



Jurnal Abmas

Media Informasi Pengabdian Kepada Masyarakat
<https://ejournal-dppm.upi.edu/abmas>



Online training for physical education teachers: Knowledge gains, assessment design gaps

Dian Budiana¹, Gifran Rihla Gifarka Latief², Achmad Syakur Fahri³, Gano Sumarno⁴

^{1,2,3,4} Universitas Pendidikan Indonesia, Bandung, Indonesia

dianbudiana@upi.edu¹

ABSTRACT

Indonesia's Deep Learning policy, in force nationally since the 2025/2026 school year, requires physical education teachers to redesign their teaching within a short policy deadline. Online training is the main way to reach a dispersed teaching workforce, yet its learning outcomes are rarely evaluated with measurable data. This community service activity aimed to establish how far a single-session online training changed teachers' knowledge of deep learning instructional design, which content domains were most resistant to change, and how teachers articulated plans for implementing authentic assessment. A university and a teachers' professional association delivered the activity. They evaluated it using an embedded mixed-methods design: 95 school teachers completed a 29-item knowledge test before and after the session and wrote implementation plans for their own schools. Knowledge increased significantly but only slightly, and only six teachers exceeded the individual reliable change threshold. Authentic assessment knowledge did not move, while only 32.6% of teachers named a concrete assessment mode in their plans. Knowledge tests may therefore overstate teachers' readiness to design assessment.

ARTICLE INFO

Article History:

Received: 23 Apr 2026

Revised: 12 Sep 2026

Accepted: 28 Sep 2026

Publish online: 14 Oct 2026

Keywords:

authentic assessment; deep learning;
mixed methods; physical education;
teacher training

Open access

Jurnal Abmas

is a peer-reviewed open-access journal

ABSTRAK

Kebijakan Pembelajaran Mendalam yang diberlakukan secara nasional sejak tahun ajaran 2025/2026 menuntut guru pendidikan jasmani untuk menata ulang praktik pembelajaran mereka dalam waktu singkat. Pelatihan daring menjadi cara utama untuk menjangkau guru yang tersebar, namun hasil belajarnya jarang dievaluasi dengan data yang terukur. Kegiatan pengabdian ini bertujuan untuk menetapkan sejauh mana pelatihan daring satu sesi mengubah pengetahuan guru tentang desain pembelajaran mendalam, domain isi mana yang paling resisten terhadap perubahan, serta bagaimana guru mengartikulasikan rencana penerapan asesmen otentik. Kegiatan diselenggarakan melalui kemitraan perguruan tinggi dan asosiasi profesi guru, serta dievaluasi dengan desain campuran tertanam: 95 guru sekolah mengerjakan tes pengetahuan 29 butir sebelum dan sesudah pelatihan serta menuliskan rencana penerapan di sekolah masing-masing. Pengetahuan meningkat secara signifikan, namun dengan magnitudo kecil, dan hanya enam guru yang melampaui ambang perubahan yang reliabel secara individual. Pengetahuan tentang asesmen otentik tidak meningkat, sementara hanya 32,6% guru yang menyebutkan bahwa mode asesmen konkret dalam rencana mereka adalah tes pengetahuan. Karena itu, hal ini berpotensi melebih-lebihkan kesiapan guru dalam merancang asesmen.

Kata Kunci: asesmen otentik; metode campuran; pelatihan guru; pendidikan jasmani; pembelajaran mendalam

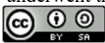
How to cite (APA Style)

Budiana, D., Latief, G. R. G., Fahri, A. S., & Sumarno, G. (2026). Online training for physical education teachers: Knowledge gains, assessment design gaps. *Jurnal Abmas*, 26(2), 291-304.

Peer review

This article underwent the journal's standard double-blind peer review, in which both reviewers and authors are anonymized during review.

Copyright



2026, Dian Budiana, Gifran Rihla Gifarka Latief, Achmad Syakur Fahri, Gano Sumarno. This an open-access is article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) <https://creativecommons.org/licenses/by-sa/4.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author, and source are credited. *Corresponding author: dianbudiana@upi.edu

INTRODUCTION

Pendidikan jasmani menempati posisi khusus di antara mata pelajaran di sekolah. Ia diharapkan mengembangkan kompetensi gerak dan literasi fisik, sekaligus mendukung pembelajaran sosial, emosional, dan kognitif, dengan efek yang diharapkan bertahan sebagai kebiasaan aktivitas fisik jauh setelah masa sekolah berakhir. Pemenuhan harapan tersebut sangat bergantung pada cara guru merancang dan melaksanakan pembelajaran, sehingga mutu guru menjadi perhatian yang berulang dalam bidang ini. Di Indonesia, perhatian itu beralasan: kajian komparatif atas pendidikan jasmani dan olahraga sekolah di Indonesia dan Türkiye mendokumentasikan keterbatasan infrastruktur serta jumlah guru berkualifikasi di sekolah negeri sebagai kendala struktural yang belum terselesaikan (Ma'mun *et al.*, 2025). Mulai tahun ajaran 2025/2026, berdasarkan Peraturan Menteri Pendidikan Dasar dan Menengah Republik Indonesia Nomor 13 Tahun 2025 tentang Pembelajaran Mendalam, Kementerian Pendidikan Dasar dan Menengah memberlakukan Pembelajaran Mendalam sebagai kerangka pedagogis pada seluruh jenjang pendidikan dasar dan menengah yang bertumpu pada tiga prinsip: berkesadaran, bermakna, dan menggembirakan.

Kerangka tersebut beririsan erat dengan agenda kebermaknaan yang berkembang dalam literatur pendidikan jasmani internasional, terutama dalam penekanan pada pembelajaran yang relevan secara personal dan menyenangkan (Fletcher & Ní Chróinín, 2021). Namun, mekanisme adopsinya berbeda tajam. Pendekatan kebermaknaan di luar Indonesia berkembang melalui kolaborasi praktisi-peneliti dalam skala kecil dan jangka panjang (Beni *et al.*, 2022; Beni *et al.*, 2023). Pendekatan Indonesia hadir sebagai kebijakan kurikulum yang ditujukan secara serentak kepada para guru di seluruh sistem pendidikan nasional. Situasi ini melahirkan persoalan implementasi yang khas: tenaga guru yang besar dan tersebar harus menguasai logika pedagogis baru dalam tenggat waktu kebijakan yang singkat. Pelatihan guru merupakan instrumen utama untuk tugas tersebut. Kajian sistematis atas 30 studi empiris terbitan 2015-2025 dalam bidang pendidikan jasmani menyimpulkan bahwa efektivitas pelatihan tidak ditentukan semata oleh format atau durasi, melainkan oleh sejauh mana program tersebut menumbuhkan keterlibatan berkelanjutan, praktik reflektif, dan relevansi kontekstual (Ilmawati *et al.*, 2026).

Komunitas belajar memperoleh perhatian khusus sebagai wahana keterlibatan semacam itu, dan tinjauan terhadap komunitas guru daring menempatkan penopangan sebagai kategori yang menentukan keberhasilannya (Dille & Røkenes, 2021; Parker *et al.*, 2022). Di negara berpendapatan rendah dan menengah, tinjauan atas 170 studi menemukan manfaat pelatihan berbasis teknologi bagi guru, tetapi bukti mengenai dampaknya terhadap praktik di kelas masih terbatas (Hennessy *et al.*, 2022). Sejumlah kajian terdahulu berdekatan dengan kegiatan yang dilaporkan di sini, namun berbeda pada beberapa poin penting. Pertama, kajian yang membahas perancangan dan validasi model pengembangan profesional bagi guru pendidikan jasmani Indonesia mencatat bahwa guru tetap menghadapi hambatan dalam mengajar meskipun telah mengikuti berbagai program pembinaan; kajian tersebut berhenti pada validasi model dan tidak mengukur apa yang benar-benar dipelajari oleh peserta (Mu'arifin & Narmaditya, 2022). Kedua, kajian ini membahas pengembangan instrumen berbasis web untuk mengukur kompetensi guru pendidikan jasmani sesuai standar nasional, tetapi pengukurannya bersifat lintas-seksional dan tidak dikaitkan dengan intervensi (Suroto *et al.*, 2022).

Ketiga, kajian ini memetakan tingkat pemahaman guru pendidikan jasmani terhadap dokumen kurikulum, termasuk prinsip penyusunan perencanaan pembelajaran, namun tanpa pengukuran sebelum dan sesudah pelatihan (Hardiyono *et al.*, 2021). Keempat, kajian tentang pengalaman guru pendidikan jasmani mengikuti pelatihan daring mulai bermunculan (Ferreira *et al.*, 2023; Lander *et al.*, 2022), dan pemanfaatan perangkat digital oleh guru pendidikan jasmani di Indonesia telah dipetakan (Nopembri *et al.*, 2022), tetapi keduanya berfokus pada persepsi, kepuasan, dan pembentukan komunitas, bukan pada hasil belajar yang dikelompokkan berdasarkan domain isi. Kelima, pada ranah asesmen, literatur mendokumentasikan bahwa asesmen tetap bermasalah dalam pendidikan jasmani (Barker *et al.*, 2025; Killian & Woods, 2021), guru berhasil menerapkan asesmen formatif dan otentik hanya dalam kondisi pendukung tertentu (Barrientos Hernán *et al.*, 2023), dan merancang aktivitas asesmen formatif jauh lebih mudah

daripada melaksanakannya (Slingerland *et al.*, 2024). Namun, kajian-kajian tersebut tidak menguji sejauh mana tes pengetahuan merepresentasikan kemampuan guru dalam merancang asesmen.

Kebaruan kegiatan pengabdian ini terletak pada penyandingan langsung dua ukuran pada orang yang sama: pengetahuan guru yang terukur melalui tes dan kemampuan merancang yang teramati melalui rencana penerapan yang disusun sendiri oleh guru. Berbeda dari kelima kelompok kajian di atas, laporan ini tidak berhenti pada validasi model, pemetaan kompetensi, maupun persepsi peserta, melainkan menguji kesetiaan tes pengetahuan sebagai ukuran hasil pelatihan, di mana dalam domain asesmen otentik, konsistensi menjadi penentu utama. Berdasarkan penelusuran penulis, penyandingan semacam ini belum pernah dilaporkan dalam konteks pelatihan guru pendidikan jasmani di Indonesia. Permasalahan pengabdian, yakni evaluasi pelatihan guru pendidikan jasmani di Indonesia, jarang menguji apakah pengetahuan yang terukur sepadan dengan kemampuan merancang, sehingga berisiko keliru dalam menetapkan domain mana yang sesungguhnya membutuhkan perhatian lebih lanjut. Kegiatan pengabdian ini bertujuan untuk: menetapkan sejauh mana pelatihan mengubah pengetahuan guru tentang desain pembelajaran mendalam; menetapkan apakah perubahan berbeda antardomain isi dan domain mana yang paling resisten; serta menetapkan bagaimana guru mengartikulasikan rencana penerapan asesmen otentik dan bagaimana kedalaman rencana tersebut berkaitan dengan pengetahuan mereka.

Literature Review

Pembelajaran Mendalam dan Kebermaknaan dalam Pendidikan Jasmani

Tiga prinsip Pembelajaran Mendalam adalah berkesadaran, bermakna, dan menggembirakan, di mana guru dituntut untuk menggeser titik tumpu pembelajaran dari penguasaan keterampilan yang terisolasi ke kualitas pengalaman murid. Terdapat dua prinsip pedagogis yang menopang penataan semacam itu: pendekatan demokratis yang memberi ruang bagi kepemilikan dan individualisasi pengalaman, serta refleksi sebagai mekanisme yang menghubungkan pengalaman masa lalu, masa kini, dan masa depan (Fletcher & Ní Chróinín, 2021). Keduanya bukan aktivitas tambahan, melainkan menuntut peninjauan ulang terhadap apa yang dianggap sebagai belajar dan, sebagai konsekuensinya, bagaimana belajar seharusnya dinilai.

Kajian tentang bagaimana guru memaknai inovasi semacam ini menunjukkan bahwa penerimaan bergantung pada proses pemaknaan dalam konteks lokal masing-masing guru, bukan semata-mata pada kejelasan rumusan kebijakan (Beni *et al.*, 2023). Perangkat pedagogis yang diperkenalkan dalam kegiatan ini adalah *Teaching Games for Understanding*, *Sport Education Model*, pembelajaran inkuiri, dan spektrum gaya mengajar, yang termasuk dalam keluarga model pedagogis dengan bukti empiris paling mapan dalam pendidikan jasmani (Arufe-Giráldez *et al.*, 2023; Fernández-Río & Iglesias, 2022). Tinjauan payung atas model-model tersebut menegaskan satu hal yang langsung relevan bagi pelatihan singkat: kekuatan sebuah model terletak pada penerapannya yang utuh, bukan pada pengenalan namanya; oleh karena itu, penguasaan nomenklatur tidak dapat diperlakukan sebagai indikator penguasaan model (Fernández-Río & Iglesias, 2022).

Pengembangan Profesional Guru Pendidikan Jasmani

Sintesis mutakhir atas pengembangan profesional berkelanjutan dalam pendidikan jasmani menyimpulkan bahwa efektivitasnya ditentukan oleh keterlibatan berkelanjutan, praktik reflektif, dan relevansi kontekstual, bukan semata-mata oleh format atau durasi (Ilmawati *et al.*, 2026). Komunitas belajar merupakan wahana yang paling sering dilaporkan untuk keterlibatan semacam itu, meskipun bukti mengenai keberlanjutan dan dampaknya terhadap pembelajaran murid masih terbatas (Parker *et al.*, 2022). Dalam format daring, tinjauan atas 52 studi empiris menempatkan penopangan sebagai kategori utama yang menentukan keberhasilan komunitas guru, dengan faktor internal guru yang berperan besar dalam interaksinya dengan materi, fasilitator, dan sesama peserta (Dille & Rokenes, 2021).

Dalam konteks reformasi kurikulum, guru melaporkan bahwa yang paling menopang mereka adalah strategi pembelajaran yang dimodelkan secara langsung selama pelatihan, kemudahan akses terhadap sumber ajar, serta peningkatan keyakinan diri (Haug & Mork, 2021). Temuan ini penting karena menempatkan pemodelan dan ketersediaan perangkat, bukan penyampaian informasi, sebagai mekanisme kerja pelatihan yang menjembatani rumusan kebijakan dan praktik di kelas. Di negara berpendapatan rendah dan menengah, tinjauan atas 170 studi menemukan bahwa manfaat pelatihan bermediasi teknologi bagi guru cukup konsisten, namun bukti mengenai keberlanjutan, efisiensi biaya, dan dampak nyata terhadap praktik di kelas masih terbatas (Hennessy *et al.*, 2022).

Asesmen Otentik dan Formatif sebagai Titik Kritis

Asesmen adalah titik di mana penataan pedagogis paling menentukan sekaligus paling sering gagal. Sintesis interpretatif kritis atas literatur asesmen pendidikan jasmani menyimpulkan bahwa persoalan asesmen sumatif berpusat pada penggunaan kriteria personal yang terinternalisasi oleh guru, sedangkan persoalan asesmen formatif berpusat pada kurangnya keakraban guru dan murid terhadap praktiknya serta kurangnya kompetensi dalam menerapkan strategi asesmen (Barker *et al.*, 2025). Tinjauan cakupan atas praktik asesmen pada pendidikan jasmani jenjang K-12 memperkuat kesimpulan tersebut dan mencatat bahwa hanya sedikit studi yang menghubungkan praktik asesmen dengan hasil belajar murid (Killian & Woods, 2021). Penelitian menemukan bahwa guru berpengalaman mampu menerapkan asesmen formatif dan otentik (Barrientos Hernán *et al.*, 2023).

Hal tersebut hanya terjadi bila pembelajaran ditata dengan asesmen sejak awal serta keanggotaan dalam komunitas praktik diidentifikasi sebagai faktor penentu dalam pengembangan literasi asesmen. Terdapat penelitian yang mendokumentasikan bahwa perpindahan dari merancang aktivitas asesmen formatif menuju pelaksanaannya menuntut pendekatan yang berkelanjutan, kolaboratif, dan didukung (Slingerland *et al.*, 2024). Secara lintas bidang, perbedaan antara asesmen sebagai, untuk, dan atas pembelajaran hanya bermakna bila ketiganya dipandang koheren satu sama lain (Schellekens *et al.*, 2021). Sementara itu, tinjauan sistematis terhadap 94 artikel tentang asesmen otentik mengidentifikasi kebutuhan pelatihan dan sumber daya yang memadai sebagai hambatan penerapan yang paling sering muncul (Vlachopoulos & Makri, 2024). Teori-teori inilah yang kemudian digunakan sebagai penguat pada bagian hasil dan pembahasan.

Kompetensi dan Kesiapan Guru Pendidikan Jasmani di Indonesia

Dalam konteks Indonesia, kompetensi guru pendidikan jasmani telah diukur menggunakan instrumen berbasis standar nasional dan terbukti dipengaruhi oleh masa kerja, jenjang, serta jenis kelamin (Suroto *et al.*, 2022). Pemahaman guru terhadap dokumen kurikulum, termasuk prinsip penyusunan perencanaan pembelajaran, dilaporkan berada pada kategori tinggi, namun dengan sebaran yang lebar (Hardiyono *et al.*, 2021). Pada masa pembelajaran jarak jauh, guru pendidikan jasmani di Indonesia yang terdokumentasi memanfaatkan perangkat digital yang sudah lazim serta mengandalkan aplikasi pesan dan sistem manajemen pembelajaran yang sederhana (Nopembri *et al.*, 2022). Ketiga temuan ini menjelaskan mengapa pelatihan daring menjadi pilihan yang masuk akal secara logistik, sekaligus mengapa hasil belajarnya perlu diukur alih-alih diasumsikan.

METHODS

Pengabdian ini menggunakan desain pra-eksperimental. Kelompok pembandingan tidak dibentuk karena kegiatan diselenggarakan sebagai pengabdian kepada masyarakat yang menjadi hak seluruh anggota mitra, sehingga penugasan acak maupun penahanan intervensi tidak dapat dibenarkan, baik secara etis maupun secara praktis.

Pelaksanaan Kegiatan

Kegiatan berupa satu sesi daring sinkron yang dilaksanakan pada 29 Juni 2026 melalui kemitraan antara FPOK UPI dan IGORNAS Jawa Barat. Pra-test dilakukan sebelum sesi pemaparan materi dan pascatest dilakukan setelah pemaparan materi. Materi mencakup alur perencanaan dari capaian pembelajaran hingga modul ajar (CP - TP - ATP - modul ajar); *backward design*; modalitas pedagogis meliputi *Teaching Games for Understanding*, *Sport Education Model*, pembelajaran inkuiri, dan spektrum gaya mengajar Mosston; penentuan materi esensial dan pembelajaran berdiferensiasi; asesmen otentik dan berbasis kinerja dengan rubrik; tiga prinsip Pembelajaran Mendalam; serta empat elemen capaian pembelajaran pendidikan jasmani.

Peserta dan Kriteria Eksklusi

Keikutsertaan bersifat sukarela dan terbuka bagi anggota mitra, tanpa insentif selain sertifikat kehadiran. Karena diselenggarakan secara daring, peserta tersebar jauh melampaui kota asal organisasi mitra. Kedua instrumen diberikan melalui formulir daring. Pretest diisi oleh 193 responden dan pascatest oleh 165 responden. Karena formulir pra-tes tidak merekam alamat surel, pencocokan dilakukan berdasarkan nama yang dinormalisasi: huruf dikecilkan, gelar akademik dihapus, serta tanda baca dan spasi ganda dibersihkan. Prosedur ini menghasilkan 129 pasangan. Normalisasi diperlukan karena banyak peserta menuliskan namanya secara tidak konsisten antara kedua sesi, terutama dengan menambahkan atau menghilangkan gelar. Setelah penerapan empat kriteria eksklusi, sampel analitik berjumlah 95 guru.

Evaluasi Kegiatan

Evaluasi kegiatan menggunakan desain campuran tertanam, yakni komponen kuantitatif berdesain satu kelompok dengan pra-test-pascatest yang dilengkapi analisis kualitatif atas rencana penerapan yang ditulis peserta pada pascatest. Komponen kuantitatif menetapkan besaran dan sebaran perubahan pengetahuan antardomain; komponen kualitatif menetapkan apa yang mampu diartikulasikan peserta sebagai praktik konkret. Integrasi dilakukan pada tahap interpretasi melalui *joint display* yang menghubungkan kedalaman rencana dengan skor pengetahuan, sesuai dengan prosedur integrasi yang direkomendasikan untuk desain campuran.

Instrumen

Tes pilihan ganda 30 butir disusun oleh tim pelaksana agar selaras dengan materi sesi, mencakup lima domain: perencanaan kurikulum (6 butir), model dan pendekatan pembelajaran (9 butir), asesmen otentik (6 butir), konsep pembelajaran mendalam (6 butir), dan capaian pembelajaran pendidikan jasmani (3 butir). Skor hasil rekonstruksi identik dengan skor yang terekam pada seluruh 193 responden pretest dan 165 responden pascatest (deviasi absolut maksimum = 0). Konsistensi internal memadai pada kedua pengukuran ($KR-20 = 0,843$ pada pretest; $0,796$ pada posttest). Tiga butir terbuka diberikan pada pascatest, meminta peserta menjelaskan makna pembelajaran mendalam dalam PJOK, menguraikan penyusunan CP, TP, ATP, dan modul ajar, serta memaparkan satu rencana nyata penerapan asesmen otentik di sekolah masing-masing.

Analisis Data

Skor gain menyimpang tajam dari normalitas (Shapiro-Wilk $W = 0,730$; $p < 0,001$), sehingga uji Wilcoxon signed-rank digunakan sebagai uji utama dengan uji t berpasangan sebagai pemeriksaan sensitivitas. Effect size dilaporkan sebagai Cohen's d dengan selang kepercayaan 95% dan sebagai korelasi rank-biserial. Karena 26% peserta mencapai skor maksimum pada pretest, selisih skor mentah terkendala langit-langit; karena itu, dilaporkan pula normalized

gain, $(g) = (\text{post} - \text{pre}) / (\text{max} - \text{pre})$, yang menyatakan perubahan sebagai proporsi dari peningkatan yang masih tersedia. Ukuran ini dipilih dengan kesadaran bahwa metrik gain apa pun tetap menyisakan ketergantungan pada skor awal, sehingga angkanya dibaca sebagai deskripsi, bukan sebagai ukuran efektivitas yang dapat dibandingkan lintas kegiatan (Christman *et al.*, 2024; Westphale *et al.*, 2021).

Dalam mencirikan perubahan pada tataran individual, Reliable Change Index dihitung menggunakan simpangan baku pra-test dan reliabilitas KR-20, mengikuti prosedur dan pertimbangan komputasi yang dirangkum dalam Blampied (2022). Subskor domain dibandingkan dengan uji Wilcoxon, perubahan per butir diuji dengan uji McNemar disertai koreksi Holm untuk 29 perbandingan, dan bias atrisi diperiksa dengan uji Mann-Whitney. Hubungan skor awal dengan gain dilaporkan secara deskriptif dan tidak ditafsirkan sebagai efektivitas diferensial, karena regresi ke rata-rata merupakan penjelasan alternatif yang memadai pada desain satu kelompok dengan langit-langit yang nyata.

Jawaban terbuka dianalisis dengan analisis isi induktif mengikuti tahapan persiapan, pengorganisasian, dan pelaporan, dengan pengodean bersama oleh dua pengode untuk butir yang meragukan (Coulston *et al.*, 2025; Vears & Gillam, 2022). Dalam butir rencana penerapan, disusun rubrik kedalaman empat level berdasarkan jumlah komponen desain pembelajaran yang dinyatakan secara eksplisit: konteks atau materi, model pembelajaran, mode asesmen, refleksi, dan tahapan. Level 0 menandai jawaban yang tidak responsif, level 1 niat generik tanpa detail operasional, level 2 metode yang disebut sebagai elaborasi terbatas, dan level 3 rencana yang terperinci. Tema isi diturunkan secara induktif dari pembacaan awal seluruh korpus, kemudian diterapkan secara sistematis.

RESULTS AND DISCUSSION

Karakteristik Peserta

Dari 193 guru yang mengerjakan pratest dan 165 yang mengerjakan pascatest, 129 dapat dicocokkan. Setelah penerapan empat kriteria eksklusi, sampel analitik berjumlah 95 guru. Peserta mengajar pada seluruh jenjang: sekolah dasar ($n = 43$; 45%), sekolah menengah atas dan kejuruan ($n = 29$; 31%), sekolah menengah pertama ($n = 21$; 22%), dan tidak terklasifikasi ($n = 2$; 2%). Sebagian besar bertugas di sekolah negeri ($n = 60$; 63%), dengan guru madrasah sebagai kelompok yang cukup besar ($n = 23$; 24%) dan sekolah swasta melengkapi sisanya ($n = 12$; 13%). Sebanyak 81 peserta (85%) berjenis kelamin laki-laki.

Nama instansi menunjukkan sebaran geografis yang luas melampaui kota asal organisasi mitra, mencakup sekolah di Papua, Kepulauan Riau, Kepulauan Bangka Belitung, Jawa Tengah, dan Jakarta. Karena sebagian besar nama instansi tidak memuat informasi lokasi, pengodean geografis yang sistematis tidak dimungkinkan, sehingga sebaran hanya dilaporkan secara deskriptif. Peserta yang berpasangan memperoleh skor pratest sedikit lebih tinggi dibandingkan dengan responden yang tidak berpasangan ($M = 25,46$ vs $23,97$; Mann-Whitney $U = 4818,5$; $p = 0,035$). Meskipun skor pascatest tidak berbeda ($M = 26,94$ vs $25,51$; $U = 2625,5$; $p = 0,130$). Seleksi ke dalam sampel berpasangan karena itu tidak sepenuhnya acak, dan hasilnya perlu dibaca sebagai berlaku bagi guru yang mengikuti kedua sesi.

Perubahan Pengetahuan

Tabel 1 menyajikan statistik deskriptif skor pengetahuan total. Skor pra-test sudah tinggi ($M = 25,96$; $SD = 3,72$ dari maksimum 29), dengan 25 peserta (26%) menjawab benar semua butir sebelum pelatihan. Skor pascatest meningkat menjadi $M = 27,02$ ($SD = 2,76$).

Tabel 1. Skor Tes Pengetahuan Sebelum dan Sesudah Pelatihan

Statistik	Nilai
Pratest	M = 25,96; SD = 3,72
Pascatest	M = 27,02; SD = 2,76
Rerata gain	1,06; IK 95% [0,48; 1,65]
Shapiro-Wilk (gain)	W = 0,730; $p < 0,001$
Wilcoxon signed-rank	W = 399,5; $p < 0,001$
Uji t berpasangan	$t(94) = 3,60$; $p < 0,001$
Cohen's dz	0,37; IK 95% [0,16; 0,58]
Rank-biserial r	0,55
Normalized gain (g)	0,35
Naik / tetap / turun	44 / 36 / 15
KR-20	0,843 (pra); 0,796 (pasca)
Skor sempurna sebelum pelatihan	25 (26%)
Perubahan reliabel (RCI = 4,09)	6 (6%) melampaui ambang

Sumber: Penelitian (2026)

Peningkatan tersebut signifikan secara statistik (Wilcoxon $W = 399,5$; $p < 0,001$; $t(94) = 3,60$; $p < 0,001$) dengan rerata gain sebesar 1,06 poin (IK 95% [0,48; 1,65]). Besaran efeknya kecil hingga sedang (Cohen's $dz = 0,37$; IK 95% [0,16; 0,58]; rank-biserial $r = 0,55$). Karena langit-langit yang nyata membatasi selisih skor mentah, normalized gain juga dihitung dan menghasilkan $\langle g \rangle = 0,35$. Angka ini dilaporkan sebagai deskripsi besaran perubahan relatif terhadap ruang peningkatan yang masih tersedia, bukan sebagai ukuran efektivitas yang dapat dibandingkan antarkegiatan. Perubahan tidak merata. Sebanyak 44 guru (46%) memperoleh skor lebih tinggi pada pascatest, 36 (38%) tetap, dan 15 (16%) memperoleh skor lebih rendah.

Penggunaan *Reliable Change Index* (SEM = 1,48; RCI = 4,09) menunjukkan bahwa hanya enam peserta (6%) menunjukkan perubahan yang melampaui ambang perubahan reliabel individual, 89 stabil, dan tidak ada yang menurun secara reliabel. Angka ini penting dilaporkan apa adanya: signifikansi pada tataran kelompok tidak otomatis berarti perubahan yang signifikan pada tataran individu. Gain berhubungan negatif dengan pengetahuan awal ($\rho = -0,609$; $p < 0,001$). Tertile terendah mengalami kenaikan rata-rata 2,68 poin, sementara tertile tengah dan atas nyaris tidak berubah ($-0,07$ dan $-0,28$). Mengingat langit-langit dan peran regresi ke rata-rata yang telah lama terdokumentasi pada desain satu kelompok, pola ini dilaporkan secara deskriptif dan tidak ditafsirkan sebagai perbedaan efektivitas pelatihan antar kelompok.

Pola Perubahan Antar Domain

Tabel 2 memilah 29 butir ke dalam lima domain isi. Pengamatan paling mencolok adalah perbedaan antara domain yang bergerak dan domain yang sama sekali tidak bergerak.

Tabel 2. Subskor Domain Sebelum dan Sesudah Pelatihan

Domain	k	pre	post	$\langle g \rangle$	p
Model dan pendekatan pembelajaran	9	0,877	0,931	0,44	$< 0,001$
Konsep pembelajaran mendalam	5	0,916	0,949	0,40	0,036
Perencanaan kurikulum	6	0,881	0,919	0,32	0,008

Domain	k	pre	post	$\langle g \rangle$	p
Capaian pembelajaran PJOK	3	0,902	0,930	0,29	0,399
Asesmen otentik	6	0,916	0,932	0,19	0,780

Sumber: Penelitian (2026)

Pengetahuan tentang model dan pendekatan pembelajaran menunjukkan perubahan terbesar ($\langle g \rangle = 0,44$; $p < 0,001$), diikuti oleh konsep pembelajaran mendalam ($\langle g \rangle = 0,40$; $p = 0,036$) dan perencanaan kurikulum ($\langle g \rangle = 0,32$; $p = 0,008$). Skor capaian pembelajaran PJOK dan, yang paling menonjol, asesmen otentik tidak berubah secara reliabel ($\langle g \rangle = 0,19$; $p = 0,780$), padahal asesmen otentik menempati porsi materi yang cukup besar. Pada tataran butir, uji McNemar menunjukkan perubahan pada beberapa butir, terbesar pada butir urutan perencanaan kurikulum (13 naik berbanding 1 turun; $p = 0,002$) dan butir definisi pendekatan pembelajaran (15 naik berbanding 3 turun; $p = 0,008$). Namun, tidak ada butir yang bertahan setelah koreksi Holm pada 29 perbandingan (p Holm terendah = 0,053). Hasil per butir, karena itu, dilaporkan sebagai temuan deskriptif, bukan konfirmatori (**Tabel 3**).

Tabel 3. Butir dengan Perubahan Terbesar

Isi butir	p pre	p post	p	p Holm
Urutan perencanaan kurikulum	0,768	0,895	0,002	0,053
Definisi pendekatan pembelajaran	0,779	0,905	0,008	0,211
Model berbasis pengalaman langsung	0,779	0,884	0,021	0,553
Alur Tujuan Pembelajaran	0,821	0,916	0,049	1,000
Penekanan model TGfU	0,863	0,947	0,021	0,553

Sumber: Penelitian (2026)

Analisis pilihan jawaban pada butir asesmen menunjukkan bahwa ketepatan sudah mendekati sempurna sebelum pelatihan pada butir konseptual: tujuan asesmen otentik (98,9% benar pada kedua pengukuran), prinsip asesmen dalam pembelajaran mendalam (96,8% menjadi 98,9%), dan tujuan asesmen formatif (93,7% pada kedua pengukuran). Satu perbedaan justru bergerak ke arah sebaliknya. Pada butir mengenai tujuan asesmen sumatif, proporsi yang memilih distraktor asesmen formatif (*“memperbaiki pembelajaran saat proses berlangsung”*) meningkat dari 8,4% menjadi 14,7%, sementara persentase jawaban benar turun dari 84,2% menjadi 82,1%.

Rencana Penerapan yang Diartikulasikan Guru

Analisis isi atas butir yang meminta peserta memaparkan satu rencana nyata penerapan asesmen otentik dan pembelajaran mendalam menghasilkan rubrik kedalaman dengan 4 level (**Tabel 4**).

Tabel 4. Kedalaman rencana penerapan yang diartikulasikan ($n = 95$)

Level	Deskripsi	n	%
0	Tidak responsif	6	6,3
1	Niat generik tanpa detail operasional	37	38,9
2	Metode disebut dengan elaborasi terbatas	32	33,7
3	Rencana terelaborasi	20	21,1

Sumber: Penelitian (2026)

Hampir separuh sampel (45%) memberikan jawaban pada level 0 atau 1. Jawaban level 1 yang khas menyebut suatu aktivitas tanpa memerinci materi, mode asesmen, atau tahapan, misalnya “*Memodifikasi permainan dan pembelajaran*” (P05) dan “*Membuat rubrik penilaian*” (P33). Sebaliknya, jawaban level 3 memadukan materi, model pembelajaran, mode asesmen, dan refleksi, seperti terlihat pada kutipan berikut.

“Saya akan menerapkan pembelajaran berbasis permainan (TGfU) dan Sport Education Model (SEM). Murid tidak hanya mempraktikkan keterampilan, tetapi juga berperan sebagai pemain, wasit, pelatih, dan pencatat skor. Penilaian dilakukan melalui observasi, rubrik keterampilan, kerja sama, sportivitas, serta refleksi diri.”
 (P63)

Sepanjang korpus, pembelajaran berbasis proyek paling sering diajukan ($n = 28$; 29,5%), diikuti oleh refleksi ($n = 27$; 28,4%), asesmen kinerja berbasis rubrik ($n = 23$; 24,2%), serta permainan yang dimodifikasi atau turnamen mini ($n = 18$; 18,9%). Hanya 31 peserta (32,6%) menyebut mode asesmen konkret dalam bentuk apa pun, dan hanya enam peserta (6,3%) menyinggung asesmen diagnostik atau formatif, yakni mekanisme yang paling membedakan asesmen otentik dari pengujian konvensional. Kasus negatif perlu dilaporkan. Enam peserta (6,3%) mengajukan rencana yang justru mempertahankan paradigma tes keterampilan dan kebugaran terisolasi yang hendak digeser oleh pelatihan, misalnya “*Tes passing bawah dalam permainan bola voli selama 30 detik atau 1 menit*” (P43) dan “*Penilaian kebugaran jasmani setelah libur sekolah*” (P86). Jawaban atas butir perencanaan kurikulum menunjukkan pola yang sejajar. Sebanyak 42 peserta (44,2%) menyebut CP, TP, dan ATP sekaligus serta mampu mereproduksi urutannya, tetapi hanya 19 peserta (20,0%) mengartikulasikan prinsip keselarasan di antara ketiganya, sebagaimana tampak pada jawaban: “*CP bersifat given disusun oleh pemerintah, guru menganalisis CP dan menurunkannya menjadi TP, kumpulan TP menjadi ATP, dari ATP tersebut kita menyusun modul ajar.*” (P08)

Integrasi Temuan Kuantitatif dan Kualitatif

Tabel 5 menyajikan *joint display* yang menghubungkan kedalaman rencana dengan ukuran kuantitatif. Kedalaman rencana berhubungan positif dengan pengetahuan pascatest ($p = 0,255$; $p = 0,013$; Kruskal-Wallis $H = 8,38$; $p = 0,039$) tetapi tidak berhubungan dengan gain ($p = -0,005$; $p = 0,961$). Artinya, kemampuan merancang pembelajaran mendalam berkaitan dengan tingkat pengetahuan yang dimiliki peserta, bukan dengan seberapa banyak pengetahuan yang bertambah melalui pelatihan.

Tabel 5. Kedalaman Rencana Berdasarkan Ukuran Kuantitatif

Kedalaman rencana	n	pre	post	gain	Domain asesmen (post)	Rerata kata
0 Tidak responsif	6	26,50	26,67	+0,17	0,92	10
1 Niat generik	37	24,78	26,11	+1,32	0,90	12
2 Metode disebut	32	26,91	27,84	+0,94	0,96	42
3 Rencana terelaborasi	20	26,45	27,50	+1,05	0,96	151

Sumber: Penelitian (2026)

Catatan. Spearman ρ (kedalaman, pascatest) = 0,255; $p = 0,013$. ρ (kedalaman, gain) = -0,005; $p = 0,961$. Kruskal-Wallis $H = 8,38$; $p = 0,039$.

Discussion

Mengapa Sebagian Domain Bergerak dan Sebagian Tidak

Pola perubahan antardomain dapat dibaca melalui mekanisme yang, dalam literatur mutakhir, dinyatakan sebagai penentu efektivitas pelatihan: keterlibatan berkelanjutan, praktik reflektif, dan relevansi kontekstual. Kegiatan ini kuat dalam relevansi kontekstual dan partisipasi kolektif, tetapi lemah dalam kedua aspek tersebut. Satu sesi sinkron tidak menyediakan ruang bagi praktik berulang, umpan balik, dan uji coba di kelas yang dalam literatur dikaitkan dengan perubahan substansial. Tinjauan sistematis yang membandingkan penyampaian tatap muka dan daring pada

pelatihan guru menyimpulkan bahwa kedua moda sama efektifnya dan bahwa yang menentukan adalah faktor psikososial serta terbentuknya komunitas komunikasi, bukan modanya (Li *et al.*, 2023).

Dalam sebuah penelitian dijelaskan juga bahwa pelatihan berdurasi pendek hanya menghasilkan peningkatan efikasi diri guru yang sederhana (Makopoulou *et al.*, 2021). Meta-analisis atas 102 studi menunjukkan efek sedang dari pelatihan daring pada tataran guru (Hedges' $g = 0,71$), sehingga medium daring itu sendiri bukanlah pembatas (Morina *et al.*, 2025); penelitian dalam bidang pendidikan jasmani juga menunjukkan bahwa format daring dapat menopang komunitas praktik bila dirancang untuk itu (Ferreira *et al.*, 2023; Lander *et al.*, 2022). Sebaliknya, pengalaman guru pendidikan jasmani dengan pembelajaran daring merekam keterbatasan yang nyata pada mata pelajaran yang bertumpu pada gerak (González-Calvo *et al.*, 2021; Mercier *et al.*, 2021), sehingga yang membatasi kegiatan ini lebih masuk akal diletakkan pada durasi dan ketiadaan kesempatan berlatih daripada pada mediumnya. Domain yang bergerak juga merupakan domain yang paling mudah dipindahkan. Pengetahuan bahwa suatu model ada, apa namanya, dan dalam urutan apa dokumen kurikulum diturunkan dapat disampaikan melalui presentasi.

Perubahan terbesar justru terletak pada urutan perencanaan, dari capaian pembelajaran hingga modul ajar, yaitu rantai prosedural yang dapat dihafal. Data kualitatif menunjukkan, bagaimanapun, bahwa pengetahuan prosedural ini tidak disertai dengan pemahaman konseptual. Sebanyak 44,2% peserta mampu menyebutkan dan mengurutkan ketiga komponen kurikulum, tetapi hanya 20,0% yang mampu mengartikulasikan prinsip keselarasan yang menghubungkan ketiganya. Peserta tampaknya mempelajari urutannya tanpa serta-merta memahami mengapa urutan itu penting. Temuan ini sejalan dengan laporan guru dalam konteks reformasi kurikulum bahwa yang paling menopang mereka adalah strategi yang dimodelkan secara langsung, bukan informasi yang disampaikan (Haug & Mork, 2021), dan bahwa pemodelan yang tepat tidak dapat disediakan dalam satu sesi presentasi.

Kesenjangan antara Pengetahuan Asesmen dan Kemampuan Merancangnya

Temuan yang paling bermakna secara substantif adalah perbedaan antara pengetahuan tentang asesmen yang terukur dan praktik asesmen yang diartikulasikan. Kinerja pilihan ganda pada domain asesmen otentik sudah mendekati langit-langit sebelum pelatihan (0,916) dan tidak berubah ($g = 0,19$; $p = 0,780$). Guru sudah mengetahui, dan tetap mengetahui, bahwa asesmen otentik bertujuan untuk menilai kemampuan nyata dalam konteks sebenarnya (98,9% benar pada kedua pengukuran). Namun, kurang dari sepertiga mampu menyebut mode asesmen konkret dalam rencananya sendiri, dan hanya 6,3% yang menyinggung asesmen formatif atau diagnostik. Konvergensi kedua sumber data ini menempatkan kesulitannya secara presisi.

Persoalannya bukan bahwa guru memiliki miskonsepsi tentang apa itu asesmen otentik, melainkan bahwa menyetujui prinsipnya tidak membekali mereka untuk mengoperasionalkannya. Pembacaan ini konsisten dengan teori yang dipaparkan dalam kajian literatur: literasi asesmen berkembang dalam kondisi pendukung tertentu dan melalui keanggotaan dalam komunitas praktik, bukan melalui pemaparan informasi (Barrientos Hernán *et al.*, 2023); persoalan asesmen dalam pendidikan jasmani tidak terselesaikan melalui penyampaian informasi semata (Barker *et al.*, 2025); dan perpindahan dari merancang menuju melaksanakan asesmen formatif menuntut dukungan berkelanjutan (Slingerland *et al.*, 2024). Kebutuhan pelatihan dan sumber daya yang memadai merupakan hambatan penerapan asesmen otentik yang paling sering muncul di berbagai bidang (Vlachopoulos & Makri, 2024).

Data tersebut menambahkan satu pengamatan spesifik: tes pengetahuan dapat melebihi-lebihkan kemampuan asesmen secara signifikan. Seandainya evaluasi hanya bersandar pada instrumen pilihan ganda, kesimpulannya adalah bahwa pengetahuan asesmen tinggi dan stabil, sehingga domain ini tidak memerlukan perhatian lebih lanjut. Data terbuka membalikkan kesimpulan tersebut. Hal ini berimplikasi metodologis bagi evaluasi kegiatan pengabdian yang menasar guru pendidikan jasmani, di mana tes pengetahuan lazim digunakan karena praktis sekaligus memperkuat catatan bahwa hanya sedikit kajian yang menghubungkan praktik asesmen dengan hasil belajar yang sesungguhnya (Killian & Woods, 2021).

Keunggulan, Kelemahan, dan Peluang Pengembangan Kegiatan

Keunggulan kegiatan ini terletak pada jangkauan dan kesesuaiannya dengan kondisi mitra: satu sesi daring menjangkau guru di setidaknya lima provinsi tanpa biaya perjalanan, dan materinya selaras dengan tenggat kebijakan yang sedang dihadapi peserta. Kelemahannya terletak pada ketiadaan ruang praktik: luaran yang dihasilkan peserta berupa rencana tertulis, bukan perangkat ajar yang telah diuji di kelas. Tingkat kesulitan pelaksanaan tergolong rendah dari sisi teknis, namun tinggi dari sisi penjaminan mutu, karena verifikasi kepesertaan dan pencocokan data hanya dapat dilakukan pada nama yang sudah dinormalisasi. Peserta memasuki pelatihan dengan pengetahuan awal yang tinggi ($M = 25,96$ dari 29; 26% sudah berada pada tingkat langit-langit).

Hal tersebut informatif, sebab menunjukkan bahwa diseminasi kebijakan Pembelajaran Mendalam telah menjangkau komunitas profesi ini dengan cukup baik pada tataran istilah dan prinsip-prinsip umum. Implikasi praktisnya, penyadaran lanjutan kemungkinan bukan kendala utama, dan penyelenggaraan yang terdiferensiasi lebih menjanjikan daripada materi yang seragam. Perlu ditegaskan bahwa ketika skor awal diukur dengan galat, peserta dengan skor terendah pada pretest secara statistik memang diharapkan bergeser mendekati rerata pada pascatest, terlepas dari adanya atau tidaknya intervensi. Peluang pengembangan ke depan terletak pada perpanjangan kegiatan menjadi rangkaian berjenjang, dengan pendampingan pascasesi dan penilaian terhadap perangkat ajar yang benar-benar digunakan di sekolah.

Keterbatasan

Beberapa keterbatasan membatasi penafsiran. Desain satu kelompok pra-test-pasca-test tidak mendukung klaim kausal; tanpa kelompok pembandingan, perubahan yang teramati tidak dapat diatribusikan pada intervensi dengan yakin, dan efek pengujian menjadi perhatian khusus karena butir yang sama diberikan dua kali pada hari yang sama. Peserta yang berpasangan memperoleh skor lebih tinggi pada pretest dibandingkan dengan yang tidak berpasangan ($p = 0,035$), sehingga temuan ini berlaku bagi guru yang mengikuti kedua sesi. Meskipun peserta berasal dari beberapa provinsi, sampel bersifat sukarela dan komposisi geografisnya tidak dapat ditetapkan secara sistematis, sehingga tidak dapat dianggap representatif secara nasional.

Instrumen menunjukkan langit-langit yang nyata dan belum divalidasi secara independen. Data pascatest dikumpulkan segera setelah sesi, sehingga data retensi tidak tersedia, dan data kualitatif merekam niat yang dinyatakan, bukan praktik yang dilakukan. Tujuh peserta dikeluarkan karena jawaban terbukanya berindikasi dibantu kecerdasan buatan generatif; deteksi ini bersifat indikatif, dan kemungkinan masih ada jawaban lain yang dibantu tanpa menampilkan penanda yang terdeteksi. Persoalan terakhir ini merupakan ancaman validitas yang relatif baru bagi evaluasi pelatihan daring.

CONCLUSION

Kegiatan pengabdian ini menetapkan tiga hal yang secara langsung menjawab ketiga tujuan yang dinyatakan pada abstrak dan di akhir pendahuluan. Pelatihan daring satu sesi mengubah pengetahuan guru secara signifikan, tetapi dengan magnitudo kecil, dan hanya enam peserta yang melampaui ambang perubahan yang dapat diandalkan secara individual. Signifikansi pada tataran kelompok tidak dapat dibaca sebagai perubahan yang berarti pada tataran individu. Perubahan tidak merata antar domain. Pengetahuan bergeser ke konten yang mudah dipindahkan, seperti nama model pembelajaran, kosakata pembelajaran mendalam, dan urutan penurunan dokumen kurikulum, tetapi tidak bergeser ke asesmen otentik. Asesmen otentik merupakan domain yang paling resisten, meskipun menempati porsi materi yang cukup besar.

Rencana penerapan yang diartikulasikan guru memperlihatkan kesenjangan yang lebih tajam daripada yang tersampaikan dalam tes: hanya 32,6% yang menyebut mode asesmen konkret dan 6,3% yang menyinggung asesmen formatif atau diagnostik, sementara kedalaman rencana lebih berhubungan dengan tingkat pengetahuan yang dimiliki, bukan dengan penambahan pengetahuan dari pelatihan. Pada kegiatan ini, kesenjangan antara apa yang diukur dalam evaluasi pelatihan dan apa yang benar-benar dapat dirancang oleh guru cukup besar sehingga dapat membalikkan kesimpulan yang akan dihasilkan oleh tes pengetahuan semata. Pelatihan daring singkat tampak sesuai untuk memperkenalkan kosakata, model, dan urutan perencanaan kepada tenaga guru yang tersebar, tetapi kurang sesuai untuk mengembangkan kemampuan merancang asesmen.

Program yang harus mengandalkan format singkat sebaiknya memperlakukannya sebagai langkah pembuka dari rangkaian yang lebih panjang, bukan sebagai intervensi utuh, serta membedakan layanan bagi peserta yang telah menguasai konsep yang diperkenalkan. Bagi praktik evaluasi pengabdian, kegiatan ini menyarankan penyesuaian yang sederhana namun berkonsekuensi: memasang ukuran pengetahuan dengan tugas perancangan singkat yang berbiaya rendah dan mampu menyingkap kesenjangan kemampuan yang tersembunyi di balik tes. Kegiatan lanjutan sebaiknya menambahkan pendampingan pasca-sesi serta pengambilan data retensi agar dapat dinilai apakah guru benar-benar mengubah praktiknya di lapangan dan di gelanggang sekolah.

AUTHOR'S NOTE

Penulis menyatakan bahwa tidak ada konflik kepentingan terkait publikasi artikel ini. Penulis menegaskan bahwa data dan isi artikel bebas dari plagiarisme. Penulis mengucapkan terima kasih kepada IGORNAS Jawa Barat atas kemitraan dalam penyelenggaraan kegiatan, serta kepada seluruh guru peserta yang bersedia mengerjakan kedua instrumen evaluasi.

REFERENCES

- Arufe-Giráldez, V., Sanmiguel-Rodríguez, A., Ramos-Álvarez, O., & Navarro-Patón, R. (2023). News of the pedagogical models in physical education—a quick review. *International Journal of Environmental Research and Public Health*, 20(3), 1-22.
- Barker, D., Tolgfors, B., & Caldeborg, A. (2025). Why assessment in physical education is still problematic: A critical interpretive synthesis of physical education assessment literature. *European Physical Education Review*, 32(3), 527-546.
- Barrientos Hernán, E. J., López-Pastor, V. M., Lorente-Catalán, E., & Kirk, D. (2023). Challenges with using formative and authentic assessment in physical education teaching from experienced teachers' perspectives. *Curriculum Studies in Health and Physical Education*, 14(2), 109-126.
- Beni, S., Fletcher, T., & Ní Chróinín, D. (2022). Teachers' engagement with professional development to support implementation of meaningful physical education. *Journal of Teaching in Physical Education*, 41(4), 570-579.
- Beni, S., Ní Chróinín, D., Fletcher, T., Bailey, J., Cariño Fraise, L., Down, M., Hamada, M., Riddick, T., Trojanovic, M., & Gross, K. (2023). Teachers' sensemaking in implementation of meaningful physical education. *Physical Education and Sport Pedagogy*, 30(6), 593-606.
- Blampied, N. M. (2022). Reliable change and the reliable change index: Still useful after all these years?. *The Cognitive Behaviour Therapist*, 15(1), 1-18.
- Christman, E., Miller, P., & Stewart, J. (2024). Beyond normalized gain: Improved comparison of physics educational outcomes. *Physical Review Physics Education Research*, 20(1), 1-17.

- Coulston, F., Lynch, F., & Vears, D. F. (2025). Collaborative coding in inductive content analysis: Why, when, and how to do it. *Journal of Genetic Counseling*, 3(4), 1-12.
- Dille, K. B., & Røkenes, F. M. (2021). Teachers' professional development in formal online communities: A scoping review. *Teaching and Teacher Education*, 105(1), 1-17.
- Fernández-Río, J., & Iglesias, D. (2022). What do we know about pedagogical models in physical education so far? An umbrella review. *Physical Education and Sport Pedagogy*, 29(2), 190-205.
- Ferreira, H. J., Gonçalves, L., & Parker, M. (2023). Physical education teachers' experiences of nurturing a community of practice online. *Journal of Teaching in Physical Education*, 42(3), 420-429.
- Fletcher, T., & Ní Chróinín, D. (2022). Pedagogical principles that support the prioritisation of meaningful experiences in physical education: Conceptual and practical considerations. *Physical Education and Sport Pedagogy*, 27(5), 455-466.
- González-Calvo, G., Barba-Martín, R. A., Bores-García, D., & Hortigüela-Alcalá, D. (2021). The (virtual) teaching of physical education in times of pandemic. *European Physical Education Review*, 28(1), 205-224.
- Hardiyono, B., Pratama, R., Hidayat, A., & Hartati, H. (2021). Physical education teachers' understanding of the 2013 curriculum. *International Journal of Human Movement and Sports Sciences*, 9(5), 841-846.
- Haug, B. S., & Mork, S. M. (2021). Taking 21st century skills from vision to classroom: What teachers highlight as supportive professional development in the light of new demands from educational reforms. *Teaching and Teacher Education*, 100(1), 1-12.
- Hennessy, S., D'Angelo, S., McIntyre, N., Koomar, S., Kreimeia, A., Cao, L., Brugha, M., & Zubairi, A. (2022). Technology use for teacher professional development in low- and middle-income countries: A systematic review. *Computers and Education*, 3(1), 1-32.
- Ilmawati, H., Julianline, T., Mahendra, A., Sucipto, S., & Gustian, U. (2026). Reconceptualizing continuing professional development in physical education: A systematic review of implementation, mechanisms, and pedagogical transformation. *Physical Education Theory and Methodology*, 26(3), 429-441.
- Killian, C. M., & Woods, A. M. (2021). Assessment practices in K-12 physical education in the United States: A scoping review of research, 2000-2020. *Research Quarterly for Exercise and Sport*, 92(2), 248-258.
- Lander, N., Lewis, S., Nahavandi, D., Amsbury, K., & Barnett, L. M. (2022). Teacher perspectives of online continuing professional development in physical education. *Sport, Education and Society*, 27(4), 434-448.
- Li, Z., Che Hassan, N., & Ab Jalil, H. (2023). The effectiveness of face-to-face versus online delivery of continuing professional development for science teachers: A systematic review. *Education Sciences*, 13(12), 1-17.
- Ma'mun, A., Tinaz, C., Anira, A., Syarifatunnisa, S., Hertem, Ö. O., Mahendra, A., & Julianline, T. (2025). Physical education and school sport in emerging nations: A comparison of Indonesia and Türkiye. *Frontiers in Sports and Active Living*, 7(1), 1-11.
- Makopoulou, K., Neville, R. D., Ntoumanis, N., & Thomas, G. (2021). An investigation into the effects of short-course professional development on teachers' and teaching assistants' self-efficacy. *Professional Development in Education*, 47(5), 780-795.
- Mercier, K., Centeio, E., Garn, A., Erwin, H., Marttinen, R., & Foley, J. (2021). Physical education teachers' experiences with remote instruction during the initial phase of the COVID-19 pandemic. *Journal of Teaching in Physical Education*, 40(2), 337-342.

- Morina, F., Fütterer, T., Hübner, N., Zitzmann, S., & Fischer, C. (2025). Effects of online teacher professional development on teacher-, classroom-, and student-level outcomes: A meta-analysis. *Computers & Education*, 228(1), 1-23.
- Mu'arifin, M., & Narmaditya, B. S. (2022). Professional development program for physical education teachers in Indonesia. *Jurnal Cakrawala Pendidikan*, 41(3), 665-675.
- Nopembri, S., Saryono, S., Muktiani, N. R., Listyarinni, A. E., & Shahril, M. I. (2022). Digital technology in physical education distance learning during pandemic: Teachers' perspective. *Jurnal Keolahragaan*, 10(1), 71-82.
- Parker, M., Patton, K., Gonçalves, L., Luguetti, C., & Lee, O. (2022). Learning communities and physical education professional development: A scoping review. *European Physical Education Review*, 28(2), 500-518.
- Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71(1), 1-15.
- Slingerland, M., Weeldenburg, G., & Borghouts, L. (2024). Formative assessment in physical education: teachers' experiences when designing and implementing formative assessment activities. *European Physical Education Review*, 30(4), 620-637.
- Suroto, S., Prakoso, B. B., Priambodo, A., Khory, F. D., & Dinata, V. C. (2022). Web-based instruments for national competences of physical education teachers. *Jurnal Cakrawala Pendidikan*, 41(3), 742-753.
- Vears, D. F., & Gillam, L. (2022). Inductive content analysis: A guide for beginning qualitative researchers. *Focus on Health Professional Education: A Multi-Professional Journal*, 23(1), 111-127.
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83(1), 1-13.
- Westphale, S., Backhaus, J., & Koenig, S. (2021). Quantifying teaching quality in medical education: the impact of learning gain calculation. *Medical Education*, 56(3), 312-320.