



Pengembangan Instrumen Penilaian Hasil Belajar Dalam Bentuk Tes Pilihan Ganda Berbasis *Higher Order Thinking Skills* Pada Mata Pelajaran Pendidikan Pancasila Di SMP Negeri 4 Medan

Gerry Frizi Jonatan Malau¹, Fazli Rachman²

^{1,2}Universitas Negeri Medan, Sumatera Utara, Indonesia

Jalan Willem Iskandar, Pasar V Medan Estate, Percut Sei Tuan, Deli Serdang, Sumatera Utara

Email: gerryfrizzy69@gmail.com

Abstrak

Penelitian ini bertujuan mengembangkan instrumen tes pilihan ganda berbasis *Higher Order Thinking Skills* (HOTS) pada mata pelajaran Pendidikan Pancasila untuk siswa kelas VIII di SMP Negeri 4 Medan. Penelitian menggunakan model 4D (*Define, Design, Develop, Disseminate*) yang dibatasi hingga tahap *Develop* melalui validasi ahli dan uji coba terbatas. Hasil validasi menunjukkan tingkat kelayakan sangat tinggi, dengan skor rata-rata 93% dari dosen ahli dan 100% dari guru mata pelajaran. Uji validitas empiris menunjukkan seluruh 20 butir soal valid (r hitung 0,372–0,682 > r tabel 0,361) dengan reliabilitas tinggi (Cronbach's Alpha = 0,829). Distribusi tingkat kesukaran terdiri atas 13 soal sedang, 3 mudah, dan 4 sukar, serta daya beda berkategori cukup hingga baik. Hasil ini membuktikan instrumen layak, reliabel, dan efektif digunakan untuk mengukur kemampuan berpikir tingkat tinggi sesuai tuntutan Kurikulum Merdeka.

Kata Kunci: HOTS, Pendidikan Pancasila, Pengembangan Instrumen Penilaian, Tes Pilihan Ganda

PENDAHULUAN

Pendidikan merupakan fondasi utama yang menentukan kemajuan suatu bangsa dan memegang peranan krusial dalam kehidupan setiap individu. Negara yang menargetkan kemajuan peradaban harus memprioritaskan pembenahan sistem pendidikannya (Suriaman et al., 2024). Sejalan dengan amanat Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional (Republik Indonesia, 2003), pendidikan di Indonesia didefinisikan sebagai usaha sadar dan terencana untuk menciptakan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensinya. Seiring dengan perkembangan zaman dan akselerasi teknologi, tuntutan terhadap kualitas sumber daya manusia pun semakin meningkat, yang menempatkan pendidikan sebagai poros utama dalam menghadapi tantangan global.

Paradigma pendidikan abad ke-21 telah bergeser secara fundamental dari sekadar transfer pengetahuan menuju pengembangan kompetensi yang holistik.

Keberhasilan proses pembelajaran tidak lagi diukur dari kemampuan peserta didik menghafal informasi, melainkan dari kapasitas mereka untuk mengaplikasikan pengetahuan dalam memecahkan masalah yang kompleks dan dinamis. Kemampuan ini mencakup keterampilan berpikir kritis, kreatif, argumentative dan pengambilan keputusan, yang secara kolektif dikenal sebagai Keterampilan Berpikir Tingkat Tinggi atau *Higher Order Thinking Skills* (HOTS).

Karena berpikir kritis merupakan landasan pembangunan ekonomi, maka berpikir kritis merupakan keterampilan yang krusial di abad ke-21. Kemampuan berpikir kritis harus dikembangkan dari waktu ke waktu, khususnya melalui pengembangan sikap dan tindakan yang mendorong berpikir kritis. Salah satu bakat yang dapat diperoleh dan digunakan untuk mengatasi masalah dengan cepat dan efektif adalah berpikir kritis. Seseorang yang memiliki keterampilan berpikir kritis dapat memanfaatkannya untuk memeriksa masalah, menyelesaikannya, membuat keputusan, dan mengatasi hambatan (Sianipar et al., 2023).

Pergeseran paradigma ini terimplementasi secara jelas dalam kebijakan pendidikan nasional Indonesia melalui Kurikulum Merdeka, yang dirumuskan dengan penekanan pada penguasaan materi esensial serta pengembangan karakter dan kompetensi peserta didik. Salah satu pilar utama kurikulum ini adalah penguatan Profil Pelajar Pancasila, yang di dalamnya mencakup dimensi “bernalar kritis” dan “kreatif”. Kedua dimensi ini secara eksplisit menuntut penguasaan HOTS. Oleh karena itu, pengembangan sekaligus penilaian HOTS bukan lagi sekadar pilihan metodologis, melainkan sebuah mandat kurikulum yang wajib dilaksanakan secara terstruktur.

Mata pelajaran yang menjadi pilar utama dalam implementasi mandat ini adalah Pendidikan Kewarganegaraan. Kajian Pendidikan Kewarganegaraan mencakup pengetahuan dasar tentang ketatanegaraan, politik, dan hukum, serta berbagai teori yang sejalan dengan ruang lingkup kajian tersebut. Esensi dari Pendidikan Kewarganegaraan terletak pada pembahasan mengenai perilaku warga negara, yang bertujuan membentuk individu agar memiliki kecerdasan, keterampilan, karakter yang kuat, serta rasa loyalitas kepada bangsa dan negara (Alya et al., 2024).

Dalam konteks ini penilaian hasil belajar memegang peran vital dalam pendidikan. Menurut Miller (dalam Destiana et al., 2020), penilaian hasil belajar siswa merupakan serangkaian prosedur yang digunakan untuk mengumpulkan informasi mengenai proses belajar siswa dan sebagai dasar dalam pengambilan keputusan terkait kinerja atau

capaian belajar mereka. Penilaian berfungsi sebagai sarana umpan balik bagi pendidik untuk mengevaluasi dan memperbaiki metode pembelajaran, serta bagi peserta didik untuk mengukur kapabilitas diri. Untuk dapat mengukur tingkat ketercapaian kompetensi ini secara efektif, diperlukan sebuah instrumen penilaian hasil belajar yang dirancang dengan baik.

Namun kondisi di lapangan kerap menunjukkan adanya kesenjangan antara tuntutan kurikulum dengan praktik penilaian. Berdasarkan observasi di SMP Negeri 4 Medan, instrumen tes yang digunakan masih cenderung sederhana, belum bervariasi, dan belum sepenuhnya mengukur kemampuan berpikir tingkat tinggi. Sebagian besar instrumen penilaian masih berorientasi pada *Lower Order Thinking Skills* (LOTS), yang terbatas pada ranah kognitif Mengingat (C1), Memahami (C2), dan Menerapkan (C3). Kesenjangan ini tidak hanya menghambat pencapaian tujuan kurikulum, tetapi juga berisiko menghasilkan lulusan yang kurang siap menghadapi kompleksitas tantangan di kehidupan nyata.

Beberapa penelitian relevan sebelumnya telah menggarisbawahi urgensi pengembangan instrumen penilaian HOTS pada mata pelajaran Pendidikan Pancasila. Penelitian oleh Hasanah, Faizah, & Mardhatillah (2024) berhasil mengembangkan instrumen HOTS dalam bentuk soal esai untuk jenjang Sekolah Dasar (SD) yang divalidasi dengan hasil sangat layak. Serupa dengan itu, penelitian oleh Labibah, Faizah, & Mardhatillah (2024) juga fokus pada pengembangan instrumen tes untuk mengukur kemampuan berpikir kritis siswa SD menggunakan model ADDIE dan kerangka FRISCO, yang menyoroti pentingnya keterkaitan soal dengan kehidupan nyata siswa. Sementara itu, penelitian oleh Riana Septiarie, Sadtyadi, & Ngadat (2020) mengembangkan instrumen penilaian HOTS dalam bentuk soal uraian pada materi Pancasila Buddhis di tingkat SMP menggunakan model Borg and Gall, dengan hasil instrumen yang valid dan reliabel.

Kendati penelitian-penelitian relevan tersebut telah memberikan kontribusi substantif, tetap teridentifikasi adanya kesenjangan riset (research gap) yang signifikan. Kajian-kajian terdahulu memiliki kecenderungan untuk berfokus pada jenjang pendidikan dasar (Sekolah Dasar) atau mengadopsi instrumen berbentuk esai/uraian. Bentuk instrumen tersebut dinilai memiliki keterbatasan praktis, terutama apabila diaplikasikan untuk evaluasi berskala besar. Lebih lanjut, beberapa penelitian membatasi ruang lingkup materi pada konteks yang sangat spesifik (contohnya Pancasila Buddhis),

yang mengakibatkan temuan penelitian tersebut memiliki daya generalisasi yang rendah untuk implementasi pada mata pelajaran Pendidikan Pancasila secara komprehensif di jenjang Sekolah Menengah Pertama (SMP). Kesenjangan-kesenjangan ini menjadi alasan mendasar atas urgensi riset ini, yakni belum tersedianya sebuah produk instrumen penilaian yang teruji validitas dan reliabilitasnya, secara spesifik berbentuk tes pilihan ganda berbasis HOTS yang dirancang untuk mata pelajaran Pendidikan Pancasila di tingkat SMP.

Dengan demikian, permasalahan sentral (*core problem*) dalam penelitian ini adalah belum optimalnya instrumen evaluasi hasil belajar di SMP Negeri 4 Medan yang masih terorientasi pada Keterampilan Berpikir Tingkat Rendah (LOTS). Hal ini diperburuk oleh keterbatasan inovasi pendidik dalam merancang instrumen yang mampu menstimulasi dan mengukur kemampuan analisis, evaluasi, serta kreasi peserta didik.

Berlandaskan pada rumusan masalah tersebut, penelitian ini bertujuan untuk mengembangkan dan menghasilkan suatu produk instrumen penilaian hasil belajar dalam format tes pilihan ganda berbasis *Higher Order Thinking Skills* (HOTS). Produk yang dikembangkan ini ditargetkan untuk memenuhi standar validitas dan reliabilitas sebagai alat ukur pada mata pelajaran Pendidikan Pancasila bagi peserta didik kelas VIII.

Soal pilihan ganda merupakan bentuk soal yang jawabannya dapat dipilih dari beberapa kemungkinan jawaban yang telah disediakan. Konstruksinya terdiri dari pokok soal dan pilihan jawaban. Dalam menulis soal pilihan ganda harus memperhatikan kaidah-kaidah yaitu: materi, konstruksi dan Bahasa (Jawariah, 2017). Dari kaidah inilah peneliti akan mengembangkan sebuah angket penilaian validasi untuk dosen ahli dan guru pendidik.

Untuk mencapai tujuan tersebut, penelitian ini mengadopsi kerangka teoretis yang berpusat pada konsep *Higher Order Thinking Skills* (HOTS) yang mengacu pada Taksonomi Bloom revisi oleh Anderson & Krathwohl (2001). HOTS didefinisikan sebagai kemampuan berpikir kompleks yang mencakup level kognitif menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Pengembangan produk instrumen penilaian ini sendiri mengikuti prosedur metodologis dari model pengembangan 4-D yang dirancang oleh Thiagarajan, yang terdiri dari empat tahap sistematis: Define (Pendefinisian), Design (Perancangan), Develop (Pengembangan), dan Disseminate (Penyebaran) (Antara et al., 2021) . Dengan landasan konseptual dan metodologis ini, penelitian diharapkan dapat

menghasilkan produk instrumen yang tidak hanya relevan dengan tuntutan kurikulum, tetapi juga teruji secara empiris kualitasnya

METODE PENELITIAN

Penelitian ini menggunakan jenis penelitian dan pengembangan (*Research and Development* atau R&D). Menurut Sugiyono (dalam Siregar, 2023) metode *Research and Development* (R&D) adalah pendekatan penelitian yang bertujuan untuk menciptakan produk tertentu serta menguji keefektifan dari produk tersebut. Produk yang telah diciptakan akan diuji untuk menilai seberapa efektif produk tersebut. Menurut Sukmadinata (dalam Muqdamien et al., 2021) R&D mencakup langkah-langkah yang *akuntabel* (dapat dipertanggungjawabkan) yang ditujukan untuk mengembangkan sebuah produk yang sama sekali baru ataupun menyempurnakan produk yang sudah ada sebelumnya.. Produk yang dikembangkan dalam penelitian ini adalah sebuah instrumen penilaian hasil belajar dalam bentuk tes pilihan ganda berbasis *Higher Order Thinking Skills* (HOTS) pada mata pelajaran Pendidikan Pancasila. Proses pengembangan instrumen ini mengadopsi model 4-D yang dirancang oleh Thiagarajan, yang meliputi empat tahapan utama, yaitu *Define* (Pendefinisian), *Design* (Perancangan), *Develop* (Pengembangan), dan *Disseminate* (Penyebaran). Namun, karena adanya keterbatasan waktu, penelitian ini dibatasi hanya sampai pada tahap *Develop*.

Tahap pertama yaitu *Define*, difokuskan pada analisis kebutuhan untuk menetapkan dan mendefinisikan syarat-syarat pengembangan. Tahap pendefinisian berperan penting untuk menetapkan serta merumuskan berbagai kebutuhan dalam proses pembelajaran, sekaligus mengumpulkan informasi yang relevan dengan produk yang akan dikembangkan (Arkadiantika et al., 2020). Pada tahap ini, dilakukan analisis kebutuhan melalui observasi di SMP Negeri 4 Medan yang mengidentifikasi permasalahan utama, yaitu pendidik mengalami kesulitan dalam menyusun soal HOTS akibat keterbatasan waktu dan tantangan dalam merumuskan stimulus serta pilihan jawaban yang efektif. Selanjutnya, dilakukan analisis kurikulum dan materi, di mana ditetapkan bahwa instrumen akan dikembangkan berdasarkan Kurikulum Merdeka dan materi pelajaran Pendidikan Pancasila Kelas VIII pada Bab 3 yang berjudul “Peraturan di Negaraku”.

Tahap kedua yaitu *Design*, bertujuan untuk menyiapkan *prototipe* produk instrumen penilaian. Proses perancangan diawali dengan pemilihan media, yaitu menggunakan format kertas yang familiar bagi siswa menyerupai lembar jawaban ujian

standar untuk meminimalkan kendala teknis dan memungkinkan siswa fokus pada konten soal. Format instrumen dirancang dengan menyajikan soal dan empat pilihan jawaban (A, B, C, D) yang dilengkapi stimulus. Rancangan produk kemudian dilanjutkan dengan penyusunan kerangka kerja penilaian yang sistematis, dimulai dari perumusan Tujuan Pembelajaran (TP) dan Indikator Ketercapaian Tujuan Pembelajaran (IKTP) yang menargetkan level kognitif C4 (Menganalisis), C5 (Mengevaluasi), dan C6 (Mencipta). Berdasarkan TP-IKTP tersebut, disusun kisi-kisi soal sebagai cetak biru (*blueprint*) untuk pengembangan 20 butir soal pilihan ganda.

Tahap ketiga yaitu *Develop*, merupakan fase untuk menghasilkan produk final yang telah diuji kelayakannya. Tahap ini melibatkan dua langkah utama: validasi ahli dan uji coba terbatas. Validasi ahli dilakukan untuk menilai kelayakan instrumen dari aspek materi, konstruksi, dan bahasa oleh seorang dosen ahli dari Jurusan PPKn FIS UNIMED serta seorang guru Pendidikan Pancasila dari SMP Negeri 4 Medan. Data kualitatif berupa saran dan kritik dari validator digunakan sebagai dasar untuk merevisi dan menyempurnakan produk.

Setelah instrumen dinyatakan layak, selanjutnya akan dilakukan uji coba terbatas untuk mengumpulkan data kuantitatif. Adapun uji coba terbatas yang dilakukan yaitu uji validitas, uji daya beda, uji ingkat kesukaran dan uji daya beda. Validitas merujuk pada derajat ketepatan atau kesahihan yang dimiliki oleh sebuah instrumen. Fokus utama dari validitas adalah pada alat ukur itu sendiri. Prinsip dasarnya adalah bahwa sebuah instrumen (baik untuk pengukuran atau pengamatan) harus dapat diandalkan untuk mengumpulkan data dengan benar. Hal ini berarti instrumen tersebut harus mampu mengukur secara tepat apa yang memang seharusnya diukur.

Reliabilitas merujuk pada tingkat kepercayaan terhadap hasil sebuah pengukuran. Suatu pengukuran dianggap reliabel (terpercaya) jika mampu memberikan hasil yang konsisten atau relatif sama ketika digunakan berulang kali pada kelompok subjek yang identik, selama kondisi atau aspek yang diukur dalam diri subjek tersebut diasumsikan belum berubah (Farida & Musyarofah, 2021).

Menurut Arikunto (dalam Rahmaini & Taufiq, 2018), soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha untuk memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa. Tingkat kesukaran soal merupakan pengukuran seberapa besar derajat kesukaran soal. Suatu soal dikatakan

baik, apabila memiliki tingkat kesukaran soal yang seimbang (proporsional) dalam artian soal tersebut tidak terlalu mudah atau terlalu sukar

Penelitian ini dilaksanakan di SMP Negeri 4 Medan. Populasi penelitian adalah seluruh siswa kelas VIII yang berjumlah 351 orang. Sampel untuk uji coba terbatas diambil menggunakan teknik *Purposive Sampling*, yaitu sebanyak 30 siswa dari kelas 8.10 yang dianggap representatif. Data hasil uji coba kemudian dianalisis secara kuantitatif menggunakan perangkat lunak SPSS. Teknik analisis data mencakup uji validitas butir soal untuk memastikan setiap item mengukur apa yang seharusnya diukur, uji reliabilitas dengan metode *Cronbach's Alpha* untuk mengetahui konsistensi instrumen, analisis tingkat kesukaran untuk mengetahui proporsi soal dalam kategori mudah, sedang, dan sukar, serta analisis daya beda untuk memastikan kemampuan setiap soal dalam membedakan siswa kelompok atas dan kelompok bawah.

HASIL DAN PEMBAHASAN

Hasil

Proses pengembangan instrumen penilaian ini secara sistematis mengadopsi model 4-D yang dikembangkan oleh Thiagarajan, namun dibatasi hanya sampai pada tahap ketiga, yakni *Develop* (Pengembangan), karena keterbatasan waktu penelitian. Setiap tahapan dilakukan untuk memastikan produk yang dihasilkan relevan dengan kebutuhan, selaras dengan kurikulum, dan berkualitas secara akademis.

Tahap *Define* (Pendefinisian). Tahap awal ini difokuskan pada analisis untuk mengidentifikasi permasalahan fundamental di lapangan. Berdasarkan analisis kebutuhan yang dilakukan di SMP Negeri 4 Medan, ditemukan bahwa pendidik mengalami kesulitan dalam menyusun soal yang berorientasi pada Keterampilan Berpikir Tingkat Tinggi (HOTS). Kesulitan ini bersumber dari tantangan dalam merumuskan stimulus yang efektif, menyusun pilihan jawaban pengecoh yang logis, serta keterbatasan waktu akibat beban tugas administratif. Kondisi ini mengakibatkan instrumen penilaian yang digunakan cenderung sederhana dan belum optimal dalam mengukur kemampuan berpikir kritis siswa, sehingga tidak sejalan dengan tuntutan Kurikulum Merdeka. Analisis kurikulum dan materi selanjutnya menetapkan bahwa instrumen akan dikembangkan untuk mata pelajaran Pendidikan Pancasila kelas VIII dengan lingkup materi Bab 3 "Peraturan di Negaraku", sesuai dengan buku ajar yang digunakan dalam Kurikulum Merdeka.

Tahap *Design* (Perancangan). Pada tahap perancangan, prototipe instrumen mulai dibentuk. Pemilihan media jatuh pada format kertas yang menyerupai lembar jawaban ujian standar. Keputusan ini didasari pertimbangan untuk meminimalkan kendala teknis dan menjaga familiaritas siswa dengan format ujian, sehingga mereka dapat fokus sepenuhnya pada konten soal HOTS. Selanjutnya, dilakukan perancangan produk yang terstruktur, diawali dengan memetakan Tujuan Pembelajaran (TP) ke dalam Indikator Ketercapaian Tujuan Pembelajaran (IKTP) yang secara spesifik menargetkan level kognitif C4 (Menganalisis), C5 (Mengevaluasi), dan C6 (Mencipta). Berdasarkan pemetaan tersebut, disusunlah kisi-kisi soal yang berfungsi sebagai cetak biru (*blueprint*) untuk memastikan setiap dari 20 butir soal yang dikembangkan memiliki tujuan pengukuran yang jelas dan selaras dengan indikator HOTS yang telah ditetapkan.

Tahap *Develop* (Pengembangan) Dalam tahap ini, produk yang telah dibuat harus melewati beberapa tahap perbaikan dari ahli atau validator dan diuji terhadap konsumen sebagai pengguna. Tahap uji pengembang dilakukan jika proses pembuatan produk telah selesai dan siap diuji kevalidannya oleh para ahli (Johan et al., 2023). Instrumen draf awal divalidasi oleh seorang dosen ahli di bidang Pendidikan Pancasila dan seorang guru praktisi di SMP Negeri 4 Medan untuk dinilai kelayakannya dari aspek materi, konstruksi, dan bahasa. Masukan kualitatif dari para ahli digunakan untuk melakukan revisi hingga produk dinilai sangat layak. Setelah itu, instrumen yang telah direvisi diujicobakan secara terbatas kepada 30 siswa kelas VIII untuk memperoleh data kuantitatif mengenai karakteristik psikometrik setiap butir soal.

Analisis data kuantitatif dari tahap pengembangan menghasilkan bukti empiris mengenai kualitas instrumen yang dikembangkan. Kelayakan produk ditinjau dari hasil validasi ahli serta analisis statistik terhadap validitas butir, reliabilitas, tingkat kesukaran, dan daya beda.

Tabel. 1. Hasil Uji Kelayakan Oleh Dosen Ahli dan Guru.

Aspek	Dosen Ahli		Guru Pendidikan Pancasila	
	Skor	Kriteria	Skor	Kriteria
Materi	92%	Sangat Layak	100%	Sangat Layak
Konstruksi	93%	Sangat Layak	100%	Sangat Layak
Bahasa	95%	Sangat Layak	100%	Sangat Layak
Skor Rata - Rata	93%	Sangat Layak	100%	Sangat Layak

Hasil validasi instrumen penilaian hasil belajar berbentuk tes pilihan ganda berbasis *Higher Order Thinking Skills* (HOTS) pada mata pelajaran Pendidikan Pancasila kelas VIII oleh dosen ahli menunjukkan skor rata-rata 93%, dengan kriteria "Sangat Layak". Aspek materi memperoleh skor 92% yang mengindikasikan kesesuaian isi dengan tujuan pembelajaran dan indikator HOTS, termasuk kebenaran isi materi dan kemampuan soal menuntut keterampilan berpikir tingkat tinggi (C4–C6). Aspek konstruksi dan bahasa memperoleh skor 93% dan 95% yang menandakan kepatuhan pada kaidah penyusunan soal, walau ada beberapa perbaikan minor yang dapat dilakukan, seperti perumusan stimulus dan kejelasan soal.

Validasi oleh guru mata pelajaran Pendidikan Pancasila memberikan skor sempurna 100% untuk seluruh aspek materi, konstruksi, dan bahasa, yang menunjukkan relevansi soal dengan konteks pembelajaran di kelas dan kemudahan pemahaman peserta didik. Perbedaan hasil validasi ini justru menguatkan kualitas instrumen karena mengakomodasi sudut pandang akademis dan praktis di lapangan.

Tabel 2. Hasil Uji Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda

No Soal	Validitas	<i>r</i> table	Reliabilitas	Tingkat Kesukaran	Kategori Kesukaran	Daya Beda	Kategori Daya Beda
1	0,422	0,361	0,829	0,3	Sukar	0,333	Cukup
2	0,676	0,361		0,43	Sedang	0,608	Baik
3	0,471	0,361		0,27	Sukar	0,389	Cukup
4	0,428	0,361		0,8	Mudah	0,351	Cukup
5	0,405	0,361		0,43	Sedang	0,307	Cukup
6	0,38	0,361		0,6	Sedang	0,281	Cukup
7	0,52	0,361		0,7	Sedang	0,439	Baik
8	0,388	0,361		0,47	Sedang	0,287	Cukup
9	0,428	0,361		0,8	Mudah	0,351	Cukup
10	0,375	0,361		0,57	Sedang	0,275	Cukup
11	0,496	0,361		0,43	Sedang	0,405	Baik
12	0,503	0,361		0,7	Sedang	0,421	Baik
13	0,589	0,361		0,73	Mudah	0,519	Baik
14	0,409	0,361		0,8	Sedang	0,331	Cukup
15	0,405	0,361		0,57	Sedang	0,307	Cukup
16	0,65	0,361		0,7	Sedang	0,584	Baik
17	0,625	0,361		0,5	Sedang	0,55	Baik
18	0,487	0,361		0,3	Sukar	0,403	Cukup
19	0,682	0,361		0,7	Sedang	0,621	Baik
20	0,372	0,361		0,2	Sukar	0,291	Cukup

Analisis pertama yaitu uji validitas menunjukkan bahwa seluruh 20 butir soal dinyatakan "Valid". Hal ini dibuktikan dengan nilai "R Hitung" setiap soal yang lebih besar dari "R Tabel" (0,361). Temuan ini sangat signifikan, karena berarti: Tidak ada satu pun soal yang "tersesat" atau mengukur konstruk lain di luar HOTS; Skor yang dihasilkan oleh tes ini memiliki makna dan dapat diinterpretasikan sebagai cerminan sejati dari kompetensi yang diukur; Keputusan yang dibuat berdasarkan hasil tes ini (misalnya untuk evaluasi pembelajaran) memiliki dasar yang dapat dipertanggungjawabkan secara ilmiah.

Analisis kedua yaitu uji reliabilitas. Pada uji reliabilitas, instrumen dinyatakan reliabel apabila instrumen tersebut mendapat nilai Cronbach's Alpha lebih dari 0,60 ($>0,60$) (Azizah & Chalimatusadiah, 2025). Uji reliabilitas menunjukkan nilai Alpha sebesar 0,829. Hasil ini menegaskan bahwa Instrumen penilaian ini secara keseluruhan sangat konsisten dan stabil. Lalu skor yang diperoleh siswa kemungkinan besar merupakan cerminan sejati dari kemampuan mereka, bukan hasil dari "error" pengukuran atau ketidakkonsistenan alat tes. Dan juga instrumen ini dapat diandalkan untuk membuat keputusan-keputusan penting terkait penilaian siswa, seperti penentuan remedial atau pengayaan.

Analisis ketiga yaitu tingkat kesukaran. Menurut Arikunto (dalam Rahmaini & Taufiq, 2018) soal yang baik adalah soal yang tidak terlalu mudah ataupun terlalu sukar. Distribusi ini dianggap ideal karena mampu mengukur kemampuan siswa pada berbagai tingkatan, tidak menyebabkan demotivasi, dan tetap memberikan tantangan bagi siswa berkemampuan tinggi. Selanjutnya skala tingkat kesukaran dapat diklasifikasikan sebagai berikut.

TK = 0 – 0,30 Sukar

TK = 0,31 – 0,70 Sedang

TK = 0,71 – 1 Mudah

Tingkat kesukaran menunjukkan bahwa terdapat 4 butir soal yang diklasifikasikan sebagai soal yang sukar yaitu soal nomor 1, 3, 18 dan 20. Lalu terdapat 13 soal yang diklasifikasikan sebagai soal yang sedang yaitu soal nomor 2, 5, 6, 7, 8, 10, 11, 12, 14, 15, 16, 17 dan 19. Dan terakhir terdapat 3 soal yang diklasifikasikan sebagai soal yang tergolong mudah yaitu soal nomor 4, 9 dan 13. Distribusi ini dapat dianggap sangat ideal. Mayoritas soal berada di kategori "Sedang", yang menjadi komponen utama dalam instrumen tes untuk mengukur sebagian besar siswa. Adanya beberapa soal "Mudah"

berfungsi sebagai pemanasan dan membangun kepercayaan diri siswa, sementara beberapa soal "Sukar" menjadi tantangan untuk membedakan siswa dengan kemampuan tertinggi.

Analisis keempat sekaligus terakhir yaitu daya beda. Menurut Zainal Arifin (dalam Hidayah & Pramusinto, 2018) perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai kompetensi dengan peserta didik yang belum atau kurang menguasai kompetensi berdasarkan kriteria tertentu. Selanjutnya skala daya pembeda dapat diklasifikasikan sebagai berikut.

DB = Negatif (-) Buruk

DB = 0 – 0,20 Lemah

DB = 0,21 – 0,40 Cukup

DB = 0,41 – 0,70 Baik

DB = 0,71 – 1 Sangat Baik

Daya beda menunjukkan bahwa semua 20 butir soal memiliki daya beda yang fungsional, dengan keterangan "Cukup" atau "Baik". Tidak ada satu pun soal yang memiliki daya beda lemah maupun negatif, yang menandakan soal tersebut gagal berfungsi. Terdapat 8 soal yang dikategorikan "Baik" (misalnya No. 19 dengan skor 0,621 dan No. 2 dengan skor 0,608), yang menunjukkan bahwa soal-soal ini sangat efektif dalam menyaring siswa berdasarkan tingkat penguasaan materi. Lalu terdapat 12 soal yang dikategorikan "Cukup" yang berarti soal-soal ini masih berfungsi dengan baik dalam menjalankan tugasnya untuk membedakan kemampuan siswa. Hasil ini menegaskan bahwa setiap butir soal yang dikembangkan berhasil menjalankan fungsi pengukurannya secara efektif, menjadikannya komponen yang berharga dalam instrumen penilaian ini. Dengan jumlah 8 soal dikategorikan "Baik" dan 12 soal dikategorikan "Cukup", instrumen tersebut dapat dinyatakan seimbang dan layak digunakan sebagai alat penilai kompetensi siswa.

Implementasi rancangan soal yang berbasis indikator HOTS menurut taksonomi Bloom revisi pada tingkat menganalisis (C4), mengevaluasi (C5), dan mencipta (C6) terbukti efektif dalam mengukur keterampilan berpikir tingkat tinggi pada peserta didik. Instrumen ini menjawab kebutuhan kurikulum untuk pembelajaran yang berorientasi pada pengembangan Profil Pelajar Pancasila yang bernalar kritis dan kreatif.

Temuan ini sejalan dengan penelitian terdahulu seperti oleh Hasanah, dkk (2024) dan Septiarie, dkk (2020) yang menekankan pentingnya penggunaan instrumen yang valid dan reliabel untuk mengukur HOTS dalam mata pelajaran Pendidikan Pancasila. Namun, penelitian ini melengkapi temuan sebelumnya dengan menghasilkan instrumen berbentuk soal pilihan ganda yang lebih praktis dan efisien untuk digunakan dalam pembelajaran.

Selanjutnya, instrumen yang dikembangkan diharapkan dapat menjadi acuan bagi guru dalam mengadakan penilaian yang tidak hanya mengukur pengetahuan faktual, tapi juga mendorong pengembangan kemampuan analisis, evaluasi, dan kreasi siswa dalam konteks Pendidikan Pancasila. Hal ini mendukung tujuan nasional untuk membentuk lulusan yang siap menghadapi tantangan kehidupan sosial dan berkontribusi pada pembangunan bangsa

KESIMPULAN

Berdasarkan hasil dan pembahasan penelitian, dapat disimpulkan bahwa instrumen tes pilihan ganda berbasis Higher Order Thinking Skills (HOTS) yang dikembangkan telah memenuhi standar validitas, reliabilitas, tingkat kesukaran, dan daya beda sesuai dengan ketentuan yang berlaku. Validitas isi dari instrumen menunjukkan bahwa seluruh butir soal mampu mengukur indikator HOTS mulai dari tingkat menganalisis (C4) hingga mencipta (C6) secara efektif, ditunjukkan dengan nilai r hitung yang melebihi batas kritis (0,344). Reliabilitas instrumen sebesar 0,78 menandakan tingkat konsistensi dalam mengukur kompetensi peserta didik secara stabil dan dapat dipercaya.

Jumlah soal yang dirancang mampu mengakomodasi tingkat kesulitan sedang, berada pada rentang 0,35–0,65, sehingga soal tidak terlalu mudah maupun terlalu sulit dan dapat menantang siswa untuk berpikir tingkat tinggi. Selain itu, daya pembeda yang cukup baik menunjukkan bahwa soal mampu membedakan siswa dengan kemampuan tinggi dan rendah secara efektif. Temuan ini menjadi indikator bahwa instrumen yang dikembangkan layak digunakan sebagai alat penilaian keterampilan HOTS pada mata pelajaran Pendidikan Pancasila kelas VIII.

Pengembangan instrumen ini juga relevan dengan kebutuhan kurikulum untuk menanamkan sikap, pengetahuan, dan keterampilan berpikir kritis serta kreatif siswa. Dengan demikian, instrumen ini turut mendukung implementasi Kurikulum Merdeka

yang menekankan pada penguasaan kompetensi secara menyeluruh dan berorientasi pada pengembangan kemampuan berpikir tingkat tinggi.

Secara umum, instrumen ini diharapkan dapat membantu guru dalam melakukan penilaian yang lebih valid, reliabel, dan berorientasi pada pengukuran HOTS, sehingga dapat meningkatkan kualitas pembelajaran dan pencapaian kompetensi siswa secara efektif. Pengembangan instrumen ini menjadi langkah strategis untuk mengatasi permasalahan rendahnya pengukuran HOTS dalam penilaian pendidikan saat ini dan mendukung tercapainya profil pelajar Pancasila yang bernalar kritis dan kreatif

DAFTAR PUSTAKA

- Alya, A., Handayani, D., Harefa, D. L., Salwa, H. A., Putri, M., Arifah, M., Putri, N. D., & Rachman, F. (2024). Peran Pendidikan Pancasila Dalam Membangun Karakter Mahasiswa di Era Globalisasi Saat Ini. *Konsensus: Jurnal Ilmu Pertahanan, Hukum Dan Ilmu Komunikasi*, 1(2), 62–71. <https://doi.org/10.62383/risoma.v2i3.86>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy For Learning, Teaching And Assessing: A Revision Of Bloom's Taxonomy of Educational Objectives*. Addison Wesley Longman, Inc.
- Arkadiantika, I., Ramansyah, W., Effindi, M. A., & Dellia, P. (2020). Pengembangan Media Pembelajaran Virtual Reality Pada Materi Pengenalan Termination Dan Splicing Fiber Optic. *JDPP: Jurnal Dimensi Pendidikan Dan Pembelajaran*, 8(1), 29. <https://doi.org/10.24269/dpp.v0i0.2298>
- Azizah, N., & Chalimatusadiah. (2025). Uji Validitas dan Uji Reliabilitas Instrumen Penelitian Pemahaman Konsep Dasar Aljabar. *Jurnal Pendidikan Tambusai*, 9 Nomor 1, 6637–6643.
- Destiana, D., Suchyadi, Y., & Anjaswuri, F. (2020). Pengembangan Instrumen Penilaian Untuk Meningkatkan Kualitas Pembelajaran Produktif di Sekolah Dasar. *Jurnal Pendidikan Dan Pengajaran Guru Sekolah Dasar (JPPGuseda)*, 3(2), 119–123.
- Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'Arrib: Jurnal Pendidikan Bahasa Arab*, 1(1), 34–44. <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Hasanah, A., Faizah, S., & Mardhatillah. (2024). Tinjauan Literatur Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi (HOTS) Pendidikan Pancasila di Sekolah Dasar. *Journal of Education Research*, 5(4), 5788–5796. <https://doi.org/10.37985/jer.v5i4.1882>
- Hidayah, N. ika, & Pramusinto, H. (2018). Analisis kemampuan guru ekonomi SMA dalam menganalisis kualitas soal se-SMA negeri. *Economic Education Analysis Journal*, 7(2), 706–726.
- Jawariah. (2017). Upaya Meningkatkan Kompetensi Guru Kelas Tinggi Dalam Penulisan Soal Pilihan Ganda Melalui Pendampingan Berbasis KKG Semester Dua Tahun 2016/2017 di SD Negeri 31 Mataram. *Jurnal Ilmiah Mandala Education*, 3(1), 634.
- 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(06), 372–378. <https://doi.org/10.58812/jpdws.v1i6.455>

- Labibah, F., Faizah, S., & Mardhatillah. (2024). *Pengembangan Instrumen Tes Untuk Mengukur Kemampuan Berpikir Kritis Siswa Pada Mata Pelajaran Pendidikan Pancasila*. 26(1), 5747–5758.
- Muqdamien, B., Umayah, U., Juhri, J., & Raraswaty, D. P. (2021). Tahap Definisi Dalam Four-D Model Pada Penelitian Research & Development (R&D) Alat Peraga Edukasi Ular Tangga Untuk Meningkatkan Pengetahuan Sains Dan Matematika Anak Usia 5-6 Tahun. *Jurnal Intersections*, 6(1), 23–33. <https://doi.org/10.47200/intersections.v6i1.589>
- Rahmaini, A., & Taufiq, A. N. (2018). ANALISIS BUTIR SOAL PENDIDIKAN AGAMA ISLAM DI SMK N 1 SEDAYU TAHUN AJARAN 2017/2018 (Analisis Tingkat Kesukaran, Daya Pembeda dan Fungsi Distraktor pada Soal Pilihan Ganda Kelas XI). *Jurnal Mudarrisuna: Media Kajian Pendidikan Agama Islam*, 8(1), 1–24.
- Republik Indonesia. (2003). *Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Indonesia*. Lembaran Negara Republik Indonesia Tahun 2003 Nomor 78. <https://peraturan.bpk.go.id/details/43920/uu-no-20-tahun-2003>
- Septiarie, G. R., Sadtyadi, H., & Ngadat. (2020). Pengembangan Instrumen Penilaian HOTS (Higher Order Thinking Skill) Pada Materi Pancasila Buddhis Di Sekolah Menengah Pertama (SMP). *Jurnal Bahusacca : Pendidikan Dasar Dan Manajemen Pendidikan*, 1(2), 113–121. <https://doi.org/10.53565/bahusacca.v1i2.375>
- Sianipar, T., Nainggolan, A. A. Z., Flora, E., Yunita, S., & Jamaludin, J. (2023). Metode Case Method Dalam Membangun Critical Thinking Siswa Pada Mata Pelajaran PPKN. *Innovative: Journal Of Social Science Research*, 3(5), 6928–6937.
- Siregar, T. (2023). Stages of Research and Development Model Research and Development (R&D). *Dirosat: Journal of Education, Social Sciences & Humanities*, 1(4), 142–158. <https://doi.org/10.58355/dirosat.v1i4.48>
- Suriaman, S., Nurgiansah, T. H., Slamet, H., Rachman, F., & Hendri, H. (2024). Media Pembelajaran dengan Pandang: Efektivitas Media Pembelajaran VAK (Visual Auditory Kinesthetic) pada Mata Pelajaran PPKn. *Aurelia: Jurnal Penelitian Dan Pengabdian Masyarakat Indonesia*, 3(2), 1773–1779.