



**PENGEMBANGAN INSTRUMEN PENILAIAN TES HOTS BERBANTUAN
WAYGROUND PADA ELEMEN TEKNIK DASAR AKTIVITAS
PERKANTORAN DI BIDANG MPLB**

Rifda Mufida Hidayati¹, Fitriana Rahmawati²

Universitas Negeri Surabaya, Indonesia^{1,2}

e-mail: rfdmh97@gmail.com

Diterima: 14/05/2026; Direvisi: 29/05/2026; Diterbitkan: 03/06/2026

ABSTRAK

Problematika rendahnya kecakapan berpikir tingkat tinggi (*Higher Order Thinking Skills/HOTS*) siswa pada pendidikan vokasi masih menjadi hambatan dalam evaluasi pembelajaran, khususnya pada elemen teknik dasar aktivitas perkantoran. Perangkat penilaian yang digunakan pendidik umumnya masih berorientasi pada kemampuan berpikir tingkat rendah dan belum mengoptimalkan media digital. Di sisi lain, riset mengenai instrumen HOTS berbasis aplikasi *Wayground* pada kompetensi Keahlian Manajemen Perkantoran dan Layanan Bisnis (MPLB) masih sangat terbatas. Penelitian *Research and Development (R&D)* menggunakan model 4D (*define, design, develop, disseminate*) ini bertujuan menghasilkan instrumen penilaian berbasis HOTS berbantuan *Wayground* yang valid, reliabel, dan interaktif. Uji coba dilakukan terhadap 40 siswa kelas X MPLB di SMK PGRI 2 Sidoarjo. Data dihimpun melalui wawancara, kuesioner validasi ahli, dan uji coba instrumen, yang kemudian dianalisis menggunakan perangkat lunak Anates V4.0.2 untuk mengukur validitas, reliabilitas, indeks kesukaran, daya pembeda, serta efektivitas pengecoh. Hasil riset menunjukkan bahwa instrumen berbasis *Wayground* ini masuk dalam kualifikasi sangat layak serta memenuhi standar validitas dan reliabilitas yang tinggi. Mayoritas butir soal memiliki tingkat kesulitan moderat dengan daya pembeda yang proporsional sehingga efektif untuk evaluasi. Sementara itu, pengukuran capaian HOTS siswa berada pada kategori baik. Dengan demikian, perangkat asesmen berbantuan *Wayground* ini direkomendasikan sebagai alternatif evaluasi pembelajaran yang interaktif, sistematis, dan mampu menstimulasi nalar kritis siswa.

Kata Kunci: *Higher Order Thinking Skills, Wayground, instrumen penilaian tes, evaluasi pembelajaran, MPLB*

ABSTRACT

The deficit in students' Higher Order Thinking Skills (HOTS) within vocational education persists as a challenge in learning evaluation, particularly in the basic office activity techniques element. Assessment tools currently used are predominantly configured toward lower-order cognition and lack of digital media optimization. Furthermore, studies examining *Wayground*-assisted HOTS assessment tools within the Office Management and Business Services (MPLB) program remain scarce. This *Research and Development (R&D)* study, utilizing the 4D model (*define, design, develop, disseminate*), aimed to develop a valid, reliable, and interactive HOTS-based assessment instrument assisted by *Wayground*. The subjects were 40 tenth-grade students of MPLB at SMK PGRI 2 Sidoarjo. Data were gathered via interviews, expert validation sheets, and trials, then analyzed using Anates V4.0.2 to evaluate validity, reliability, difficulty level, discrimination index, and distractor effectiveness. The outcomes demonstrated that the *Wayground*-assisted tool achieved a highly feasible designation and fulfilled valid and



reliable criteria. Most items exhibited moderate difficulty with good discrimination power, making them effective for evaluation. Meanwhile, the measurement of students' HOTS fell into the good category. Consequently, this Wayground-backed framework is recommended as an interactive and systematic alternative to optimize evaluation while nurturing students' critical thinking.

Keywords: *Higher Order Thinking Skills, Wayground, test assessment instrument, learning evaluation, MPLB*

PENDAHULUAN

Akselerasi pendidikan pada abad ke-21 mengarahkan orientasi pembelajaran agar tidak hanya berfokus pada penguasaan ranah teoretis-akademik, melainkan juga pada pembekalan kecakapan hidup yang relevan dengan dinamika zaman. Transformasi instruksional modern dirancang untuk menguatkan penalaran kritis, daya kreativitas, kecakapan berkomunikasi, serta sinergi kolaboratif sebagai modalitas fundamental bagi peserta didik dalam mengarungi disrupsi teknologi dan pergeseran sosial (Narassati et al., 2021). Desakan global ini mengharuskan institusi pendidikan merekayasa sistem pembelajaran yang akomodatif terhadap stimulasi kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS). Konstruk HOTS melampaui batas kognisi konvensional yang sekadar mengandalkan memori atau hafalan; dimensi ini menuntut aktivitas mental yang lebih kompleks mencakup dekonstruksi analitis, penilaian kritis (*evaluating*), pemecahan masalah secara solutif, hingga penciptaan gagasan-gagasan baru yang orisinal (Widana, 2017).

Pengarusutamaan nalar kritis ini menjadi pilar utama dalam filosofi Kurikulum Merdeka yang menempatkan siswa sebagai subjek sentral pembelajaran guna menyongsong kompetensi masa depan. Nasution et al. (2024) menggarisbawahi bahwa rekonstruksi instrumen asesmen yang bernuansa HOTS mendesak untuk dikembangkan agar evaluasi pendidikan mampu mendeteksi kedalaman berlogika siswa secara komprehensif. Sejalan dengan pandangan tersebut, Muchtar et al. (2024) menegaskan bahwa draf instrumen evaluasi yang didesain secara adaptif berbasis HOTS terbukti efektif menstimulasi daya analitis dan ketajaman pemecahan masalah selama proses belajar. Pada ranah pendidikan vokasi, urgensi penguasaan HOTS menjadi kian mutakhir. Lulusan sekolah kejuruan tidak lagi cukup dibekali kompetensi teknis (*hard skills*) yang bersifat mekanis, melainkan dituntut memiliki fleksibilitas kognitif untuk mengambil keputusan strategis, menformulasikan solusi taktis, serta tanggap terhadap pergeseran industri (Januariawan et al., 2020).

Kebutuhan mendesak mengenai revitalisasi HOTS di Indonesia dikonfirmasi oleh potret empiris hasil survei *Programme for International Student Assessment* (PISA) edisi 2022, yang merilis fakta bahwa indeks berpikir analitis dan kreatif pelajar Indonesia masih tertinggal dalam klaster rendah dibanding performa global (OECD, 2023). Koreksi fundamental dari data makro tersebut mengirimkan pesan bahwa ekosistem pembelajaran dan pola evaluasi domestik harus dirombak lewat sentuhan inovasi edukatif. Asesmen tidak boleh lagi diposisikan secara sempit sebagai alat penghakiman nilai akhir (*assessment of learning*), melainkan harus difungsikan sebagai katalisator perbaikan mutu instruksional (*assessment for/as learning*) (Shofiah et al., 2023). Oleh karena itu, perkawinan antara teknologi digital dan instrumen pengujian menjadi alternatif solutif untuk mengoptimalkan efisiensi serta akurasi asesmen berbasis HOTS.

Pemanfaatan instrumen digital dalam evaluasi memberikan peluang besar bagi pendidik untuk memformulasikan media penilaian yang dinamis, interaktif, dan rekreatif. Menurut Azizah dan Hidayat (2024), aplikasi asesmen digital yang mengadopsi karakteristik permainan edukatif terbukti mampu mendongkrak keterlibatan psikologis (*engagement*) siswa selama



ujian. Sentuhan teknologi dalam kegiatan evaluasi juga diidentifikasi mampu mencairkan ketegangan kelas, menciptakan iklim belajar yang aktif, serta merangsang motivasi belajar (Aini et al., 2025). Salah satu platform mutakhir yang potensial diintegrasikan dalam fungsi ini adalah *Wayground*. Media ini memfasilitasi guru untuk mendesain butir soal secara interaktif dengan diversifikasi format yang adaptif terhadap target kurikulum. Keunggulan sistem penilaian otomatis (*automated scoring*) pada platform ini juga menyederhanakan beban kerja administratif guru menjadi lebih efisien. Selaras dengan studi Pitoyo dan Asib (2020), pengkondisian asesmen bermuatan gamifikasi memberikan dampak linier pada retensi dan partisipasi aktif siswa. Pola ini diperkuat oleh Jong dan Tacoh (2024) yang menyatakan bahwa insersi kuis interaktif sukses memicu motivasi belajar melalui penciptaan atmosfer kompetisi akademis yang sehat dan menyenangkan.

Namun demikian, potret kontras ditemukan berdasarkan hasil wawancara dengan guru kelas X Program Keahlian Manajemen Perkantoran di SMK PGRI 2 Sidoarjo, di mana perancangan instrumen HOTS masih membentur barikade hambatan teknis. Pendidik kerap mengalami dilema dalam memilih diksi operasional yang presisi, sehingga butir soal sering kali terjebak dalam bias makna yang multitafsir dan membingungkan siswa. Problem ini diperparah oleh resistensi siswa yang enggan menelaah teks stimulus yang panjang, padahal karakteristik soal HOTS menuntut keandalan dalam membaca konteks secara utuh. Selain itu, variasi format soal masih monoton pada model pilihan ganda konvensional, isian pendek, benar-salah, atau esai, sementara pemanfaatan tipe soal modern seperti penalaran sebab-akibat serta pilihan ganda kompleks belum pernah dijamah. Kesenjangan ini menunjukkan perlunya restrukturisasi perangkat evaluasi HOTS, khususnya dalam diversifikasi bentuk tes dan estetika visual. Variasi format soal sangat krusial diimplementasikan agar instrumen penilaian mampu memotret spektrum kognitif siswa secara multi-dimensi, baik dari sudut pandang analisis maupun logika penalaran (Listiani & Rachmawati, 2022).

Di samping aspek variasi formal, objektivitas hasil pengukuran mensyaratkan standarisasi kualitas psikometrik soal yang ketat, mencakup pemenuhan kriteria validitas, indeks reliabilitas, derajat kesulitan yang proporsional, daya pembeda yang tajam, serta fungsi pengecoh (*distractor*) yang bekerja efektif (Saputri et al., 2023). Analisis mendalam terhadap kualitas butir soal merupakan tahapan imperatif dalam riset pengembangan guna menjamin kelayakan alat ukur sebelum dioperasikan secara massal (Sijabat et al., 2024). Dengan demikian, pengembangan instrumen penilaian berbasis HOTS ini tidak sekadar mengejar kebaruan bungkus digital, melainkan tetap menjaga kualitas konstruk substantif dari soal yang diproduksi.

Penetapan elemen teknik dasar aktivitas perkantoran sebagai lokus kajian didasarkan pada posisi strategis kompetensi ini sebagai materi fondasional pada Program Keahlian Manajemen Perkantoran dan Layanan Bisnis (MPLB) di SMK. Elemen ini memuat materi esensial mengenai keterampilan administrasi, tata pelayanan, pola komunikasi bisnis, serta tata kelola kerja perkantoran yang menjadi bekal awal sebelum siswa melangkah ke tingkat kejuruan yang lebih spesifik. Unit materi ini bersinggungan langsung dengan kecakapan pemberian layanan secara efisien, komunikatif, dan profesional sesuai standar dunia usaha dan dunia industri (DUDI). Muslikhin et al. (2025) menekankan bahwa performa pelayanan prima memiliki kontribusi determinan dalam mengeskalasi kepuasan klien serta membentuk reputasi profesional lembaga. Di sisi lain, Wulandari dan Lina (2025) memaparkan bahwa mutu pelayanan sangat dipengaruhi oleh kelenturan komunikasi, kecepatan respons, dan sensitivitas empati terhadap pelanggan. Atas dasar itu, elemen teknik dasar aktivitas perkantoran dinilai



sangat ideal diangkat sebagai objek pengembangan tes HOTS karena mampu merefleksikan profil keterampilan nyata yang dibutuhkan siswa dalam kompetensi kerja.

Berangkat dari untaian latar belakang tersebut, riset mengenai inovasi pengembangan instrumen penilaian berbasis HOTS berbantuan platform *Wayground* pada elemen teknik dasar aktivitas perkantoran mendesak untuk diwujudkan demi mendukung pergeseran sistem evaluasi yang adaptif terhadap roh Kurikulum Merdeka. Kehadiran produk ini diharapkan dapat memberikan asistensi praktis bagi guru dalam menyusun bank soal yang variatif dan bermutu, meningkatkan level keterlibatan siswa dalam suasana ujian, serta mempercepat pematangan kemampuan berpikir tingkat tinggi siswa vokasi. Oleh karenanya, riset ini diarahkan untuk mengonstruksi instrumen penilaian berbasis *Higher Order Thinking Skills* (HOTS) berbantuan media *Wayground* bagi kelas X MPLB di SMK PGRI 2 Sidoarjo, sekaligus menguliti kelayakan teknis perangkat dari dimensi validitas, reliabilitas, tingkat kesukaran, daya pembeda, serta respons empiris peserta didik terhadap pemanfaatan platform tersebut.

METODE PENELITIAN

Desain penelitian ini menerapkan metode *Research and Development* (R&D) yang diorientasikan untuk merekayasa produk instrumen asesmen berbasis *Higher Order Thinking Skills* (HOTS) berbantuan platform *Wayground* pada elemen teknik dasar aktivitas perkantoran bagi siswa kelas X MPLB di SMK PGRI 2 Sidoarjo. Pendekatan R&D diintegrasikan dalam studi ini karena kemampuannya dalam memformulasikan produk edukatif melalui tahapan sistematis yang teruji validitas serta fungsionalitasnya (Widodo, 2021). Prosedur pengembangan yang diadopsi adalah model 4D, yang memayungi empat fase utama: *define* (pendefinisian), *design* (perancangan), *develop* (pengembangan), dan *disseminate* (penyebaran) (Johan et al., 2023). Kerangka kerja ini dinilai sangat ideal karena memiliki sistematika operasional yang terstruktur dan selaras untuk mengakomodasi rancang bangun perangkat evaluasi berbasis kemampuan berpikir tingkat tinggi (Hamzah et al., 2022).

Pada pelaksanaannya, fase *define* diaktualisasikan melalui identifikasi kebutuhan lapangan, pemetaan karakteristik psikologis dan kognitif peserta didik, analisis substansi konsep materi, penelaahan tugas, serta perumusan indikator tujuan pembelajaran. Langkah berikutnya adalah fase *design*, yang mencakup kodifikasi kisi-kisi instrumen, sinkronisasi fitur-fitur pada media *Wayground*, penetapan diversifikasi format soal, hingga penyusunan draf awal (*prototype*) instrumen penilaian. Memasuki fase *develop*, serangkaian uji justifikasi dilakukan melalui validasi oleh trias pakar yang meliputi ahli materi, ahli asesmen, serta ahli kebahasaan guna menjamin pemenuhan kualifikasi kelayakan teoretis. Setelah perbaikan dilakukan, fase *disseminate* direalisasikan dalam skala terbatas kepada 40 peserta didik kelas X Program Keahlian Manajemen Perkantoran di SMK PGRI 2 Sidoarjo untuk menguji performa empiris produk di kelas nyata.

Data yang dihimpun dalam riset ini mengombinasikan data kuantitatif dan kualitatif. Representasi data kuantitatif bersumber dari kalkulasi numerik uji validitas, koefisien reliabilitas, indeks tingkat kesukaran, daya pembeda, serta pemetaan efektivitas fungsi pengecoh (*distractor analysis*) butir soal, yang seluruhnya diproses dengan mengoperasikan perangkat lunak komputasi Anates V4.0.2 karena kepraktisannya dalam menyajikan output statistik secara menyeluruh (Puteri et al., 2025). Di sisi lain, data kualitatif disintesis dari hasil wawancara mendalam serta catatan korektif pada lembar penilaian para validator. Angka pada lembar validasi ditakar menggunakan instrumen kuesioner berskala Likert rentang 1–5 (Wahjusaputri & Purwanto, 2022). Output persentase skor dari hasil tabulasi data validasi

tersebut selanjutnya dikonversikan dan diinterpretasikan merujuk pada standar kriteria kelayakan produk yang dipaparkan pada Tabel 1.

Tabel 1. Kriteria Penilaian Kelayakan

Persentase Kriteria	
0%–20%	Sangat Tidak Layak
21%–40%	Tidak Layak
41%–60%	Cukup Layak
61%–80%	Layak
81%–100%	Sangat Layak

Sumber: Wahjusaputri dan Purwanto (2022)

Berdasarkan Tabel 1, tingkat kelayakan produk ditentukan melalui persentase hasil validasi ahli. Semakin tinggi persentase yang diperoleh, maka kualitas instrumen yang dikembangkan semakin layak untuk digunakan dalam proses evaluasi pembelajaran.

Nilai hasil pengerjaan soal HOTS digunakan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik dengan rumus berikut:

$$P = \frac{\text{Skor peserta didik}}{\text{Skor maksimum}} \times 100$$

Hasil perolehan nilai selanjutnya diklasifikasikan ke dalam kategori kemampuan berpikir tingkat tinggi sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Interpretasi Kemampuan Berpikir Tingkat Tinggi

Nilai Peserta Didik Tingkat Kemampuan	
81–100	Sangat Baik
61–80	Baik
41–60	Cukup
21–40	Kurang
<20	Sangat Kurang

Sumber: Purbaningrum (2017)

Berdasarkan Tabel 2, kemampuan berpikir tingkat tinggi peserta didik diinterpretasikan berdasarkan rentang skor yang diperoleh setelah mengerjakan instrumen HOTS. Kategori tersebut digunakan untuk mempermudah analisis tingkat kemampuan berpikir kritis dan analitis peserta didik pada elemen teknik dasar aktivitas perkantoran.

HASIL DAN PEMBAHASAN

Hasil

Proses Pengembangan Instrumen Penilaian Tes Berbasis HOTS Berbantuan Wayground pada Elemen Teknik Dasar Aktivitas Perkantoran di Bidang Manajemen Perkantoran dan Layanan Bisnis

Tahap *define* (pendefinisian) diawali dengan analisis awal dan analisis akhir untuk mengidentifikasi kondisi pembelajaran di sekolah. Hasil analisis menunjukkan bahwa SMK PGRI 3 Sidoarjo telah menerapkan Kurikulum Merdeka dengan sistem asesmen yang mengarah pada kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skills / HOTS*) pada level kognitif C4 sampai C6. Meskipun demikian, proses pembelajaran masih menghadapi kendala berupa rendahnya minat membaca peserta didik dalam memahami stimulus soal. Kondisi



tersebut menyebabkan sebagian peserta didik mengalami kesulitan dalam menganalisis konteks permasalahan yang disajikan pada soal HOTS.

Analisis karakteristik peserta didik menunjukkan bahwa siswa kelas X lebih responsif terhadap pembelajaran berbasis digital yang bersifat kontekstual, visual, dan interaktif. Selanjutnya, analisis konsep dan tujuan pembelajaran dilakukan dengan mengacu pada capaian pembelajaran elemen teknik dasar aktivitas perkantoran yang meliputi pelayanan prima, teknik pelayanan prima, layanan pelanggan, serta prosedur dan instruksi kerja. Hasil analisis tugas memperlihatkan bahwa variasi bentuk soal HOTS yang digunakan guru masih terbatas pada soal pilihan ganda sederhana, sedangkan bentuk soal pilihan ganda kompleks dan sebab-akibat belum banyak digunakan. Oleh karena itu, diperlukan pengembangan instrumen penilaian tes berbasis HOTS yang lebih variatif dan sesuai dengan capaian pembelajaran pada Kurikulum Merdeka.

Tahap *design* (perancangan) dilakukan melalui penyusunan standar tes dan pengembangan kisi-kisi instrumen sebagai pedoman penyusunan soal. Kisi-kisi disusun secara sistematis agar setiap butir soal memiliki kesesuaian dengan indikator, materi, tujuan pembelajaran, serta level kognitif HOTS. Instrumen penilaian yang dikembangkan terdiri atas 30 butir soal dengan variasi bentuk soal pilihan ganda, pilihan ganda kompleks, dan sebab-akibat pada level kognitif C4, C5, dan C6. Media yang digunakan dalam implementasi instrumen adalah aplikasi Wayground yang dapat diakses secara daring melalui gawai maupun komputer. Pemilihan media tersebut dilakukan untuk mendukung pelaksanaan evaluasi pembelajaran yang lebih interaktif, fleksibel, dan sesuai dengan karakteristik peserta didik pada era digital.

Tahap *develop* (pengembangan) meliputi validasi ahli, revisi produk, uji coba instrumen, dan analisis kualitas butir soal. Validasi dilakukan oleh ahli materi, ahli bahasa, ahli asesmen, serta pendidik. Hasil validasi menunjukkan bahwa instrumen memperoleh persentase kelayakan sebesar 88% dari ahli materi, 96% dari ahli bahasa, 95% dari ahli asesmen, dan 93% dari pendidik. Seluruh hasil validasi tersebut termasuk dalam kategori sangat layak. Saran dari validator digunakan sebagai dasar revisi instrumen, terutama pada aspek kejelasan konteks soal, penyederhanaan bahasa, penggunaan istilah yang lebih komunikatif, dan penambahan unsur visual pada beberapa soal HOTS.

Setelah direvisi, instrumen diujicobakan kepada 40 peserta didik kelas X MPLB SMK PGRI 2 Sidoarjo. Uji coba dilakukan menggunakan 30 butir soal dengan sistem penskoran 1 untuk jawaban benar dan 0 untuk jawaban salah. Hasil analisis butir soal menunjukkan bahwa 27 soal dinyatakan valid dan 3 soal tidak valid. Koefisien reliabilitas instrumen sebesar 0,84 termasuk dalam kategori tinggi, sehingga menunjukkan bahwa instrumen memiliki tingkat konsistensi yang baik. Berdasarkan tingkat kesukaran, diperoleh 7 soal berkategori mudah dan 23 soal berkategori sedang. Pada analisis daya pembeda diperoleh 10 soal berkategori baik sekali, 14 soal berkategori baik, 3 soal cukup, 2 soal jelek, dan 1 soal bernilai negatif sehingga disarankan untuk dibuang. Selain itu, sebagian besar pengecoh telah berfungsi efektif, meskipun masih terdapat beberapa opsi jawaban yang perlu direvisi karena kurang mampu menarik perhatian peserta didik.

Tahap terakhir dalam model 4-D adalah *disseminate* (penyebaran). Pada penelitian ini, penyebaran dilakukan secara terbatas, yaitu hanya pada peserta didik kelas X Jurusan Manajemen Perkantoran SMK PGRI 2 Sidoarjo yang berjumlah 40 orang. Penyebaran terbatas dilakukan untuk mengetahui keterlaksanaan penggunaan instrumen penilaian sebelum diterapkan secara lebih luas.

Kelayakan Instrumen Penilaian Tes Berbasis HOTS Berbantuan Wayground pada Elemen Teknik Dasar Aktivitas Perkantoran di Bidang Manajemen Perkantoran dan Layanan Bisnis

Kelayakan instrumen penilaian tes berbasis HOTS berbantuan Wayground dianalisis melalui validasi ahli, penilaian pendidik, serta pengujian kualitas butir soal yang meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Tahapan tersebut dilakukan untuk memastikan bahwa instrumen yang dikembangkan memenuhi kriteria kualitas instrumen penilaian yang baik sehingga layak digunakan sebagai alat evaluasi pembelajaran.

Tabel 3. Hasil Validasi Para Ahli

No	Jenis Validator	Persentase Skor	Kategori Kelayakan
1	Ahli Materi	88%	Sangat Layak
2	Ahli Bahasa	96%	Sangat Layak
3	Ahli Asesmen	95%	Sangat Layak
4	Penilaian Pendidik	93%	Sangat Layak
Rata-Rata Skor		93%	Sangat Layak

Berdasarkan Tabel 3, hasil validasi menunjukkan bahwa instrumen penilaian tes berbasis HOTS memperoleh rata-rata persentase kelayakan sebesar 93% dengan kategori sangat layak. Hasil tersebut menunjukkan bahwa instrumen telah memenuhi aspek materi, kebahasaan, konstruksi asesmen, dan kesesuaian penggunaan dalam pembelajaran.

Validasi ahli materi memperoleh persentase sebesar 88% dengan kategori sangat layak. Penilaian meliputi kesesuaian soal dengan capaian pembelajaran, indikator soal, materi, dan karakteristik HOTS sesuai taksonomi Bloom revisi. Validator memberikan masukan berupa penambahan penjelasan konteks pada beberapa soal dan penyederhanaan istilah agar lebih sesuai dengan karakteristik peserta didik.

Validasi ahli bahasa memperoleh persentase sebesar 96% dengan kategori sangat layak. Aspek yang dinilai mencakup kejelasan bahasa, penggunaan kalimat komunikatif, kesesuaian dengan kaidah kebahasaan, serta ketepatan struktur kalimat. Masukan yang diberikan meliputi penyederhanaan kalimat, penyesuaian istilah asing, dan perbaikan redaksi soal agar lebih mudah dipahami oleh peserta didik.

Sementara itu, validasi ahli asesmen memperoleh persentase sebesar 95% dengan kategori sangat layak. Penilaian difokuskan pada konstruksi soal, kesesuaian karakteristik HOTS, penggunaan bahasa baku, dan kejelasan penyajian stimulus berupa tabel maupun gambar. Validator menyarankan penambahan unsur visual pada beberapa soal HOTS agar konteks permasalahan lebih jelas dan mudah dipahami peserta didik.

Penilaian pendidik memperoleh persentase sebesar 93% dengan kategori sangat layak. Pendidik menilai bahwa instrumen telah sesuai dengan indikator pembelajaran dan mampu melatih kemampuan berpikir kritis peserta didik. Selain itu, instrumen dinilai dapat digunakan sebagai alternatif alat evaluasi pada pembelajaran MPLB. Meskipun demikian, pendidik menyarankan adanya perbaikan pada ketepatan penggunaan bahasa agar soal lebih komunikatif.

Analisis validitas butir soal dilakukan menggunakan perangkat lunak Anates V4.0.2 dengan nilai batas signifikansi sebesar 0,393 berdasarkan jumlah responden sebanyak 40 peserta didik. Hasil analisis menunjukkan bahwa dari 30 butir soal yang diujikan, sebanyak 27 butir soal dinyatakan valid, sedangkan 3 butir soal tidak valid, yaitu nomor 10, 21, dan 23. Butir soal nomor 10 memiliki nilai koefisien korelasi sebesar 0,253 sehingga berada di bawah batas



signifikansi. Sementara itu, butir soal nomor 21 dan 23 memiliki nilai korelasi negatif masing-masing sebesar -0,004 dan -0,066 sehingga tidak memenuhi kriteria validitas.

Hasil analisis reliabilitas menunjukkan koefisien reliabilitas sebesar 0,84 dengan kategori tinggi. Hasil tersebut menunjukkan bahwa instrumen memiliki tingkat konsistensi yang baik sehingga layak digunakan sebagai alat evaluasi pembelajaran.

Analisis tingkat kesukaran menunjukkan bahwa instrumen didominasi oleh soal berkategori sedang. Terdapat 7 butir soal berkategori mudah, yaitu nomor 1, 5, 8, 9, 11, 13, dan 27, sedangkan 23 butir soal lainnya termasuk kategori sedang. Dominasi soal dengan tingkat kesukaran sedang menunjukkan bahwa instrumen mampu mengukur kemampuan peserta didik secara lebih proporsional.

Pada analisis daya pembeda diperoleh 10 butir soal berkategori baik sekali, 14 butir soal berkategori baik, 3 butir soal berkategori cukup, 2 butir soal berkategori jelek, dan 1 butir soal bernilai negatif. Hasil tersebut menunjukkan bahwa sebagian besar soal telah mampu membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Namun demikian, soal dengan kategori cukup dan jelek perlu direvisi, sedangkan soal dengan nilai negatif disarankan untuk tidak digunakan.

Analisis efektivitas pengecoh menunjukkan bahwa sebagian besar alternatif jawaban pada setiap soal telah berfungsi dengan baik. Hal tersebut terlihat dari adanya pilihan jawaban selain kunci yang dipilih oleh lebih dari 5% peserta didik. Kondisi ini menunjukkan bahwa pengecoh mampu menarik perhatian peserta didik yang belum memahami materi secara optimal. Akan tetapi, masih terdapat beberapa pengecoh yang belum berfungsi efektif karena dipilih oleh kurang dari 5% peserta didik, yaitu pada soal nomor 1, 3, 4, 5, 9, 13, 16, 23, 25, dan 26. Oleh karena itu, opsi jawaban tersebut perlu direvisi agar lebih menarik dan relevan.

Keterampilan Berpikir Tingkat Tinggi Peserta Didik Melalui Pengerjaan Instrumen Penilaian Tes

Kemampuan berpikir tingkat tinggi peserta didik dianalisis berdasarkan hasil pengerjaan instrumen penilaian tes berbasis HOTS. Analisis dilakukan untuk mengetahui distribusi kemampuan peserta didik dalam memahami, menganalisis, mengevaluasi, dan memecahkan masalah yang disajikan pada soal.

Tabel 4. Distribusi Kemampuan Berpikir Tingkat Tinggi Peserta Didik

Kategori	Jumlah Peserta Didik	Persentase
Sangat Baik	10	25%
Baik	12	30%
Cukup	12	30%
Kurang	6	15%

Berdasarkan Tabel 4, kemampuan berpikir tingkat tinggi peserta didik didominasi oleh kategori baik dan sangat baik dengan persentase sebesar 55%. Hasil tersebut menunjukkan bahwa sebagian besar peserta didik telah mampu memahami dan menyelesaikan soal HOTS dengan cukup baik. Peserta didik pada kategori ini umumnya mampu menganalisis informasi, menentukan solusi permasalahan, serta memberikan jawaban berdasarkan penalaran yang logis.

Meskipun demikian, persentase peserta didik pada kategori cukup dan kurang masih mencapai 45%. Kondisi tersebut menunjukkan bahwa kemampuan berpikir tingkat tinggi peserta didik belum berkembang secara merata. Sebagian peserta didik masih mengalami kesulitan dalam memahami stimulus soal, menghubungkan konsep, serta menentukan jawaban



berdasarkan analisis yang mendalam. Oleh karena itu, diperlukan upaya pembelajaran yang lebih intensif dan berkelanjutan untuk meningkatkan kemampuan HOTS peserta didik secara menyeluruh.

Pembahasan

Konstruksi perangkat penilaian tes berbasis *Higher Order Thinking Skills* (HOTS) dengan mengintegrasikan platform *Wayground* dalam penelitian ini mengadopsi cetak biru model pengembangan 4D yang memayungi tahapan *define, design, develop, dan disseminate*. Pemilihan kerangka kerja ini didasarkan pada keunggulannya dalam menyajikan panduan sistematis guna mengonstruksi produk evaluasi pendidikan yang valid serta akomodatif. Implementasi model 4D dalam rekayasa perangkat pembelajaran terbukti andal karena menyediakan ruang untuk proses evaluasi berkala dan perbaikan berkelanjutan (*formative feedback loop*) pada setiap fasenya sebelum instrumen diujicobakan secara riil kepada subjek sasaran (Johan dkk., 2023).

Fase awal diaktualisasikan melalui tahap *define* atau pendefinisian, yang diproyeksikan untuk memetakan urgensi kebutuhan di lapangan serta mengidentifikasi rintangan sistemik dalam dinamika evaluasi pembelajaran. Merujuk pada data hasil wawancara bersama pendidik pengampu mata pelajaran Dasar-Dasar Manajemen Perkantoran di SMK PGRI 2 Sidoarjo, terkonfirmasi bahwa institusi tersebut telah mengadopsi Kurikulum Merdeka di seluruh tingkat kelas. Kendati demikian, potret aktual menunjukkan bahwa format asesmen yang dioperasikan masih didominasi oleh ragam soal pilihan ganda konvensional, bentuk benar-salah, serta ulasan studi kasus linier. Diversifikasi instrumen HOTS yang lebih kompleks, seperti format pilihan ganda kompleks maupun tipe penalaran sebab-akibat, belum diimplementasikan secara optimal. Kondisi riil ini mempertegas bahwa perangkat evaluasi yang tersedia belum sepenuhnya mampu mendeteksi maupun menstimulasi kapasitas penalaran tingkat tinggi siswa.

Hasil identifikasi terhadap karakteristik peserta didik menunjukkan sebuah kecenderungan di mana siswa kelas X Program Keahlian Manajemen Perkantoran mengindikasikan level partisipasi yang lebih intensif ketika proses pembelajaran maupun evaluasi dikemas lewat media digital interaktif. Temuan empiris ini mengindikasikan adanya preferensi kuat dari generasi digital asli (*digital natives*) terhadap ekosistem pengujian berbasis teknologi yang mampu menghadirkan pengalaman evaluasi yang lebih menantang dan kontekstual. Aini dkk. (2025) memaparkan bahwa asimilasi teknologi dalam aktivitas instruksional berkontribusi linier terhadap peningkatan keterlibatan psikologis peserta didik, mengingat media digital interaktif memiliki fleksibilitas untuk menciptakan atmosfer pembelajaran yang inovatif dan partisipatif. Indikator ini menguatkan tesis bahwa pelibatan platform digital ke dalam skema evaluasi berkorelasi positif dengan eskalasi motivasi dan retensi siswa.

Di samping pemetaan profil peserta didik, fase *define* juga melibatkan analisis mendalam terhadap struktur konsep serta indikator tujuan pembelajaran. Fokus materi dalam pengembangan perangkat tes ini dispesifikasikan pada elemen teknik dasar aktivitas perkantoran, yang mencakup pilar pelayanan prima (*excellent service*), manajemen layanan pelanggan, serta standardisasi prosedur kerja perkantoran. Rumpun materi ini dipilih karena memiliki densitas yang kaya akan problem kognitif yang membutuhkan ketajaman berpikir analitis dan formulasi pemecahan masalah. Dalam domain pendidikan vokasi, konsep pelayanan prima tidak lagi sekadar diposisikan sebagai hafalan teoretis, melainkan menuntut kecakapan siswa dalam menguraikan problem pelayanan yang dinamis serta menformulasikan keputusan taktis yang presisi sesuai realitas industri. Muslikhin dkk. (2025) menegaskan bahwa



kompetensi pelayanan prima memegang peranan krusial dalam peningkatan kualitas layanan karena mengonstruksi kemampuan individu untuk menguraikan ekspektasi pelanggan sekaligus mengambil keputusan krusial secara tepat dalam situasi darurat.

Melangkah pada fase berikutnya, yaitu *design* (perancangan), aktivitas difokuskan pada kodifikasi kisi-kisi instrumen tes yang diselaraskan dengan capaian pembelajaran (CP) dan tujuan pembelajaran (TP) pada elemen teknik dasar aktivitas perkantoran. Penyusunan matriks kisi-kisi ini berfungsi sebagai jangkar teoritis agar setiap butir soal memiliki korelasi linier dengan indikator kompetensi serta taksonomi kognitif level HOTS. Perangkat evaluasi yang divalidasi ini melingkupi 30 butir soal yang memuat variasi format berupa pilihan ganda reguler, pilihan ganda kompleks, dan tipe penalaran sebab-akibat. Diversifikasi struktur soal tersebut dirancang secara sengaja untuk menguji kapasitas siswa dalam aspek dekonstruksi informasi (analisis), penaksiran situasi (evaluasi), hingga penemuan solusi atas problematika yang disajikan.

Wahana digital yang diposisikan sebagai basis implementasi instrumen ini adalah platform *Wayground*. Pengintegrasian media berbasis awan ini didasarkan pada kapabilitasnya untuk menyajikan sistem evaluasi yang responsif, interaktif, serta praktis diakses melalui gawai portabel milik peserta didik. Pemanfaatan perangkat lunak kuis interaktif ini dinilai efektif mereduksi kecemasan ujian (*test anxiety*) sekaligus menstimulasi gairah kompetisi akademis yang sehat. Jong dan Tacoh (2024) menjelaskan bahwa utilisasi aplikasi kuis interaktif mampu mengeskalasi motivasi belajar siswa disebabkan oleh kemampuannya mengubah ritme evaluasi menjadi lebih rekreatif dan tidak menjemukan. Hasil ini selaras dengan kesimpulan ilmiah Pitoyo dan Asib (2020) yang mengemukakan bahwa asesmen bermuatan gamifikasi (*gamified assessment*) memberikan implikasi positif terhadap eskalasi keterlibatan serta pengalaman belajar yang mendalam bagi peserta didik.

Formulasi instrumen berbasis HOTS dalam penelitian ini juga menjadi representasi nyata dari amanat Kurikulum Merdeka yang menekankan pada penguatan nalar kritis, orisinalitas kreatif, dan kemandirian memecahkan masalah. Widana (2017) memaparkan bahwa HOTS merupakan jembatan kognitif tingkat lanjut yang mengintegrasikan proses analisis, evaluasi, hingga penciptaan solusi atas sebuah kompleksitas masalah. Atas dasar itu, insersi variasi tipe soal sekelas pilihan ganda kompleks dan penalaran sebab-akibat dinilai sangat relevan untuk membedah kedalaman kapasitas HOTS siswa secara lebih komprehensif dibandingkan dengan model tes konvensional.

Bila ditinjau dari perspektif teori tes klasik (*Classical Test Theory*), kualitas psikometrik dari format soal pilihan ganda kompleks dan sebab-akibat tetap dapat diukur reliabilitasnya melalui indikator validitas, koefisien reliabilitas, tingkat kesukaran, serta indeks daya pembeda. Sumaryanta (2021) menguraikan bahwa derajat kelayakan sebuah perangkat tes dapat diuji secara ilmiah melalui analisis kuantitatif pada tiap-tiap butir soalnya. Konsekuensinya, instrumen yang dikonstruksi tidak hanya mengejar aspek estetika atau variasi digital semata, melainkan tetap memprioritaskan validitas empiris dari setiap butir soal guna menjamin keakuratan data hasil pengukuran. Brataningrum dan Farisayah (2025) menambahkan bahwa mutu butir soal dalam payung teori tes klasik ditentukan oleh sensitivitas soal tersebut dalam mendeteksi gradasi tingkat penguasaan materi di antara peserta didik serta konsistensi metriknya dalam mengukur kompetensi target.

Tahapan selanjutnya dikristalisasikan dalam fase *develop* (pengembangan), yang mencakup pengujian oleh tim justifikasi ahli, revisi produk berdasarkan umpan balik, uji coba lapangan, serta pembedahan kualitas psikometrik butir soal. Proses validasi melibatkan sinergi dari trias pakar (ahli materi, ahli bahasa, ahli asesmen) serta praktisi pendidikan untuk menakar



derajat kelayakan produk sebelum didistribusikan pada subjek siswa. Hasil penilaian tim validator mengonfirmasi bahwa instrumen pengujian ini masuk dalam kualifikasi sangat layak, yang mengindikasikan bahwa produk ini telah memenuhi standar validitas isi, konstruksi bahasa, arsitektur soal, serta kesesuaian dengan indikator HOTS.

Secara teoretis, persentase validitas yang diraih dari ahli materi menyentuh angka 88%, diikuti ahli bahasa sebesar 96%, ahli asesmen mencapai 95%, dan dari sudut pandang penilaian praktisi pendidikan (guru) sebesar 93%. Akumulasi skor tinggi ini mengonfirmasi bahwa instrumen yang didesain telah konvergen dengan capaian pembelajaran dan tuntutan indikator HOTS. Realitas data ini sejalan dengan tesis Maulidia dan Pahlevi (2020) yang memproklamasikan bahwa sebuah perangkat penilaian HOTS dinyatakan memenuhi syarat kelayakan operasional apabila mampu mengintegrasikan keselarasan substansi materi, kokohnya konstruksi soal, dan penggunaan artikulasi bahasa yang komunikatif.

Dari total 30 butir soal yang diuji secara empiris, hasil analisis validitas menunjukkan bahwa 27 butir soal dinyatakan valid, sementara 3 butir soal dikategorikan tidak valid. Anshari dkk. (2024) menerangkan bahwa sebuah butir soal memiliki kualifikasi valid apabila koefisien korelasi yang dihasilkan melebihi ambang batas statistik pada r -tabel. Di sisi lain, Murniatie dkk. (2025) menekankan bahwa butir soal yang berhasil memenuhi standar validitas berhak digunakan sebagai alat pengumpul data utama karena terbukti presisi dalam memotret kompetensi yang dibidik. Dengan demikian, mayoritas butir soal dalam penelitian ini telah lolos sensor validitas sehingga sangat reliabel untuk diaplikasikan dalam evaluasi kelas.

Perangkat evaluasi yang direkayasa ini juga mencatatkan indeks reliabilitas yang tinggi dengan koefisien sebesar 0,84. Perolehan nilai ini menjadi bukti bahwa instrumen memiliki tingkat stabilitas dan konsistensi yang sangat baik dalam mengukur kompetensi siswa. Augustia dkk. (2025) menjabarkan bahwa instrumen pengujian dengan tingkat reliabilitas tinggi menjamin stabilitas keluaran data meskipun dioperasikan secara berulang pada subjek yang sama dengan kondisi yang setara. Data ini memperkuat posisi instrumen penilaian HOTS berbantuan *Wayground* sebagai alat ukur pembelajaran yang bermutu tinggi.

Berdasarkan pemetaan tingkat kesukaran, mayoritas butir soal mendominasi pada rentang kategori sedang (moderat). Proporsi soal berkategori sedang yang dominan ini menandakan bahwa instrumen yang dikembangkan mampu menakar kompetensi siswa secara seimbang dan tidak bias. Saputri dkk. (2023) menyebutkan bahwa butir soal dengan derajat kesulitan moderat jauh lebih efektif dalam mendeteksi perbedaan kompetensi antarsiswa dibandingkan dengan tipe soal yang terlalu mudah atau terlampau sukar. Pola temuan ini juga selaras dengan konseptualisasi Syam dkk. (2026) yang menjelaskan bahwa struktur tes yang didominasi oleh soal kategori sedang mengindikasikan kualitas instrumen yang ideal karena mampu menyajikan deskripsi kemampuan peserta didik secara objektif.

Dilihat dari indeks daya pembeda, sebagian besar butir soal bertengger pada klasifikasi baik dan sangat baik. Temuan statistik ini membuktikan bahwa soal yang disusun memiliki ketajaman untuk memisahkan antara kelompok siswa berkemampuan tinggi (*upper group*) dan kelompok siswa berkemampuan rendah (*lower group*). Kurniawan dkk. (2022) menguraikan bahwa fungsi daya pembeda ditujukan untuk mengukur sensitivitas soal dalam mendeteksi disparitas tingkat penguasaan akademis siswa. Oleh karena itu, soal dengan daya pembeda tinggi dinilai lebih fungsional dalam mengawal objektivitas evaluasi. Kendati demikian, riset ini tetap mendeteksi keberadaan beberapa butir soal berkategori cukup, jelek, bahkan bernilai negatif, yang langsung direspons dengan tindakan rekonstruksi atau eliminasi demi menjaga performa instrumen agar tetap optimal.



Analisis terhadap efektivitas pengecoh (*distractor analysis*) menunjukkan bahwa mayoritas opsi jawaban salah telah menjalankan fungsinya dengan optimal karena dipilih oleh lebih dari 5% total peserta uji coba. Indikator ini membuktikan bahwa pengecoh yang didesain memiliki daya tarik (*attractiveness*) untuk mengalihkan perhatian siswa yang belum menguasai konsep materi secara matang. Sari dkk. (2022) menjelaskan bahwa fungsi pengecoh yang ideal adalah alternatif jawaban salah yang dikonstruksi secara logis sehingga mampu menguji konsistensi siswa yang belum tuntas menguasai kompetensi. Namun, riset ini juga mengidentifikasi adanya beberapa opsi pengecoh yang tidak efektif akibat dipilih kurang dari 5% peserta. Ambarwati dan Ismiyati (2021) menyatakan bahwa opsi pengecoh yang memiliki daya pikat rendah atau terlalu kentara kesalahannya wajib direvisi guna mendongkrak kualitas intrinsik soal.

Fase final dalam siklus 4D adalah *disseminate* atau penyebaran. Dalam konteks penelitian ini, aktivitas deseminasi direalisasikan dalam skala terbatas kepada peserta didik kelas X Program Keahlian Manajemen Perkantoran di SMK PGRI 2 Sidoarjo. Kebijakan pembatasan jangkauan deseminasi ini diambil sebagai kompromi atas keterbatasan waktu pelaksanaan serta efisiensi sumber daya penelitian. Walakin, instrumen yang telah teruji ini menyimpan potensi besar untuk diadaptasi secara lebih luas dalam skala makro pada riset mendatang, sekaligus menjadi opsi perangkat evaluasi berbasis HOTS yang relevan bagi institusi sekolah menengah kejuruan secara general.

Profil keterampilan berpikir tingkat tinggi siswa dianalisis secara presisi berdasarkan lembar jawaban hasil pengerjaan instrumen HOTS berbasis *Wayground* yang telah dikembangkan. Output analisis data mengonfirmasi bahwa mayoritas peserta didik berada pada klaster kategori baik dan sangat baik dengan akumulasi persentase mencapai 55%. Fakta empiris ini mengindikasikan bahwa sebagian besar siswa telah memiliki kemampuan kognitif untuk mengurai informasi, melakukan komparasi kritis, serta mengevaluasi problem-problem kompleks yang disajikan dalam stimulus soal. Listiani dan Rachmawati (2022) mengutarakan bahwa manifestasi HOTS berakar pada kapasitas individu dalam menganalisis situasi, mengevaluasi fakta, hingga merumuskan pemecahan atas masalah yang rumit. Dengan demikian, kelompok siswa pada kategori baik dan sangat baik ini dapat dinyatakan telah memiliki kematangan logika berpikir tingkat tinggi yang memadai.

Meskipun menunjukkan tren positif, riset ini juga merekam fakta bahwa masih terdapat sekitar 45% peserta didik yang tertahan pada kategori cukup dan kurang. Data kontras ini memberikan sinyal kuat bahwa sebaran kemampuan HOTS di antara siswa belum merata secara berkeadilan. Saraswati dan Agustika (2020) menguraikan bahwa fluktuasi kecakapan berpikir tingkat tinggi pada tiap-tiap individu bersifat personal karena dipengaruhi oleh stimulasi latar belakang karakteristik personal serta rekam jejak pengalaman belajar yang dialami sebelumnya. Realitas ini menegaskan perlunya skema pembelajaran yang konsisten dan berkelanjutan dalam melatih ketajaman nalar kritis, salah satunya melalui pembiasaan paparan soal-soal HOTS secara periodik.

Secara menyeluruh, konklusi dari riset ini menegaskan bahwa konstruksi instrumen penilaian berbasis HOTS berbantuan platform *Wayground* efektif diposisikan sebagai alternatif sistem evaluasi modern yang andal dalam memetakan kapasitas berpikir tingkat tinggi peserta didik. Temuan ini sekaligus memperkuat bangunan teori dari riset-riset terdahulu yang memproklamasikan bahwa asimilasi perangkat asesmen digital bermuatan kuis interaktif terbukti ampuh mendongkrak mutu evaluasi pendidikan sekaligus menjadi instrumen strategis untuk mengakselerasi keterampilan abad ke-21 pada diri peserta didik (Narassati dkk., 2021; Mulyawan & Dwijayanti, 2024).



KESIMPULAN

Riset ini diorientasikan untuk mengonstruksi instrumen asesmen bermuatan *Higher Order Thinking Skills* (HOTS) terintegrasi platform *Wayground* pada elemen teknik dasar aktivitas perkantoran, sekaligus memetakan derajat kelayakan teknis perangkat dan memprofilkan kapasitas berpikir tingkat tinggi peserta didik. Berdasarkan hasil analisis data, dapat disimpulkan bahwa perangkat penilaian yang dikembangkan ini memiliki kualifikasi yang sangat layak untuk dioperasikan dalam ekosistem evaluasi pembelajaran. Kelayakan tersebut didasarkan pada terpenuhinya standarisasi psikometrik yang ketat, mencakup pemenuhan aspek validitas konstruk, koefisien reliabilitas yang tinggi, indeks tingkat kesukaran yang proporsional, daya pembeda yang tajam, serta berfungsinya opsi pengecoh secara efektif. Pengintegrasian media *Wayground* terbukti mampu mentransformasikan skema evaluasi tradisional menjadi lebih interaktif, sistematis, dan adaptif terhadap karakteristik kompetensi abad ke-21.

Di samping itu, output pengukuran empiris mengonfirmasi bahwa profil kecakapan HOTS peserta didik secara umum berada pada kategori cukup baik. Penyerapan instrumen evaluasi bernuansa HOTS ini secara nyata menstimulasi keterlibatan aktif siswa dalam mengurai informasi (analisis), menakar argumentasi (evaluasi), hingga memformulasikan jalan keluar atas problem-problem kontekstual pada materi teknik dasar aktivitas perkantoran.

Secara praktis, penelitian ini memberikan implikasi strategis bahwa adopsi instrumen penilaian berbasis digital bernalar kritis berpotensi besar menjadi opsi evaluasi mutakhir di lingkungan Sekolah Menengah Kejuruan (SMK), khususnya pada Program Keahlian Manajemen Perkantoran dan Layanan Bisnis (MPLB). Kehadiran produk ini dapat memberikan asistensi bagi para pendidik dalam menyusun bank soal berkualitas yang tidak lagi sekadar menguji ingatan jangka pendek siswa, melainkan berfokus pada penguatan nalar kritis, ketajaman analisis, dan fleksibilitas pemecahan masalah yang selaras dengan ekspektasi dunia usaha dan dunia industri (DUDI) di era disrupsi modern.

DAFTAR PUSTAKA

- Aini, I. Q., Noraziza, Nufus, A., & Agustin, K. N. (2025). Meningkatkan keterlibatan siswa dengan teknologi sebagai alat pembelajaran interaktif. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 10(4), 167–186. <https://doi.org/10.23969/jp.v10i4.37234>
- Ambarwati, Y. F., & Ismiyati, I. (2021). Analisis butir soal pilihan ganda ulangan akhir semester genap mata pelajaran kearsipan. *Measurement in Educational Research*, 1(2), 64–75. <https://doi.org/10.1177/074193258300400305>
- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis validitas dan reliabilitas butir soal sumatif akhir semester ganjil mata pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 964–975. <https://doi.org/10.31004/edukatif.v6i1.5931>
- Augustia, A. D., Agustia, C. N., Azzahra, D., Hakim, L., & Pratiwi, V. (2025). Analisis validitas dan reliabilitas soal pilihan ganda dengan menggunakan software Anates pada mata pelajaran perpajakan. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250–265. <https://doi.org/10.37481/jmeh.v5i1.1165>
- Azizah, N., & Hidayat, R. (2024). Pengembangan asesmen digital berbasis game edukatif Quizizz pada mata pelajaran marketing the quizz educational game based digital assessment development for marketing class. *Journal of Innovation and Teacher*



- Professionalism*, 2(2), 195–209. <https://doi.org/10.17977/um084v2i22024p195-209>
- Brataningrum, N. P., & Farisiyah, U. (2025). Karakteristik butir soal materi penggabungan badan usaha pada matakuliah akuntansi keuangan lanjutan I. *Measurement in Educational Research*, 5(1), 16–31. <https://doi.org/10.33292/meter.v5i1.391>
- Fitriani, N. (2021). Pengecoh soal pelatihan kewaspadaan kegawatdaruratan maternal dan neonatal. *Paedagoria: Jurnal Kajian, Penelitian dan Pengembangan Kependidikan*, 12(1), 199–205. <https://doi.org/10.31764/paedagoria.v12i2.4956>
- Hamzah, H., Hamzah, M. I., & Zulkifli, H. (2022). Systematic literature review on the elements of metacognition-based higher order thinking skills (HOTS) teaching and learning modules. *Sustainability*, 14(2), 1–15. <https://doi.org/10.3390/su14020813>
- Januariawan, I. W., Wisnu, I. K., Wijaya, B., & Supadmini, N. K. (2020). Pengembangan keterampilan berpikir tingkat tinggi melalui pendekatan open-ended. *Jurnal Ilmu Pendidikan*, 3(2), 125–139. <https://doi.org/10.37329/cetta.v3i2.444>
- Johan, J. R., Iriani, T., & Maulana, A. (2023). Penerapan model Four-D dalam pengembangan media video keterampilan mengajar kelompok kecil dan perorangan. *Jurnal Pendidikan West Science*, 1(6), 372–378. <https://doi.org/10.58812/jpdws.v1i6.455>
- Jong, A., & Tacoh, Y. T. B. (2024). Pemanfaatan aplikasi Quizizz untuk meningkatkan motivasi belajar siswa. *Jurnal Dimensi Pendidikan dan Pembelajaran*, 12(1), 131–147. <http://journal.umpo.ac.id/index.php/dimensi/index>
- Khaerudin. (2016). Teknik penskoran tes obyektif model pilihan ganda. *Jurnal Madaniyah*, 2(9), 183–200. <https://journal.stitpemalang.ac.id/index.php/madaniyah/article/download/27/14/>
- Kurniawan, A., Febrianti, A., Hardianti, T., Ichsan, Desy, & Risan, R. (2022). *Evaluasi pembelajaran* (A. Yanto, Ed.). PT Global Eksekutif Teknologi.
- Listiani, W., & Rachmawati, R. (2022). Transformasi taksonomi Bloom dalam evaluasi pembelajaran berbasis HOTS. *Jurnal Jendela Pendidikan*, 2(3), 397–402. <https://doi.org/10.57008/jjp.v2i03.266>
- Maulidia, F., & Pahlevi, T. (2020). Pengembangan instrumen penilaian tes soal pilihan ganda berbasis HOTS pada mata pelajaran administrasi umum jurusan OTKP SMK Negeri 1 Lamongan. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, 8, 136–145. <https://doi.org/10.26740/jpap.v8n1.p136-145>
- Muchtar, Z., Gusfa, M., Dibiyanti, R. E., Sutiani, A., & Sinaga, M. (2024). Pengembangan instrumen evaluasi untuk mengukur keterampilan berpikir tingkat tinggi pada materi kesetimbangan kimia. *Jurnal Ilmiah Ilmu Pendidikan*, 7(8), 8149–8155. <https://doi.org/10.54371/jiip.v7i8.4901>
- Mulyawan, A., & Dwijayanti, R. (2024). Pengembangan asesmen berbasis Kurikulum Merdeka elemen 6 mata pelajaran dasar-dasar pemasaran berbantuan Quizizz di SMK Negeri Mojoagung. *Jurnal Pendidikan Tata Niaga (JPTN)*, 12(2), 141–152. <https://doi.org/10.26740/jptn.v12n2.p141-152>
- Murniatie, I. U., Wahyudi, D. A., Siswiyanti, F., Hidayah, L., & Wicaksono, H. (2025). Kualitas soal ujian akhir semester pada mata pelajaran bahasa Indonesia di sekolah menengah atas. *Indonesian Language Education and Literature*, 11(1), 163–174. <https://doi.org/10.24235/ileal.v11i1.20889>
- Muslikhin, U., Azizah, S. N., Hidayatullah, M. T., & Mu'alimin. (2025). Pelayanan prima dalam pendidikan: Tinjauan sistematis tentang strategi meningkatkan kepuasan



- pelanggan dan mutu layanan. *Jurnal Manajemen dan Pendidikan Agama Islam*, 3(5), 190–199. <https://doi.org/10.61132/jmpai.v3i5.1476>
- Narassati, N. A., Saleh, R., & Arthur, R. (2021). Pengembangan alat evaluasi berbasis HOTS menggunakan aplikasi Quizizz pada mata pelajaran mekanika teknik dalam pembelajaran jarak jauh. *Jurnal Pendidikan Teknik Sipil*, 3(2), 169–180. <https://doi.org/10.21831/jpts.v3i2.43919>
- Nasution, R. M., Hasruddin, & Harahap, F. (2024). Pengembangan instrumen penilaian hasil belajar berbasis HOTS pada materi biologi semester ganjil kelas XI SMA. *Didaktika: Jurnal Kependidikan*, 13(3), 3513–3522. <https://doi.org/10.58230/27454312.919>
- OECD. (2023). *PISA 2022 results (Vol. I): The state of learning and equity in education*. OECD Publishing.
- Opi, N. (2019). Analisis Kuantitatif Soal Ulangan Harian Pada Materi Struktur Dan Fungsi Jaringan Tumbuhan Di Kelas VIII SMP Negeri 24 Padang. *Jurnal Bioconecta*, 4(2), 29-37. <http://dx.doi.org/10.22202/bc.2018.v4i2.2947>
- Pitoyo, M. D., & Asib, A. (2020). Gamification-based assessment: The washback effect of Quizizz on students' learning in higher education. *International Journal of Language Education*, 4(1), 1–10. <https://doi.org/10.26858/ijole.v4i2.8188>
- Purbaningrum, K. A. (2017). Kemampuan berpikir tingkat tinggi siswa SMP dalam pemecahan masalah matematika ditinjau dari gaya belajar. *Jurnal Penelitian dan Pembelajaran Matematika*, 10(2), 40–49. <https://doi.org/10.30870/jppm.v10i2.2029>
- Puteri, N. D., Putricia, N. D., Arsyanti, K., Hakim, L., & Pratiwi, V. (2025). Analisis kelayakan butir soal pada handout akuntansi keuangan dana kas kecil berbasis e-book menggunakan software Anates. *JIBEMA: Jurnal Ilmu Bisnis, Ekonomi, Manajemen, dan Akuntansi*, 2(3), 205–214. <https://doi.org/10.62421/jibema.v2i3.108>
- Saputri, H. A., Zuhijrah, Larasati, N. J., & Shaleh. (2023). Analisis instrumen assesmen: Validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 9(5), 2986–2995. <https://doi.org/10.36989/didaktik.v9i5.2268>
- Saraswati, P. M. S., & Agustika, G. N. S. (2020). Kemampuan berpikir tingkat tinggi dalam menyelesaikan soal HOTS mata pelajaran matematika. *Jurnal Ilmiah Sekolah Dasar*, 4(2), 257–269. <https://doi.org/10.23887/jisd.v4i2.25336>
- Sari, V. N. I., Utomo, A. P. Y., & Sumarwati. (2022). Kualitas soal bahasa Indonesia di SMP Muhammadiyah 1 Pontianak: Analisis butir soal. *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 11(2), 112–119. <https://journal.unnes.ac.id/sju/jpbsi/article/view/58091>
- Shofiah, S., Bachtiar, E., Kd, D. P., Syahropi, H., & Zaman, N. (2023). *Dasar-dasar evaluasi pembelajaran* (Sarwadi, Ed.). PT Mifandi Mandiri Digital.
- Sijabat, M. P., Hutabarat, K., Sitorus, L., & Syahrial. (2024). Analisis soal tes hasil belajar siswa soal berstandar nasional bahasa Indonesia kelas 5 sekolah dasar. *Jurnal Basicedu*, 8(2), 1265–1277. <https://doi.org/10.31004/basicedu.v8i2.7354>
- Sitorus, D. S., & Santoso, T. N. B. (2022). Pemanfaatan Quizizz sebagai media pembelajaran berbasis game pada masa pandemi Covid-19. *Scholaria: Jurnal Pendidikan dan Kebudayaan*, 12(2), 81–88. <https://doi.org/10.24246/j.js.2022.v12.i2.p81-88>
- Sumaryanta. (2021). *Teori tes klasik & teori respon butir: Konsep & contoh penerapannya*. CV Confident.



- Syam, A. N. F., Majid, A. F., Mania, S., & Baharuddin. (2026). Analisis kualitas instrumen tes berbasis etnomatematika pada motif batik lontara untuk mengukur kemampuan berpikir kritis siswa. *Papanda Journal of Mathematics and Sciences Research (PJMSR)*, 5(1), 185–204. <https://doi.org/10.56916/pjmsr.v5i1.3104>
- Wahjusaputri, S., & Purwanto, A. (2022). *Statistika pendidikan: Teori dan aplikasi*. CV Bintang Semesta Media.
- Widana, I. W. (2017). Higher order thinking skills assessment (HOTS). *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 3(1), 32-44. <https://doi.org/10.21009/jisae.v3i1.4859>
- Widayanti, W., Bistari, & Suparjan. (2021). Analisis butir soal pilihan ganda penilaian tengah semester pada pembelajaran tematik kelas V Sekolah Dasar Negeri 39 Pontianak Kota. *Jurnal DIDIKA: Wahana Ilmiah Pendidikan Dasar*, 7(2), 279–296. <https://doi.org/10.29408/didika.v7i2.4370>
- Widodo, B. S. (2021). *Metode penelitian pendidikan pendekatan sistematis & komprehensif*. Eiga Media.
- Wulandari, S. T., & Lina, H. N. (2025). Pengaruh pelayanan prima oleh customer service terhadap loyalitas pelanggan Trans Icon Mall Surabaya. *The Commercium*, 9(2), 36–47. <https://doi.org/10.26740/tc.v9i2.66495>