



**PENGEMBANGAN INSTRUMEN PENILAIAN BAHASA INDONESIA  
BERBANTUAN *GENERATIVE ARTIFICIAL INTELLIGENCE (GENAI)*  
UNTUK MENGUKUR KEMAMPUAN KOLABORASI  
PESERTA DIDIK KELAS IV DI SD**

**Insan Esa Azyfah<sup>1</sup>, Deri Anggraini<sup>2</sup>**

Universitas PGRI Yogyakarta<sup>1,2</sup>

e-mail: [ieazyfah@gmail.com](mailto:ieazyfah@gmail.com)

Diterima: 1/5/2026; Direvisi: 8/5/2026; Diterbitkan: 18/5/2026

**ABSTRAK**

Penelitian ini dilatarbelakangi oleh belum tersedianya instrumen penilaian yang mampu mengukur kemampuan kolaborasi peserta didik pada pembelajaran Bahasa Indonesia di sekolah dasar. Penilaian yang dilakukan selama ini masih berfokus pada aspek kognitif sehingga aspek kolaborasi belum terukur secara objektif. Penelitian ini bertujuan untuk mengembangkan instrumen penilaian berbantuan *Generative Artificial Intelligence (GenAI)* menggunakan rubrik penilaian yang dikemas dalam bentuk buku untuk mengukur kemampuan kolaborasi peserta didik kelas IV di SD Negeri Jarakan Yogyakarta. Metode yang digunakan adalah penelitian dan pengembangan (*Research and Development*) dengan model 4-D yang meliputi tahap *define, design, development, dan disseminate*. Subjek penelitian terdiri dari 46 peserta didik untuk uji coba instrumen dan 13 peserta didik untuk uji coba terbatas. Teknik pengumpulan data menggunakan observasi, wawancara, dan angket, sedangkan analisis data dilakukan secara kuantitatif dan kualitatif. Hasil penelitian menunjukkan bahwa instrumen penilaian yang dikembangkan memenuhi kriteria sangat layak berdasarkan validasi ahli dengan persentase 100%, valid secara empiris dengan nilai signifikansi  $< 0,05$  dan koefisien korelasi  $\geq 0,60$ , serta memiliki reliabilitas sangat tinggi dengan nilai *Cronbach's Alpha* sebesar 0,979. Selain itu, produk instrumen penilaian dinyatakan sangat praktis berdasarkan respon guru dengan persentase 100%. Dengan demikian, instrumen penilaian berbantuan GenAI yang dikembangkan layak digunakan untuk mengukur kemampuan kolaborasi peserta didik secara objektif.

**Kata Kunci:** *Instrumen Penilaian, Gen AI, Kemampuan Kolaborasi*

**ABSTRACT**

This research is motivated by the lack of assessment instruments capable of measuring students' collaborative abilities in Indonesian language learning in elementary schools. Current assessments have focused on cognitive aspects, thus rendering collaboration unmeasured objectively. This study aims to develop an assessment instrument assisted by Generative Artificial Intelligence (GenAI) using an assessment rubric packaged in book form to measure the collaborative abilities of fourth-grade students at Jarakan Elementary School, Yogyakarta. The method used is Research and Development (R&D) with a 4-D model that includes the stages of define, design, development, and disseminate. The research subjects consisted of 46 students for the instrument trial and 13 students for the limited trial. Data collection techniques used observation, interviews, and questionnaires, while data analysis was conducted quantitatively and qualitatively. The results of the study indicate that the developed assessment instrument meets the criteria of very feasible based on expert validation with a percentage of 100%, is empirically valid with a significance value of  $< 0.05$  and a correlation coefficient of



$\geq 0.60$ , and has very high reliability with a Cronbach's Alpha value of 0.979. In addition, the assessment instrument product is declared very practical based on teacher responses with a percentage of 100%. Thus, the developed GenAI-assisted assessment instrument is suitable for use to objectively measure students' collaboration abilities.

**Keywords:** *Assessment Instruments, Genai, Collaboration Capabilities*

## PENDAHULUAN

Abad ke-21 ditandai dengan kemajuan teknologi digital yang sangat pesat serta keterbukaan informasi yang mengubah cara manusia berinteraksi secara mendasar di seluruh dunia. Era revolusi industri 4.0 membawa paradigma baru melalui peningkatan konektivitas dan kecerdasan buatan yang berdampak langsung pada transformasi sektor pendidikan nasional saat ini. Peserta didik kini dituntut menguasai berbagai kecakapan esensial seperti kreativitas, pemecahan masalah, komunikasi, serta penguatan karakter yang menjadi modal utama dalam menghadapi tantangan masa depan (Louvette & Budiyo, 2026; Maleni et al., 2025). Jenjang sekolah dasar memiliki posisi strategis sebagai fondasi awal untuk menanamkan kesiapan belajar yang kokoh bagi anak didik. Dalam ekosistem ini, peran guru telah bergeser secara signifikan dari sekadar pusat informasi menjadi fasilitator yang bertugas menciptakan lingkungan belajar yang aktif dan kondusif bagi pengembangan potensi siswa. Hal ini selaras dengan kebutuhan dunia modern yang memerlukan individu yang tidak hanya cerdas secara akademik tetapi juga memiliki fleksibilitas sosial yang tinggi. Oleh karena itu, penyiapan kurikulum yang adaptif serta metode pengajaran yang berpusat pada siswa menjadi sebuah keniscayaan guna mencetak generasi unggul yang mampu beradaptasi dengan dinamika global yang terus berubah secara masif setiap saat di lingkungan pendidikan dasar (Andani & Arifin, 2026; Louvette & Budiyo, 2026; Pardiyanan et al., 2026).

Kolaborasi muncul sebagai salah satu kompetensi utama yang harus dikembangkan sejak dini karena menjadi sarana bagi peserta didik untuk saling membantu dalam menguasai materi secara kolektif. Aktivitas pembelajaran yang melibatkan partisipasi aktif memungkinkan terjadinya pertukaran ide, proses mendengarkan pendapat orang lain, serta pengambilan keputusan bersama secara konstruktif dan bermartabat. Kemampuan ini bukan sekadar bekerja dalam kelompok biasa, melainkan mencakup indikator bekerja secara produktif, berkontribusi aktif, serta menunjukkan fleksibilitas dalam mencapai kesepakatan bersama yang menguntungkan semua pihak. Selain itu, sikap tanggung jawab terhadap tugas dan rasa menghargai kontribusi anggota lain menjadi kunci terciptanya kerja sama yang positif di dalam lingkungan kelas (Ibrahim et al., 2022; Ilma, 2023; Thohiroh & Suryadi, 2026). Pembiasaan bekerja secara sinergis diharapkan mampu mengikis perilaku individualistis yang sering menjadi penghambat dalam perkembangan sosial anak di sekolah. Dengan adanya standar indikator yang jelas, pendidik dapat memantau sejauh mana kemampuan kerja sama ini berkembang secara objektif dan terukur. Pengembangan aspek afektif ini sangat penting untuk membentuk pribadi yang peduli dan mampu bersosialisasi dengan baik di lingkungan masyarakat yang lebih luas, sehingga tercipta harmoni sosial yang berawal dari kebiasaan positif di sekolah dasar secara berkelanjutan (Hartono & Warsiyah, 2025).

Realitas di lapangan menunjukkan adanya kesenjangan yang cukup nyata antara target ideal tersebut dengan kondisi senyatanya di lingkungan persekolahan dasar saat ini. Berdasarkan hasil observasi di SD Negeri Jarakan Yogyakarta pada 30 Oktober 2025, ditemukan fakta bahwa kemampuan kolaborasi siswa kelas 4 masih perlu dioptimalkan secara serius melalui pendekatan yang sistematis. Selama proses kerja kelompok berlangsung, hanya sebagian kecil peserta didik yang terlihat aktif berpartisipasi, sementara sisanya cenderung pasif



dan belum terlibat secara maksimal dalam penyelesaian tugas bersama. Interaksi yang terjalin belum menunjukkan adanya pertukaran gagasan yang berkualitas, serta pembentukan kelompok yang kurang merata turut menghambat efektivitas kerja sama yang harmonis. Kondisi ini diperburuk oleh sistem penilaian yang selama ini masih sangat dominan pada aspek kognitif melalui soal pilihan ganda atau uraian dari buku teks semata. Guru belum memiliki instrumen penilaian yang spesifik, valid, dan reliabel untuk mengukur kemampuan kolaborasi secara sistematis dan objektif di dalam kelas. Penilaian yang hanya mengandalkan pengamatan umum tanpa alat ukur yang jelas mengakibatkan gambaran mengenai kemampuan afektif siswa menjadi kurang akurat bagi kemajuan belajar mereka di masa depan yang akan datang.

Menghadapi tantangan evaluasi yang kompleks tersebut, integrasi inovasi teknologi digital dalam proses penilaian atau *assessment* menjadi solusi yang sangat relevan dan adaptif di era digital. Dunia pendidikan saat ini mulai memanfaatkan potensi besar dari *Generative Artificial Intelligence* atau GenAI yang mampu menghasilkan konten otomatis seperti teks, gambar, hingga video berdasarkan instruksi tertentu. Pemanfaatan teknologi ini membuka peluang bagi pendidik untuk merancang instrumen penilaian yang lebih menarik dan efisien tanpa harus memiliki keahlian teknis desain yang terlalu mendalam (Amri et al., 2025; Berg, 2024; Tanberkan et al., 2024; Yan et al., 2024). Namun, penggunaan kecerdasan buatan ini tetap harus berpedoman pada prinsip etika dan panduan resmi dari pemerintah agar terhindar dari risiko negatif seperti plagiarisme. GenAI diposisikan sebagai mitra strategis yang membantu manusia dalam memproses informasi dan mempercepat pengembangan perangkat evaluasi yang lebih modern bagi sekolah. Kesadaran penuh dari pendidik untuk tetap mengendalikan sistem otomatis sangat diperlukan agar integritas akademik dan keaslian ide tetap terjaga dengan baik. Dengan pemanfaatan yang bijak, teknologi ini dapat mentransformasi metode penilaian konvensional berbasis kertas menjadi sistem digital yang lebih interaktif dan responsif terhadap perkembangan zaman, sehingga proses pengumpulan data mengenai perkembangan siswa dapat dilakukan secara praktis.

Inovasi dalam penelitian ini difokuskan pada pengembangan instrumen penilaian bahasa Indonesia berbantuan GenAI untuk mengukur kemampuan kolaborasi siswa kelas 4 di SD Negeri Jarakan pada tahun ajaran 2025/2026. Produk ini dirancang sebagai mitra teknis peneliti dengan memanfaatkan berbagai *platform* cerdas seperti *DreaminaAI* untuk visualisasi gambar, *RunwayAI* untuk pengembangan video pembelajaran, serta *Google AI Studio* untuk narasi suara otomatis. Seluruh elemen multimedia tersebut diintegrasikan ke dalam sebuah buku penilaian sistematis melalui aplikasi *Canva* berbasis kecerdasan buatan yang dilengkapi fitur penskoran otomatis secara instan. Nilai kebaruan dari studi ini terletak pada penciptaan perangkat evaluasi yang tidak hanya valid dan reliabel secara ilmiah, tetapi juga memiliki daya tarik visual yang tinggi serta kepraktisan penggunaan di berbagai perangkat digital. Tujuan akhirnya adalah menyediakan alat ukur yang mampu mengidentifikasi tingkat kerja sama peserta didik secara objektif dalam konteks pembelajaran bahasa yang dinamis. Melalui pendekatan ini, proses penilaian diharapkan tidak lagi menjadi beban administratif yang berat, melainkan sebuah aktivitas evaluasi yang fleksibel, inovatif, dan mampu memberikan kontribusi nyata bagi peningkatan kualitas interaksi sosial siswa sekolah dasar.

## METODE PENELITIAN

Penelitian ini menerapkan metode *Research and Development* melalui model pengembangan 4-D yang mencakup tahap pendefinisian, perancangan, pengembangan, serta penyebaran secara sistematis. Lokasi pelaksanaan studi bertempat di SD Negeri Jarakan Yogyakarta dengan melibatkan 46 peserta didik pada tahap uji coba instrumen serta 13 siswa

kelas 4 untuk uji coba terbatas. Fase awal dimulai dengan kegiatan pendefinisian guna mengidentifikasi kebutuhan kurikulum, analisis tugas memirsa video instruksi arah, serta penjabaran 5 indikator kolaborasi utama seperti kontribusi aktif dan tanggung jawab. Memasuki tahap perancangan, peneliti mengonstruksi struktur instrumen penilaian non-tes yang dikemas dalam format buku digital maupun cetak untuk mempermudah guru. Proses kreatif pembuatan elemen visual, audio, serta tabel dalam buku tersebut memanfaatkan ragam teknologi *Generative Artificial Intelligence* meliputi *DreaminaAI*, *RunwayAI*, *GammaAI*, dan *GoogleAI Studio*. Peneliti bertindak sebagai instrumen kunci yang mengintegrasikan hasil kecerdasan buatan tersebut ke dalam kisi-kisi dan rubrik penilaian unjuk kerja secara sistematis guna memastikan alat ukur bersifat adaptif, futuristik, serta menarik bagi pengguna di sekolah dasar.

Tahap pengembangan dijalankan melalui proses validasi ahli yang melibatkan pakar materi serta evaluasi pembelajaran guna menjamin kualitas isi produk secara objektif dan mendalam. Hasil pengujian menunjukkan bahwa instrumen penilaian ini memperoleh tingkat kelayakan sempurna sebesar 100% berdasarkan penilaian para ahli melalui 3 kali tahap revisi produk. Setelah dinyatakan valid secara isi, dilakukan pengujian empiris untuk mengukur tingkat akurasi serta konsistensi 15 butir rubrik di dalam kelas. Data kuantitatif menunjukkan bahwa seluruh indikator memiliki nilai signifikansi di bawah 0,05 dengan koefisien korelasi di atas 0,60, sementara tingkat reliabilitas mencapai angka sangat tinggi sebesar 0,979 pada hitungan *Cronbach's Alpha*. Selain itu, pengujian tingkat kepraktisan menghasilkan persentase respon guru sebesar 100% yang menunjukkan kemudahan penggunaan alat ukur ini dalam situasi pembelajaran nyata sehari-hari. Analisis data kualitatif digunakan untuk menyaring masukan perbaikan teknis seperti penyesuaian kontras warna rubrik serta sinkronisasi kecepatan suara pada video simulasi tugas. Akhirnya, tahap penyebaran dilakukan dengan menyerahkan buku instrumen digital kepada pihak sekolah sebagai mitra pendukung pembelajaran Bahasa Indonesia.

## HASIL DAN PEMBAHASAN

### Hasil

Hasil penelitian ini berupa pengembangan produk instrumen penilaian non-tes dalam bentuk rubrik kemampuan kolaborasi yang dikemas menjadi format buku berbantuan *Generative Artificial Intelligence (GenAI)*. Instrumen ini dirancang untuk mengukur kemampuan kolaborasi peserta didik kelas IV SD dalam pembelajaran Bahasa Indonesia. Proses pengembangannya dilakukan dengan mengacu pada model 4D, setiap tahap-tahap pengembangannya disajikan sebagai berikut.

#### 1. Tahap pendefinisian (*define*)

Tahap *define* merupakan langkah awal untuk mengidentifikasi kebutuhan dan permasalahan aktual pembelajaran di lapangan melalui empat tahapan utama: analisis awal, analisis tugas, analisis konsep, dan analisis tujuan pembelajaran. Pada analisis awal di SD Negeri Jarakan Yogyakarta, ditemukan bahwa kemampuan kolaborasi peserta didik kelas IV belum optimal dan sistem penilaian masih berfokus pada ranah kognitif. Ketiadaan alat ukur objektif membuat pengembangan instrumen penilaian afektif khusus kolaborasi sangat diperlukan. Selanjutnya, analisis tugas merumuskan kegiatan unjuk kerja kelompok mata pelajaran Bahasa Indonesia materi petunjuk arah, di mana siswa memirsa video instruksi, berdiskusi, lalu mempraktikkan penelusuran rute di lingkungan sekolah. Analisis konsep menjabarkan kemampuan kolaborasi menjadi lima indikator operasional—bekerja produktif, berkontribusi aktif, menunjukkan fleksibilitas, bertanggung jawab, dan menghargai—sebagai dasar penyusunan rubrik. Terakhir, analisis tujuan pembelajaran memfokuskan turunan capaian

pembelajaran pada aktivitas kerja sama kelompok dan pemecahan masalah agar seluruh kompetensi tersebut dapat terukur secara sistematis melalui instrumen rubrik yang dikembangkan oleh peneliti.

## 2. Tahap perancangan (*design*)

Pada tahap *design* atau perancangan berisi kegiatan untuk membuat rancangan terhadap produk yang telah ditetapkan. Peneliti merancang produk kedalam tiga bagian. Bagian awal terdiri dari kata pengantar, profil pengembang, daftar isi, dan petunjuk penggunaan. Bagian isi terdiri dari, kisi-kisi, rubrik, prosedur penskoran, dashboard guru, kategori penilaian, penilaian unjuk kerja, prosedur pelaksanaan unjuk kerja, dan video unjuk kerja. Bagian akhir terdiri dari rekomendasi, etika penggunaan *GenAI*, glosarium, dan daftar pustaka. Bagian-bagian tersebut kemudian dituangkan menjadi bentuk buku. Elemen-elemen yang ada dalam format buku instrumen penilaian ini, dihasilkan dari berbagai jenis ragam *GenAI*, diantaranya adalah *DreaminaAI* digunakan untuk menghasilkan gambar, *RunwayAI* digunakan untuk menghasilkan video, *GammaAI* digunakan untuk menghasilkan tabel, serta *GoogleAI Studio* digunakan untuk menghasilkan suara.

## 3. Tahap pengembangan (*development*)

Pada tahap *development* atau pengembangan, peneliti melakukan proses validasi instrumen penilaian kepada para ahli diantaranya adalah ahli materi dan ahli evaluasi pembelajaran. Setelah instrumen penilaian dinyatakan layak atau valid secara isi, peneliti melakukan uji coba diantaranya adalah uji coba instrumen untuk mengetahui validitas secara empiris dan reliabilitas, kemudian setelah dinyatakan valid dan reliabel, butir instrumen digunakan kembali pada saat uji coba terbatas. Pada uji coba terbatas ini, peneliti juga memberikan angket respon kepada guru untuk mengetahui kepraktisan pada produk instrumen penilaian yang dikembangkan.

### a. Validasi ahli (*ekspert judgment*)

#### 1) Hasil validasi ahli evaluasi pembelajaran

Validasi oleh ahli evaluasi pembelajaran dilakukan untuk menguji kelayakan atau validitas isi terhadap produk yang dikembangkan yaitu instrumen penilaian. Kegiatan ini dilakukan dengan memberikan lembar angket validasi kepada satu orang ahli evaluasi pembelajaran yaitu dosen Universitas PGRI Yogyakarta. Proses validasi dilaksanakan sebanyak 3 kali. Validasi pertama, validator memberikan saran dan masukan terhadap produk instrumen penilaian yang dikembangkan, kemudian peneliti melakukan perbaikan. Validasi kedua, juga terdapat saran dan perbaikan dari validator, kemudian peneliti melakukan perbaikan kembali. Setelah itu pada validasi ketiga produk instrumen penilaian dinyatakan layak untuk digunakan. Pengisian angket validasi oleh ahli evaluasi hanya dilakukan satu kali, setelah semua proses revisi telah dilakukan. Hasil validasi ahli evaluasi pembelajaran tersebut disajikan pada tabel berikut.

**Tabel 1. Hasil Validasi Ahli Evaluasi**

Aspek	Skor Perolehan	Presentase	Keterangan
Tampilan	36	100%	Sangat layak
Konstruksi	28	100%	Sangat layak
Pemrograman	24	100%	Sangat layak
<b>Jumlah</b>	<b>88</b>	<b>100%</b>	<b>Sangat layak</b>

Berdasarkan tabel 1 hasil validasi ahli evaluasi, aspek tampilan memperoleh skor 36 dengan presentase 100% dan kriteria “sangat layak”. Aspek konstruksi memperoleh skor 28 dengan presentase 100% dan kriteria “sangat layak”. Serta aspek pemrograman memperoleh

skor 24 dengan presentase 100% dan kriteria “sangat layak. Secara keseluruhan, produk instrumen penilaian memperoleh skor total 88 dengan presentase 100% dan kriteria “sangat layak”. Dengan demikian, dapat disimpulkan bahwa produk instrumen penilaian yang dikembangkan telah memenuhi syarat kelayakan untuk digunakan dalam tahap uji coba.

## 2) Hasil validasi ahli materi

Validasi oleh ahli materi dilakukan untuk menguji kelayakan atau validitas isi terhadap produk yang dikembangkan yaitu instrumen penilaian dari aspek materi. Kegiatan ini dilakukan dengan memberikan lembar angket validasi kepada satu orang ahli materi, yaitu guru kelas IV SD Negeri Jarakan Yogyakarta. Proses validasi dilaksanakan sebanyak 2 kali. Validasi pertama, validator memberikan saran dan masukan terhadap produk instrumen penilaian yang dikembangkan, kemudian peneliti melakukan perbaikan. Setelah itu pada validasi kedua, produk instrumen penilaian dinyatakan layak untuk digunakan. Pengisian angket validasi oleh ahli materi hanya dilakukan satu kali, setelah proses revisi telah dilakukan. Hasil validasi ahli materi pembelajaran tersebut disajikan pada tabel berikut.

**Tabel 2. Hasil Validasi Ahli Materi**

Aspek	Skor Perolehan	Presentase	Keterangan
Materi	84	100%	Sangat layak
Bahasa	16	100%	Sangat layak
<b>Jumlah</b>	100	100%	Sangat layak

Berdasarkan tabel 2 hasil validasi ahli materi, aspek materi memperoleh skor 84 dengan presentase 100% dan kriteria “sangat layak”. Serta aspek bahasa memperoleh skor 16 dengan presentase 100% dan kriteria “sangat layak”. Secara keseluruhan, produk instrumen penilaian memperoleh skor total 100 dengan presentase 100% dan kriteria “sangat layak”. Dengan demikian, dapat disimpulkan bahwa produk instrumen penilaian yang dikembangkan telah memenuhi syarat kelayakan untuk digunakan dalam tahap uji coba.

## b. Revisi produk

Tahap revisi produk dilakukan berdasarkan saran ahli evaluasi pembelajaran dan ahli materi untuk menyempurnakan instrumen penilaian. Validasi ahli evaluasi pembelajaran dari dosen Universitas PGRI Yogyakarta berlangsung tiga tahap. Pada tahap pertama, peneliti merombak total antarmuka website GenAI agar tidak menyerupai Word atau Excel serta membuat rubrik dapat diklik secara otomatis. Tahap kedua berfokus pada perubahan warna rubrik menjadi lebih cerah, penambahan penanda warna saat diklik, dan penghapusan kata hubung ganda pada subindikator agar penilaian tidak bermakna ambigu. Pada tahap ketiga, seluruh produk instrumen penilaian akhirnya dinyatakan sangat layak digunakan. Sementara itu, proses validasi ahli materi oleh guru kelas empat SD Negeri Jarakan dilaksanakan dalam dua tahap. Revisi awal secara khusus difokuskan pada perbaikan video tugas unjuk kerja yang berbasis Runway AI, yakni dengan mempercepat tempo suara agar peserta didik tidak merasa bosan ketika memirsa, serta menyesuaikan konsistensi visual agar tayangan sepenuhnya berbentuk tiga dimensi. Pada tahap kedua, instrumen dinyatakan valid. .

## c. Uji validitas empiris

Setelah produk instrumen penilaian dinyatakan layak dari segi isi berdasarkan penilaian ahli materi dan ahli evaluasi, langkah selanjutnya adalah peneliti melakukan uji coba instrumen yang melibatkan 46 peserta didik kelas IV A dan IV B SD Negeri Jarakan Yogyakarta. Pada tahap ini, peneliti melaksanakan pembelajaran dengan materi petunjuk arah menggunakan stimulus berupa video instruksi yang harus disimak oleh peserta didik. Peserta didik mencatat informasi yang diperoleh dari video, kemudian menggunakan catatan tersebut sebagai panduan

dalam mempraktikkan penentuan arah dari titik awal hingga titik akhir secara berkelompok. Selama proses praktik berlangsung, guru melakukan penilaian terhadap kemampuan kolaborasi peserta didik menggunakan rubrik yang telah dikembangkan. Data hasil penilaian tersebut selanjutnya digunakan untuk analisis validitas empiris dan reliabilitas.

d. Analisis data hasil uji validitas empiris (uji coba instrumen)

1) Hasil uji validitas empiris

Uji validitas empiris pada penelitian ini menggunakan bantuan perangkat lunak SPSS versi 25. Tingkat validitas setiap indikator pada rubrik penilaian dapat diketahui melalui nilai signifikansi (*Sig. 2-tailed*) yang dihasilkan dari output SPSS. Kriteria pengambilan keputusan validitas indikator didasarkan pada taraf signifikansi 5% (0,05). Apabila nilai signifikansi lebih kecil dari 0,05 (*Sig. < 0,05*), maka indikator pada rubrik penilaian dinyatakan valid. Sebaliknya, apabila nilai signifikansi lebih besar atau sama dengan 0,05 (*Sig.  $\geq$  0,05*), maka indikator dinyatakan tidak valid dan perlu dilakukan perbaikan (Hauteas et al., 2017). Selain itu, untuk melihat tingkat validitas butir dapat dilihat pada output *pearson correlation*, suatu butir dinyatakan memiliki koefisien *pearson correlation* minimal berada pada kategori tinggi (baik), yaitu  $\geq 0,60$ . Adapun hasil uji validitas empiris dapat dilihat pada tabel berikut.

**Tabel 3. Hasil Analisis Validitas Empiris**

Kode Butir Rubrik	<i>Pearson Correlation</i>	Nilai Sig.	Keterangan
Sub Indikator (SI1)	0.928	0.000	Valid
Sub Indikator (SI2)	0.859	0.000	Valid
Sub Indikator (SI3)	0.785	0.000	Valid
Sub Indikator (SI4)	0.875	0.000	Valid
Sub Indikator (SI5)	0.776	0.000	Valid
Sub Indikator (SI6)	0.897	0.000	Valid
Sub Indikator (SI7)	0.911	0.000	Valid
Sub Indikator (SI8)	0.897	0.000	Valid
Sub Indikator (SI9)	0.852	0.000	Valid
Sub Indikator (SI10)	0.868	0.000	Valid
Sub Indikator (SI11)	0.922	0.000	Valid
Sub Indikator (SI12)	0.898	0.000	Valid
Sub Indikator (SI13)	0.914	0.000	Valid
Sub Indikator (SI14)	0.888	0.000	Valid
Sub Indikator (SI15)	0.886	0.000	Valid

Berdasarkan tabel 3 hasil uji validitas empiris, seluruh butir rubrik menunjukkan nilai signifikansi (*Sig.*) sebesar 0.000, yang berarti (*Sig. < 0,05*), sehingga dinyatakan valid. Selain itu, nilai *pearson correlation* pada setiap butir ( $\geq 0,60$ ) berada pada kategori tinggi (baik). Dengan demikian, seluruh butir rubrik yang dikembangkan dinyatakan valid dan layak digunakan.

2) Hasil uji reliabilitas

Setelah dilakukan uji validitas, maka langkah selanjutnya adalah melakukan uji reliabilitas. Uji reliabilitas ini bertujuan untuk mengetahui tingkat konsistensi instrumen dalam mengukur konstruk yang sama apabila digunakan dalam kondisi yang serupa. Instrumen dikatakan memiliki reliabilitas yang tinggi apabila rubrik yang dibuat mempunyai hasil yang konsisten. Untuk melakukan uji reliabilitas, peneliti menggunakan perangkat lunak *SPSS Statistic 25*. Instrumen dinyatakan reliabel apabila nilai *Cronbach's Alpha*  $> 0,70$  dan apabila

nilai *Cronbach's Alpha* < 0,70 maka instrumen dinyatakan tidak reliabel. Hasil uji reliabilitas instrumen kolaborasi dapat dilihat pada tabel berikut.

**Tabel 4. Hasil Uji Reliabilitas**

<i>Reliability Statistics</i>	
<i>Cronbach's Alpha</i>	<i>N of Items</i>
0.979	15

Berdasarkan tabel 4 hasil uji reliabilitas menggunakan *Cronbach's Alpha* dengan bantuan *SPSS Statistic 25*, diperoleh nilai sebesar 0,979. Nilai tersebut lebih besar dari 0,50 (*Cronbach's Alpha* > 0,70), sehingga instrumen penilaian untuk mengukur kolaborasi dinyatakan reliabel. Nilai *Cronbach's Alpha* sebesar 0,979 menunjukkan bahwa instrumen memiliki tingkat reliabilitas sangat tinggi, yang berarti butir-butir indikator kolaborasi dalam instrumen memiliki konsistensi internal yang sangat baik. Dengan demikian, instrumen ini layak digunakan sebagai alat ukur dalam penelitian karena mampu memberikan hasil yang stabil dan dapat dipercaya.

e. Hasil analisis uji coba terbatas (kepraktisan)

Setelah pelaksanaan uji coba terbatas selesai, peneliti mengumpulkan dan menganalisis hasil penilaian yang diberikan oleh guru melalui angket kepraktisan. Analisis tersebut dilakukan untuk mengetahui tingkat kepraktisan instrumen penilaian yang telah dikembangkan, ditinjau dari aspek kemudahan penggunaan dan kebermanfaatannya dalam mendukung proses penilaian di kelas. Adapun hasil angket kepraktisan tersebut dapat dilihat pada tabel berikut.

**Tabel 5. Presentase Hasil Angket Respon Guru (Kepraktisan)**

Aspek	Skor Perolehan	Presentase	Keterangan
Kemudahan penggunaan	56	100%	Sangat praktis
Kebermanfaatan	24	100%	Sangat praktis
<b>Jumlah</b>	80	100%	Sangat praktis

Berdasarkan tabel 5 hasil angket respon guru, aspek kemudahan penggunaan memperoleh skor 56 dengan presentase 100% dan kriteria "sangat praktis". Serta aspek kebermanfaatannya memperoleh skor 24 dengan presentase 100% dan kriteria "praktis". Secara keseluruhan, produk instrumen penilaian memperoleh skor total 80 dengan presentase 100% dan kriteria "sangat praktis". Dengan demikian, dapat disimpulkan bahwa produk instrumen penilaian yang dikembangkan telah memenuhi syarat kepraktisan untuk digunakan dalam proses penilaian untuk mengukur kemampuan kolaborasi.

### **Pembahasan**

Pengembangan instrumen penilaian kemampuan kolaborasi ini melalui prosedur sistematis yang mencakup identifikasi kebutuhan hingga penyebaran produk secara terbatas kepada pengguna akhir. Pada fase pengujian kualitas melalui penilaian para ahli, didapatkan hasil yang menunjukkan tingkat kelayakan yang sangat tinggi bagi instrumen tersebut. Ahli evaluasi memberikan penilaian pada 3 komponen utama yang mencakup aspek tampilan dengan angka 36, dimensi konstruksi sebesar 28, serta bagian pemrograman pada angka 24. Akumulasi nilai dari ahli evaluasi ini mencapai angka 88 yang menempatkan produk pada standar kelayakan maksimal bagi pengguna. Sementara itu, tinjauan dari ahli materi menunjukkan angka 84 untuk substansi isi dan 16 untuk ketepatan penggunaan bahasa dengan total nilai keseluruhan mencapai 100. Pencapaian angka-angka tersebut memberikan implikasi bahwa rancangan instrumen dalam bentuk buku ini telah memenuhi kriteria teknis serta teoretis yang dibutuhkan dalam dunia pendidikan. Struktur instrumen yang terbagi dalam bagian awal, inti, dan akhir memudahkan proses pemetaan kemampuan siswa secara terorganisir. Keselarasan antara materi dengan tujuan pembelajaran yang telah didefinisikan sebelumnya menjadi kunci



utama validitas isi produk yang dikembangkan ini. Hasil ini juga merefleksikan bahwa setiap butir soal telah melewati proses kurasi yang ketat (Chasani et al., 2023; Nabeta & Syolendra, 2026; Sarirah & Andromeda, 2026).

Analisis lebih lanjut mengenai validitas empiris dilakukan dengan menguji coba instrumen tersebut kepada subjek sasaran guna mendapatkan data statistik yang akurat bagi penguji. Melalui pengolahan data menggunakan perangkat lunak statistik, diperoleh fakta bahwa seluruh butir penilaian memiliki nilai signifikansi sebesar 0,000. Angka tersebut jauh di bawah standar kesalahan 0,05 yang berarti setiap item memiliki ketepatan tinggi dalam mengukur parameter kolaborasi di sekolah. Koefisien korelasi *pearson correlation* yang dihasilkan pada setiap butir instrumen berada pada rentang di atas 0,60 yang menunjukkan derajat hubungan yang kuat dengan total skor penilaian. Implikasi dari temuan ini adalah instrumen mampu membedakan tingkat kemampuan siswa secara objektif tanpa adanya bias pengukuran yang merugikan bagi hasil evaluasi (Utama et al., 2020; Wirawan et al., 2024). Validitas empiris ini memberikan kepastian bahwa butir-butir penilaian yang dirancang dalam tahap pengembangan memiliki kualitas yang baik secara matematis bagi peneliti. Data ini juga menunjukkan bahwa tidak ada butir yang harus digugurkan karena seluruhnya telah memenuhi kriteria signifikansi yang ditetapkan secara baku. Kualitas instrumen yang teruji secara statistik ini menjamin bahwa informasi yang dikumpulkan melalui alat ukur tersebut memiliki derajat kebenaran yang tinggi untuk keperluan evaluasi mandiri.

Aspek reliabilitas menjadi faktor penentu berikutnya dalam menilai stabilitas instrumen penilaian kolaborasi yang dikembangkan melalui model *Four-D Models* ini secara komprehensif bagi pembaca. Berdasarkan hasil analisis data terhadap 15 butir instrumen, diperoleh nilai *cronbach's alpha* sebesar 0,979 yang menunjukkan tingkat konsistensi yang sangat luar biasa dalam pengukuran. Angka ini jauh melampaui batas minimal reliabilitas yang biasanya ditetapkan pada angka 0,70 untuk sebuah instrumen penelitian pendidikan yang berkualitas. Dengan nilai sebesar 0,979, instrumen ini dapat memberikan hasil pengukuran yang stabil meskipun digunakan dalam waktu yang berbeda pada subjek yang memiliki karakteristik serupa. Hal ini memberikan jaminan bagi para guru bahwa data yang dihasilkan bukan merupakan faktor kebetulan melainkan cerminan nyata dari kemampuan peserta didik yang dinilai. Stabilitas instrumen ini sangat penting mengingat penilaian kolaborasi bersifat dinamis dan membutuhkan alat ukur yang tidak mudah dipengaruhi oleh berbagai variabel pengganggu lainnya. Kekuatan reliabilitas ini didukung oleh struktur instrumen yang sistematis dalam bentuk buku yang memandu proses penilaian secara bertahap bagi pengguna. Konsistensi internal yang tinggi mengindikasikan bahwa setiap butir instrumen bekerja secara harmonis untuk mengukur satu konstruk tunggal yang sama secara tuntas (Akbar et al., 2024; Cahyono et al., 2021; Martins & Riyanto, 2020).

Kepraktisan merupakan elemen krusial yang menentukan apakah sebuah instrumen penilaian dapat diterapkan dengan mudah oleh pendidik di lingkungan sekolah yang nyata secara langsung. Melalui angket respons yang diberikan kepada guru sebagai pengguna utama, diperoleh gambaran objektif mengenai kemudahan penggunaan dan kebermanfaatan produk di lapangan. Komponen kemudahan penggunaan memperoleh angka sebesar 56, sementara aspek kebermanfaatan instrumen dalam proses evaluasi mencatatkan angka 24 bagi tenaga pendidik. Total skor praktikalitas yang diraih mencapai angka 80 yang menunjukkan bahwa produk ini sangat efisien untuk dioperasikan dalam kegiatan rutin di ruang kelas. Implikasi praktis dari temuan ini adalah guru tidak memerlukan pelatihan teknis yang sangat rumit untuk memahami mekanisme penilaian yang ditawarkan dalam buku instrumen tersebut. Kemudahan akses dan tata letak yang jelas membantu pengajar menghemat waktu dalam melakukan observasi



terhadap interaksi antar siswa selama diskusi berlangsung di sekolah (Ardyanti et al., 2026; Efriani et al., 2023; Nugroho et al., 2021; Putri et al., 2021). Instrumen ini terbukti fungsional dalam menjembatani kebutuhan guru akan alat ukur yang praktis namun tetap memiliki dasar ilmiah yang kuat bagi penilaian. Kebermanfaatan yang dirasakan mencakup aspek kejelasan rubrik dan prosedur skoring yang memudahkan guru dalam memberikan umpan balik secara cepat.

Meskipun hasil penelitian menunjukkan kualitas yang sangat baik dari segi validitas, reliabilitas, serta kepraktisan, terdapat beberapa keterbatasan yang perlu diperhatikan oleh peneliti. Tahap penyebaran produk atau *disseminate* masih dilakukan dalam ruang lingkup yang terbatas sehingga generalisasi hasil pada populasi yang lebih luas memerlukan kehati-hatian secara mendalam. Selain itu, instrumen ini dikemas dalam bentuk fisik buku yang mungkin memerlukan adaptasi lebih lanjut jika ingin diterapkan dalam sistem pembelajaran berbasis digital atau *online*. Keterbatasan lainnya mencakup fokus pengukuran yang hanya berpusat pada kemampuan kolaborasi sehingga diperlukan pengembangan tambahan untuk mengukur aspek keterampilan abad 21 lainnya secara bersamaan. Namun secara keseluruhan, produk ini memberikan kontribusi nyata bagi pengembangan alat evaluasi pendidikan yang lebih objektif dan terukur bagi semua pihak. Rekomendasi untuk penelitian selanjutnya adalah melakukan uji coba pada jenjang pendidikan yang berbeda atau mengonversi instrumen ini ke dalam format aplikasi digital yang canggih. Dengan nilai reliabilitas mencapai 0,979 dan skor validitas yang tinggi, instrumen ini menjadi standar baru dalam penilaian afektif siswa. Keberhasilan pengembangan ini membuktikan bahwa pendekatan model *define, design, development*, dan *disseminate* sangat efektif dalam menghasilkan produk berkualitas tinggi.

#### UCAPAN TERIMAKASIH

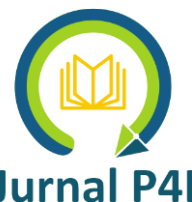
Penulis menyampaikan terimakasih yang sebesar-besarnya kepada Universitas PGRI Yogyakarta atas dukungan pendanaan yang diberikan melalui Anggaran LPPM Tahun 2025-2026. Bantuan ini sangat berarti bagi penulis karena mendukung pelaksanaan dan penyelesaian penelitian ini secara optimal. Semoga dukungan yang diberikan dapat terus mendorong pengembangan ilmu pengetahuan, peningkatan kualitas penelitian, serta memberikan manfaat yang luas bagi masyarakat.

#### KESIMPULAN

Pengembangan instrumen penilaian kemampuan kolaborasi yang dilakukan secara sistematis telah menghasilkan produk yang sangat layak digunakan dalam dunia pendidikan. Berdasarkan penilaian para ahli serta pengujian di lapangan, instrumen ini terbukti memiliki tingkat ketepatan yang tinggi dalam mengukur parameter kerja sama antarsiswa tanpa adanya bias yang merugikan bagi para pengguna. Konsistensi internal yang sangat baik menjamin bahwa data yang dihasilkan mencerminkan kemampuan peserta didik yang sebenarnya secara stabil dan akurat bagi penguji. Selain itu, aspek kemudahan penggunaan membuat alat ukur ini sangat fungsional bagi para pendidik untuk diterapkan dalam kegiatan rutin tanpa memerlukan prosedur teknis yang rumit. Secara keseluruhan, keberhasilan model pengembangan tersebut memberikan kontribusi nyata bagi terciptanya alat evaluasi yang objektif, sistematis, dan mampu menjembatani kebutuhan guru akan instrumen penilaian afektif berkualitas.

#### DAFTAR PUSTAKA

Akbar, S., Darungan, T. S., & Rahma, D. S. (2024). Instrument development for study programs excellence evaluation based on Kirkpatrick evaluation model. *Jurnal Pendidikan*



*Kedokteran Indonesia: The Indonesian Journal of Medical Education*, 13(2), 106.

<https://doi.org/10.22146/jpki.91960>

Amri, H., Yohanes, R. A., Rapsanjani, H., Rahim, A., & Nurfahrunnisa, A. (2025). Pelatihan pengembangan instrumen evaluasi pembelajaran dengan menggunakan chatbot AI. *JUAN: Jurnal Pengabdian Nusantara*, 2(4), 25–32.

<https://doi.org/10.63545/juan.v2.i4.182>

Andani, F., & Arifin, Z. (2026). Pengaruh media pembelajaran interaktif terhadap minat belajar dan pemahaman siswa pada mata pelajaran matematika. *SCIENCE: Jurnal Inovasi Pendidikan Matematika Dan IPA*, 6(1), 177–188.

<https://doi.org/10.51878/science.v6i1.9361>

Ardyanti, Y., Faradila, A., Rahmawati, A., Restuningtyas, N., K, I. I., & Arisanti, F. (2026). Pembelajaran di luar kelas: Refleksi tentang pengembangan keterampilan akademik dan keterampilan sosial siswa melalui KKL kolaboratif. *LEARNING: Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, 6(2), 1242–1254.

<https://doi.org/10.51878/learning.v6i2.7669>

Berg, G. van den. (2024). Generative AI and educators: Partnering in using open digital content for transforming education. *Open Praxis*, 16(2), 130–141.

<https://doi.org/10.55982/openpraxis.16.2.640>

Cahyono, S. M., Kartawagiran, B., & Mahmudah, F. N. (2021). Construct exploration of teacher readiness as an assessor of vocational high school competency test. *European Journal of Educational Research*, 10(3), 1471–1485. <https://doi.org/10.12973/eu-er.10.3.1471>

Chasani, A., Nasir, M., & Erviyenni, E. (2023). Development of interactive learning media for H5P-based elasticity materials in the mobilizing school curriculum. *Jurnal Penelitian Pendidikan IPA*, 9(4), 2089–2096.

<https://doi.org/10.29333/jppipa.v9i4.3109>

Efriani, A., Zulkardi, Z., Putri, R. I. I., & Aisyah, N. (2023). Belajar bilangan dengan konteks memasak kue bola ubi. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, 7(5), 5289–5302. <https://doi.org/10.31004/obsesi.v7i5.5025>

Hartono, H., & Warsiyah, W. (2025). Implementasi kegiatan Jumat Amal dalam menumbuhkan karakter kepedulian sosial bagi peserta didik tingkat sekolah dasar. *BUDAI: Multidisciplinary Journal of Islamic Studies*, 4(1), 59–74.

<https://doi.org/10.30659/budai.4.1.59-74>

Ibrahim, D. S. M., Aswasulaskin, A., Ramdhani, S., Mukti, H., & Agustina, B. W. (2022). Peran guru dalam membentuk karakter siswa melalui pembelajaran kooperatif. *Jurnal DIDIKA: Wahana Ilmiah Pendidikan Dasar*, 8(1), 102–113.

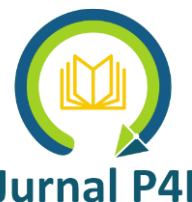
<https://doi.org/10.29408/didika.v8i1.5834>

Ilma, Z. (2023). Peningkatan keterampilan kolaborasi dengan pendekatan berdiferensiasi berbantuan e-modul matematika. *Jurnal Pendidikan Guru Profesional*, 1(2), 225–243. <https://doi.org/10.26877/jpgp.v1i2.232>

Louvette, R. H., & Budiyanto, M. (2026). Peningkatan keterampilan berpikir kritis melalui penerapan model problem based learning berbasis STEM. *LEARNING: Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, 6(2), 1346–1357.

<https://doi.org/10.51878/learning.v6i2.10017>

Maleni, L., Pardini, A., Kristian, D., Iswandi, W., Yudisman, A., Hidayat, T., & Rifa'i, R. (2025). Mempersiapkan siswa untuk masa depan: Literasi AI sebagai keterampilan



- abad 21. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 6375–6379. <https://doi.org/10.31004/riggs.v4i2.1587>
- Martins, M. A. J., & Riyanto, S. (2020). The effect of user experience on customer satisfaction on Netflix streaming services in Indonesia. *International Journal of Innovative Science and Research Technology (IJISRT)*, 5(7), 573–577. <https://doi.org/10.38124/ijisrt20jul545>
- Nabeta, E., & Syolendra, D. F. (2026). Pengembangan instrumen tes berbasis wordwall pada materi asam basa fase F SMA/MA. *SCIENCE: Jurnal Inovasi Pendidikan Matematika Dan IPA*, 6(1), 233–243. <https://doi.org/10.51878/science.v6i1.9377>
- Nugroho, W., Yudha, R. P., Sundari, S., & Praja, H. N. (2021). Analisis instrumen asesmen unjuk kerja pada pembelajaran PJOK di sekolah dasar kota Cirebon. *Gelombang Olahraga: Jurnal Pendidikan Jasmani Dan Olahraga (JPJO)*, 4(2), 126–141. <https://doi.org/10.31539/jpjo.v4i2.1795>
- Pardiyawan, P., Sugiantoro, S., Ali, M., Afdal, M., Khaeruddin, K., & Victorynie, I. (2026). Rekonseptualisasi diklat guru dalam pengembangan profesional berkelanjutan: Analisis literatur. *LEARNING: Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, 6(2), 962–984. <https://doi.org/10.51878/learning.v6i2.9689>
- Putri, R. D., Herpratiwi, H., & Rosidin, U. (2021). Pengembangan instrumen asesmen kinerja berbasis literasi sains pada pembelajaran tematik terpadu peserta didik kelas V sekolah dasar. *Jurnal Basicedu*, 5(6), 5946–5952. <https://doi.org/10.31004/basicedu.v5i6.1688>
- Sarirah, T., & Andromeda, A. (2026). Validitas dan praktikalitas E-LKM berbasis PBL berbantuan wizer.me pada materi termokimia untuk meningkatkan kemampuan berpikir kritis murid SMA/MA. *SCIENCE: Jurnal Inovasi Pendidikan Matematika Dan IPA*, 6(1), 507–519. <https://doi.org/10.51878/science.v6i1.9524>
- Tanberkan, H., Özer, M., & Gelbal, S. (2024). Impact of artificial intelligence on assessment and evaluation approaches in education. *DergiPark*. <https://doi.org/10.5281/zenodo.14016103>
- Thohiroh, U., & Suryadi, Y. (2026). Pengaruh model cooperative learning tipe jigsaw dalam pembelajaran IPS terhadap sikap kepedulian sosial peserta didik kelas VII SMP. *SOCIAL: Jurnal Inovasi Pendidikan IPS*, 6(1), 517–528. <https://doi.org/10.51878/social.v6i1.9732>
- Utama, C., Sajidan, S., Nurkamto, J., & Wiranto, W. (2020). The instrument development to measure higher-order thinking skills for pre-service biology teacher. *International Journal of Instruction*, 13(4), 833–848. <https://doi.org/10.29333/iji.2020.13451a>
- Wirawan, N., Nindiasari, H., & Sukirwan, S. (2024). Pengembangan instrumen literasi numerasi konten budaya Banten materi perbandingan untuk siswa SMP. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 8(2), 1148–1161. <https://doi.org/10.31004/cendekia.v8i2.2634>
- Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *arXiv*. <https://doi.org/10.48550/arxiv.2408.12143>