampuan berpikir kritis, pemecahan masalah, kreativitas, dan pengambilan keputusan peserta didik (OECD, 2019). Oleh sebab itu, peningkatan kompetensi guru dalam mengembangkan instrumen penilaian menjadi prasyarat penting bagi implementasi pembelajaran yang berorientasi pada *Deep Learning*.

Paradigma *Deep Learning* menuntut asesmen yang mampu mengungkap kualitas penalaran peserta didik, bukan sekadar ketepatan jawaban akhir. Fullan et al. (2018) menegaskan bahwa pembelajaran bermakna hanya dapat diwujudkan apabila asesmen mampu memberikan informasi yang kaya mengenai proses berpikir peserta didik sehingga hasil penilaian dapat dimanfaatkan sebagai dasar perbaikan pembelajaran. Dalam konteks tersebut, pengembangan format asesmen objektif kompleks, seperti *Multiple Choice Multiple Answer* (MCMA) dan *Multiple Choice with Category* (MCC), yang dipadukan dengan strategi penskoran parsial, menjadi salah satu alternatif untuk meningkatkan kualitas informasi hasil asesmen karena mampu mengidentifikasi tingkat penguasaan konsep secara lebih rinci dibandingkan tes pilihan ganda konvensional (Haladyna & Rodriguez, 2013).

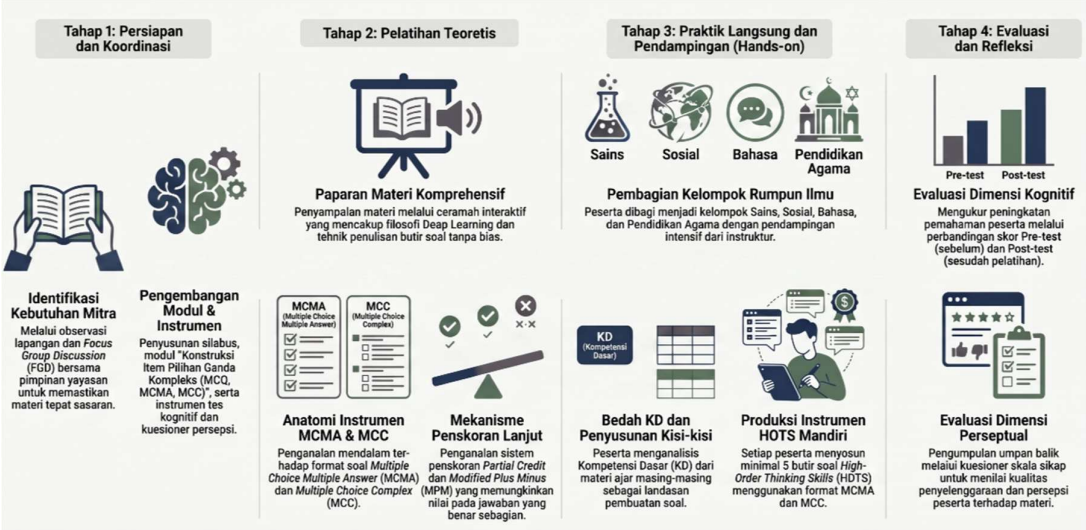
Pelaksanaan pemberdayaan guru dalam kegiatan ini menggunakan pendekatan pelatihan partisipatif yang selaras dengan teori *Experiential Learning* dari Kolb (1984). Teori ini menjelaskan bahwa pembelajaran orang dewasa berlangsung secara optimal melalui siklus pengalaman konkret, refleksi, konseptualisasi, dan eksperimen aktif. Oleh karena itu, pelatihan tidak hanya berupa penyampaian konsep, tetapi juga melibatkan praktik penyusunan instrumen, diskusi, pemberian umpan balik, serta pendampingan sehingga peserta memperoleh pengalaman langsung dalam mengembangkan instrumen penilaian berbasis *Deep Learning*. Pendekatan tersebut terbukti lebih efektif dalam meningkatkan kompetensi profesional guru dibandingkan pelatihan yang hanya berorientasi pada ceramah (Darling-Hammond et al., 2017).

Berdasarkan landasan konseptual tersebut, kegiatan Pengabdian kepada Masyarakat ini memandang pemberdayaan guru sebagai proses pengembangan kapasitas yang berkelanjutan melalui pelatihan partisipatif, praktik, dan pendampingan. Pendekatan ini diharapkan tidak hanya meningkatkan pengetahuan guru mengenai asesmen berbasis *Deep Learning*, tetapi juga menghasilkan perubahan kompetensi yang tercermin dalam kemampuan mengembangkan instrumen penilaian yang lebih autentik, valid, dan mampu mengukur kemampuan berpikir tingkat tinggi peserta didik secara lebih komprehensif.

3. METODE PELAKSANAAN

Kegiatan Pengabdian kepada Masyarakat ini menggunakan pendekatan pelatihan partisipatif (*participatory training*), pendampingan terstruktur (*mentoring*), praktik langsung (*hands-on training*), dan evaluasi berkelanjutan seperti pada Gambar 1. Pendekatan tersebut dipilih karena berbagai penelitian mutakhir menunjukkan bahwa pengembangan profesional orang dewasa menjadi lebih efektif ketika peserta terlibat secara aktif dalam proses pembelajaran, memperoleh kesempatan untuk mempraktikkan keterampilan baru dalam kerja nyata, serta mendapatkan dukungan dan tindak lanjut selama proses implementasi (Ansyari et al., 2022; Sims et al., 2021). Selain itu, kajian terbaru mengenai transfer hasil pelatihan menegaskan bahwa *active learning*, mentoring, dan pembelajaran yang relevan dengan tugas profesional berkontribusi signifikan terhadap keberhasilan *knowledge transfer* dari kegiatan pelatihan ke praktik kerja sehari-hari (Nīmante et al., 2025; Siliņa-Jasjukeviča et al., 2025). Dengan demikian, peserta tidak hanya berperan sebagai penerima informasi secara pasif, tetapi sebagai subjek aktif yang mengonstruksi dan menerapkan pengetahuan baru untuk menyelesaikan permasalahan yang dihadapi dalam lingkungan kerjanya. Sasaran kegiatan ini adalah 30 orang guru yang mewakili berbagai bidang studi (seksama lintas disiplin ilmu) di lingkungan Yayasan Al Ashriyyah Nurul Iman Islamic Boarding School, Parung, Bogor. Program ini dirancang dengan durasi yang proporsional untuk menjembatani kesenjangan

kompetensi guru dari tataran teoretis konseptual hingga kemampuan praksis teknis dalam menyusun instrumen penilaian berbasis HOTS yang mutakhir.



Gambar 1. Tahapan kegiatan PKM

Tahapan pelaksanaan kegiatan dibagi ke dalam empat tahap utama yang saling berkesinambungan.

1. Tahap Persiapan dan Koordinasi.
   Pada tahap ini, tim pengabdi melakukan observasi lapangan dan *Focus Group Discussion* (FGD) dengan pimpinan yayasan untuk melakukan identifikasi kebutuhan mitra secara presisi. Berdasarkan hasil FGD, tim mengembangkan silabus pelatihan, merancang modul "Konstruksi Item Pilihan Ganda Kompleks (MCQ, MCMA dan MCC)", serta menyusun instrumen evaluasi kegiatan (berupa tes kognitif dan kuesioner persepsi).
2. Tahap Pelaksanaan Pelatihan Teoretis.
   Pada tahap ini, tim pengabdi memberikan paparan materi komprehensif melalui metode ceramah interaktif dan tanya jawab. Materi yang disampaikan meliputi esensi dan filosofi *Deep Learning*, karakteristik instrumen asesmen yang baik, serta teknik penulisan butir soal konvensional yang terhindar dari bias. Fokus utama pada sesi ini adalah pengenalan mendalam terhadap anatomi dan keunggulan format MCMA dan MCC. Pada sesi ini juga diuraikan mekanisme penskoran tingkat lanjut yang menantang paradigma lama, yaitu penskoran *Partial Credit* dan *Modified Plus Minus* (MPM), yang memungkinkan pemberian skor pada jawaban yang sebagian benar.
3. Tahap Praktik Langsung dan Pendampingan (*Hands-on*)
   Peserta pelatihan dibagi ke dalam kelompok-kelompok kecil berdasarkan serumpun keilmuan (kelompok Sains, Sosial, Bahasa, dan Pendidikan Agama). Setiap kelompok didampingi secara intensif oleh instruktur dari tim pengabdi. Peserta ditugaskan untuk membawa materi ajar masing-masing, kemudian membedah Kompetensi Dasar (KD) dan menyusun kisi-kisi soal. Selanjutnya, peserta langsung menyusun minimal 5 butir instrumen soal HOTS mandiri yang mengaplikasikan format MCMA dan MCC.
4. Tahap Evaluasi dan Refleksi
   Tahap terakhir ditujukan untuk mengukur tingkat efektivitas, keberhasilan, dan dampak pelatihan. Evaluasi dilakukan secara komprehensif melalui dua dimensi: kognitif dan perseptual. Dimensi kognitif diukur melalui instrumen tes (*pre-test* yang diberikan sebelum pelatihan dimulai dan *post-test* yang diberikan di akhir kegiatan). Dimensi

perseptual diukur melalui penyebaran kuesioner skala sikap pasca-pelatihan untuk menjaring umpan balik terkait penyelenggaraan acara.

Data kuantitatif yang terkumpul dari seluruh instrumen evaluasi dianalisis secara deskriptif menggunakan perangkat lunak statistik. Untuk mengetahui signifikansi peningkatan kompetensi kognitif peserta, dilakukan analisis perbandingan rata-rata nilai awal dan akhir. Besaran peningkatan efektivitas pembelajaran dihitung menggunakan rumus *Normalized Gain* (N-Gain) dari Hake, dengan kategori: Tinggi (N-Gain > 0,7), Sedang (0,3 ≤ N-Gain ≤ 0,7), dan Rendah (N-Gain < 0,3). Sementara itu, tingkat kepuasan dan keterterimaan materi oleh peserta dianalisis menggunakan tendensi sentral (rata-rata) dari skor respons kuesioner berskala Likert 1 hingga 5. Hal ini penting untuk mengevaluasi kualitas narasumber, relevansi materi, dan kemanfaatan yang dirasakan mitra pasca intervensi secara holistik.

4. HASIL DAN PEMBAHASAN

Kegiatan Pengabdian kepada Masyarakat (PkM) di Yayasan Al Ashriyyah Nurul Iman Islamic Boarding School diikuti oleh 30 guru dan berlangsung dengan baik hingga seluruh rangkaian kegiatan selesai. Tingginya tingkat kehadiran dan partisipasi aktif peserta selama pelatihan menunjukkan adanya kebutuhan nyata terhadap peningkatan kompetensi dalam pengembangan instrumen penilaian. Selama sesi diskusi, guru mengemukakan berbagai permasalahan yang dihadapi dalam praktik asesmen, seperti kecenderungan peserta didik menebak jawaban, kesulitan menyusun soal berbasis *Higher Order Thinking Skills* (HOTS), serta keterbatasan dalam memberikan penilaian yang mampu mengakomodasi pemahaman parsial siswa. Kondisi tersebut memperlihatkan bahwa guru memerlukan penguatan kompetensi asesmen yang lebih sesuai dengan karakteristik pembelajaran abad ke-21. Temuan ini sejalan dengan Riyadi et al. (2022) dan Otaya et al. (2025) yang menyatakan bahwa pelatihan berbasis praktik mampu meningkatkan kapasitas guru dalam mengembangkan instrumen penilaian, sekaligus memperkuat motivasi mereka untuk melakukan perbaikan pembelajaran secara berkelanjutan.

Pada tahap pelatihan, peserta diperkenalkan pada pengembangan instrumen asesmen berbasis *Deep Learning* melalui pemanfaatan format *Multiple Choice Multiple Answer* (MCMA) dan *Multiple Choice with Category* (MCC). Berbeda dengan format *Multiple Choice Single Answer* (MCSA) yang hanya memiliki satu jawaban benar, MCMA mengharuskan peserta didik mengevaluasi setiap opsi jawaban secara independen, sedangkan MCC menuntut kemampuan mengelompokkan sejumlah pernyataan ke dalam kategori tertentu. Selanjutnya, peserta mengikuti praktik penyusunan instrumen dengan pendampingan intensif mulai dari penyusunan stimulus kontekstual, perancangan distraktor yang efektif, hingga penerapan strategi penskoran.



Gambar 2. Pembukaan kegiatan Pelatihan


Gambar 3. Suasana Pendampingan Praktik Konstruksi Instrumen Asesmen

Gambar 3 memperlihatkan proses pendampingan secara langsung dalam penyusunan instrumen, yang memungkinkan peserta memperoleh umpan balik terhadap kualitas stimulus, distraktor, dan strategi penskoran yang dikembangkan. Selama proses tersebut, sebagian besar guru menunjukkan perubahan dalam cara mengembangkan soal, dari yang sebelumnya didominasi oleh pertanyaan berbasis ingatan menuju penyusunan soal yang menuntut kemampuan analisis dan evaluasi. Pendampingan juga menunjukkan bahwa kesulitan utama guru bukan terletak pada penguasaan materi, melainkan pada kemampuan menyusun distraktor yang logis dan mampu membedakan tingkat penguasaan peserta didik. Temuan ini mendukung pendapat Haladyna dan Rodriguez (2013) bahwa kualitas asesmen tidak hanya ditentukan oleh isi materi, tetapi juga oleh kualitas konstruksi butir soal dan fungsi setiap alternatif jawaban. Selain itu, penggunaan format MCMA dan MCC sejalan dengan rekomendasi Beiting-Parrish

et al. (2020) dan Hastuti et al. (2026) yang menyatakan bahwa format soal objektif kompleks lebih efektif dalam mengurangi peluang *guessing* serta memberikan informasi yang lebih akurat mengenai kemampuan berpikir peserta didik.

Efektivitas pelatihan dievaluasi menggunakan tes yang terdiri atas 10 soal dengan skor maksimum 10. Rata-rata skor peserta meningkat dari 5,50 pada pre-test menjadi 9,23 pada post-test sebagaimana disajikan pada Tabel 1. Rata-rata skor peserta meningkat dari 5,50 pada *pre-test* menjadi 9,23 pada *post-test*. Hasil ini menunjukkan bahwa pelatihan berhasil meningkatkan penguasaan konsep guru mengenai pengembangan instrumen asesmen berbasis *Deep Learning*.

Tabel 1. Perbandingan Skor Pre-test dan Post-test Peserta Pelatihan

Aspek Statistik	Pre-test	Post-test	Peningkatan (N-Gain)
Nilai Rata-Rata	5,50	9,23	0,84
Nilai Minimum	3,00	6,50	-
Nilai Maksimum	7,00	10,00	-
Standar Deviasi	1,12	0,85	0,12

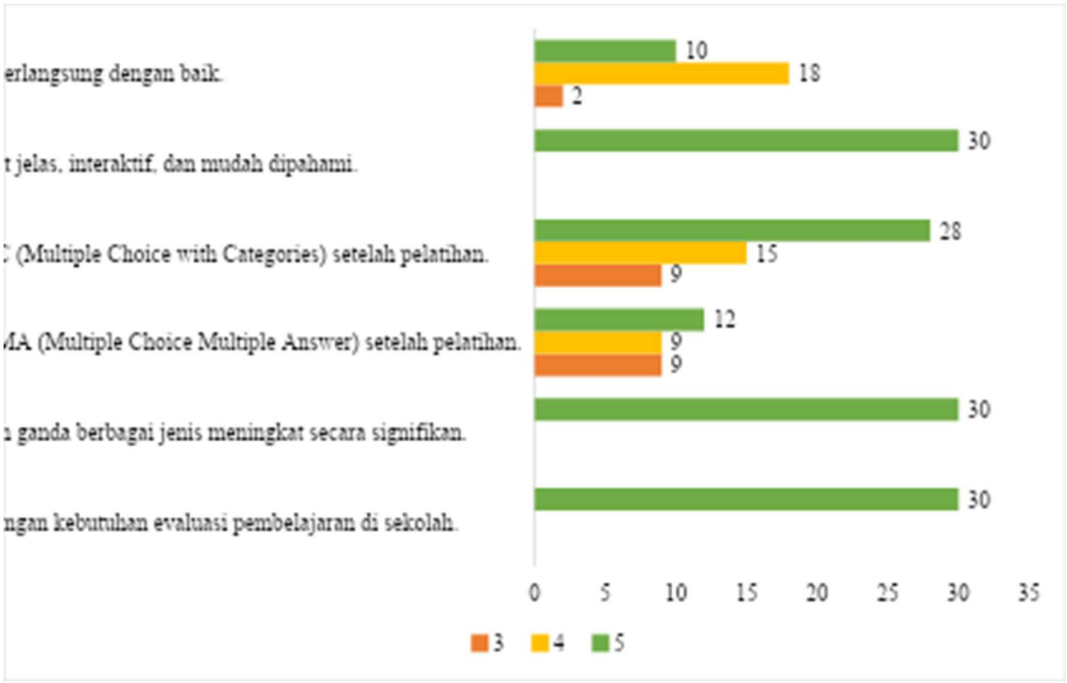
Selanjutnya, analisis menggunakan *Normalized Gain* (N-Gain) menghasilkan nilai rata-rata sebesar 0,84 yang termasuk dalam kategori tinggi. Distribusi kategori N-Gain (Tabel 2) menunjukkan bahwa sebanyak 22 guru (73,33%) berada pada kategori tinggi, tujuh guru (23,33%) pada kategori sedang, dan hanya satu guru (3,33%) pada kategori rendah.

Tabel 2. Distribusi Kategori N-Gain

Kategori N-Gain	Kriteria	Frekuensi	Persentase
Tinggi	N-Gain $\geq$ 0,70	22	73,33%
Sedang	0,30 $\leq$ N-Gain < 0,70	7	23,33%
Rendah	N-Gain < 0,30	1	3,33%
Jumlah		30	100,00%

Peningkatan tersebut menunjukkan bahwa model pelatihan yang mengombinasikan penyampaian konsep, praktik penyusunan instrumen, dan pendampingan memberikan dampak yang efektif terhadap peningkatan kompetensi peserta. Temuan ini memperkuat hasil penelitian Joyce dan Showers (2002) yang menunjukkan bahwa pelatihan guru akan memberikan dampak optimal apabila disertai praktik dan *coaching*, bukan hanya ceramah. Hasil tersebut juga sejalan dengan Darling-Hammond dan Adamson (2014), Asterhan and Lefstein (2024), serta Sims et al. (2025) yang menegaskan bahwa pengembangan profesional guru lebih efektif apabila dilaksanakan secara kolaboratif, kontekstual, dan berbasis pengalaman langsung (*experiential learning*). Dengan demikian, peningkatan kompetensi yang diperoleh peserta tidak hanya mencerminkan pemahaman konseptual, tetapi juga kemampuan mengaplikasikan pengetahuan tersebut dalam penyusunan instrumen penilaian.

Evaluasi persepsi peserta yang dilakukan pada akhir kegiatan juga menunjukkan respons yang sangat positif. Gambar 4 menunjukkan bahwa seluruh aspek pelatihan memperoleh respons positif dari peserta, terutama pada relevansi materi dan peningkatan pemahaman. Namun, tingkat kepercayaan diri dalam menyusun soal MCC masih lebih rendah dibandingkan MCMA, yang mengindikasikan perlunya pendampingan lanjutan.



Gambar 4. Hasil Evaluasi perseptual peserta

Aspek relevansi materi terhadap tugas profesional guru, peningkatan pemahaman konsep asesmen, dan kompetensi narasumber memperoleh skor rata-rata tertinggi. Sementara itu, tingkat kepercayaan diri peserta dalam menyusun soal MCMA memperoleh skor rata-rata 4,10, sedangkan penyusunan MCC memperoleh skor 3,90. Perbedaan tersebut menunjukkan bahwa guru memandang penyusunan MCC relatif lebih kompleks dibandingkan MCMA karena memerlukan kemampuan mengembangkan kategori yang konsisten serta mempertimbangkan hubungan antarkonsep. Kondisi ini dapat dijelaskan melalui *Cognitive Load Theory*, yang menyatakan bahwa tugas dengan tingkat kompleksitas konseptual lebih tinggi akan meningkatkan beban kognitif, terutama bagi individu yang baru mempelajari suatu keterampilan (Sweller, 2011; Zou et al., 2025). Oleh karena itu, pendampingan lanjutan masih diperlukan agar guru semakin terampil mengembangkan format soal tersebut.

Selain meningkatkan kompetensi teknis, kegiatan ini juga mendorong perubahan paradigma guru terhadap fungsi asesmen. Peserta mulai memandang asesmen bukan sekadar sebagai alat pemberian nilai (*assessment of learning*), tetapi sebagai bagian dari proses pembelajaran yang berfungsi mendiagnosis kesulitan belajar serta memberikan umpan balik bagi peserta didik (*assessment for learning*). Pengenalan strategi penskoran *Partial Credit* dan *Modified Plus Minus* (MPM) semakin memperkuat pemahaman tersebut karena memungkinkan pemberian penghargaan terhadap pemahaman parsial siswa, berbeda dengan sistem penskoran dikotomis yang hanya membedakan jawaban benar dan salah. Pendekatan ini sejalan dengan konsep *assessment fairness* yang dikemukakan oleh Masters (2016) dan Wu, Tam, and Jen (2016), bahwa sistem penskoran parsial mampu menghasilkan informasi yang lebih adil dan diagnostik mengenai kemampuan peserta didik.

Meskipun demikian, pelaksanaan kegiatan masih menghadapi beberapa keterbatasan. Waktu praktik yang tersedia belum cukup untuk mengembangkan instrumen pada seluruh mata pelajaran secara mendalam. Selain itu, beberapa guru, terutama yang telah lama menggunakan sistem penilaian konvensional, masih memerlukan waktu untuk beradaptasi dengan penyusunan soal objektif kompleks dan strategi penskoran MPM. Oleh karena itu, keberlanjutan program melalui komunitas belajar guru atau *Musyawarah Guru Mata Pelajaran*

(MGMP) internal sekolah menjadi penting agar proses pendampingan dapat terus berlangsung. Pendekatan tersebut sejalan dengan rekomendasi Ambarita, Meilani, and Wanasida (2022) dan Septiand et al. (2025) yang menekankan bahwa pengembangan kompetensi guru akan lebih berkelanjutan apabila didukung oleh komunitas belajar profesional yang memungkinkan terjadinya kolaborasi, refleksi, dan berbagi praktik baik secara berkesinambungan.

5. KESIMPULAN

Kegiatan Pengabdian kepada Masyarakat ini menunjukkan bahwa pelatihan penyusunan instrumen penilaian berbasis *Deep Learning* dapat meningkatkan kompetensi guru dalam mengembangkan asesmen yang lebih sesuai dengan tuntutan pembelajaran abad ke-21. Peningkatan tersebut ditunjukkan oleh kenaikan skor rata-rata peserta dari 5,50 menjadi 9,23 pada tes dengan skor maksimum 10, yang menunjukkan adanya peningkatan penguasaan materi setelah pelatihan. Hasil tersebut juga didukung oleh nilai rata-rata *Normalized Gain* (N-Gain) sebesar 0,84 yang termasuk dalam kategori tinggi. Selain meningkatkan pemahaman konseptual, kegiatan ini juga memberikan pengalaman praktis kepada guru dalam menyusun instrumen berbentuk *Multiple Choice Multiple Answer* (MCMA) dan *Multiple Choice with Category* (MCC), serta menerapkan strategi penskoran *Modified Plus Minus* (MPM) dan *Partial Credit*. Hasil evaluasi peserta juga menunjukkan bahwa materi pelatihan dinilai relevan dengan kebutuhan guru dan mendukung pengembangan praktik asesmen yang lebih autentik serta mampu memberikan informasi yang lebih komprehensif mengenai tingkat penguasaan peserta didik.

Untuk mendukung keberlanjutan program, sekolah disarankan memfasilitasi kegiatan pengembangan profesional secara berkelanjutan melalui forum *Musyawarah Guru Mata Pelajaran* (MGMP) atau komunitas belajar internal sebagai wadah untuk berdiskusi, melakukan telaah sejawat (*peer review*), dan menyempurnakan instrumen penilaian yang dikembangkan. Selain itu, pelatihan lanjutan dapat diarahkan pada implementasi instrumen MCMA dan MCC dalam sistem asesmen berbasis digital, seperti *Computer-Based Test* (CBT) atau *Learning Management System* (LMS), sehingga proses penskoran parsial dapat dilakukan secara lebih efisien. Pengembangan program berikutnya juga dapat mencakup penyusunan asesmen autentik berbasis proyek dan instrumen penilaian kinerja sebagai pelengkap asesmen objektif dalam mendukung evaluasi pembelajaran yang lebih komprehensif.

6. UCAPAN TERIMA KASIH

Tim pengabdi menyampaikan terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Negeri Jakarta atas dukungan yang diberikan sehingga kegiatan Pengabdian kepada Masyarakat ini dapat terlaksana dengan baik. Ucapan terima kasih juga disampaikan kepada Ketua Yayasan Al Ashriyyah Nurul Iman Islamic Boarding School, Bapak Sonif Mahfud, M.Ag., beserta jajaran manajemen sekolah atas kerja sama dan dukungan selama pelaksanaan kegiatan. Apresiasi turut diberikan kepada seluruh guru peserta yang telah berpartisipasi secara aktif dalam setiap tahapan kegiatan, sehingga program ini dapat berjalan sesuai dengan tujuan yang telah ditetapkan.

7. DAFTAR PUSTAKA

Ambarita, E., Meilani, Y. F. C. P., & Wanasida, A. S. (2022). Peranan Musyawarah Guru Mata Pelajaran (MGMP) Sebagai Organisasi Pembelajar terhadap Peningkatan Kompetensi Guru Bidang Studi. *SUKMA: Jurnal Pendidikan*, *6*(2), 227–243.

Ansyari, M. F., Groot, W., & De Witte, K. (2022). A systematic review and meta-analysis of data use professional development interventions. *Journal of Professional Capital and Community*, *7*(3), 256–289. https://doi.org/10.1108/JPCC-09-2021-0055

Asterhan, C. S. C., & Lefstein, A. (2024). The search for evidence-based features of effective teacher professional development: A critical analysis of the literature. *Professional Development in Education*, *50*(1), 11–23.

Beiting-Parrish, M., Verkuilen, J., McCluskey, S., Everson, H., & Wladis, C. (2020). Multiple Answer Multiple Choice Items: A Problematic Item Type? *The Annual Meeting of the Psychometric Society*, 347–357.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Ascd.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Ascd.

Darling-Hammond, L., & Adamson, F. (2014). *Beyond the bubble test: How performance assessments support 21st century learning*. John Wiley & Sons.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. http://learningpolicyinstitute.org/product/effective-teacher-professional-development-report

Fullan, M., Quinn, J., & Mceachen, J. (2018). *Deep learning: Engage the world, change the world*. Corwin Press. https://smansamendobarat.id/pluginfile.php/668/coursecat/description/02 _ deep learning.pdf

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and Validating Test Items* (1st Editio). Routledge. https://doi.org/10.4324/9780203850381

Hastuti, N. S., Sarifah, I., & Wahyuni, L. D. (2026). Tinjauan Sistematis Kebutuhan Asesmen Koneksi Matematis Berbasis Integrasi MCMA dengan PCM. *J-PiMat: Jurnal Pendidikan Matematika*, *8*(1), 2171–2182.

Joyce, B., & Showers, B. (2002). *Student Achievement through Staff Development*. Association for Supervision and Curriculum Development.

Kolb, D. A. (1984). *xperiential learning: Experience as the source of learning and development*. Prentice Hall.

Masters, G. N. (2016). Partial Credit Model. In *Handbook of Item Response Theory*. Chapman and Hall/CRC.

Nīmante, D., Kokare, M., Baranova, S., & Surikova, S. (2025). Transferring Results of Professional Development into Practice: A Scoping Review. In *Education Sciences* (Vol. 15, Issue 1, p. 95). https://doi.org/10.3390/educsci15010095

OECD. (2019). Future of education and skills 2030: OECD learning compass 2030. *OECD Publishing*. https://www.oecd.org/en/about/projects/future-of-education-and-skills-2030.html

Otaya, L. G., Yahiji, K., Rasimin, R., Rahmawati, R., Imtihan, N., & Muhamad, N. B. (2025). Assessment of teachers' ability to develop HOTS-based test items: A GRM analysis. *Jurnal Ilmiah Peuradeun*, *13*(3), 1995–2018.

Padmarani, K., Luriawati, D., & Wagiran, W. (2026). Evaluasi dan Rekonstruksi Butir Soal Pilihan Ganda sebagai Asesmen Formatif dalam Pembelajaran Berbasis Deep Learning. *JURNAL MULTIDISIPLIN ILMU AKADEMIK*, *3*(2), 541–550.

Riyadi, Madani, F., Apriliani, L., & Siva, A. N. (2022). Peningkatan kompetensi guru melalui penyusunan instrumen penilaian berbasis HOTS. *Sarwahita*, *19*, 568–582.

Riyadi Riyadi, Margono, G., Avianti, R. A., Nahdatul, R., Raihasabiq, M., & Islami, A. R. (2021). Peningkatan kualitas guru produktif SMK melalui penyusunan instrumen penilaian kinerja praktik berbasis HOTS. *Prosiding Seminar Nasional Pengabdian Kepada Masyarakat*, *2*, SNPPM2021SH-9.

Riyadi, Tyassmadi, A. T., Avianti, R. A., Cahyaningsih, A., Alvaritsi, M., & Kevin, M. (2020). Peningkatan Mutu Pembelajaran Melalui Pelatihan Penyusunan Instrumen Higher Order Tihinking Skill (HOTS) Bagi Guru Produktif SMK Negeri Di Jakarta Timur. *Prosiding Seminar Nasional Pengabdian Kepada Masyarakat*, *1*, SNPPM2020P-31. https://journal.unj.ac.id/unj/index.php/snppm/article/view/19623

Septiand, W. R., Hariyati, N., Mustaji, M., Khamidi, A., Amalia, K., & Nursalim, M. (2025). Implementasi MGMP sebagai strategi penguatan kompetensi guru di Community Learning Center (CLC). *JIIP-Jurnal Ilmiah Ilmu Pendidikan*, *8*(6), 6681–6689.

Siliņa-Jasjukeviča, G., Lūsēna-Ezera, I., Iliško, D., & Surikova, S. (2025). Promoting effective vocational education and training teacher's professional development and its transfer to practice: A systematic review. *Education Sciences*, *15*(5), 596.

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., Van Herwegen, J., & Anders, J. (2025). Effective teacher professional development: New theory and a meta-analytic test. *Review of Educational Research*, *95*(2), 213–254.

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Van Herwegen, J., & Anders, J. (2021). What are the Characteristics of Teacher Professional Development that Increase Pupil Achievement? A systematic review and meta-analysis. *Education Endowment Foundation*. https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/teacherprofessional-development-characteristics

Stosich, E. L. (2016). Building teacher and school capacity to teach to ambitious standards in high-poverty schools. *Teaching and Teacher Education*, *58*, 43–53. https://doi.org/https://doi.org/10.1016/j.tate.2016.04.010

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Elsevier.

Wu, M., Tam, H. P., & Jen, T.-H. (2016). Educational measurement for applied researchers. *Theory into Practice*, *136*.

Zou, L., Zhang, Z., Mavilidi, M., Chen, Y., Herold, F., Ouwehand, K., & Paas, F. (2025). The synergy of embodied cognition and cognitive load theory for optimized learning. *Nature Human Behaviour*, *9*(5), 877–885.