Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri,
ISSN Cetak : 2477-5673 ISSN Online : 2614-722X
Volume 12 Nomor 04, Desember 2026

PENERAPAN ASESMEN FORMATIF INDIVIDUAL BERBASIS HTML *OFFLINE* UNTUK MEMETAKAN POLA PILIHAN JAWABAN KELIRU MURID PPKn: PENELITIAN TINDAKAN KELAS

Ahmad Arif[1], Mila Nirmala Sari Hasibuan[2]
[1,2]Pendidikan Profesi Guru, Universitas Labuhanbatu
([1]arifahmad4109@gmail.com), ([2]milanirmalasari7@gmail.com)

ABSTRACT

Group quizzes in Pancasila and Civic Education (PPKn) offered limited evidence of individual achievement because some students dominated answer selection. This descriptive classroom action research examines an offline HTML-based individual formative assessment in Cycle I (version 2) and Cycle II (revised version) in class XI-8 at SMA Negeri 1 Rantau Selatan. Cycle I (version 2) yielded final scores for 30 students (mean = 18.1/20), but not scores by attempt. In Cycle II (revised version), 36 students completed three attempts using the same 20 case-based items. An audit of 108 response files confirmed 20 responses per file and agreement between stored and recalculated scores. Mean scores were 13.44 (SD = 4.00), 14.14 (SD = 3.26), and 15.47 (SD = 3.59); the first-to-third attempt difference was 2.03 points (95% CI [0.60, 3.45]). First-attempt item difficulty ranged from 0.47 to 0.83, KR-20 was 0.78, and all distractors were selected at least once. Categories M2 and M6 were most frequent, while M1 had the highest frequency relative to its number of distractors. Because students did not complete the optional reasoning field, M1–M6 indicate response patterns, not confirmed misconceptions. Repeated exposure to identical items and immediate feedback may have contributed to score changes; therefore, the increase cannot be interpreted as proof of conceptual learning or intervention effectiveness.

Keywords: formative assessment, offline HTML, response patterns, Pancasila and Civic Education

ABSTRAK

Kuis kelompok dalam pembelajaran PPKn belum menggambarkan capaian individual karena sebagian anggota lebih dominan menentukan jawaban. PTK deskriptif ini mengkaji penerapan asesmen formatif individual berbasis HTML *offline* pada Siklus I (versi 2) dan Siklus II (versi revisi) di kelas XI-8 SMA Negeri 1 Rantau Selatan, perubahan skor dalam tiga percobaan, serta karakteristik butir dan pengecoh pada percobaan pertama. Pada Siklus I (versi 2), tersedia skor akhir 30 murid (rata-rata 18,1/20), tetapi skor tiap percobaan tidak tercatat. Pada Siklus II (versi revisi), 36 murid mengerjakan tiga percobaan dengan 20 butir kasus yang sama. Audit 108 berkas memastikan setiap berkas memuat 20 respons dan skor tersimpan sesuai penghitungan ulang. Rata-rata P1, P2, dan P3 berturut-turut 13,44 (SD = 4,00), 14,14 (SD = 3,26), dan 15,47 (SD = 3,59); selisih P1–P3 sebesar 2,03 poin (IK 95% [0,60–3,45]). Indeks kesukaran P1 berkisar 0,47–0,83, KR-20 sebesar 0,78, dan seluruh pengecoh terpilih setidaknya sekali. M2 dan M6 paling sering dipilih, sedangkan M1 memiliki rasio pilihan tertinggi setelah memperhitungkan jumlah pengecohnya. Kolom alasan tidak diisi sehingga M1–M6 hanya menunjukkan pola respons, bukan miskonsepsi terkonfirmasi. Pengulangan butir identik dan umpan balik langsung mungkin turut memengaruhi kenaikan skor; hasil

ini tidak membuktikan peningkatan pemahaman konseptual atau efektivitas tindakan.

Kata Kunci: asesmen formatif, HTML *offline*, pola pilihan jawaban, PPKn

A. Pendahuluan

Dalam Program Pengalaman Lapangan di kelas XI-8 SMA Negeri 1 Rantau Selatan, pembelajaran PPKn dilaksanakan melalui beberapa tahap, mulai dari ceramah dan hafalan, *Gallery Walk*, hingga kuis HTML *offline* berkelompok sebagai kondisi awal (tiga kelompok, masing-masing 12 murid). Pengamatan pada kuis kelompok menunjukkan satu hingga tiga murid per kelompok cenderung menentukan jawaban, sementara anggota lain mengikuti. Skor kelompok karena itu belum menggambarkan penguasaan individual; kontribusi tiap murid juga tidak dicatat sehingga dominasi tersebut tidak dapat diukur. Dalam refleksi informal, murid mengusulkan asesmen individual dengan soal lebih menantang. Selain itu, berdasarkan pengamatan penulis, sekitar 40% murid mengalami kesulitan mengakses asesmen daring karena keterbatasan jaringan atau perangkat. Kedua kondisi tersebut menjadi dasar pengembangan asesmen individual yang dapat digunakan tanpa koneksi internet.

Asesmen formatif menyediakan informasi tentang perkembangan belajar untuk menentukan tindak lanjut, bukan sekadar menghasilkan nilai. Umpan balik membantu murid mengenali kesenjangan capaian dan tujuan pembelajaran serta membantu guru merancang langkah berikutnya (Black & Wiliam, 1998; Hattie & Timperley, 2007). Dalam PPKn, asesmen perlu mengukur penalaran tentang hubungan nilai dasar Pancasila, penjabarannya dalam peraturan, dan penerapannya pada kasus, bukan hanya hafalan. Praktik asesmen formatif Pendidikan Pancasila juga masih perlu diperkuat melalui umpan balik yang mendorong refleksi murid (Zai et al., 2026).

Pada Siklus I (versi 2), tindakan diwujudkan melalui asesmen formatif individual berbasis HTML *offline*. Instrumen memuat soal kasus pada jenjang kognitif C4 dan C5, umpan balik langsung untuk setiap pilihan jawaban, serta pengelompokan pengecoh dengan kode M1–M6.

Berkas HTML berukuran sekitar 40 KB dapat diunduh satu kali dan digunakan tanpa koneksi internet.

Asesmen digital telah dikembangkan, antara lain melalui Wordwall di sekolah dasar (Akbar et al., 2025) dan untuk belajar mandiri pada materi statistika SMA (Adi et al., 2026). Namun, platform daring bergantung pada jaringan dan perangkat, dan penekanan berlebihan pada kecepatan dapat membatasi penalaran konseptual (Nurulfathimah et al., 2026). Identifikasi miskonsepsi memerlukan informasi tambahan, seperti alasan jawaban atau tingkat keyakinan (Kaningtyas et al., 2025; Nisa et al., 2024; Sumarna et al., 2026). Berdasarkan pustaka yang ditelaah, kajian asesmen formatif Pendidikan Pancasila masih terbatas. Penelitian ini mengisi ruang tersebut melalui penerapan asesmen individual *offline* dan analisis pola pilihan jawaban keliru, dengan interpretasi yang dibatasi karena alasan jawaban tidak tersedia.

Penelitian ini menjawab tiga pertanyaan: (1) bagaimana asesmen formatif individual berbasis HTML *offline* diterapkan pada Siklus I (versi 2) dan Siklus II (versi revisi); (2) bagaimana pola perubahan skor murid dalam tiga percobaan pada Siklus II (versi revisi); dan (3) bagaimana kualitas butir serta distribusi pola pilihan jawaban keliru pada percobaan pertama Siklus II (versi revisi). Istilah pola pilihan jawaban keliru merujuk pada kecenderungan murid memilih pengecoh yang dikelompokkan ke dalam kode M1–M6. Kode tersebut merupakan kategori pilihan yang dirancang dalam instrumen, bukan bukti langsung adanya miskonsepsi tertentu pada murid.

Secara praktis, penelitian ini menawarkan contoh asesmen individual *offline* dan pemanfaatan hasilnya untuk merancang tindak lanjut. Secara akademik, penelitian ini menyajikan data penerapan asesmen formatif digital dalam PPKn beserta batas interpretasinya.

B. Metode Penelitian

Desain dan subjek penelitian

Penelitian ini merupakan PTK deskriptif dengan model Kemmis dan McTaggart (Kemmis et al., 2014; McTaggart, 1988), meliputi perencanaan, tindakan, observasi, dan refleksi. Pendekatan deskriptif digunakan untuk memaparkan pelaksanaan dan data, bukan menguji

efektivitas melalui kelas pembanding atau inferensi kausal.

Siklus I (versi 2) dan Siklus II (versi revisi) merupakan dua tahap tindakan individual yang dianalisis. Tahap pembelajaran sebelumnya meliputi ceramah dan hafalan serta *Gallery Walk*, sedangkan kuis HTML *offline* berkelompok digunakan sebagai kondisi awal. Pada kondisi awal tersebut, dominasi satu sampai tiga murid per kelompok tercatat secara naratif, dan murid meminta asesmen individual dengan soal lebih menantang. Pada Siklus I (versi 2), yang dilaksanakan pada 16 Oktober 2026, tindakan berupa asesmen individual HTML *offline* versi 2; data yang tersedia hanya skor akhir sehingga refleksinya menuntut pencatatan percobaan, batas waktu, dan pengacakan. Pada Siklus II (versi revisi), yang dilaksanakan pada 23 Oktober 2026, versi revisi diterapkan dengan tiga percobaan dalam dua jam pelajaran. Refleksi tahap ini mengarah pada kebutuhan bank soal paralel, alasan wajib, analisis tiap percobaan, dan telaah ahli ulang.

Subjek penelitian adalah 36 murid kelas XI-8. Pada Siklus I (versi 2), 30 murid hadir dan enam tidak hadir karena sakit. Pada Siklus II (versi revisi), seluruh 36 murid mengikuti tiga percobaan pada 23 Oktober 2026.

Tindakan pada Siklus I (versi 2) dan Siklus II (versi revisi)

Tindakan menggunakan satu berkas HTML berisi 20 soal pilihan ganda berbasis kasus pada materi Unit 2 buku PPKn kelas XI, disusun mengacu pada modul soal keterampilan berpikir tingkat tinggi (Direktorat Pembinaan Sekolah Menengah Atas, 2019). Butir mencakup Sila 1 (3), Sila 2 (5), Sila 3 (2), Sila 4 (4), Sila 5 (4), nilai dasar (1), dan nilai Pancasila (1), dengan 11 butir C4 dan 9 butir C5. Kunci tersebar seimbang (masing-masing lima butir A–D).

Setiap pilihan memunculkan umpan balik dan penjelasan. Pengecoh dikelompokkan ke enam kode: M1 nilai dasar dianggap dapat diganti (3 pengecoh); M2 tataran nilai tertukar (13); M3 hak dianggap mutlak tanpa kewajiban (6); M4 persoalan dianggap bukan tanggung jawab negara (11); M5 solusi mengorbankan hak atau tidak menyentuh akar masalah (12); dan M6 kasus dihubungkan dengan pasal atau sila yang tidak relevan (15). Kolom alasan bersifat opsional. Halaman hasil

memberi rekomendasi pengayaan (skor ≥16), penguatan (12–15), dan remedial (di bawah 12). Pengulangan merupakan kesempatan mengerjakan kembali soal dengan umpan balik, bukan intervensi remedial terpisah yang efektivitasnya diuji.

Versi 2 pada Siklus I (versi 2) belum membatasi waktu atau mengacak butir dan pilihan. Sistem menyimpan identitas, skor akhir, jawaban per butir, alasan, dan waktu pengiriman; pengulangan dipantau guru. Versi Siklus II (versi revisi) menambahkan batas 45 menit, pengacakan butir dan pilihan saat pengulangan (urutan P1 tetap), riwayat skor, persentase, durasi, dan status waktu tiap percobaan, serta menu rekap guru untuk membaca berkas JSON secara *offline*. Walau sistem tidak membatasi pengulangan, pelaksanaan kelas dibatasi tiga percobaan karena alokasi dua jam pelajaran; skor tertinggi dilaporkan sebagai nilai murid.

Telaah instrumen

Instrumen ditelaah oleh guru pamong PPKn dan dosen pembimbing lapangan (penulis kedua) menggunakan 11 pernyataan berskala 0–4 yang mencakup materi, konstruksi, bahasa, umpan balik, dan media HTML. Penelaahan pada 6 Oktober 2026 menghasilkan persentase dan indeks Aiken's V. Catatan pada dua lembar validasi mencakup aspek tampilan, pengawasan menyontek, rekap pengulangan, *timer*, dan pengacakan. Karena salah satu penelaah merupakan penulis kedua dan telaah dilakukan sebelum revisi Siklus II (versi revisi), hasilnya bukan validasi independen atas versi akhir aplikasi.

Pengumpulan dan analisis data

Data Siklus I (versi 2) berasal dari berkas hasil akhir, pemantauan guru atas pengulangan, catatan pengamatan pada kondisi awal berupa kuis kelompok, refleksi informal, dan dokumentasi video. Siklus II (versi revisi) menghasilkan 108 berkas JSON (36 murid × 3 percobaan). Setelah normalisasi koma penutup pada ekspor—tanpa mengubah jawaban atau skor—seluruh berkas terbaca. Ketiga percobaan memuat 36 identitas yang sama, setiap berkas berisi 20 respons, dan skor tersimpan cocok dengan penghitungan ulang pada 108 berkas. Analisis butir dan pengecoh menggunakan P1, sedangkan perubahan skor menggunakan ketiga percobaan.

Data dianalisis secara deskriptif melalui rata-rata, simpangan baku, median, rentang, dan distribusi skor. Perubahan antarpercobaan diringkas dengan selisih rata-rata berpasangan dan interval kepercayaan (IK) 95%, tanpa uji hipotesis dan tanpa tafsir kausal. Kualitas butir pada P1 (n = 36) dianalisis melalui indeks kesukaran, daya pembeda (kelompok atas dan bawah masing-masing 10 murid), korelasi butir-total terkoreksi, KR-20, dan fungsi pengecoh. Indeks kesukaran 0,31–0,70 dikategorikan sedang dan di atas 0,70 mudah (Arikunto, 2018); daya pembeda 0,20–0,39 cukup dan 0,40 ke atas baik; pengecoh dianggap berfungsi bila dipilih sekurang-kurangnya satu murid. Dominasi kelompok pada kondisi awal (kuis kelompok) tidak diukur kuantitatif sehingga pengurangannya tidak diuji.

Indikator analisis meliputi: (1) ketuntasan individual dengan skor minimal 16 (80%); (2) ketercapaian butir, yaitu sekurang-kurangnya 80% murid menjawab benar; dan (3) kualitas butir berdasarkan indeks kesukaran, daya pembeda, dan fungsi pengecoh. Identitas murid disamarkan. Penelitian dilaksanakan dengan sepengetahuan guru pamong dan pihak sekolah.

C. Hasil Penelitian dan Pembahasan

Kondisi awal

Pada kondisi awal (kuis kelompok), dominasi satu hingga tiga murid dalam setiap kelompok hanya tercatat secara naratif. Skor kelompok ditulis di papan tulis tanpa dokumentasi, sehingga tidak tersedia data kuantitatif yang dapat digunakan sebagai pembanding.

Hasil Siklus I (versi 2)

Pada Siklus I (versi 2), 30 murid memperoleh rata-rata skor akhir 18,1/20 (SD = 2,3; median = 19; rentang = 8–20); 28 murid (93,3%) mencapai minimal 16. Versi 2 tidak merekam skor tiap percobaan, sehingga skor tersebut merupakan hasil akhir setelah kesempatan mengulang, bukan skor percobaan pertama. Enam murid tidak hadir. Dengan demikian, hasil Siklus I (versi 2) hanya menggambarkan peserta yang hadir dan tidak dapat dibandingkan langsung dengan skor per percobaan Siklus II (versi revisi).

Telaah instrumen oleh ahli

Rata-rata penilaian guru pamong 3,73 dan dosen pembimbing 3,82;

rata-rata keseluruhan 3,77 (94,3%) dengan Aiken's V 0,94 (rentang 0,75–1,00). Aspek materi memperoleh V 0,97, konstruksi soal 0,96, bahasa 1,00, dan umpan balik serta media HTML terendah (rata-rata 3,50; V = 0,88). Catatan pada dua lembar validasi mencakup tampilan, pengawasan menyontek, rekap pengulangan, *timer*, dan pengacakan. Hasil ini merujuk pada instrumen sebelum revisi Siklus II (versi revisi), bukan validasi independen versi akhir.

Hasil Siklus II (versi revisi)

Rata-rata skor meningkat dari 13,44 pada P1 menjadi 14,14 pada P2 dan 15,47 pada P3 (Tabel 1). Murid yang mencapai minimal 16 berturut-turut berjumlah 12, 11, dan 19. Berdasarkan skor tertinggi dari tiga percobaan, 23 murid (63,9%) mencapai 16; 11 murid memperoleh 12–15 dan dua murid di bawah 12.

Tabel 1. Skor Asesmen pada Tiap Percobaan, Siklus II (versi revisi) (n = 36)

Ukuran	P1	P2	P3	Maks.
Rata-rata	13,4	14,1	15,5	16,5
SD	4,0	3,3	3,6	3,0
Rentang	6–19	5–20	10–20	10–20
Skor ≥16 (n)	12	11	19	23
Skor ≥16 (%)	33,3	30,6	52,8	63,9
Skor 12–15 (n)	13	22	9	11
Skor <12 (n)	11	3	8	2

Keterangan: Maks. = skor tertinggi dari tiga percobaan.

Selisih rata-rata P3–P1 berdasarkan skor belum dibulatkan adalah +2,03 poin (IK 95% 0,60–3,45). Selisih P2–P1 sebesar +0,69 (IK 95% −0,28 hingga 1,67), sedangkan P3–P2 +1,33 (IK 95% 0,37–2,30). Skor meningkat pada 23 murid (63,9%), tetap pada empat (11,1%), dan menurun pada sembilan (25%); empat murid turun sekurang-kurangnya tiga poin, dengan penurunan terbesar delapan poin. Kenaikan terpusat pada murid berskor awal rendah: dari 11 murid kategori remedial pada P1, rata-rata naik 5,4 poin dan lima murid mencapai skor 16 atau lebih pada P3, sedangkan 12 murid yang sudah mencapai 16 sejak P1 rata-rata turun 0,6 poin dan empat di antaranya tidak lagi mencapai ambang itu. Skor awal berkorelasi negatif dengan selisih skor (r = −0,62). Sebanyak 13 murid tidak pernah mencapai skor 16 pada ketiga percobaan.

Pada P1, waktu pengerjaan rata-rata 24,3 menit dari batas 45 menit (rentang 18,1–33,4); tidak ada murid kehabisan waktu atau membiarkan butir kosong. Kolom alasan tidak diisi. Indeks kesukaran 0,47–0,83 (rata-rata 0,67): 12 butir sedang dan delapan mudah; tiga butir dijawab benar oleh ≥80% murid. Proporsi benar pada

butir C5 lebih rendah daripada C4 (63,3% dan 70,5%). Proporsi terendah terdapat pada butir Sila 2 (57,8%) dan nilai dasar (58,3%). Butir tersulit ialah 7 (47,2%), 19 (50,0%), 8 (52,8%), 2 dan 20 (58,3%), serta 16 (61,1%). Daya pembeda 14 butir baik dan enam cukup; KR-20 0,78. Lima butir (8, 10, 11, 14, 15) memiliki korelasi butir-total terkoreksi di bawah 0,20.

Dari 720 respons P1, terdapat 236 jawaban keliru dan semuanya dapat dipetakan ke kode pengecoh (Tabel 2). M2 dan M6 paling sering muncul secara keseluruhan. Setelah jumlah pengecoh per kategori diperhitungkan, M1 memiliki rasio pilihan tertinggi (5,67); 11 dari 17 pilihannya muncul pada butir 2 dan enam pada butir 20.

Tabel 2. Frekuensi Kategori Pengecoh pada Percobaan Pertama, Siklus II (versi revisi) (236 jawaban keliru)

Kode	Pengecoh	Dipilih (%)	Rasio	% Siklus I (versi 2)
M2	13	56 (23,7)	4,31	30,4
M6	15	53 (22,5)	3,53	25,0
M4	11	48 (20,3)	4,36	16,1
M5	12	43 (18,2)	3,58	10,7
M3	6	19 (8,1)	3,17	7,1
M1	3	17 (7,2)	5,67	10,7
Total	60	236 (100)	–	100,0

Keterangan: Rasio = frekuensi dibagi jumlah pengecoh; % Siklus I (versi 2) = persentase kategori pada hasil akhir Siklus I (versi 2), berdasarkan 30 murid.

Hasil pengamatan terhadap indikator

Tabel 3 merangkum hasil pengamatan terhadap setiap indikator. Karena target ketuntasan klasikal tidak ditetapkan sebelum tindakan, penelitian ini tidak menyimpulkan keberhasilan klasikal. Skor 16 atau lebih teramati pada 63,9% murid (skor tertinggi), sedangkan 13 murid tidak pernah mencapainya. Lima butir dan tujuh pengecoh perlu ditinjau, dan perubahan skor dilaporkan secara deskriptif, bukan sebagai indikator tunggal efektivitas.

Tabel 3. Hasil Pengamatan terhadap Indikator pada Siklus I (versi 2) dan Siklus II (versi revisi)

Indikator	Siklus I (versi 2)	Siklus II (versi revisi)
Skor ≥16 (individual)	28/30 murid (93,3%)	P1 12; P3 19; tertinggi 23 dari 36 (33,3%; 52,8%; 63,9%)
Butir benar ≥80% murid	19/20 butir	3/20 butir (P1)
Kualitas butir (kesukaran 0,31–0,70; daya pembeda ≥0,20; pengecoh berfungsi)	Hanya hasil akhir: semua butir mudah; 25 pengecoh tidak dipilih	12/20 butir sedang; daya pembeda ≥0,20 pada 20 butir; 60 pengecoh dipilih (tujuh hanya sekali)
Perubahan skor P1–P3	Tidak tersedia	+2,03 poin (IK 95% 0,60–3,45)

Pembahasan

Asesmen individual menghasilkan informasi capaian per

murid yang tidak tersedia dari kuis kelompok kondisi awal (kuis kelompok). Temuan ini sejalan dengan fungsi asesmen formatif sebagai dasar bagi guru dan murid untuk menentukan langkah belajar berikutnya (Black & Wiliam, 1998). Umpan balik pada setiap pilihan juga menyediakan penjelasan atas respons, meskipun pemanfaatannya untuk memperbaiki pemahaman belum dapat dipastikan karena alasan jawaban tidak diisi.

Kenaikan rata-rata dari 13,44 pada P1 menjadi 15,47 pada P3 sejalan dengan *testing effect*, yakni peningkatan performa setelah peserta mengerjakan kembali butir yang sama (Adesope et al., 2017; Butler & Roediger, 2008; Yang et al., 2021). Namun, *retest effect* dan paparan berulang dapat meningkatkan skor tanpa perubahan setara pada penguasaan konsep (Scharfen et al., 2018). Karena ketiga percobaan memakai 20 butir identik dan disertai umpan balik langsung, kenaikan skor tidak dapat dipisahkan dari efek pengulangan, ingatan terhadap jawaban, atau pemanfaatan umpan balik.

Perubahan skor tidak merata. Selisih P1–P2 relatif kecil dan IK-nya mencakup nol, sedangkan P2–P3 lebih besar. Kenaikan pada murid berskor awal rendah mungkin juga dipengaruhi regresi ke rata-rata (Barnett et al., 2005). Sebaliknya, sembilan murid mengalami penurunan dan 13 tidak pernah mencapai 16. Karena itu, rata-rata kelas saja tidak cukup untuk menentukan tindak lanjut; guru perlu memeriksa pola respons individual.

Umpan balik merupakan salah satu faktor yang mungkin mendukung hasil percobaan berulang. Pada tes pilihan ganda, umpan balik korektif dapat memperkuat manfaat pengujian dan mengurangi dampak pilihan keliru (Butler & Roediger, 2008); secara lebih luas, pengaruh umpan balik bergantung pada isi dan cara penyampaiannya (Hattie & Timperley, 2007). Akan tetapi, penelitian ini tidak memisahkan pengaruh umpan balik dari pengulangan butir, dan kolom alasan yang kosong membatasi penilaian apakah murid memahami penjelasan atau sekadar mengganti pilihan.

Skor akhir Siklus I (versi 2) (rata-rata 18,1) tidak dapat dibandingkan langsung dengan Siklus II (versi revisi) karena versi 2 tidak merekam skor tiap percobaan dan kondisi pengerjaan

berbeda. Pada Siklus I (versi 2), 19 dari 20 butir dijawab benar oleh setidaknya 80% peserta dan 25 pengecoh tidak dipilih. Pada P1 Siklus II (versi revisi), kesukaran lebih bervariasi (0,47–0,83) dan seluruh pengecoh terpilih. Perbedaan ini dapat berkaitan dengan revisi instrumen, peserta yang hadir, atau kondisi pengerjaan; data tidak cukup untuk menentukan penyebabnya.

M2 dan M6 merupakan pola pilihan terbanyak pada Siklus I (versi 2) dan P1 Siklus II (versi revisi). M1 memiliki frekuensi total terendah, tetapi rasio pilihan tertinggi setelah memperhitungkan jumlah pengecoh; pilihannya terkonsentrasi pada butir 2 dan 20, yang menuntut pemahaman bahwa nilai dasar Pancasila tetap sedangkan penjabarannya dapat berubah. Pilihan terhadap pengecoh dapat menunjukkan bahwa opsi tersebut dianggap masuk akal meskipun keliru (Gierl et al., 2017), tetapi frekuensi M2 dan M6 juga dipengaruhi jumlah pengecohnya yang lebih banyak (13 dan 15). Kode M1–M6 disusun dari karakteristik opsi, bukan diagnosis proses berpikir; konfirmasi miskonsepsi memerlukan alasan, tingkat keyakinan, pertanyaan lanjutan, atau wawancara (Kaningtyas et al., 2025; Nisa et al., 2024; Sumarna et al., 2026), sedangkan kolom alasan tidak diisi.

KR-20 sebesar 0,78 menunjukkan konsistensi internal yang cukup pada P1, tetapi lima butir dengan korelasi butir-total terkoreksi di bawah 0,20 perlu ditelaah. Seluruh 60 pengecoh sempat dipilih, tetapi tujuh hanya sekali. Ambang 5% peserta sering dipakai untuk menyatakan pengecoh berfungsi (Haladyna & Downing, 1993; Tarrant et al., 2009); pada n = 36, ambang itu setara minimal dua pemilih, sehingga tujuh pengecoh tersebut lebih tepat disebut lemah atau belum optimal. Revisi perlu memperhatikan konteks kasus, redaksi, dan kewajaran opsi. Analisis hanya memakai P1 dari satu kelas sehingga pengujian pada kelompok lebih luas dan bank soal paralel masih diperlukan.

Keterbatasan penelitian meliputi: (1) hanya melibatkan satu kelas tanpa pembanding; (2) kontribusi individual dan skor kuis kelompok pada kondisi awal tidak terdokumentasi sistematis; (3) enam murid tidak memiliki data Siklus I (versi 2) dan skor per percobaannya tidak tersimpan; (4) analisis butir Siklus II (versi revisi) hanya menggunakan P1; (5) butir

identik dan umpan balik langsung membuat pengaruh pengulangan tidak dapat dipisahkan; serta (6) kolom alasan tidak diisi sehingga kode M1–M6 tidak dapat dipastikan sebagai miskonsepsi. Selain itu, telaah Aiken's V dilakukan sebelum revisi Siklus II (versi revisi) dan melibatkan salah satu penulis.

Refleksi akhir dan tindak lanjut

Refleksi akhir Siklus II (versi revisi) mengarahkan tindak lanjut kepada 13 murid yang tidak pernah mencapai skor 16 dan sembilan murid yang skornya menurun. Pendampingan perlu menelusuri alasan pemilihan jawaban dan memperkuat konsep terkait, bukan sekadar meminta murid mengulang soal yang sama. Guru juga perlu meninjau butir berkorelasi rendah dan pengecoh yang jarang dipilih. Rencana tindak lanjut yang belum dilaksanakan mencakup soal paralel, perekaman alasan atau tingkat keyakinan, dan dokumentasi kontribusi individual selama pembelajaran kelompok, dengan indikator yang ditetapkan sebelum tindakan, misalnya proporsi murid yang mencapai skor 16 pada percobaan pertama bank soal paralel.

D. Kesimpulan

Sebagai PTK deskriptif, penelitian ini menunjukkan bahwa: (1) asesmen formatif individual berbasis HTML *offline* dapat diterapkan pada pembelajaran PPKn dan menghasilkan skor per murid tanpa koneksi internet; keterbatasan pencatatan Siklus I (versi 2) mendorong penambahan riwayat percobaan, batas waktu, dan pengacakan pada Siklus II (versi revisi); (2) rata-rata skor Siklus II (versi revisi) meningkat dari 13,44 pada P1 menjadi 15,47 pada P3, dengan selisih 2,03 poin (IK 95% 0,60–3,45), tetapi kenaikan ini tidak membuktikan peningkatan pemahaman karena butir identik dan umpan balik digunakan berulang; dan (3) pada P1, KR-20 sebesar 0,78, indeks kesukaran 0,47–0,83, serta seluruh pengecoh terpilih. M2 dan M6 paling sering dipilih, sedangkan M1 memiliki rasio pilihan tertinggi setelah memperhitungkan jumlah pengecoh. Kode M1–M6 menggambarkan pola respons, bukan diagnosis miskonsepsi.

Guru dapat menggunakan skor dan pola respons individual untuk menentukan penguatan, terutama bagi murid yang belum mencapai 16 atau mengalami penurunan.

Penelitian lanjutan perlu memakai bank soal paralel, menganalisis respons pada setiap percobaan, merekam alasan atau keyakinan murid, dan mendokumentasikan kontribusi individual agar perubahan skor serta pola jawaban dapat ditafsirkan lebih kuat.

DAFTAR PUSTAKA

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Adi, F. Z. A., Pasani, C. F., & Noorbaiti, R. (2026). Pengembangan asesmen formatif berbasis SLOA pada materi statistika kelas XI SMA. *Jurnal Pendidikan*, *27*(1), 26–39. https://doi.org/10.52850/jpn.v27i1.25882

Akbar, A., Hairida, H., Priyadi, A. T., Halidjah, S., & Safrianty, N. (2025). Pengembangan asesmen formatif berbasis Wordwall materi kalimat majemuk setara di kelas 4 SDN 03 Pontianak Selatan. *Jurnal PGSD: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *18*(2), 87–94. https://doi.org/10.33369/pgsd.18.2.87-94

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

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. https://doi.org/10.3758/MC.36.3.604

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, *87*(6), 1082–1116. https://doi.org/10.3102/0034654317726529

Haladyna, T. M., & Downing, S. M. (1993). How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement*, *53*(4), 999–1010. https://doi.org/10.1177/0013164493053004013

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of*

Educational Research, *77*(1), 81–112. https://doi.org/10.3102/00346543 0298487

Kaningtyas, D. Y. K., Rokhman, F., & Suminar, T. (2025). Pengembangan instrumen tes diagnostik bentuk two-tier multiple choice (TTMC) pada mata pelajaran Sejarah di SMA. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, *10*(3), 221–234. https://doi.org/10.23969/jp.v10i03 .31218

Kemmis, S., & McTaggart, R. (1988). *The action research planner* (3rd ed.). Deakin University.

Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer. https://doi.org/10.1007/978-981- 4560-67-2

Nisa, A. K., Ismail, I., & Adnan, A. (2024). Development of a diagnostic test based on three-tier multiple choice to identify student misconceptions on circulatory system material. *Jurnal Pembelajaran Dan Biologi Nukleus*, *10*(2), 364–378. https://doi.org/10.36987/jpbn.v10i 2.5410

Nurulfathimah, A., Adawiyyah, K. G., Resnawati, C., & Nana. (2026). Efektivitas asesmen formatif digital berbasis Kahoot dalam mendukung pembelajaran mendalam terhadap pemahaman konsep IPA SMP. *STRATEGY: Jurnal Inovasi Strategi dan Model Pembelajaran*, *6*(3), 966–975. https://doi.org/10.51878/strategi. v6i3.11213

Scharfen, J., Peters, J. M., & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, *67*, 44–66. https://doi.org/10.1016/j.intell.201 8.01.003

Sumarna, A. N. A., Sunaengsih, C., & Nugraha, R. G. (2026). Analisis penggunaan tes diagnostik three tier untuk mengidentifikasi miskonsepsi siswa kelas IV SD. *PIJAR: Jurnal Pendidikan dan Pengajaran*, *4*(3), 1081–1092. https://doi.org/10.58540/pijar.v4i3 .1711

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education*, *9*, 40. https://doi.org/10.1186/1472- 6920-9-40

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul00003 09

Zai, S. O., Lase, I. W., Harefa, A. R., & Hura, D. (2026). Analisis pelaksanaan asesmen formatif

dalam pembelajaran Pendidikan Pancasila Kurikulum Merdeka di kelas V UPTD SD Negeri 071082 Lawelu. *SOCIAL: Jurnal Inovasi Pendidikan IPS*, *6*(4), 3255–3263. https://doi.org/10.51878/social.v6i4.15884