MISOOL: Jurnal Pendidikan Dasar
ISSN 2964-691X
Volume 8 Issue 2: 2026 https://doi.org/10.47945/misool.v8i2.3351

Changes in Elementary School Teachers' Competency Scores After In-On-In Deep Learning Pedagogy Training in Lhokseumawe, Indonesia

Julrissani[a*] | Sarah Fazilla[a] | Kamaruzzaman[b] | Isfiana Nursari[a]

[a]Pendidikan Guru Madrasah Ibtidaiyah, Universitas Islam Negeri Sultanah Nahrasiyah Lhokseumawe, Indonesia
[b]National Research University Higher School of Economics, Rusia

Correspondence: jjulrissani@gmail.com

Article History:
Received: 15/09/2026
Accepted: 21/11/2026

Keywords:
Deep Learning Pedagogy, Teacher Competence, Teacher Professional Development, Merdeka Curriculum, Elementary School Teachers.

Abstract
This study examined changes in elementary school teachers' competency test scores on Deep Learning Pedagogy (*Pembelajaran Mendalam*) within Indonesia's Merdeka Curriculum following a four-month professional development program in Lhokseumawe City. Deep Learning Pedagogy is a pedagogical approach and is distinct from deep learning in artificial intelligence. Teacher competence was operationalized as conceptual and pedagogical knowledge of Deep Learning Pedagogy, measured with the training provider's competency test 30 items; score range 0–100). A one-group pretest–posttest design was employed with 26 teachers from 13 elementary schools who completed an In-On-In (in-service, on-the-job, in-service) training program from September to December 2025. Data were analyzed using descriptive statistics, the Shapiro–Wilk test, a paired-samples *t*-test, Cohen's d_z with 95% confidence intervals, and normalized gain (N-Gain). Test scores showed a statistically significant increase from 70.81 (*SD* = 17.84) to 84.08 (*SD* = 7.78), with a mean difference of 13.27 points (95% CI [6.98, 19.56]; $t(25) = 4.34$, $p < .001$; $d_z = 0.85$, 95% CI [0.39, 1.30]). Individual changes varied: 11.5% of participants showed high, 57.7% moderate, and 30.8% low N-Gain. Product and attitude–participation scores recorded during the training were very high but are reported as indicators of engagement rather than as independent evidence of change in competence. Because the design lacked a control group, the increase cannot be causally attributed to the training; testing effects, maturation, and history cannot be ruled out. The findings offer preliminary positive evidence that warrants controlled evaluations and classroom-observation studies.

© 2026 The Authors. This open access article is distributed under a (CC-BY License)

How to Cite:
Julrissani, Julrissani, S. F., Kamaruzzaman, & Julrissani, I. N. (2026). Changes in Elementary School Teachers' Competency Scores After In-On-In Deep Learning Pedagogy Training in Lhokseumawe, Indonesia. MISOOL: Jurnal Pendidikan Dasar, 8(2), 301–316. https://doi.org/10.47945/misool.v8i2.3351

PENDAHULUAN

Kurikulum Merdeka diluncurkan secara resmi pada 11 Februari 2022 setelah diujicobakan sejak tahun 2021 melalui Program Sekolah Penggerak (Direktorat Sekolah Dasar, 2022). Kurikulum ini diposisikan sebagai bagian dari kebijakan pemulihan pembelajaran pascapandemi COVID-19 (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2022), dengan karakteristik yang lebih fleksibel, lebih sederhana, dan berpusat pada peserta didik, serta memberi ruang bagi guru untuk menyesuaikan pembelajaran dengan kebutuhan peserta didik (Direktorat Sekolah Dasar, 2022). Pada Oktober 2024, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) direstrukturisasi menjadi tiga kementerian, yaitu Kementerian Pendidikan Dasar dan Menengah (Kemendikdasmen); Kementerian Pendidikan Tinggi, Sains, dan Teknologi; serta Kementerian Kebudayaan (Republik Indonesia, 2024). Kemendikdasmen kemudian menempatkan peningkatan mutu pembelajaran sebagai prioritas, dengan argumen bahwa berbagai indikator capaian belajar belum menunjukkan perbaikan yang memadai (Rahmawati et al., 2025).

Salah satu indikator yang sering dirujuk adalah hasil *Programme for International Student Assessment* (PISA). Pada PISA 2022, skor rata-rata peserta didik Indonesia adalah 366 untuk matematika, 359 untuk membaca, dan 383 untuk sains, atau turun 12–13 poin dibandingkan dengan PISA 2018 (Tabel 1), meskipun peringkat Indonesia naik lima hingga enam posisi (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2023; OECD, 2023). Skor tersebut termasuk yang terendah sejak Indonesia pertama kali

berpartisipasi dalam PISA pada tahun 2000, sehingga tren capaian dalam dua dekade terakhir belum menunjukkan peningkatan yang berarti (Aziz et al., 2024). Kenaikan peringkat yang disertai penurunan skor juga menunjukkan bahwa peringkat bersifat relatif terhadap negara peserta lain. Penurunan skor pada periode yang sama terjadi pula di banyak negara, termasuk pada rata-rata negara OECD, dan dikaitkan antara lain dengan gangguan pembelajaran selama pandemi (OECD, 2023). Karena PISA mengukur peserta didik berusia 15 tahun, data ini hanya menjadi bukti tidak langsung tentang mutu pembelajaran di sekolah dasar.

Tabel 1. Skor Rata-Rata PISA Indonesia Tahun 2018 dan 2022

Domain	Indonesia 2018	Indonesia 2022	Perubahan	Rata-rata OECD 2022
Membaca	371	359	−12	476
Matematika	379	366	−13	472
Sains	396	383	−13	485

Sumber: OECD (2023).

Hasil PISA merupakan indikator pada tingkat sistem yang dipengaruhi oleh banyak faktor, antara lain latar belakang sosial-ekonomi, kesempatan belajar, ketersediaan sumber belajar, dukungan sekolah, dan gangguan pembelajaran selama pandemi (OECD, 2023). Kajian di Indonesia juga mengaitkan rendahnya literasi sains dengan pembelajaran yang belum kontekstual, rendahnya kemampuan membaca, penggunaan bahan ajar yang kurang tepat, miskonsepsi peserta didik, serta sarana dan prasarana yang belum memadai (Suparya et al., 2022; Yusmar & Fadilah, 2023), di samping kompetensi guru yang masih perlu ditingkatkan dan beban administratif yang tinggi (Alfaruqi & Nurwahidah, 2025; Altariusta, 2025). Karena dipengaruhi banyak faktor, skor PISA tidak dapat dijadikan bukti bahwa kurangnya pelatihan guru merupakan penyebab, atau bahwa pelatihan tertentu merupakan solusi langsung, atas rendahnya capaian peserta didik. Dalam kerangka kebijakan, Kemendikdasmen menekankan bahwa pendekatan pembelajaran yang kurang efektif, terbatasnya ruang kreativitas guru, dan rendahnya kemampuan berpikir kritis peserta didik merupakan persoalan yang perlu dijawab (Rahmawati et al., 2025). Keberhasilan kebijakan perbaikan pembelajaran semacam itu bergantung pada kapasitas guru untuk menerjemahkannya ke dalam praktik kelas. Oleh karena itu, perubahan kompetensi guru selama mengikuti program pengembangan profesional perlu didokumentasikan secara empiris sebelum dampaknya diuji dengan desain terkontrol.

Sebagai respons kebijakan, Kemendikdasmen menetapkan Pembelajaran Mendalam sebagai pendekatan pembelajaran dalam kurikulum melalui Peraturan Menteri Pendidikan Dasar dan Menengah Nomor 13 Tahun 2025 (Kementerian Pendidikan Dasar dan Menengah, 2025). Pembelajaran Mendalam bukan kurikulum baru, melainkan pendekatan yang memuliakan dengan menekankan penciptaan suasana belajar dan proses pembelajaran yang berkesadaran, bermakna, dan menggembirakan melalui olah pikir, olah hati, olah rasa, dan olah raga secara holistik dan terpadu (Suyanto et al., 2025). Ketiga prinsip tersebut, yaitu berkesadaran (*mindful*), bermakna (*meaningful*), dan menggembirakan (*joyful*), dioperasionalkan melalui tiga pengalaman belajar, yaitu memahami, mengaplikasi, dan merefleksi, serta empat komponen kerangka pembelajaran, yaitu praktik pedagogis, kemitraan pembelajaran, lingkungan pembelajaran, dan pemanfaatan digital (Anwar & Sodik, 2025; Suyanto et al., 2025). Sejumlah kajian konseptual menempatkan Pembelajaran Mendalam sebagai upaya transformasi menuju pembelajaran yang bermutu dan bermakna (Nurhakim et al., 2025; Rahmandani et al., 2025), sekaligus mengidentifikasi berbagai tantangan implementasinya di satuan pendidikan (Paramita et al., 2025).

Secara akademik, istilah pembelajaran mendalam berakar pada pembedaan antara pendekatan belajar mendalam (*deep approach*) dan dangkal (*surface approach*), yakni orientasi belajar untuk memahami makna, bukan sekadar mereproduksi informasi (Marton & Säljö, 1976). Gagasan serupa tampak dalam konsep *deeper learning*, yaitu proses yang memungkinkan individu mentransfer pengetahuan dan keterampilan yang dipelajari ke situasi baru (National Research Council, 2012), serta dalam kerangka *New Pedagogies for Deep Learning* yang merancang pembelajaran melalui empat elemen, yakni praktik pedagogis, kemitraan belajar, lingkungan belajar, dan pemanfaatan teknologi digital (Fullan et al., 2018). Keempat komponen kerangka pembelajaran dalam kebijakan Indonesia sejalan dengan keempat elemen tersebut. Makna pedagogis ini berbeda secara mendasar dari *deep learning* dalam kecerdasan artifisial, yang

merujuk pada metode pembelajaran mesin berbasis jaringan saraf tiruan berlapis banyak (LeCun et al., 2015). Untuk menghindari ambiguitas, artikel ini menggunakan istilah Pembelajaran Mendalam (*deep learning pedagogy*) secara konsisten untuk merujuk pada pendekatan pedagogis sebagaimana dirumuskan dalam kebijakan Kemendikdasmen.

Gagasan pembelajaran mendalam telah diadopsi dalam berbagai inisiatif internasional, tetapi bukti empiris mengenai dampaknya perlu dibaca secara hati-hati. Studi kuasi-eksperimental di Amerika Serikat menemukan bahwa peserta didik pada sekolah-sekolah jaringan *deeper learning* memperoleh skor yang lebih tinggi pada tes berbasis PISA yang mengukur penguasaan konten inti dan pemecahan masalah kompleks, serta memiliki tingkat kelulusan yang lebih tinggi dibandingkan dengan peserta didik berkarakteristik serupa di sekolah pembanding (Zeiser et al., 2014). Namun, kajian lapangan yang mendalam juga menunjukkan bahwa praktik pembelajaran mendalam jarang ditemukan secara konsisten, bahkan di sekolah yang secara eksplisit mengusungnya, karena praktik tersebut menuntut kapasitas guru dan dukungan organisasi yang tinggi (Mehta & Fine, 2019). Dengan demikian, rujukan terhadap praktik di berbagai negara dalam naskah akademik Pembelajaran Mendalam (Suyanto et al., 2025) lebih tepat dibaca sebagai landasan konseptual kebijakan, bukan sebagai bukti bahwa pendekatan tersebut pasti berhasil ketika diterapkan dalam konteks Indonesia. Keberhasilan implementasinya sangat bergantung pada kesiapan guru.

Dalam konteks ini, kompetensi guru menjadi penghubung antara kebijakan dan praktik. Undang-Undang Nomor 14 Tahun 2005 tentang Guru dan Dosen menetapkan empat kompetensi guru, yaitu kompetensi pedagogik, kepribadian, sosial, dan profesional (Republik Indonesia, 2005). Secara teoretis, kompetensi dapat dipandang sebagai kontinum yang bergerak dari disposisi (pengetahuan, keyakinan, dan motivasi), melalui keterampilan situasional (mempersepsi, menafsirkan, dan mengambil keputusan dalam situasi pembelajaran), hingga kinerja yang dapat diamati di kelas (Blömeke et al., 2015). Kerangka ini penting karena perubahan pada satu komponen, misalnya pengetahuan, tidak otomatis berarti perubahan kinerja di kelas. Literatur pengembangan profesional guru menunjukkan bahwa program yang efektif cenderung berfokus pada konten, melibatkan pembelajaran aktif, koheren dengan kebijakan dan kebutuhan sekolah, berlangsung berkelanjutan, dan mendorong partisipasi kolektif (Darling-Hammond et al., 2017; Desimone, 2009; Garet et al., 2001). Model interkoneksi pertumbuhan profesional guru memandang perubahan guru sebagai proses yang tidak linear, yang menghubungkan masukan eksternal, pengetahuan dan keyakinan pribadi, eksperimen dalam praktik, serta konsekuensi yang dirasakan melalui penerapan dan refleksi (Clarke & Hollingsworth, 2002). Guskey (2002) bahkan menekankan bahwa perubahan keyakinan guru sering kali baru terjadi setelah guru melihat dampak perubahan praktik terhadap hasil belajar peserta didik. Dengan demikian, dalam penelitian ini Pembelajaran Mendalam merupakan pendekatan yang hendak diterapkan dalam Kurikulum Merdeka, pelatihan merupakan bentuk pengembangan profesional guru, dan kompetensi guru merupakan hasil terdekat (*proximal outcome*) yang diharapkan berubah sebelum perubahan praktik kelas dan hasil belajar peserta didik dapat terjadi.

Untuk menyiapkan guru, Kemendikdasmen menyelenggarakan pelatihan Pembelajaran Mendalam secara nasional, antara lain melalui Balai Guru dan Tenaga Kependidikan (BGTK) di tingkat provinsi. Pelatihan dirancang dengan pola In-On-In, yaitu *in-service learning* 1 (IN-1) untuk memahami konsep, *on-the-job learning* (ON) untuk menerapkan rancangan pembelajaran di sekolah dengan pemantauan dan pendampingan, serta *in-service learning* 2 (IN-2) untuk berbagi hasil dan merefleksikan penerapan. Pelatihan berlangsung sekitar empat bulan dengan kombinasi kegiatan sinkron dan asinkron serta diskusi kelompok terfokus. Rancangan ini secara konseptual memuat sejumlah ciri pengembangan profesional yang efektif, tetapi perubahan kompetensi guru yang menyertainya dalam konteks daerah belum banyak dikaji.

Sejumlah studi terbaru melaporkan bahwa pelatihan atau pendampingan Pembelajaran Mendalam meningkatkan pemahaman dan kompetensi pedagogik guru (Atmojo et al., 2025; Muqtada et al., 2026; Rustyawati & Fauzi, 2025; Susanti et al., 2026; Susilana et al., 2025). Akan tetapi, bukti tersebut masih memiliki beberapa keterbatasan. Pertama, kelima studi tersebut merupakan laporan kegiatan pengabdian kepada masyarakat berskala kecil, dengan 20 sampai 75 peserta dari satu sekolah, dua sekolah, satu kecamatan, atau satu kota, dan salah satunya berupa intervensi dua hari (Susilana et al., 2025), sehingga belum menggambarkan program nasional berjangka panjang sebagaimana dilaksanakan oleh BGTK. Kedua,

hasil umumnya dilaporkan secara deskriptif, misalnya sebagai rata-rata skor *pretest* dan *posttest*, N-Gain, atau persentase peningkatan, tanpa interval kepercayaan ataupun ukuran efek terstandar. Ketiga, hasil cenderung dilaporkan pada tingkat kelompok sehingga variasi perubahan antarguru belum terungkap. Keempat, studi-studi tersebut dilaksanakan di Jawa dan Kepulauan Riau, sehingga pelaksanaan pelatihan Pembelajaran Mendalam bagi guru SD di Aceh, termasuk di Kota Lhokseumawe, belum tergambarkan dalam bukti tersebut.

Berdasarkan kesenjangan tersebut, penelitian ini mengkaji perubahan skor tes kompetensi guru SD yang mengikuti pelatihan Pembelajaran Mendalam berpola In-On-In selama empat bulan di Kota Lhokseumawe. Kontribusi penelitian ini ada tiga. Pertama, penelitian ini menyajikan data dari satu kelas pelaksanaan program nasional di tingkat daerah, bukan dari lokakarya singkat. Kedua, penelitian ini melaporkan besaran perubahan secara transparan melalui selisih rata-rata, interval kepercayaan 95%, ukuran efek, dan perubahan individual sehingga heterogenitas respons peserta dapat dinilai. Ketiga, penelitian ini menafsirkan temuan dalam kerangka teori pengembangan profesional guru dan kompetensi sebagai kontinum, sehingga batas inferensinya, baik akibat desain satu kelompok maupun akibat penggunaan skor tes untuk menggambarkan praktik kelas, dapat dinyatakan secara eksplisit. Pertanyaan penelitian ini adalah (1) apakah terdapat perbedaan yang signifikan antara skor tes kompetensi guru sebelum dan sesudah pelatihan; (2) seberapa besar perubahan tersebut dan bagaimana variasinya antarpeserta; dan (3) bagaimana indikator proses pelatihan, yaitu produk serta sikap dan keaktifan, menggambarkan keterlibatan peserta selama pelatihan. Temuan penelitian diharapkan bermanfaat bagi BGTK, dinas pendidikan, dan sekolah dalam merancang tindak lanjut pengembangan profesional guru yang lebih terdiferensiasi.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain praeksperimen *one-group pretest–posttest* (O_1 X O_2), yaitu pengukuran awal (O_1), pemberian perlakuan berupa pelatihan (X), dan pengukuran akhir (O_2) pada kelompok yang sama (Campbell & Stanley, 1963). Data yang dianalisis merupakan data penilaian rutin pelatihan. Desain ini digunakan karena pelatihan merupakan program nasional yang diikuti oleh seluruh peserta dalam kelas pelatihan, sehingga peneliti tidak dapat membentuk kelompok kontrol yang tidak menerima pelatihan. Konsekuensinya, desain ini hanya memungkinkan estimasi perubahan skor dalam kelompok, bukan estimasi pengaruh kausal pelatihan. Tanpa kelompok pembanding, perubahan skor dapat pula dipengaruhi oleh efek pengujian (*testing*), yaitu pengaruh pengalaman mengerjakan *pretest* terhadap skor *posttest*; maturasi, yaitu perkembangan alami guru selama empat bulan; sejarah (*history*), yaitu paparan terhadap informasi Pembelajaran Mendalam dari sumber lain pada periode yang sama; instrumentasi, karena *pretest* dan *posttest* menggunakan butir yang berbeda sehingga perbedaan tingkat kesukaran kedua bentuk tes dapat ikut memengaruhi selisih skor; atrisi, karena hanya peserta yang menyelesaikan seluruh tahap dan memiliki skor lengkap yang dianalisis; regresi ke arah rerata, yang terutama memengaruhi pola perubahan individual; serta faktor eksternal lainnya (Knapp, 2016; Marsden & Torgerson, 2012; Shadish et al., 2002). Oleh karena itu, artikel ini menggunakan istilah perubahan atau peningkatan skor, sedangkan istilah efektivitas dan kontribusi pelatihan hanya digunakan secara terbatas dan dengan kualifikasi.

Penelitian dilaksanakan pada kelas pelatihan Pembelajaran Mendalam jenjang SD kelas Rendah dan Tinggi yang diselenggarakan oleh BGTK Provinsi Aceh di Kota Lhokseumawe pada September–Desember 2025. Penulis pertama merupakan salah satu fasilitator yang ditugaskan pada pelatihan tersebut. Peserta pelatihan adalah guru perwakilan sekolah dasar di lingkup Kota Lhokseumawe yang dipilih dengan kategori penerima BOS kinerja. Partisipan penelitian adalah seluruh guru dalam kelas tersebut yang memenuhi kriteria inklusi (sampel jenuh). Dengan demikian, sampel bersifat nonprobabilitas dan tidak dimaksudkan untuk mewakili seluruh guru SD di Kota Lhokseumawe. Kriteria inklusi analisis adalah (a) terdaftar sebagai peserta kelas; (b) mengikuti tahap IN-1, ON, dan IN-2; dan (c) memiliki skor *pretest* dan *posttest* yang lengkap. Dari 26 guru dari 13 sekolah memenuhi kriteria dan dianalisis.

Pelatihan mengikuti struktur nasional berpola In-On-In dengan durasi sekitar empat bulan 212 JP (Gambar 1; Tabel 2). Tahap IN-1 dilaksanakan pada 19 September 2025. Tahap ini berisi materi umum (kebijakan dan orientasi kegiatan), materi inti tentang Pembelajaran Mendalam, serta materi penunjang

berupa penyusunan rencana tindak lanjut (RTL) dan evaluasi. Tahap ON berlangsung selama 1 Bulan di sekolah masing-masing. Pada tahap ini peserta melaksanakan RTL melalui empat kegiatan, yaitu merancang pembelajaran, mengimplementasikan dan melakukan asesmen pembelajaran, merefleksikan pembelajaran, serta melakukan inkuiri kolaboratif di kelompok kerja satuan pendidikan dan bersama sesama peserta, disertai pemantauan dan pendampingan selama proses tersebut. Tahap IN-2 dilaksanakan pada 6–9 Desember 2025 di Kota Lhokseumawe dan berisi kegiatan berbagi hasil implementasi, refleksi, serta penyusunan RTL untuk implementasi berkelanjutan, termasuk rencana berbagi di Kelompok Kerja Guru (KKG). Secara konseptual, ketiga tahap tersebut mencerminkan tiga pengalaman belajar dalam Pembelajaran Mendalam, yaitu memahami (IN-1), mengaplikasi (ON), dan merefleksi (IN-2).

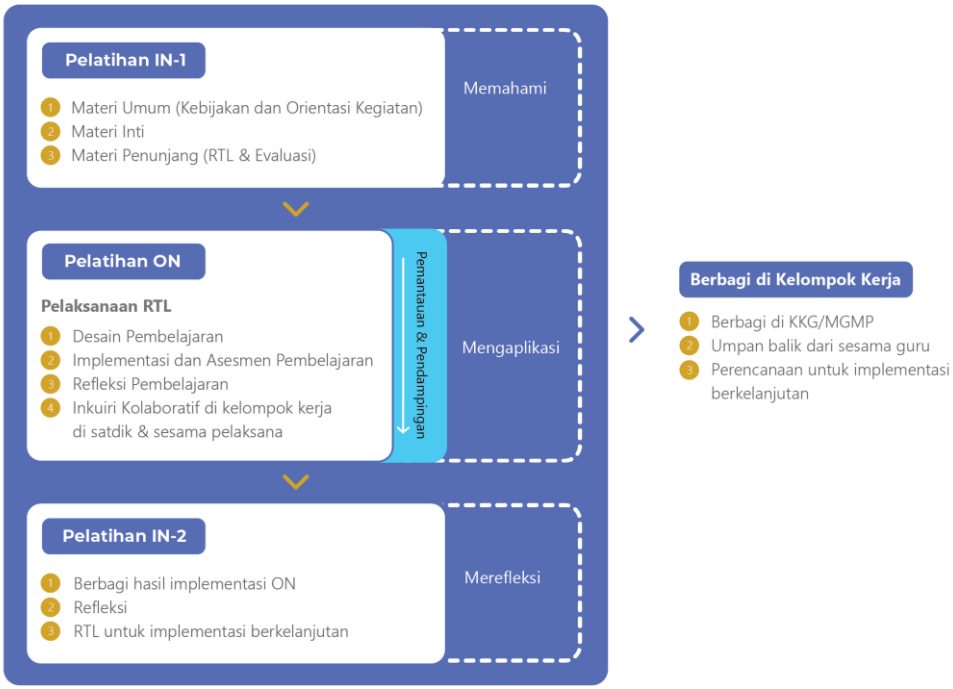
Fasilitator berperan menyampaikan materi dan memfasilitasi diskusi kelompok terfokus pada tahap IN-1 dan IN-2 serta memantau dan mendampingi pelaksanaan RTL pada tahap ON. Keterlaksanaan program (*fidelity of implementation*) diupayakan melalui penggunaan modul, jadwal, dan instrumen penilaian standar dari penyelenggara serta pemantauan penyelesaian tugas pada tahap ON. Namun, keterlaksanaan tidak diukur dengan instrumen *fidelity* terstruktur, misalnya daftar periksa observasi sesi, sehingga variasi kualitas pelaksanaan tidak dapat dianalisis (O'Donnell, 2008).

Tabel 2. Struktur dan Aktivitas Program Pelatihan Pembelajaran Mendalam

Tahap	Waktu dan moda	Kegiatan utama	Output
IN-1 (*in-service learning* 1): memahami	19 September 2025; Daring dan Luring 60 JP	Materi umum (kebijakan dan orientasi kegiatan); materi inti; materi penunjang (RTL dan evaluasi)	*Pretest*; RTL
ON (*on-the-job learning*): mengaplikasi	19 Oktober – 19 November 2025; Luring	Pelaksanaan RTL: desain pembelajaran; implementasi dan asesmen pembelajaran; refleksi pembelajaran; inkuiri kolaboratif di kelompok kerja satuan pendidikan dan bersama sesama peserta; pemantauan dan pendampingan	Produk
IN-2 (*in-service learning* 2): merefleksi	26 November – 4 Desember 2025; Luring	Berbagi hasil implementasi ON; refleksi; RTL untuk implementasi berkelanjutan	*Posttest* (tes akhir)

Instrumen utama penelitian ini adalah tes kompetensi Pembelajaran Mendalam yang diberikan sebelum dan sesudah pelatihan. Dalam penelitian ini, kompetensi guru didefinisikan secara operasional sebagai pengetahuan konseptual dan pedagogis guru tentang Pembelajaran Mendalam sebagaimana diukur oleh tes tersebut. Dalam kerangka kompetensi sebagai kontinum (Blömeke et al., 2015), tes ini terutama mengukur komponen disposisi kognitif, bukan kinerja mengajar yang teramati di kelas. Tes terdiri atas 30 butir Pilihan Ganda dan skornya dinyatakan pada skala 0–100. Tes disusun oleh kemendikdasmen pusat sebagai instrumen standar pelatihan dan dikerjakan melalui lms guru kemdendikdasmen pada awal pelatihan (*pretest*) dan pada akhir pelatihan (*posttest*), dengan waktu pengerjaan 45 menit. *Pretest* dan *posttest* menggunakan butir soal yang sama dengan urutan yang diacak. Bukti validitas isi bertumpu pada kesesuaian butir dengan materi pelatihan yang ditetapkan penyusun instrumen. Penelitian ini tidak melakukan validasi konstruk, analisis butir, ataupun penyetaraan (*equating*) kedua bentuk tes dikarenakan sudah disusun oleh lembaga pusat kemendikdasment.

Selain tes kompetensi, penelitian ini menggunakan catatan penilaian pelatihan yang terdiri atas tiga komponen nilai akhir sesuai ketentuan penyelenggara, yaitu produk (bobot 60%), tes akhir (20%), serta sikap dan keaktifan (20%). Komponen tes akhir merupakan skor *posttest* yang sama dengan yang digunakan dalam analisis utama. Produk mencakup keluaran yang dihasilkan peserta selama pelatihan, antara lain rancangan dan media pembelajaran, sedangkan sikap dan keaktifan menggambarkan partisipasi peserta selama kegiatan. Kedua komponen tersebut dinilai oleh penyelenggara melalui sistem penilaian pelatihan dengan instrumen standar, bukan oleh peneliti. Instrumen penilaian tersebut dikembangkan oleh penyelenggara, dan penelitian ini tidak mengestimasi reliabilitas antarpenilai. Karena merupakan bagian dari sistem penilaian pelatihan yang sama dan tidak divalidasi secara terpisah sebagai ukuran kompetensi, data ini diperlakukan sebagai indikator proses dan keterlibatan, bukan sebagai bukti independen atas perubahan kompetensi.



Gambar 1. Struktur Model Pelatihan Pembelajaran Mendalam Berpola In-On-In

Sumber: bahan pelatihan dari penyelenggara.

Data penelitian ini dianalisis menggunakan IBM SPSS Statistics dengan taraf signifikansi 0,05 (dua sisi). Statistik deskriptif, yaitu rata-rata, simpangan baku, nilai minimum, dan nilai maksimum, digunakan untuk meringkas skor *pretest*, *posttest*, dan indikator proses pelatihan. Karena uji-*t* sampel berpasangan mengasumsikan normalitas distribusi selisih skor, normalitas skor selisih (*posttest* − *pretest*) diuji dengan uji Shapiro–Wilk, yang direkomendasikan untuk sampel kurang dari 50 (Ghasemi & Zahediasl, 2012). Perbedaan skor sebelum dan sesudah pelatihan diuji dengan uji-*t* sampel berpasangan. Selain nilai *p*, penelitian ini melaporkan selisih rata-rata dengan interval kepercayaan 95% serta ukuran efek Cohen's d_z, yaitu selisih rata-rata dibagi simpangan baku skor selisih. Interval kepercayaan 95% untuk d_z dihitung dengan distribusi *t* nonsentral, dan koreksi bias sampel kecil dilaporkan sebagai Hedges' g_z (Lakens, 2013). Ambang konvensional 0,2, 0,5, dan 0,8 untuk efek kecil, sedang, dan besar (Cohen, 1988) hanya digunakan sebagai patokan kasar karena ambang tersebut dikembangkan untuk perbandingan antarkelompok. Selain itu, ukuran efek dari desain satu kelompok dengan tes yang selaras dengan materi pelatihan cenderung lebih besar daripada ukuran efek dari studi terkontrol (Kraft, 2020; Lakens, 2013). Perubahan skor setiap peserta dilaporkan dan disajikan secara grafis untuk menunjukkan heterogenitas respons. Sebagai analisis sensitivitas, perbedaan skor juga diuji dengan uji peringkat bertanda Wilcoxon dan dengan analisis ulang tanpa peserta berskor *pretest* ekstrem.

Besarnya peningkatan relatif terhadap ruang peningkatan yang tersedia dianalisis dengan *normalized gain* (N-Gain) sesuai ketentuan Hake (1998) yang menyebutkan N-Gain dengan simbol g. Nilai *g* dikategorikan tinggi ($g \geq 0{,}70$), sedang ($0{,}30 \leq g < 0{,}70$), dan rendah ($g < 0{,}30$) (Hake, 1998). N-Gain digunakan karena memperhitungkan perbedaan ruang peningkatan antarpeserta: kenaikan 10 poin memiliki makna yang berbeda bagi peserta dengan skor awal 40 dan 85. Namun, N-Gain memiliki sejumlah karakteristik matematis yang membatasi penafsirannya. Pertama, nilainya tidak terdefinisi apabila skor *pretest* sama dengan skor maksimum dan menjadi tidak stabil ketika skor *pretest* mendekati maksimum karena penyebutnya sangat kecil. Kedua, nilainya dapat negatif dan tidak simetris: batas atasnya 1, sedangkan penurunan skor yang kecil pada peserta berskor awal tinggi dapat menghasilkan nilai negatif yang besar (Marx & Cummings, 2007). Ketiga, N-Gain berkorelasi dengan skor awal dan dapat bias terhadap kelompok dengan skor awal tertentu, sehingga tidak dapat menggantikan ukuran efek (Coletta & Phillips, 2005; Nissen et al., 2018). Keempat, kategori Hake dikembangkan untuk rata-rata kelas pada tes konseptual fisika, bukan untuk mengklasifikasikan individu. Oleh karena itu, N-Gain dalam penelitian ini dilaporkan

dalam dua bentuk, yaitu *gain* rata-rata kelas ⟨*g*⟩ yang dihitung dari rata-rata skor kelompok (Hake, 1998) dan distribusi N-Gain individual, dan hanya digunakan secara deskriptif untuk menggambarkan heterogenitas peningkatan, bukan sebagai uji efektivitas. Rata-rata skor *pretest* juga dibandingkan antar kategori N-Gain, dan hubungan antara skor *pretest* dan N-Gain dihitung dengan korelasi Spearman. Karena skor *pretest* terdapat dalam rumus N-Gain, korelasi negatif keduanya dapat muncul akibat keterkaitan matematis dan regresi ke arah rerata, sehingga hanya ditafsirkan secara deskriptif.

Data penelitian berasal dari catatan penilaian rutin pelatihan melalui LSM yang diisi oleh fasilitator. Data skor dianalisis dan dilaporkan tanpa mencantumkan identitas peserta. Karena penulis pertama juga berperan sebagai fasilitator, potensi bias dikurangi dengan menggunakan skor yang dihasilkan oleh sistem penilaian penyelenggara dan dengan melakukan analisis data setelah pelatihan dan penilaian selesai.

HASIL

Bagian ini menyajikan hasil analisis skor tes kompetensi 26 guru sebelum dan sesudah pelatihan, diikuti distribusi N-Gain individual dan indikator proses pelatihan. Indikator proses berfungsi menggambarkan keterlibatan peserta, bukan sebagai ukuran independen kompetensi.

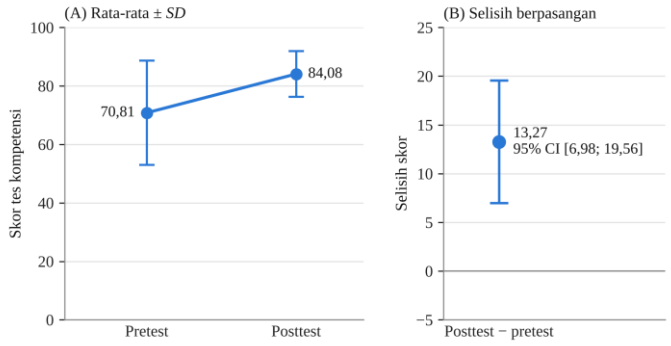
Tabel 3. Statistik Deskriptif Skor Tes Kompetensi *Pretest* dan *Posttest* (*N* = 26)

Pengukuran	*N*	Minimum	Maksimum	*M*	*SE*	*SD*
Pretest	26	26	93	70,81	3,50	17,84
Posttest	26	67	97	84,08	1,52	7,78

Sumber: Hasil pengolahan data SPSS (2026).

Tabel 3 menunjukkan bahwa rata-rata skor tes kompetensi meningkat dari 70,81 (*SD* = 17,84) pada *pretest* menjadi 84,08 (*SD* = 7,78) pada *posttest*, dengan selisih rata-rata 13,27 poin. Rentang skor menyempit dari 26–93 menjadi 67–97. Penyempitan sebaran ini perlu ditafsirkan secara hati-hati. Penurunan simpangan baku tidak serta-merta berarti bahwa kompetensi peserta menjadi lebih merata, karena pola serupa dapat dihasilkan oleh karakteristik pengukuran, yaitu (a) efek plafon, karena skor maksimum tes adalah 100 sehingga skor peserta berkemampuan tinggi terkumpul di dekat batas atas; (b) perbedaan galat pengukuran antara kedua waktu, misalnya bila *pretest* dikerjakan dengan motivasi yang lebih rendah atau dalam format yang belum familier sehingga lebih banyak jawaban bersifat tebakan; dan (c) kemungkinan perbedaan tingkat kesukaran antara bentuk tes *pretest* dan *posttest*. Korelasi skor *pretest* dan *posttest* yang moderat (*r* = 0,49) menunjukkan bahwa posisi relatif peserta hanya sebagian stabil. Regresi ke arah rerata tidak memperkecil sebaran skor kelompok bila galat pengukuran pada kedua waktu setara, tetapi memengaruhi pola perubahan individual: peserta dengan skor *pretest* ekstrem cenderung bergerak ke arah rerata pada pengukuran berikutnya (Barnett et al., 2005). Penyempitan sebaran lebih tepat dibaca sebagai berkurangnya perbedaan skor tes antarpeserta, bukan sebagai bukti langsung pemerataan kompetensi.

Pada tingkat individual, sebanyak 20 dari 26 peserta (76,9%) memperoleh skor posttest yang lebih tinggi daripada skor pretest, 1 peserta memperoleh skor yang sama, dan 5 peserta mengalami penurunan skor (Gambar 2). Kenaikan terbesar terjadi pada peserta dengan skor pretest rendah: delapan peserta dengan skor pretest di bawah 65 seluruhnya meningkat, dengan kenaikan 13,1 hingga 53,8 poin (tertinggi A1, dari 26,4 menjadi 80,2). Sebaliknya, kelima peserta yang skornya turun memiliki skor pretest yang sudah tinggi (76,9-93,4), dengan penurunan yang relatif kecil, yaitu 3,3 hingga 9,9 poin.



Gambar 2. Rerat skor *Pretest* dan *Posttest* (± *SD*) dan Rerata selisih serta Interval kepercayaan 95% (*N* = 26)

Selanjutnya dilakukan Uji Normalitas Skor Selisih yang berfungsi untuk menentukan apakah distribusi perbedaan skor sebelum dan sesudah perlakuan mengikuti sebaran normal sebelum dilakukan uji statistik lanjutan

Tabel 4. Hasil Uji Normalitas Skor Selisih (*Posttest − Pretest*)

Variabel	Kolmogorov–Smirnov[a]	*df*	*P*	Shapiro–Wilk (*W*)	*df*	*p*	Keputusan
Skor selisih	0,16	26	0,082	0,95	26	0,221	Tidak menyimpang dari normal

Catatan. [a]Dengan koreksi signifikansi Lilliefors. Sumber: Hasil pengolahan data SPSS (2026).

Uji normalitas dilakukan terhadap skor selisih (*posttest − pretest*) karena asumsi uji-*t* sampel berpasangan berlaku pada distribusi selisih, bukan pada masing-masing skor. Hasil uji Shapiro–Wilk menunjukkan $W = 0{,}95$ dan $p = 0{,}221$ (Tabel 4). Karena $p > 0{,}05$, distribusi skor selisih tidak menyimpang secara signifikan dari distribusi normal sehingga uji-*t* sampel berpasangan dapat digunakan. Hasil uji Kolmogorov–Smirnov ($p = 0{,}082$) mengarah pada kesimpulan yang sama, tetapi interpretasi utama mengacu pada uji Shapiro–Wilk karena ukuran sampel kurang dari 50 (Ghasemi & Zahediasl, 2012).

Selain itu, Hasil uji-*t* sampel berpasangan menunjukkan bahwa rata-rata skor *posttest* lebih tinggi secara signifikan daripada rata-rata skor *pretest.* Berikut Hasil Uji-t Sampel Berpasangan dan Ukuran Efek

Tabel 5. Hasil Uji-*t* Sampel Berpasangan dan Ukuran Efek ($N = 26$)

Perbandingan	M_{diff}	SD_{diff}	95% CI M_{diff}	*t*	*df*	*p*	d_z [95% CI]	g_z
Posttest − pretest	13,27	15,58	[6,98; 19,56]	4,34	25	< 0,001	0,85 [0,39; 1,30]	0,83

Catatan. M_{diff} = rata-rata selisih; SD_{diff} = simpangan baku skor selisih; CI = *confidence interval* (interval kepercayaan); $d_z = M_{diff}/SD_{diff}$ (Lakens, 2013), dengan interval kepercayaan berbasis distribusi *t* nonsentral; $g_z = d_z$ setelah koreksi Hedges untuk bias sampel kecil. Nilai *t* dilaporkan dalam nilai absolut; pada output SPSS, $t = -4{,}342$ karena selisih dihitung sebagai *pretest − posttest*. Sumber: Hasil pengolahan data SPSS (2026).

Hasil uji-*t* sampel berpasangan menunjukkan bahwa rata-rata skor *posttest* lebih tinggi secara signifikan daripada rata-rata skor *pretest*, $t(25) = 4{,}34$, $p < 0{,}001$ (Tabel 5). Rata-rata selisih sebesar 13,27 poin memiliki interval kepercayaan 95% [6,98; 19,56]; artinya, data selaras dengan rata-rata peningkatan pada populasi antara sekitar 7 dan 20 poin. Ukuran efek Cohen's d_z sebesar 0,85 (95% CI [0,39; 1,30]) dan Hedges' g_z sebesar 0,83 melampaui ambang 0,8 yang lazim disebut besar (Cohen, 1988), meskipun ambang tersebut dikembangkan untuk perbandingan antarkelompok (Lakens, 2013). Lebarnya interval kepercayaan menunjukkan bahwa estimasi besaran perubahan belum presisi, yaitu berkisar dari efek kecil–sedang hingga sangat besar. Selain itu, d_z distandardisasi dengan simpangan baku skor selisih sehingga tidak dapat dibandingkan secara langsung dengan ukuran efek antarkelompok dari studi terkontrol (Kraft, 2020). Analisis sensitivitas dengan uji peringkat bertanda Wilcoxon ($z = -3{,}64$; $p < 0{,}001$) dan analisis ulang tanpa dua peserta berskor *pretest* 26 ($n = 24$; $M_{diff} = 10{,}42$; 95% CI [5,25; 15,59]; $t(23) = 4{,}17$; $p < 0{,}001$; $d_z = 0{,}85$) menghasilkan kesimpulan yang sama. Hasil ini menunjukkan adanya peningkatan skor tes kompetensi yang signifikan setelah guru mengikuti pelatihan. Namun, karena tidak ada kelompok kontrol, peningkatan tersebut tidak dapat ditafsirkan sebagai besarnya pengaruh kausal pelatihan.

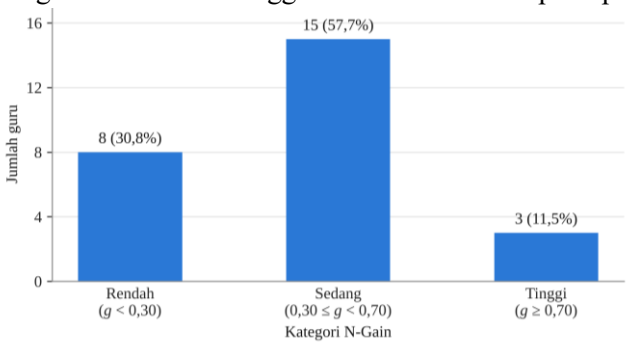
Analisis lanjutan terhadap nilai N-Gain dilakukan dengan mengelompokkan capaian peserta didik ke dalam kategori rendah, sedang, dan tinggi guna menelaah variasi peningkatan belajar serta karakteristik skor *pretest* dan *posttest* di tiap kelompok. Pendekatan ini membantu mengidentifikasi pola sebaran hasil belajar—termasuk kelompok dengan peningkatan optimal maupun kelompok yang memerlukan perhatian lebih lanjut—sebelum diuraikan lebih mendalam pada bagian pembahasan. Rekapitulasi distribusi kategori N-Gain dan rata-rata skor per kategori berdasarkan hasil pengolahan data SPSS ditunjukkan pada Tabel berikut

Tabel 6. Distribusi Kategori N-Gain dan Rata-Rata Skor per Kategori

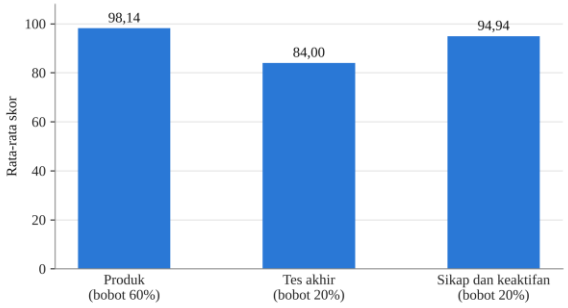
Kategori	Kriteria	*f*	%	*M pretest*	*M posttest*
Rendah	$g < 0{,}30$	8	30,8	85,13	82,25
Sedang	$0{,}30 \le g < 0{,}70$	15	57,7	66,67	83,60
Tinggi	$g \ge 0{,}70$	3	11,5	53,33	91,33
Total		26	100,0	70,81	84,08

Sumber: Hasil pengolahan data SPSS (2026).

Gain rata-rata kelas yang dihitung dari rata-rata skor kelompok adalah $\langle g \rangle = (84{,}08 - 70{,}81)/(100 - 70{,}81) = 0{,}45$, yang termasuk kategori sedang (Hake, 1998). Rata-rata N-Gain individual adalah 0,28 (*SD* = 0,53; rentang −1,29 sampai 0,93; median 0,46). Rata-rata ini lebih rendah daripada $\langle g \rangle$ karena beberapa peserta berskor *pretest* tinggi memperoleh N-Gain negatif yang besar akibat penyebut yang kecil, misalnya −1,29 untuk penurunan 9 poin dari skor *pretest* 93, sedangkan median N-Gain individual lebih mendekati $\langle g \rangle$. Distribusi kategori pada Tabel 6 dan Gambar 3 menunjukkan bahwa 15 peserta (57,7%) berada pada kategori sedang, 8 peserta (30,8%) pada kategori rendah, dan 3 peserta (11,5%) pada kategori tinggi. Pola ini menunjukkan bahwa peningkatan tidak terjadi secara seragam. Untuk menelaah apakah kategori rendah mencerminkan kecilnya peningkatan atau terbatasnya ruang peningkatan, rata-rata skor *pretest* dibandingkan antarkategori. Peserta pada kategori rendah memiliki rata-rata skor *pretest* 85,13 (rentang 77–93), sedangkan peserta pada kategori sedang dan tinggi masing-masing 66,67 dan 53,33. Sebaliknya, rata-rata skor *posttest* kategori rendah (82,25) hampir sama dengan kategori sedang (83,60), sedangkan kategori tinggi mencapai 91,33 (Tabel 6). Korelasi Spearman antara skor *pretest* dan N-Gain adalah $\rho = -0{,}67$ ($p < 0{,}001$). Karena kategori Hake dikembangkan untuk rata-rata kelas dan N-Gain individual peka terhadap skor awal, distribusi ini digunakan untuk menggambarkan variasi respons peserta, bukan untuk menilai efektivitas pelatihan.



Gambar 3. Distribusi kategori N-gain peserta pelatihan



Gambar 4. Rerata skor komponen penilaian pelatihan

Penilaian keberhasilan pelatihan ditinjau melalui tiga komponen yang memiliki bobot berbeda terhadap nilai akhir, yakni produk sebagai komponen dominan (60%), tes akhir (20%), dan sikap serta keaktifan (20%). Analisis deskriptif melalui rentang skor, rata-rata, dan standar deviasi digunakan untuk melihat kekuatan capaian sekaligus sebaran variasi skor pada masing-masing aspek penilaian. Rincian statistik deskriptif tiap komponen penilaian pelatihan berdasarkan hasil pengolahan data SPSS ditunjukkan pada Tabel berikut.

Tabel 9. Statistik Deskriptif Komponen Penilaian Pelatihan (*N* = 26)

Komponen (bobot)	*N*	Minimum	Maksimum	*M*	*SD*	Kategori
Produk (60%)	26	93,00	100,00	98,14	2,59	Sangat baik
Tes akhir (20%)[a]	26	67,00	96,70	84,00	7,79	Baik
Sikap dan keaktifan (20%)	26	89,94	99,59	94,94	3,32	Sangat baik

Sumber: Hasil pengolahan data SPSS (2026).

Tabel 9 dan Gambar 4 menyajikan tiga komponen nilai akhir pelatihan. Data ini dilaporkan sebagai indikator proses dan keterlibatan, bukan sebagai bukti independen atas perubahan kompetensi, karena ketiga komponen merupakan bagian dari sistem penilaian pelatihan yang sama dan tidak divalidasi secara terpisah sebagai ukuran kompetensi guru. Komponen tes akhir (*M* = 84,00) merupakan skor *posttest* yang telah dianalisis pada Tabel 3 sehingga tidak memberikan bukti tambahan di luar analisis utama. Nilai produk (*M* = 98,14; rentang 93,00–100,00) berada sangat dekat dengan skor maksimum. Pola ini menunjukkan efek plafon sehingga penilaian produk kurang mampu membedakan kualitas produk antarpeserta. Nilai sikap dan keaktifan (*M* = 94,94; *SD* = 3,32) menunjukkan tingkat partisipasi yang tinggi. Secara teoretis, produk pelatihan seperti rancangan pembelajaran merupakan artefak yang paling dekat dengan keterampilan perencanaan dalam konstruk kompetensi guru, sedangkan sikap dan keaktifan lebih mencerminkan keterlibatan dalam proses pelatihan (Blömeke et al., 2015; Guskey, 2000). Karena dinilai dalam sistem penilaian pelatihan tanpa estimasi reliabilitas antarpenilai, kedua komponen ini hanya dapat digunakan untuk

menggambarkan keterlaksanaan dan keterlibatan, bukan untuk menggantikan pengukuran kompetensi yang tervalidasi.

Penelitian ini menemukan bahwa skor tes kompetensi guru SD tentang Pembelajaran Mendalam meningkat secara signifikan setelah guru mengikuti pelatihan berpola In-On-In selama empat bulan, dengan besaran perubahan yang melampaui ambang konvensional efek besar tetapi diestimasi kurang presisi. Temuan ini perlu ditempatkan secara proporsional. Ukuran efek dari desain satu kelompok yang menggunakan tes selaras dengan materi pelatihan umumnya lebih besar daripada ukuran efek dari studi terkontrol yang menggunakan instrumen terstandar (Kraft, 2020). Sebagai gambaran, meta-analisis studi kausal tentang *coaching* guru melaporkan efek rata-rata sekitar 0,49 simpangan baku terhadap praktik pembelajaran dan 0,18 simpangan baku terhadap capaian belajar peserta didik (Kraft et al., 2018). Karena berbeda dalam desain, hasil yang diukur, dan cara standardisasi, angka-angka tersebut tidak dapat dibandingkan secara langsung dengan ukuran efek dalam penelitian ini. Temuan ini lebih tepat dibaca sebagai indikasi bahwa pengetahuan guru tentang Pembelajaran Mendalam berubah selama periode pelatihan, bukan sebagai estimasi besarnya dampak pelatihan.

Beberapa mekanisme dapat menjelaskan peningkatan skor tersebut. Pertama, struktur In-On-In memuat siklus perubahan guru sebagaimana dijelaskan dalam model interkoneksi pertumbuhan profesional guru (Clarke & Hollingsworth, 2002). Masukan konseptual pada tahap IN-1 (domain eksternal) memperbarui pengetahuan guru (domain personal), penerapan rancangan pembelajaran pada tahap ON memberi kesempatan bereksperimen dalam praktik (domain praktik), dan refleksi pada tahap IN-2 menghubungkan pengalaman tersebut dengan konsekuensinya bagi peserta didik (domain konsekuensi). Kedua, rancangan pelatihan memuat beberapa ciri pengembangan profesional yang efektif, yaitu fokus pada konten yang spesifik, pembelajaran aktif melalui penyusunan produk dan diskusi, koherensi dengan kebijakan kurikulum, durasi berkelanjutan, serta partisipasi kolektif melalui inkuiri kolaboratif di kelompok kerja (Darling-Hammond et al., 2017; Desimone, 2009). Meta-analisis terbaru juga menunjukkan bahwa program yang memadukan mekanisme membangun pengetahuan, memotivasi guru, mengembangkan teknik mengajar, dan menanamkan praktik cenderung lebih efektif (Sims et al., 2025). Ketiga, pelatihan ini bersifat kongruen secara pedagogis: guru mengalami sendiri urutan memahami, mengaplikasi, dan merefleksi yang kelak diharapkan mereka terapkan bagi peserta didik. Meskipun demikian, mekanisme ini merupakan penjelasan teoretis yang masuk akal, bukan hasil pengujian dalam penelitian ini, karena penelitian tidak mengukur proses-proses tersebut secara langsung.

Penjelasan alternatif juga perlu dipertimbangkan secara serius. Karena *pretest* dan *posttest* menggunakan butir yang berbeda, peningkatan skor kecil kemungkinannya berasal dari hafalan butir. Namun, pengalaman mengerjakan *pretest* tetap dapat membuat peserta lebih akrab dengan format dan cakupan tes, dan perbedaan tingkat kesukaran kedua bentuk tes tidak dapat dikesampingkan karena penyetaraan keduanya tidak diuji dalam penelitian ini. Keselarasan tes dengan materi pelatihan juga dapat menghasilkan peningkatan yang lebih mencerminkan penguasaan materi pelatihan daripada kompetensi yang dapat ditransfer ke situasi lain. Selama periode September–Desember 2025, guru juga dapat memperoleh informasi tentang Pembelajaran Mendalam dari sumber lain, seperti sosialisasi dinas, KKG, atau platform daring, sehingga faktor sejarah tidak dapat dikesampingkan. Regresi ke arah rerata juga perlu diperhitungkan (Barnett et al., 2005; Marsden & Torgerson, 2012). Karena peserta tidak dipilih berdasarkan skor *pretest*, regresi ke arah rerata lebih berpengaruh pada pola perubahan individual daripada pada rata-rata kelompok: peserta berskor *pretest* sangat rendah cenderung memperoleh skor lebih tinggi, sedangkan peserta berskor sangat tinggi cenderung memperoleh skor lebih rendah pada pengukuran berikutnya. Karena desain penelitian tidak dapat memisahkan pengaruh pelatihan dari faktor-faktor tersebut, peningkatan skor dalam penelitian ini konsisten dengan pengaruh pelatihan, tetapi tidak membuktikannya.

Temuan yang perlu mendapat perhatian adalah bahwa hampir sepertiga peserta berada pada kategori N-Gain rendah. Dua kelompok penjelasan perlu dibedakan. Penjelasan pertama bersifat pengukuran: bagi peserta dengan skor *pretest* tinggi, penyebut N-Gain sangat kecil sehingga fluktuasi beberapa poin, termasuk akibat galat pengukuran dan regresi ke arah rerata, dapat menghasilkan N-Gain rendah atau bahkan negatif (Barnett et al., 2005; Marx & Cummings, 2007). Data penelitian ini sejalan dengan penjelasan tersebut:

kedelapan peserta pada kategori rendah memiliki rata-rata skor *pretest* tertinggi (85,13; rentang 77–93), lima di antaranya mengalami penurunan skor dan satu memperoleh skor yang sama, tetapi rata-rata skor *posttest* mereka (82,25) hampir sama dengan rata-rata kategori sedang (83,60). Korelasi negatif antara skor *pretest* dan N-Gain ($\rho = -0{,}67$) juga mengarah pada pola yang sama, meskipun sebagian korelasi tersebut dapat muncul akibat keterkaitan matematis kedua variabel. Penjelasan kedua bersifat substantif: perbedaan pengetahuan awal, pengalaman menerapkan Kurikulum Merdeka, masa kerja, kesiapan digital untuk mengikuti kegiatan asinkron, beban kerja selama tahap ON, dan dukungan sekolah dapat memengaruhi besarnya perubahan skor guru (Desimone, 2009; Kennedy, 2016). Karena jumlah peserta kecil dan faktor-faktor tersebut tidak diukur, penelitian ini tidak dapat menyimpulkan faktor penyebabnya. Namun, heterogenitas ini menunjukkan bahwa pelatihan dengan format yang seragam kemungkinan tidak diikuti perubahan yang sama pada semua guru.

Dibandingkan dengan studi terdahulu, temuan ini searah dengan laporan bahwa pelatihan dan pendampingan Pembelajaran Mendalam meningkatkan pemahaman dan kompetensi pedagogik guru (Atmojo et al., 2025; Muqtada et al., 2026; Rustyawati & Fauzi, 2025; Susanti et al., 2026; Susilana et al., 2025). Atmojo et al. (2025) melaporkan peningkatan kemampuan pedagogik sebesar 45% pada 20 guru SD di Surakarta, Muqtada et al. (2026) melaporkan kenaikan rata-rata skor pengetahuan dari 42,5 menjadi 81,5 pada 24 guru dari dua SD di Magelang, dan Rustyawati dan Fauzi (2025) melaporkan perbaikan pemahaman konsep serta kemampuan menyusun rencana pembelajaran mendalam dan modul ajar pada 75 peserta di satu madrasah. Susilana et al. (2025) melaporkan N-Gain kategori sedang (0,43–0,56) pada enam aspek Pembelajaran Mendalam setelah pelatihan dua hari bagi 50 guru SD; *gain* rata-rata kelas dalam penelitian ini ($\langle g \rangle = 0{,}45$) berada pada rentang yang serupa, meskipun instrumen dan durasi pelatihannya berbeda. Penelitian ini memperluas bukti tersebut dalam tiga hal. Pertama, program yang dikaji adalah program nasional berdurasi empat bulan dengan pendampingan di sekolah, bukan lokakarya singkat atau kegiatan di satu sekolah. Kedua, besaran perubahan dilaporkan dengan interval kepercayaan dan ukuran efek sehingga ketidakpastian estimasi terlihat. Ketiga, analisis tingkat individual menunjukkan bahwa perubahan skor tidak seragam dan bahwa hampir sepertiga peserta berada pada kategori N-Gain rendah, suatu pola yang tidak tampak dalam laporan yang hanya menyajikan rata-rata kelompok atau persentase ketuntasan. Di sisi lain, penelitian ini memiliki kelemahan yang sama dengan studi-studi tersebut, yaitu tidak memiliki kelompok pembanding dan tidak mengamati praktik kelas. Literatur internasional juga mengingatkan bahwa efek program pengembangan profesional sangat bervariasi dan tidak sedikit yang lemah ketika dievaluasi dengan desain yang ketat (Kennedy, 2016), serta bahwa praktik pembelajaran mendalam sulit diwujudkan secara konsisten di kelas (Mehta & Fine, 2019). Dengan demikian, akumulasi bukti tentang pelatihan Pembelajaran Mendalam di Indonesia masih berada pada tahap menjanjikan dan belum dapat dianggap mapan.

Durasi pelatihan selama empat bulan memberi ruang bagi guru untuk menerapkan rancangan, menerima umpan balik, dan merefleksikan praktik. Literatur memang menempatkan durasi yang berkelanjutan sebagai salah satu ciri pengembangan profesional yang efektif (Darling-Hammond et al., 2017; Desimone, 2009; Garet et al., 2001), tetapi bukti tentang durasi tidak konsisten. Kennedy (2016) tidak menemukan bahwa program dengan jumlah jam yang lebih banyak secara otomatis lebih efektif. Durasi kemungkinan penting sejauh memungkinkan terjadinya siklus penerapan dan umpan balik, bukan karena lamanya waktu itu sendiri. Karena penelitian ini tidak membandingkan pelatihan empat bulan dengan pelatihan berdurasi lebih singkat, klaim bahwa pola pelatihan ini lebih efektif daripada pelatihan singkat tidak dapat diajukan. Klaim tersebut lebih tepat dirumuskan sebagai hipotesis untuk penelitian selanjutnya, yaitu bahwa pelatihan berpola In-On-In dengan pendampingan di sekolah menghasilkan peningkatan kompetensi yang lebih besar dan lebih bertahan lama dibandingkan dengan lokakarya singkat.

Peningkatan skor tes juga tidak dapat langsung ditafsirkan sebagai perubahan praktik pembelajaran di kelas. Dalam kerangka kompetensi sebagai kontinum (Blömeke et al., 2015), tes yang digunakan terutama mengukur disposisi kognitif, sedangkan kinerja mengajar menuntut transformasi pengetahuan ke dalam keputusan nyata di kelas. Dalam lima tingkat evaluasi pengembangan profesional menurut Guskey (2000), data penelitian ini berada pada tingkat reaksi dan keterlibatan peserta (indikator proses) serta tingkat pembelajaran peserta (tes kompetensi). Data tersebut belum menjangkau dukungan organisasi, penggunaan

pengetahuan dan keterampilan baru di kelas, ataupun hasil belajar peserta didik. Dokumentasi produk pembelajaran (Gambar 5) dan sesi berbagi praktik baik (Gambar 6) menunjukkan adanya penerapan awal dan diseminasi, tetapi tidak merupakan bukti sistematis perubahan praktik kelas. Oleh karena itu, apakah peningkatan skor kompetensi diikuti oleh perubahan praktik pembelajaran dan hasil belajar peserta didik masih perlu diuji melalui observasi kelas dan pengukuran tindak lanjut.



Gambar 5. Produk hasil kreasi peserta pelatihan



Gambar 6. Sesi sharing praktik baik antar peserta

Temuan ini memiliki beberapa implikasi praktis yang perlu dibaca secara tentatif. Bagi BGTK dan dinas pendidikan, heterogenitas perubahan skor menunjukkan perlunya tindak lanjut yang terdiferensiasi, misalnya pendampingan berbasis sekolah atau *coaching* bagi guru yang skor *posttest*-nya masih rendah; meta-analisis studi kausal menunjukkan bahwa bentuk dukungan ini berdampak positif terhadap praktik pembelajaran (Kraft et al., 2018). Sasaran tindak lanjut sebaiknya ditentukan berdasarkan skor akhir, bukan semata-mata berdasarkan N-Gain, karena N-Gain rendah juga dapat muncul pada guru yang skor awalnya sudah tinggi. Bagi penyelenggara pelatihan, nilai produk yang mendekati skor maksimum menunjukkan perlunya rubrik yang lebih diskriminatif dan penilaian oleh lebih dari satu penilai agar data penilaian pelatihan dapat dimanfaatkan sebagai bukti mutu. Bagi sekolah dan KKG, rencana tindak lanjut yang disusun pada tahap IN-2 dapat menjadi titik awal siklus inkuiri kolaboratif yang berkelanjutan, sehingga perubahan pengetahuan guru berpeluang diterjemahkan ke dalam praktik.

Penelitian ini memiliki sejumlah keterbatasan. Pertama, desain *one-group pretest–posttest* tidak memungkinkan atribusi kausal karena tidak ada kelompok kontrol. Kedua, sampel terdiri atas 26 guru dari satu kelas pelatihan di satu kota yang tidak dipilih secara acak sehingga generalisasi temuan terbatas. Ketiga, informasi validitas dan reliabilitas tes terbatas, dan kesetaraan bentuk tes *pretest* dan *posttest* tidak diuji sehingga sebagian selisih skor dapat berasal dari perbedaan tingkat kesukaran. Keempat, penulis pertama berperan ganda sebagai fasilitator dan peneliti; meskipun skor dihasilkan oleh sistem penilaian penyelenggara, potensi bias dalam penafsiran tidak dapat sepenuhnya dikesampingkan. Kelima, penelitian tidak mencakup pengukuran tindak lanjut ataupun observasi praktik kelas, sehingga keberlanjutan perubahan dan transfernya ke praktik pembelajaran tidak dapat diketahui. Keenam, N-Gain individual tidak stabil di dekat skor maksimum dan dipengaruhi regresi ke arah rerata sehingga kategori peningkatan perlu ditafsirkan secara hati-hati.

KESIMPULAN

Penelitian ini menunjukkan bahwa skor tes kompetensi guru SD tentang Pembelajaran Mendalam meningkat secara signifikan setelah guru mengikuti pelatihan berpola In-On-In selama empat bulan di Kota Lhokseumawe, dengan selisih rata-rata 13,27 poin (95% CI [6,98; 19,56]) dan ukuran efek $d_z = 0,85$ (95% CI [0,39; 1,30]). Perubahan tersebut tidak merata: sebagian besar peserta berada pada kategori N-Gain sedang, tetapi hampir sepertiga peserta berada pada kategori rendah. Indikator proses pelatihan menunjukkan keterlibatan dan penyelesaian tugas yang tinggi, tetapi tidak dapat diperlakukan sebagai bukti independen atas perubahan kompetensi.

Temuan ini memberikan indikasi positif mengenai potensi pelatihan Pembelajaran Mendalam dalam meningkatkan kompetensi guru, tetapi belum dapat dijadikan dasar kesimpulan kausal. Keterbatasan utama penelitian ini adalah desain *one-group pretest–posttest* tanpa kelompok kontrol, sehingga pengaruh pelatihan tidak dapat dipisahkan dari efek pengujian, maturasi, dan sejarah, sementara regresi ke arah rerata dapat

memengaruhi pola perubahan individual. Selain itu, pengukuran yang bertumpu pada tes tidak mencerminkan secara langsung kemampuan guru dalam praktik kelas. Secara praktis, penyelenggara pelatihan dan dinas pendidikan disarankan menyediakan pendampingan lanjutan yang terdiferensiasi bagi guru yang skor akhirnya masih rendah. Penelitian selanjutnya disarankan (1) menggunakan desain kuasi-eksperimental dengan kelompok pembanding, misalnya guru yang dijadwalkan mengikuti pelatihan pada angkatan berikutnya, atau desain eksperimen bila memungkinkan; (2) melibatkan sampel yang lebih besar dan beragam dari beberapa daerah; (3) menggunakan instrumen dengan bukti validitas dan reliabilitas yang dilaporkan, termasuk bentuk tes paralel yang telah disetarakan; (4) melakukan pengukuran tindak lanjut beberapa bulan setelah pelatihan; serta (5) mengobservasi praktik pembelajaran dan menganalisis hasil belajar peserta didik untuk mengetahui apakah peningkatan skor kompetensi benar-benar ditransfer ke praktik kelas.

DAFTAR PUSTAKA

Alfaruqi, A. Z., & Nurwahidah, N. (2025). Reflection on Indonesia's PISA scores and the 2024 Madrasah Teacher Competency Assessment results: Challenges in enhancing teacher competence. *Jurnal Pendidikan IPS*, *15*(1), 11–19. https://doi.org/10.37630/jpi.v15i1.2559

Altariusta, N. (2025). *Pengaruh beban kerja dan administrasi terhadap kinerja guru dengan work-life balance*. *7*(4).

Anwar, M., & Sodik, H. (2025). Kerangka konseptual pembelajaran mendalam (deep learning) dan implementasinya dalam pendidikan di Indonesia. *Tafhim Al-'Ilmi: Jurnal Pendidikan dan Pemikiran Islam*, *17*(1), 69–95. https://doi.org/10.37459/tafhim.v17i01.340

Atmojo, I. R. W., Muzzazinah, M., Ekawati, E. Y., Triastuti, R., Isnantyo, F. D., Sukarno, S., & Ramadian, R. K. (2025). Pelatihan implementasi pendekatan pembelajaran deep learning untuk meningkatkan kompetensi pedagogik guru SD di Kota Surakarta. *Jurnal Pengabdian UNDIKMA*, *6*(1), 123. https://doi.org/10.33394/jpu.v6i1.14507

Aziz, E. A., Khak, M. A., Nasrullah, R., Asmarini, P., Solihah, A., Prihartono, W., & Aziz, M. (2024). *Memperkuat literasi Indonesia: Menuju bangsa yang maju dan bermartabat* (Risalah Kebijakan No. 3). Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Coletta, V. P., & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*, *73*(12), 1172–1182. https://doi.org/10.1119/1.2117109

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://doi.org/10.54300/122.311

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Direktorat Sekolah Dasar. (2022, February 12). *Luncurkan Kurikulum Merdeka, Mendikbudristek: Ini lebih fleksibel!* https://ditsd.kemendikdasmen.go.id/artikel/detail/luncurkan-kurikulum-merdeka-mendikbudristek-ini-lebih-fleksibel

Fullan, M., Quinn, J., & McEachen, J. (2018). *Deep learning: Engage the world, change the world*. Corwin.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945. https://doi.org/10.3102/00028312038004915

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. https://doi.org/10.1119/1.18809

Kementerian Pendidikan Dasar dan Menengah. (2025). *Peraturan Menteri Pendidikan Dasar dan Menengah Republik Indonesia Nomor 13 Tahun 2025 tentang Perubahan atas Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 12 Tahun 2024 tentang Kurikulum pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah*.

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2022). *Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 262/M/2022 tentang Perubahan atas Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 56/M/2022 tentang Pedoman Penerapan Kurikulum dalam Rangka Pemulihan Pembelajaran*. https://bpmpntb.kemendikdasmen.go.id/upload/unduhan/20220712080500.pdf

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2023, December 5). *Peringkat Indonesia pada PISA 2022 naik 5–6 posisi dibanding 2018* [Siaran pers]. https://www.kemendikdasmen.go.id/siaran-pers/8990-peringkat-indonesia-pada-pisa-2022-naik-5-6-posisi-dibanding

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Knapp, T. R. (2016). Why is the one-group pretest–posttest design still used? *Clinical Nursing Research*, *25*(5), 467–472. https://doi.org/10.1177/1054773816666280

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education*, *38*(5), 583–616. https://doi.org/10.1080/03054985.2012.731208

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology*, *46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, *75*(1), 87–91. https://doi.org/10.1119/1.2372468

Mehta, J., & Fine, S. (2019). *In search of deeper learning: The quest to remake the American high school*. Harvard University Press.

Muqtada, M. R., Agustyaningrum, N., Normawati, A., Hadi, N. S. K., Nailufar, U., & Oktavia, F. (2026). Pelatihan pembelajaran mendalam bagi guru SD di Magelang sebagai transformasi pembelajaran Kurikulum Merdeka. *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, *10*(1).

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. The National Academies Press. https://doi.org/10.17226/13398

Nissen, J. M., Talbot, R. M., Nasim Thompson, A., & Van Dusen, B. (2018). Comparison of normalized gain and Cohen's *d* for analyzing gains on concept inventories. *Physical Review Physics Education Research*, *14*(1), 010115. https://doi.org/10.1103/PhysRevPhysEducRes.14.010115

Nurhakim, H. Q., Isnan, I. R., Harsing, H., Supiana, S., & Zakiah, Q. Y. (2025). Inovasi kurikulum dan teknologi pembelajaran (deep learning). *EduTeach: Jurnal Edukasi dan Teknologi Pembelajaran*, *6*(2), 134–143. https://doi.org/10.37859/eduteach.v6i02.9487

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. https://doi.org/10.3102/0034654307313793

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Paramita, A. A. J., Ariani, K. B., Mahadewi, N. P. D., & Werang, B. R. (2025). Tantangan implementasi pembelajaran mendalam sebagai pendekatan dalam Kurikulum Merdeka. *DIKSI: Jurnal Kajian Pendidikan dan Sosial*, *6*(4), 757–769. https://doi.org/10.53299/diksi.v6i4.3030

Rahmandani, F., Hamzah, M. R., Handayani, T., & Kurniawan, M. W. (2025). Integrasi pembelajaran mendalam (deep learning) dalam mewujudkan pembelajaran yang bermutu dan bermakna bagi peserta didik. *Jurnal Sosial Humaniora dan Pendidikan*, *4*(3), 769–781. https://doi.org/10.55606/inovasi.v4i3.4896

Rahmawati, Y., Mu'ti, A., Suyanto, S., & Herianingtyas, N. L. R. (2025). Pembelajaran mendalam: Transformasi pembelajaran menuju pendidikan bermutu. *Jurnal Penelitian Kebijakan Pendidikan*, *18*(1). https://doi.org/10.24832/jpkp.v18i1.1281

Republik Indonesia. (2005). *Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 tentang Guru dan Dosen* (Lembaran Negara Republik Indonesia Tahun 2005 No. 157).

Republik Indonesia. (2024). *Peraturan Presiden Republik Indonesia Nomor 139 Tahun 2024 tentang Penataan Tugas dan Fungsi Kementerian Negara Kabinet Merah Putih Periode Tahun 2024–2029*.

Rustyawati, D., & Fauzi, M. (2025). Peningkatan kompetensi pedagogik guru melalui pelatihan kurikulum deep learning di MI Hidayatun Najah. *Eastasouth Journal of Impactive Community Services*, *4*(1), 102–110. https://doi.org/10.58812/ejimcs.v4i01.419

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

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. https://doi.org/10.3102/00346543231217480

Suparya, I. K., Suastra, I. W., & Arnyana, I. B. P. (2022). Rendahnya literasi sains: Faktor penyebab dan alternatif solusinya. *Jurnal Ilmiah Pendidikan Citra Bakti*, *9*(1), 153–166. https://doi.org/10.38048/jipcb.v9i1.580

Susanti, Dwinata, A., Elvi, M., Putri, N. H. S., Izzati, N., & Fera, M. (2026). Pendampingan pengembangan perencanaan pembelajaran mendalam bagi guru SMP. *JMM (Jurnal Masyarakat Mandiri)*, *10*(1).

Susilana, R., Ikanubun, L. E., Amelia, D., Hadiapurwa, A., Dewi, L., Setiawan, B., Anisa, D., & Sipayung, A. T. (2025). Pelatihan pengembangan desain pembelajaran mendalam (deep learning) untuk guru sekolah dasar. *Publikasi Pendidikan*, *15*(3), 1055. https://doi.org/10.70713/publikan.v15i3.76539

Suyanto, Mubarak, A. Z., Suryadi, B., Darmawan, C., Wahyudin, D., Qodir, D. A., Iskandar, H., Wiyono, H. T., Ismah, Kadir, Yuliati, K., Maskur, Faradella, N. E., Herianingtyas, N. L. R., Waspodo, R. M., Purwahida, R., Indrajit, R. E., Yazid, R., Oktaviana, S., … Damarjati, T. (2025). *Naskah akademik pembelajaran mendalam menuju pendidikan bermutu untuk semua*. Pusat Kurikulum dan Pembelajaran, Badan Standar, Kurikulum, dan Asesmen Pendidikan, Kementerian Pendidikan Dasar dan Menengah.

Yusmar, F., & Fadilah, R. E. (2023). Analisis rendahnya literasi sains peserta didik Indonesia: Hasil PISA dan faktor penyebab. *LENSA (Lentera Sains): Jurnal Pendidikan IPA*, *13*(1), 11–19. https://doi.org/10.24929/lensa.v13i1.283

Zeiser, K. L., Taylor, J., Rickles, J., Garet, M. S., & Segeritz, M. (2014). *Evidence of deeper learning outcomes* (Findings from the Study of Deeper Learning: Opportunities and Outcomes, Report No. 3). American Institutes for Research.