



Facilitating the Utilization and Refinement of the Sains Masemba Application to Support Science Assessment Readiness at MTsN 9 Bantul

Ari Cahya Mawardi^{✉1}, Sinta Herahmawati², Anggit Damaz Istoko³, Hauriyatul Illiyyin⁴, Diana Wahyu Eka Rossalita⁵

^{1, 3, 4, 5} UIN Sunan Kalijaga Yogyakarta, Sleman, Indonesia

² MTsN 9 Bantul, Bantul, Indonesia

ABSTRACT

Design/methods/approach – The program employed Participatory Action Research through two cycles. Cycle I involved 49 ninth-grade students and one science teacher on 13–15 June 2026. Cycle II implemented application version 18 and involved 59 students and three teachers on 23–24 June 2026. Data were collected using four-point Likert questionnaires, open- and semi-open-ended questions, usage observations, and application documentation. The analysis included descriptive statistics, internal consistency testing, a Wilcoxon signed-rank test involving 49 students who participated in both cycles, and descriptive thematic coding.

Findings – Students' mean response increased from 3.31, or 82.86% of the maximum score, in Cycle I to 3.81, or 95.17%, in Cycle II. Positive responses increased from 94.90% to 99.83%. Paired data showed an increase from 3.31 to 3.82, with 44 students improving, five remaining unchanged, and none declining. The difference was statistically significant ($p < 0.001$; $r = 0.87$). Teacher responses increased descriptively from 85.00% to 98.33%. Performance, access, and connectivity concerns decreased from 26.53% to 8.47%.

Research implications/limitations – The findings represent perceived usability rather than learning achievement. Future evaluations should include technical logs, task-completion data, and paired learning outcomes.

Originality/value – This program demonstrates how participatory feedback and iterative application refinement can improve digital science assessment experiences in a madrasah context.

OPEN ACCESS

ARTICLE HISTORY

Received: 29-06-2026

Accepted: 22-07-2026

KEYWORDS

Sains Masemba application; digital science assessment; Participatory Action Research; user experience;

Introduction

Transformasi digital dalam pendidikan memperluas fungsi teknologi dari penyampaian materi menuju pengelolaan latihan, asesmen, umpan balik, dan pengambilan keputusan pembelajaran. Agar teknologi memberikan manfaat pedagogis, pengembangannya perlu mengintegrasikan pengetahuan teknologi, pedagogi, dan konten agar aplikasi tidak berhenti sebagai alat administrasi, tetapi benar-benar mendukung tujuan belajar (Mishra & Koehler, 2006). Selain itu, Kerangka kompetensi digital pendidik juga menempatkan asesmen, analisis bukti belajar, dan pemberian umpan balik sebagai bagian penting dari praktik profesional guru (Redecker, 2017).

Kebutuhan tersebut relevan bagi persiapan Tes Kemampuan Akademik Daerah (TKAD) di MTsN 9 Bantul. Dalam pelaksanaannya, tryout IPA bagi siswa kelas IX menggunakan 40 soal dan perangkat komputer madrasah, sedangkan tryout bagi guru IPA melibatkan tiga guru untuk memperkuat kualitas soal dan pembelajaran (Kantor Kementerian Agama Kabupaten Bantul, 2025a, 2025b). Konteks ini menunjukkan perlunya sistem yang dapat menghubungkan materi, kisi-kisi, bank soal, paket tryout, pembahasan, progres, dan riwayat hasil dalam satu alur yang mudah digunakan.

Diagnosis awal terhadap 155 siswa menunjukkan bahwa nilai IPA memiliki rata-rata 30,90 dan median 30,00 serta merupakan nilai terendah dibandingkan tiga mata pelajaran lain yang dianalisis. Data tersebut digunakan sebagai dasar identifikasi kebutuhan pengembangan, bukan untuk menyimpulkan bahwa rendahnya capaian IPA disebabkan oleh satu faktor tertentu. Pada tingkat yang lebih luas, skor sains Indonesia dalam PISA 2022 sebesar 383 masih berada di bawah rata-rata OECD sebesar 485 (OECD, 2023). Oleh karena itu, pengembangan sistem asesmen digital yang mampu mengidentifikasi kesenjangan penguasaan konsep dan memberikan umpan balik yang mendukung perbaikan pembelajaran tetap menjadi kebutuhan yang relevan.

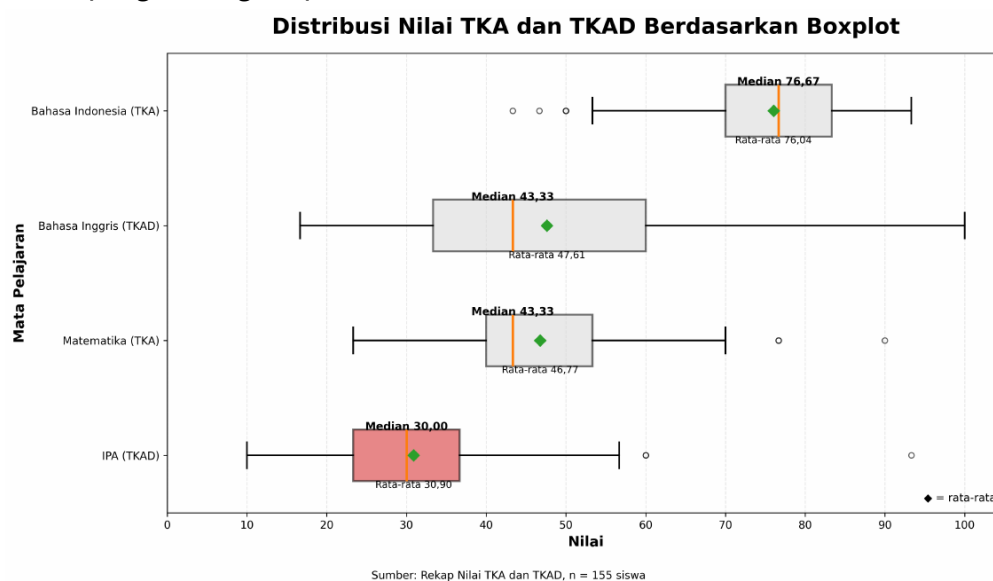
Asesmen formatif menggunakan bukti belajar untuk mengidentifikasi tingkat penguasaan siswa, memberi umpan balik yang dapat ditindaklanjuti, serta menyesuaikan strategi pembelajaran berikutnya (Black & Wiliam, 2009; Bennett, 2011; Shute, 2008). Teknologi berpotensi mendukung proses tersebut melalui pengumpulan data dan penyajian umpan balik secara lebih cepat, tetapi manfaat pedagogisnya bergantung pada kualitas desain sistem, kesiapan pengguna, dan pemanfaatan informasi yang dihasilkan dalam proses pembelajaran. Dalam konteks asesmen digital, penerapan *learning analytics* juga berpotensi membantu guru memahami perkembangan belajar siswa. Namun, kajian pada sekolah menengah menunjukkan bahwa implementasinya masih memerlukan pengembangan desain, kesiapan pengguna, serta mekanisme pemanfaatan data yang sesuai dengan konteks sekolah (de Sousa et al., 2021).

Kualitas sistem, kualitas informasi, kemudahan navigasi, serta persepsi terhadap manfaat aplikasi merupakan faktor yang memengaruhi kepuasan pengguna dan niat untuk terus menggunakan aplikasi pembelajaran (Johannsen et al., 2023). Namun, aplikasi seluler yang tampak mudah digunakan juga tidak selalu menghasilkan perubahan belajar apabila pengguna tidak memperoleh panduan, tidak melihat kebutuhan penggunaan, atau hanya memakai aplikasi secara terbatas (Baars et al., 2022). Temuan tersebut menegaskan perlunya pengujian lapangan yang tidak hanya menghitung skor penerimaan, tetapi juga menelusuri kendala akses, keterlihatan menu, keterbacaan, dan kebermaknaan informasi hasil yang disajikan.

Sebagai upaya menjawab kebutuhan tersebut, dikembangkan aplikasi Sains Masemba, yaitu platform berbasis web yang mendukung kegiatan belajar, latihan, dan tryout IPA melalui portal bagi super admin, guru, siswa, dan orang tua. Aplikasi ini dirancang untuk mengintegrasikan materi pembelajaran, kisi-kisi, bank soal, pelaksanaan tryout, pembahasan, pemantauan progres belajar, serta riwayat hasil dalam satu sistem yang terintegrasi.

Pengujian pada Siklus I menunjukkan bahwa aplikasi memperoleh respons awal yang positif, namun masih ditemukan beberapa kendala, antara lain kinerja sistem saat memuat halaman, proses login, navigasi pada perangkat seluler, penyajian konten yang terlalu panjang, tampilan antarmuka, notifikasi, serta kejelasan penyajian riwayat hasil belajar. Masukan yang diperoleh dari guru dan siswa kemudian digunakan sebagai dasar penyempurnaan aplikasi secara iteratif hingga menghasilkan versi v18 yang selanjutnya dievaluasi kembali pada Siklus II. Pendekatan ini tidak hanya menilai tingkat penerimaan pengguna, tetapi juga memanfaatkan umpan balik pengguna sebagai dasar penyempurnaan aplikasi secara berkelanjutan sehingga pengembangan sistem lebih sesuai dengan kebutuhan nyata di lingkungan madrasah.

Kegiatan ini bertujuan: (1) membandingkan respons siswa dan guru terhadap penggunaan aplikasi Sains Masemba pada Siklus I dan Siklus II; (2) menilai perubahan pengalaman penggunaan pada siswa yang mengikuti kedua siklus; (3) membandingkan tema masukan responden; dan (4) merumuskan agenda penyempurnaan lanjutan. Kontribusi kegiatan terletak pada penggunaan data kuantitatif dan kualitatif secara berulang untuk menghubungkan evaluasi pengguna dengan keputusan pengembangan aplikasi dalam konteks madrasah.



Gambar 1. Distribusi nilai TKA dan TKAD berdasarkan data diagnosis awal ($n = 155$ siswa)

Methods

1. Desain Kegiatan

Kegiatan ini menggunakan pendekatan Participatory Action Research (PAR) karena menempatkan guru dan siswa sebagai mitra yang terlibat secara aktif dalam penggunaan aplikasi, pemberian umpan balik, identifikasi kendala, serta penyusunan rekomendasi penyempurnaan aplikasi. Melalui siklus diagnosis, tindakan, observasi, dan refleksi, PAR memungkinkan proses pengembangan dilakukan secara berkelanjutan berdasarkan pengalaman dan kebutuhan pengguna (Baum et al., 2006; Kemmis et al., 2014).

Penelitian dirancang dalam dua siklus yang saling berkesinambungan. Siklus I bertujuan mengidentifikasi kekuatan, kelemahan, dan berbagai kendala penggunaan aplikasi sebagai dasar evaluasi awal. Hasil refleksi pada siklus ini digunakan untuk menyempurnakan aplikasi hingga menghasilkan versi v18 yang kemudian dievaluasi kembali pada Siklus II. Perbandingan hasil antarsiklus dilakukan pada tingkat kelompok, sedangkan analisis perubahan individual hanya dilakukan terhadap 49 siswa yang dapat dipadankan berdasarkan identitas dan kelas pada kedua siklus pengumpulan data.



Gambar 2. Alur dua siklus PAR pada pengujian dan penyempurnaan aplikasi Sains Masemba

2. Lokasi, Waktu, dan Partisipan

Kegiatan dilaksanakan di MTsN 9 Bantul, Kabupaten Bantul, Daerah Istimewa Yogyakarta. Siklus I berlangsung pada 13–15 Juni 2026, sedangkan Siklus II berlangsung pada 23–24 Juni 2026. Data dianalisis secara agregat dan nama individual tidak ditampilkan.

Tabel 1. Komposisi partisipan pada Siklus I dan Siklus II

Siklus	Kelompok	Jumlah	Waktu respons
Siklus I	Siswa IX A	1	13–15 Juni 2026
Siklus I	Siswa IX B	20	13–15 Juni 2026
Siklus I	Siswa IX E	28	13–15 Juni 2026
Siklus I	Guru IPA	1	13 Juni 2026
Siklus II	Siswa IX A	4	23–24 Juni 2026
Siklus II	Siswa IX B	24	23–24 Juni 2026
Siklus II	Siswa IX E	31	23–24 Juni 2026
Siklus II	Guru IPA	3	24 Juni 2026

a. *Prosedur Dua Siklus*

Tabel 2. Tahapan kegiatan PAR dalam dua siklus

Tahap	Aktivitas utama	Luaran
Diagnosis	Menelaah capaian IPA, kebutuhan TKAD, perangkat, akses, dan alur aplikasi.	Peta kebutuhan penggunaan.
Perencanaan Siklus I	Menyiapkan akun, materi, bank soal, paket tryout, petunjuk, dan angket.	Skenario uji awal.
Aksi Siklus I	Siswa login, belajar, berlatih, mengikuti tryout, mengirim jawaban, dan membaca hasil; guru memantau.	Data respons dan observasi.
Refleksi Siklus I	Menganalisis skor serta masukan tentang performa, navigasi, konten, visual, dan riwayat nilai.	Prioritas revisi.
Iterasi v18	Mengarahkan penyempurnaan pada area masalah yang ditemukan dalam Siklus I.	Versi aplikasi untuk uji ulang.
Aksi Siklus II	Mengulangi penggunaan aplikasi dan pengisian angket dengan indikator yang dapat dibandingkan.	Data uji ulang siswa dan guru.
Refleksi akhir	Membandingkan kedua siklus dan merumuskan pengembangan berikutnya.	Agenda penyempurnaan berbasis bukti.

3. Instrumen dan Pengumpulan Data

Angket Sains Masemba untuk siswa memuat sepuluh pernyataan tertutup dan tiga pertanyaan terbuka atau semiterbuka. Item tertutup mengukur kemandirian login, keterlihatan menu, keterbacaan melalui telepon seluler, pemahaman ceklis latihan, motivasi melalui progres, kejelasan petunjuk tryout, kemudahan mengerjakan dan mengirim jawaban, akses nilai dan pembahasan, manfaat riwayat nilai, serta kesediaan menggunakan aplikasi kembali. Angket guru menggunakan sepuluh item yang menilai pemahaman menu, penggunaan mandiri, pengelolaan konten dan soal, pembuatan tryout, monitoring progres, pemahaman riwayat nilai, pemanfaatan hasil, dan niat penggunaan rutin.

Skala Likert menggunakan empat pilihan skor, yaitu 1 sampai 4. Skor 3 dan 4 diklasifikasikan sebagai respons positif. Pertanyaan terbuka atau semiterbuka meminta fitur yang disukai, kendala yang dialami, dan saran perbaikan. Observasi penggunaan serta dokumentasi antarmuka digunakan untuk menafsirkan respons angket dan menyusun prioritas teknis maupun pedagogis.

4. Analisis Data

Data kuantitatif dianalisis menggunakan rata-rata item, persentase terhadap skor maksimum, proporsi jawaban positif, median, dan rentang antarkuartil. Konsistensi internal sepuluh item siswa dihitung dengan koefisien Cronbach alpha sebagai informasi tambahan. Alpha tidak diperlakukan sebagai satu-satunya dasar kualitas instrumen karena item mencakup beberapa dimensi penggunaan dan nilai alpha dipengaruhi oleh jumlah item serta variasi skor (Tavakol & Dennick, 2011).

Perbandingan inferensial dilakukan hanya pada 49 siswa yang teridentifikasi mengikuti kedua siklus. Skor komposit setiap siswa dihitung sebagai rata-rata sepuluh item. Uji Wilcoxon signed-rank digunakan karena data berasal dari pasangan yang sama dan berskala ordinal-komposit. Besaran efek dihitung dengan $r = \frac{|Z|}{\sqrt{n}}$ pada pasangan yang memiliki selisih tidak nol. Data sepuluh siswa tambahan pada Siklus II tetap disertakan dalam statistik deskriptif, tetapi tidak dimasukkan dalam uji berpasangan.

Jawaban terbuka dikodekan melalui analisis isi tematik deskriptif. Tema kendala disejajarkan antarsiklus agar dapat dibandingkan, yaitu tidak ada kendala berarti, performa/akses/koneksi, konten/teks, navigasi/learnability, notifikasi, dan belum mengeksplorasi. Kejelasan prosedur pengodean dan hubungan antara tema dengan bukti respons dijaga agar interpretasi tidak melebihi data (Braun & Clarke, 2021). Visualisasi utama menggunakan tabel tema dan diagram batang persentase karena frekuensi tema mempertahankan konteks analitis dan memungkinkan perbandingan antarsiklus.

Result

1. Gambaran Aplikasi yang Diuji

Aplikasi Sains Masemba merupakan platform web dengan portal super admin, guru, siswa, dan orang tua. Struktur aplikasi menghubungkan materi dan tujuan pembelajaran dengan kisi-kisi, bank soal, paket tryout, jawaban, pembahasan, hasil, dan monitoring. Hubungan tersebut dimaksudkan agar guru dapat menelusuri indikator pembelajaran sampai hasil pengerjaan siswa, sedangkan siswa memperoleh alur belajar, latihan, tryout, dan umpan balik dalam satu platform.

Tabel 3. Fitur utama aplikasi Sains Masemba yang menjadi objek pengujian

Kelompok fitur	Fungsi utama	Kontribusi
Pengelolaan konten	Topik, materi, ringkasan, dan tujuan pembelajaran.	Menata bahan belajar IPA.
Kisi-kisi dan bank soal	Indikator, level kognitif, kesulitan, stimulus, opsi, kunci, bobot, dan pembahasan.	Menyusun soal berbasis indikator.
Variasi soal	Pilihan ganda, pilihan ganda kompleks, dan benar-salah.	Mendukung variasi asesmen.
Tryout dan penilaian	Durasi, aturan, jadwal, status, exact match, dan partial no penalty.	Menjalankan simulasi asesmen.
Monitoring dan hasil	Status pengerjaan, skor, pembahasan, warning, progres, dan riwayat nilai.	Mendukung umpan balik dan tindak lanjut.
Import dan akses	Import Excel, sesi login, profil, navigasi seluler, dan deployment daring.	Mempercepat pengelolaan dan akses.

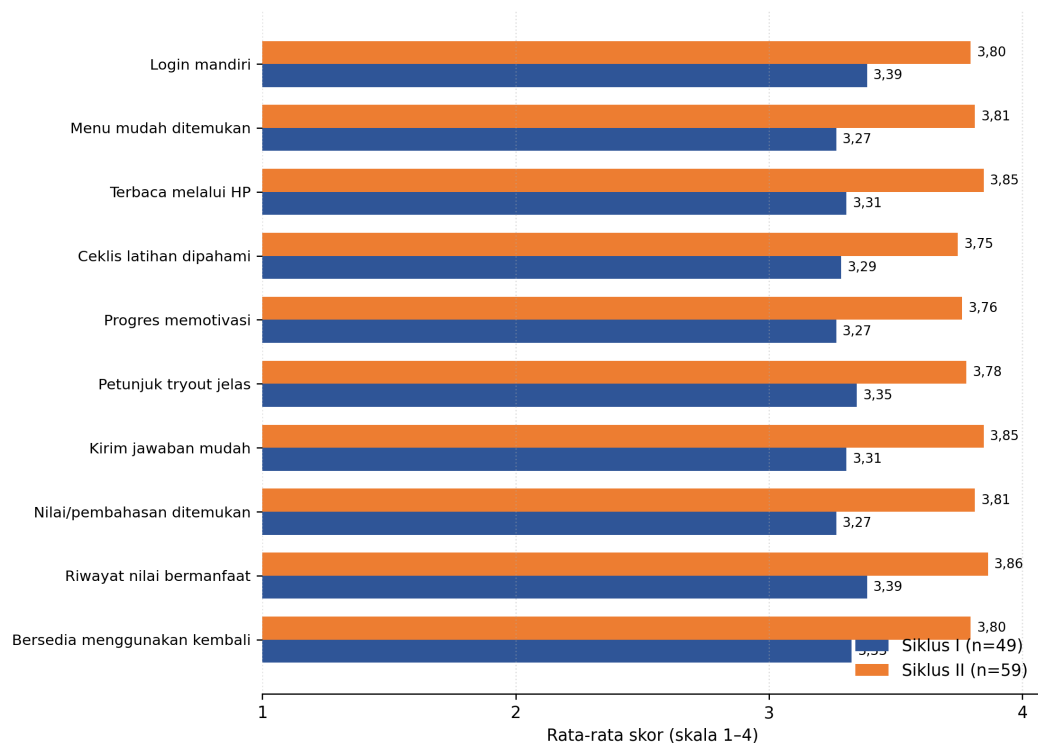
2. Perbandingan Respons Siswa

Seluruh respons siswa memiliki data lengkap pada sepuluh item tertutup. Rata-rata komposit kelompok meningkat dari 3,31 pada Siklus I menjadi 3,81 pada Siklus II. Peningkatan terjadi pada seluruh indikator. Skor tertinggi Siklus II terdapat pada manfaat riwayat nilai sebesar 3,86, sedangkan keterbacaan melalui telepon seluler dan kemudahan mengirim jawaban masing-masing mencapai 3,85.

Tabel 4. Perbandingan rata-rata respons siswa pada setiap indikator

No.	Indikator	Siklus I	Siklus II	Δ
1	Login mandiri	3,39	3,80	+0,41
2	Menu mudah ditemukan	3,27	3,81	+0,55
3	Terbaca melalui HP	3,31	3,85	+0,54
4	Ceklis latihan dipahami	3,29	3,75	+0,46
5	Progres memotivasi	3,27	3,76	+0,50
6	Petunjuk tryout jelas	3,35	3,78	+0,43
7	Kirim jawaban mudah	3,31	3,85	+0,54
8	Nilai/pembahasan ditemukan	3,27	3,81	+0,55
9	Riwayat nilai bermanfaat	3,39	3,86	+0,48

10	Bersedia menggunakan kembali	3,33	3,80	+0,47
—	Rata-rata keseluruhan	3,31	3,81	+0,49



Gambar 3. Perbandingan rata-rata indikator respons siswa pada Siklus I dan Siklus II

Tabel 5. Ringkasan hasil siswa pada tingkat kelompok dan pasangan yang sama

Ukuran	Siklus I	Siklus II	Analisis berpasangan
Jumlah siswa	49	59	49 siswa dipadankan
Rata-rata komposit	3,31	3,81	3,31 → 3,82
Persentase skor maksimum	82,86%	95,17%	+12,31 poin persentase
Jawaban positif (skor 3–4)	465/490 (94,90%)	589/590 (99,83%)	+4,93 poin persentase
Median komposit	3,40	3,80	3,40 → 3,80
Rentang antarkuartil	3,00–3,80	3,70–3,90	3,70–4,00 pada Siklus II
Cronbach alpha	0,934	0,593	Dibaca hati-hati karena variasi Siklus II sempit
Perubahan individu	—	—	44 meningkat; 5 tetap; 0 menurun
Uji Wilcoxon	—	—	$p < 0,001$; $r = 0,87$

Pada 49 siswa yang mengikuti kedua siklus, rata-rata komposit meningkat sebesar 0,51 poin, dari 3,31 menjadi 3,82. Sebanyak 44 siswa menunjukkan skor lebih tinggi, lima siswa tidak berubah, dan tidak terdapat siswa dengan skor lebih rendah. Uji Wilcoxon menunjukkan perbedaan yang bermakna ($p < 0,001$) dengan besaran efek $r = 0,87$. Besaran efek tersebut menunjukkan perubahan yang kuat pada respons penggunaan dalam sampel berpasangan, tetapi tidak dapat ditafsirkan sebagai efek terhadap hasil belajar.

Nilai Cronbach alpha turun dari 0,934 pada Siklus I menjadi 0,593 pada Siklus II. Penurunan ini terjadi bersamaan dengan konsentrasi jawaban yang sangat tinggi pada skor 3 dan 4; hanya satu dari 590 jawaban Siklus II berada pada skor 2. Homogenitas dan efek plafon mengurangi variasi antarpeserta sehingga dapat menurunkan korelasi antarpoin. Oleh sebab itu, interpretasi Siklus II lebih ditekankan pada pola item, perubahan berpasangan, dan bukti kualitatif daripada menjadikan alpha sebagai penentu tunggal.

3. Perbandingan Respons Guru

Tabel 6. Perbandingan deskriptif respons guru

Indikator guru	Siklus I (n=1)	Siklus II (n=3)	Pola
Pemahaman fungsi menu	4,00	4,00	Tetap sangat tinggi
Penggunaan tanpa bantuan	4,00	4,00	Tetap sangat tinggi
Pengelolaan materi, kisi-kisi, dan soal	3,00	4,00	Meningkat
Pembuatan paket tryout	4,00	4,00	Tetap sangat tinggi
Ceklis latihan untuk monitoring	4,00	4,00	Tetap sangat tinggi
Tampilan topik dan centang progres	3,00	3,67	Meningkat
Menu nilai tryout	3,00	4,00	Meningkat
Kejelasan riwayat nilai	2,00	3,67	Peningkatan terbesar
Hasil untuk tindak lanjut	4,00	4,00	Tetap sangat tinggi
Kesediaan menggunakan rutin	3,00	4,00	Meningkat
Rata-rata	3,40	3,93	85,00% → 98,33%

Satu guru pada Siklus I memperoleh skor 34 dari 40 atau 85,00%. Pada Siklus II, tiga guru memperoleh total 118 dari 120 atau 98,33%. Peningkatan paling jelas terdapat pada kejelasan riwayat nilai, pengelolaan konten dan soal, menu nilai, serta kesediaan menggunakan aplikasi secara rutin.

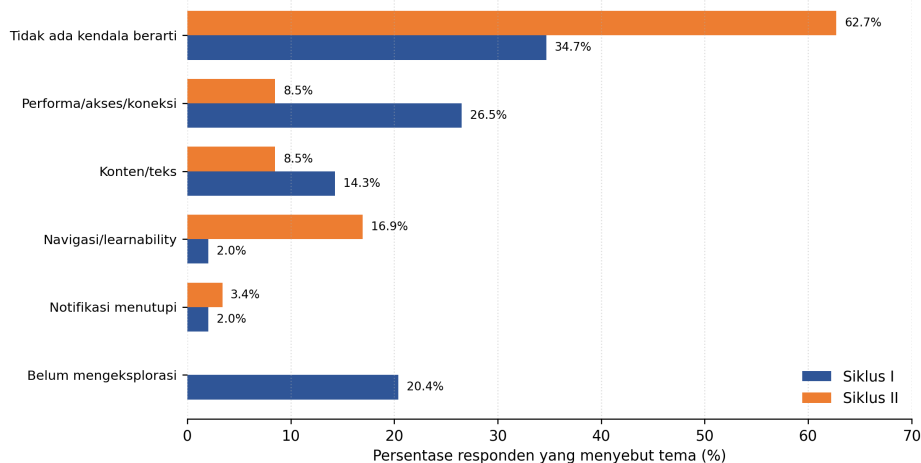
Guru yang sama dan dapat dipadankan pada kedua siklus meningkat dari 34 menjadi 40. Temuan ini berguna sebagai bukti kasus perkembangan pengalaman pengguna, tetapi tidak diuji secara inferensial karena hanya berasal dari satu pasangan guru. Penambahan dua guru pada Siklus II memperluas sudut pandang, namun jumlah tersebut masih terlalu kecil untuk generalisasi.

4. Masukan Responden dan Perubahan Tema Kendala

Tabel 7. Perbandingan tema utama kendala yang disebutkan siswa

Tema dominan	Siklus I	Siklus II	Perubahan
Tidak ada kendala berarti	17 (34,69%)	37 (62,71%)	+28,02 pp
Performa, akses, atau koneksi	13 (26,53%)	5 (8,47%)	-18,06 pp
Konten atau teks	7 (14,29%)	5 (8,47%)	-5,82 pp

Navigasi atau learnability	1 (2,04%)	10 (16,95%)	+14,91 pp
Notifikasi menutupi tampilan	1 (2,04%)	2 (3,39%)	+1,35 pp
Belum mengeksplorasi fitur	10 (20,41%)	0 (0,00%)	-20,41 pp



Gambar 4. Perubahan persentase tema kendala siswa antar-siklus

Proporsi siswa yang menyatakan tidak mengalami kendala berarti meningkat dari 34,69% menjadi 62,71%. Tema performa, akses, atau koneksi turun dari 26,53% menjadi 8,47%, sedangkan tema beban konten atau teks turun dari 14,29% menjadi 8,47%. Pola tersebut mendukung penilaian bahwa hambatan utama pada Siklus I berkurang pada pengujian v18.

Tema navigasi atau learnability meningkat dari 2,04% menjadi 16,95%. Kenaikan tersebut tidak langsung berarti navigasi memburuk. Pada Siklus II, masalah yang disebutkan lebih spesifik, yaitu menu bawah perlu digeser pada layar kecil dan beberapa ikon membutuhkan pembiasaan. Pergeseran tema dapat menunjukkan bahwa setelah masalah performa dan akses berkurang, pengguna lebih mampu mengidentifikasi hambatan antarmuka yang lebih rinci.

Fitur yang paling sering disukai pada Siklus II adalah navigasi bawah dan dashboard modern, masing-masing disebut oleh 10 siswa, diikuti menu belajar dan tryout yang lebih mudah ditemukan serta proses pengerjaan tryout, masing-masing sembilan siswa. Ceklis progres disebut delapan siswa, riwayat nilai dan pembahasan tujuh siswa, serta materi dan latihan enam siswa.

Saran Siklus II beralih dari perbaikan mendasar menuju pengayaan fitur. Sebelas siswa meminta notifikasi latihan baru; masing-masing sepuluh siswa mengusulkan mode untuk koneksi terbatas serta pembahasan dan contoh yang lebih beragam; sembilan siswa mengusulkan gambar atau ilustrasi dan pemeliharaan kecepatan; sedangkan lima siswa meminta tombol seluler diperbesar dan jumlah latihan ditambah. Temuan tersebut memperlihatkan bahwa kebutuhan berikutnya mencakup stabilitas teknis, akses inklusif, dan kedalaman pedagogis.

5. Agenda Penyempurnaan Setelah Siklus II

Tabel 8. Agenda pengembangan aplikasi berdasarkan refleksi Siklus II

Urutan	Area	Tindakan lanjutan	Indikator pemeriksaan
Prioritas 1	Stabilitas dan koneksi	Catat waktu muat, gagal login, respons tombol, perangkat, dan jaringan.	Waktu muat dan kegagalan menurun.

Urutan	Area	Tindakan lanjutan	Indikator pemeriksaan
Prioritas 2	Navigasi seluler	Perbesar target sentuh; tambah indikator geser, label ikon, dan onboarding.	Fungsi utama ditemukan tanpa bantuan.
Prioritas 3	Riwayat nilai dan analitik	Tampilkan tanggal, percobaan, tren, topik lemah, dan rekomendasi.	Data dapat diterjemahkan menjadi tindak lanjut.
Prioritas 4	Konten dan pembahasan	Seimbangkan stimulus, variasi soal, contoh, ilustrasi, dan pembahasan.	Konten terbaca dan mengoreksi miskonsepsi.
Prioritas 5	Notifikasi dan keterlibatan	Atur notifikasi agar tidak menutup soal; sediakan pengendalian pengguna.	Notifikasi informatif dan tidak mengganggu.
Prioritas 6	Bukti efektivitas	Tambahkan hasil tryout berpasangan, log penggunaan, dan uji keberhasilan tugas.	Klaim didukung bukti perilaku dan hasil belajar.

Discussion

Perbandingan dua siklus menunjukkan perbaikan yang konsisten pada seluruh indikator respons siswa. Kenaikan rata-rata kelompok sebesar 0,49 poin dan peningkatan berpasangan sebesar 0,51 poin memperlihatkan bahwa pengalaman menggunakan v18 dinilai lebih baik daripada versi pada Siklus I. Konsistensi perubahan juga terlihat dari meningkatnya skor pada 44 dari 49 siswa, tanpa adanya siswa yang mengalami penurunan skor. Meskipun demikian, hasil tersebut harus ditempatkan sebagai bukti penerimaan, kemudahan, dan pengalaman penggunaan, bukan sebagai bukti kausal peningkatan penguasaan IPA.

Perbaikan terbesar muncul pada keterlihatan menu dan kemudahan menemukan nilai serta pembahasan. Temuan ini selaras dengan model keberhasilan aplikasi pembelajaran yang menempatkan kualitas sistem, kualitas informasi, kepuasan, dan niat penggunaan kembali sebagai komponen yang saling berhubungan (Johannsen et al., 2023). Rata-rata kesediaan menggunakan kembali mencapai 3,80 pada Siklus II, namun, tingginya niat untuk menggunakan kembali belum tentu diikuti oleh penggunaan aktual apabila aplikasi belum terintegrasi dengan kebutuhan belajar maupun praktik pembelajaran sehari-hari. Penelitian Baars et al. (2022) menunjukkan bahwa aplikasi yang dinilai cukup ramah pengguna tetap dapat digunakan secara terbatas apabila dukungan tidak terintegrasi dengan kebutuhan belajar dan praktik pembelajaran.

Penurunan keluhan performa, akses, dan konektivitas merupakan hasil penting karena reliabilitas sistem membentuk prasyarat bagi aktivitas pedagogis. Ketika pemuatan lambat atau proses login gagal, siswa kehilangan waktu, perhatian, dan kepercayaan terhadap platform. Data Siklus II mengindikasikan perbaikan, tetapi belum tersedia pengukuran waktu muat dan log kegagalan yang objektif. Kesimpulan mengenai peningkatan performa karena itu harus tetap dibatasi pada persepsi responden. Siklus berikutnya perlu merekam waktu respons, perangkat, browser, kekuatan jaringan, dan titik kegagalan.

Kenaikan tema navigasi pada Siklus II memperlihatkan nilai analitis data kualitatif. Rata-rata item menu meningkat tajam, tetapi sebagian pengguna layar kecil masih harus menggeser menu dan mempelajari makna ikon. Skor agregat yang tinggi dapat menyembunyikan friksi pada kelompok tertentu. Prinsip PAR memungkinkan masukan minoritas diperlakukan sebagai sumber desain, sehingga ukuran target sentuh, label ikon, indikator geser, dan onboarding singkat perlu diuji melalui tugas penggunaan yang terstruktur.

Riwayat nilai memperoleh skor siswa tertinggi pada Siklus II, dan guru juga memberikan penilaian lebih baik terhadap kejelasan informasi tersebut. Fungsi ini berpotensi mendukung asesmen formatif apabila riwayat skor tidak hanya menampilkan angka, tetapi membantu mengidentifikasi topik lemah, perkembangan antarpelaksanaan, dan tindakan berikutnya. Asesmen formatif memerlukan umpan balik yang spesifik dan dapat digunakan untuk memperbaiki proses belajar (Black & Wiliam, 2009; Shute, 2008). Kajian learning analytics di sekolah menengah juga menekankan potensi data untuk memberi umpan balik dan mendukung intervensi, tetapi penerapannya harus relevan dengan konteks sekolah (de Sousa et al., 2021).

Peningkatan skor guru belum cukup apabila data aplikasi tidak diterjemahkan menjadi keputusan pedagogis. Penelitian pada 1.059 guru sekolah menengah menunjukkan bahwa ketersediaan teknologi perlu disertai keyakinan positif, keterampilan literasi data, dan praktik penggunaan data untuk pembelajaran (Michos et al., 2023). Oleh karena itu, Sains Masemba perlu menyajikan informasi yang ringkas, dapat ditelusuri sampai indikator soal, dan dilengkapi panduan tindak lanjut. Pelatihan guru sebaiknya tidak hanya menjelaskan operasi menu, tetapi juga latihan membaca pola skor, mengenali miskonsepsi, dan menyusun remediasi.

Saran mengenai pembahasan lebih beragam, ilustrasi, dan tambahan latihan menunjukkan bahwa kualitas pengalaman pengguna berkaitan dengan kualitas substansi. Tampilan modern dapat meningkatkan ketertarikan awal, tetapi keberlanjutan penggunaan membutuhkan materi yang relevan, menantang, mudah dipahami, serta dilengkapi umpan balik yang bermakna. Integrasi teknologi, pedagogi, dan konten dalam TPACK menuntut keputusan desain yang menjaga keseimbangan antara kelengkapan fitur, beban kognitif, dan tujuan pembelajaran (Mishra & Koehler, 2006).

Nilai alpha Siklus II yang lebih rendah perlu dipahami bersama efek plafon (*ceiling effect*) Hampir seluruh jawaban berada pada kategori positif sehingga variasi antarresponden menjadi sangat kecil. Selain itu, sepuluh item mencakup beberapa aspek berbeda, mulai dari akses, navigasi, motivasi, hingga niat penggunaan. Tavakol dan Dennick (2011) mengingatkan bahwa alpha dipengaruhi oleh struktur konstruk dan karakteristik item. Instrumen berikutnya perlu diuji melalui validasi isi, analisis faktor apabila sampel memadai, dan penambahan ukuran perilaku agar evaluasi tidak hanya bergantung pada konsistensi internal.

Kekuatan utama kegiatan terletak pada hubungan langsung antara refleksi pengguna dan iterasi produk. Siklus I tidak berhenti pada pelaporan skor, tetapi menghasilkan agenda revisi yang diuji kembali pada Siklus II. Proses tersebut sesuai dengan PAR yang menempatkan partisipan sebagai sumber pengetahuan praktis dalam tindakan dan refleksi (Baum et al., 2006; Kemmis et al., 2014). Kontribusi yang dihasilkan bersifat kontekstual: aplikasi menunjukkan kelayakan penggunaan yang makin baik di MTsN 9 Bantul, sementara efektivitas terhadap hasil belajar masih memerlukan desain evaluasi yang berbeda.

Kegiatan memiliki beberapa keterbatasan. Pertama, jumlah guru sangat kecil, yaitu satu orang pada Siklus I dan tiga orang pada Siklus II, sehingga hasil guru bersifat deskriptif dan tidak dapat digeneralisasi. Kedua, Siklus II menambahkan sepuluh siswa baru; oleh karena itu, uji perubahan individual dibatasi pada 49 siswa yang benar-benar dapat dipadankan.

Ketiga, data utama berasal dari persepsi diri dan belum didukung log penggunaan terstruktur, durasi penyelesaian, keberhasilan setiap tugas, atau data gangguan teknis otomatis. Keempat, respons Siklus II menunjukkan efek plafon dan homogenitas tinggi yang dapat dipengaruhi oleh pengalaman penggunaan yang lebih baik, format jawaban, suasana pelaksanaan, atau kecenderungan memberikan respons positif.

Kelima, kegiatan tidak menggunakan pretest-posttest hasil belajar atau kelompok pembanding. Hasil tidak dapat digunakan untuk menyimpulkan bahwa Sains Masemba meningkatkan penguasaan IPA atau kesiapan TKAD. Keenam, pengodean masukan terbuka menggunakan tema dominan per respons sehingga satu respons yang memuat beberapa isu mungkin tidak seluruhnya tercermin dalam frekuensi utama. Penelitian berikutnya perlu melakukan pengodean ganda, menyimpan kutipan anonim yang representatif, dan menguji kesepakatan pengode apabila pendekatan analisis isi kuantitatif digunakan.

Conclusion

Pengujian dua siklus melalui PAR menunjukkan bahwa aplikasi Sains Masemba versi v18 memperoleh respons penggunaan yang lebih positif dibandingkan versi Siklus I. Rata-rata respons siswa meningkat dari 3,31 menjadi 3,81, proporsi jawaban positif meningkat dari 94,90% menjadi 99,83%, dan analisis 49 siswa berpasangan menunjukkan peningkatan yang bermakna dengan besaran efek kuat. Respons guru juga meningkat secara deskriptif dari 85,00% menjadi 98,33%.

Perbaikan paling terlihat pada keterlihatan menu, keterbacaan seluler, kemudahan mengerjakan dan mengirim jawaban, akses nilai dan pembahasan, serta manfaat riwayat nilai. Masukan responden menunjukkan bahwa keluhan performa, akses, dan konektivitas menurun, sementara kebutuhan berikutnya bergeser menuju penyempurnaan navigasi pada layar kecil, pembahasan yang lebih beragam, mode koneksi terbatas, ilustrasi, notifikasi, dan pengayaan latihan.

Hasil mendukung kelayakan awal Sains Masemba sebagai sarana latihan dan asesmen IPA dalam konteks MTsN 9 Bantul, tetapi belum membuktikan peningkatan hasil belajar. Tahap selanjutnya perlu mengintegrasikan log teknis, uji tugas, analitik topik, produk guru, dan hasil tryout berpasangan agar manfaat penggunaan dapat dinilai pada tingkat sistem, praktik pengajaran, dan capaian siswa.

Declarations

Author contribution statement

Penulis pertama berperan dalam perencanaan kegiatan, pengembangan aplikasi, pelaksanaan uji, analisis data, dan penyusunan artikel. Penulis kedua berperan dalam dokumentasi, pengumpulan data, dan pengolahan hasil. Penulis ketiga, ke empat, dan ke lima berperan dalam validasi instrumen, refleksi kegiatan, dan penyuntingan naskah.

Funding statement

Kegiatan ini tidak menerima pendanaan khusus dari lembaga publik, komersial, atau nirlaba.

Data availability statement

Data pendukung meliputi angket respons siswa dan guru, dokumentasi kegiatan, serta dokumentasi teknis aplikasi. Data individual tidak ditampilkan untuk melindungi identitas partisipan. Data agregat dapat tersedia dari penulis korespondensi berdasarkan permintaan yang wajar dan sesuai dengan ketentuan etika. Izin pelaksanaan penelitian telah diperoleh dari guru dan siswa. Guru dan siswa yang terlibat telah diberikan penjelasan mengenai tujuan, prosedur, serta penggunaan data yang dikumpulkan dalam penelitian ini. Keikutsertaan partisipan bersifat sukarela dan persetujuan telah diperoleh sebelum proses pengumpulan data dilakukan. Identitas dan informasi pribadi seluruh partisipan dijaga kerahasiaannya. Data dianalisis dan disajikan secara agregat hanya untuk kepentingan penelitian dan pengembangan pendidikan tanpa mengungkapkan informasi yang dapat digunakan untuk mengidentifikasi partisipan secara individual.

Declaration of interests statement

Penulis menyatakan tidak memiliki konflik kepentingan yang dapat memengaruhi pelaksanaan kegiatan dan penulisan artikel ini.

References

Amemasor, S. K., Oppong, S. O., Ghansah, B., & Essel, D. D. (2025). A systematic review on the impact of teacher professional development on digital instructional integration and teaching practices. *Frontiers in Education*, 10, 1541031. <https://doi.org/10.3389/feduc.2025.1541031>

- Baars, M., Khare, S., & Ridderstap, L. (2022). Exploring students' use of a mobile application to support their self-regulated learning processes. *Frontiers in Psychology*, 13, 793002. <https://doi.org/10.3389/fpsyg.2022.793002>
- Baum, F., MacDougall, C., & Smith, D. (2006). Participatory action research. *Journal of Epidemiology and Community Health*, 60(10), 854–857. <https://doi.org/10.1136/jech.2004.028662>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- de Sousa, E. B. G., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B., & Gašević, D. (2021). Applications of learning analytics in high schools: A systematic literature review. *Frontiers in Artificial Intelligence*, 4, 737891. <https://doi.org/10.3389/frai.2021.737891>
- Johannsen, F., Knipp, M., Loy, T., Mirbabaie, M., Möllmann, N. R. J., Voshaar, J., et al. (2023). What impacts learning effectiveness of a mobile learning app focused on first-year students? *Information Systems and e-Business Management*, 21, 629–673. <https://doi.org/10.1007/s10257-023-00644-0>
- Kantor Kementerian Agama Kabupaten Bantul. (2025a, 23 Oktober). Guru IPA MTsN 9 Bantul ikuti Try Out TKAD Dikpora 1 untuk tingkatan kualitas soal dan pembelajaran.
- Kantor Kementerian Agama Kabupaten Bantul. (2025b, 23 Oktober). Siswa kelas IX MTsN 9 Bantul akhiri Try Out TKAD Dikpora 1 dengan ujian IPA.
- 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>
- Michos, K., Schmitz, M. L., & Petko, D. (2023). Teachers' data literacy for learning analytics: A central predictor for digital data use in upper secondary schools. *Education and Information Technologies*, 28, 14453–14471. <https://doi.org/10.1007/s10639-023-11772-y>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. Publications Office of the European Union. <https://doi.org/10.2760/159770>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>