



JURNAL RISET DAN INOVASI PENDIDIKAN SAINS (JRIPS)

Vol. 5 No. 2 (2026) pp. 218-229

<http://jurnal.umb.ac.id/index.php/JRIPS/>

p-ISSN: 2809-5200 e-ISSN: 2809-5219

Pengembangan Instrumen Penilaian *Higher Order Thinking Skills* untuk Materi Bentuk Molekul Berdasarkan Teori VSEPR

Annisa Jovania Rizky Ermansyah^{1*}, Afrida², Asmiyunda³, Epinur⁴, Kriswantoro⁵

^{1*2,3,4,5} Pendidikan Kimia, Universitas Jambi

*Corresponden Author : annisajovaniaa@gmail.com

ABSTRAK

Kemampuan *Higher Order Thinking Skills* (HOTS) menjadi kompetensi penting dalam pembelajaran abad ke-21, namun instrumen penilaian kimia yang mampu mengukur kemampuan tersebut secara komprehensif masih terbatas, khususnya pada materi bentuk molekul berdasarkan teori VSEPR. Penelitian ini bertujuan mengembangkan instrumen penilaian HOTS yang valid, reliabel, serta memiliki karakteristik butir soal yang baik pada materi bentuk molekul berdasarkan teori VSEPR di SMA. Penelitian menggunakan metode *Research and Development* (R&D) dengan model pengembangan 4D yang meliputi tahap *define, design, develop, dan disseminate*. Uji coba dilakukan terhadap 29 peserta didik kelas XI MIPA SMAS Adhyaksa 1 Kota Jambi. Data diperoleh melalui lembar validasi ahli dan hasil uji coba instrumen, kemudian dianalisis menggunakan validitas isi Aiken's V, validitas butir Pearson Product Moment, reliabilitas Alpha Cronbach, tingkat kesukaran, dan daya beda. Hasil penelitian menghasilkan instrumen penilaian HOTS berbentuk soal esai yang terdiri atas 5 soal dengan 15 sub soal pada level kognitif C4, C5, dan C6. Instrumen memperoleh rata-rata nilai validitas isi sebesar 0,88 dengan kategori valid dan koefisien reliabilitas Alpha Cronbach sebesar 0,735 yang menunjukkan reliabilitas tinggi. Analisis butir menunjukkan 9 butir valid, 1 butir mudah, 5 butir sedang, dan 9 butir sukar, serta daya beda terdiri atas 2 butir berkategori baik, 9 cukup, dan 4 kurang. Dengan demikian, instrumen yang dikembangkan layak digunakan untuk mengukur kemampuan HOTS peserta didik pada materi bentuk molekul berdasarkan teori VSEPR serta berpotensi mendukung pelaksanaan penilaian yang lebih berkualitas dalam pembelajaran kimia.

Kata Kunci: HOTS, Instrumen Penilaian, Bentuk Molekul, Teori VSEPR

PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi pada abad ke-21 menuntut peserta didik memiliki kemampuan *Higher Order Thinking Skills* (HOTS), yaitu kemampuan menganalisis, mengevaluasi, dan mencipta untuk menyelesaikan

berbagai permasalahan secara kritis dan kreatif (Thornhill-Miller et al., 2023). Kemampuan tersebut sejalan dengan Taksonomi Bloom revisi yang menempatkan analisis (C4), evaluasi (C5), dan kreasi (C6) sebagai tingkat kognitif tertinggi (Anderson & Krathwohl, 2001). Sejalan dengan itu, Kurikulum Merdeka menekankan pentingnya asesmen yang mampu mengukur kemampuan berpikir tingkat tinggi. Namun, hasil *Programme for International Student Assessment* (PISA) menunjukkan bahwa kemampuan berpikir tingkat tinggi peserta didik Indonesia masih relatif rendah sehingga diperlukan instrumen penilaian yang valid dan reliabel untuk mengukur HOTS secara tepat.

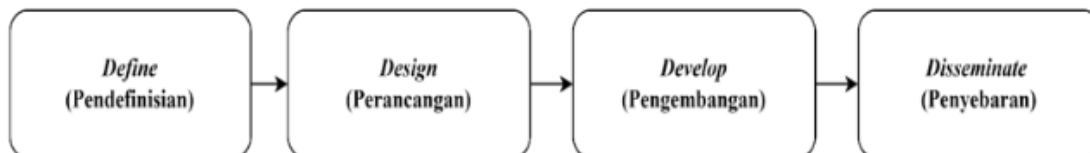
Pada pembelajaran kimia, materi bentuk molekul berdasarkan teori *Valence Shell Electron Pair Repulsion* (VSEPR) merupakan materi yang bersifat abstrak dan memerlukan kemampuan analisis serta penalaran tingkat tinggi (Effendy (2016); Petrucci et al. (2011)). Hasil observasi di SMAS Adhyaksa 1 Kota Jambi menunjukkan bahwa instrumen penilaian yang digunakan guru masih didominasi soal pada level kognitif rendah (C1–C3) dan berbentuk pilihan ganda sehingga belum mampu mengukur HOTS secara optimal. Temuan tersebut sejalan dengan penelitian Lestari et al. (2023) yang menyatakan bahwa soal evaluasi kimia di SMA masih didominasi kategori *Lower Order Thinking Skills* (LOTS).

Berbagai penelitian telah mengembangkan instrumen penilaian HOTS pada berbagai materi kimia, seperti ikatan kimia, asam–basa, hidrolisis garam, laju reaksi, dan elektrolit–non-elektrolit dengan hasil validitas dan reliabilitas yang baik (Afriani et al. (2018); Muchlis & Andromeda (2020); L. N. Azmi et al. (2021); Sinaga & Yusuf (2023); Azizah & Silfianah (2024)). Namun, pengembangan instrumen HOTS yang secara khusus ditujukan pada materi bentuk molekul berdasarkan teori VSEPR masih terbatas. Kesenjangan tersebut menunjukkan perlunya pengembangan instrumen yang memenuhi aspek validitas, reliabilitas, tingkat kesukaran, dan daya beda sehingga mampu mengukur kemampuan HOTS peserta didik secara komprehensif.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan instrumen penilaian HOTS pada materi bentuk molekul berdasarkan teori VSEPR menggunakan model pengembangan 4D (*define, design, develop, dan disseminate*). Penelitian ini difokuskan untuk menjawab pertanyaan mengenai proses pengembangan instrumen, tingkat validitas dan reliabilitas, serta karakteristik butir soal berdasarkan tingkat kesukaran dan daya beda. Hasil penelitian diharapkan dapat menyediakan instrumen penilaian yang layak digunakan dalam evaluasi pembelajaran kimia serta menjadi referensi bagi pengembangan instrumen HOTS pada materi kimia lainnya.

METODE PENELITIAN

Penelitian ini menggunakan metode Penelitian dan Pengembangan (R&D) untuk mengembangkan instrumen penilaian *Higher Order Thinking Skills* (HOTS) pada bahan bentuk molekuler berdasarkan teori *Valence Shell Electron Pair Repulsion* (VSEPR). Prosedur pengembangan mengadopsi model 4D yang diusulkan oleh Thiagarajan, Semmel, dan Semmel, yang terdiri dari empat tahap: mendefinisikan, merancang, mengembangkan, dan menyebarluaskan.



Gambar 1. Alur model pengembangan 4D.

Tahap *define* dilakukan untuk mengidentifikasi kebutuhan pengembangan instrumen penilaian HOTS melalui analisis awal-akhir (*front-end analysis*), analisis karakteristik peserta didik (*learner analysis*), analisis konsep (*concept analysis*), analisis tugas (*task analysis*), dan analisis tujuan pembelajaran (*learning objective analysis*). Tahap ini bertujuan memperoleh informasi mengenai kondisi pembelajaran, karakteristik peserta didik, serta kesesuaian materi bentuk molekuler berdasarkan teori VSEPR dengan indikator HOTS yang akan diukur. Tahap *design* meliputi penyusunan kisi-kisi instrumen, pengembangan butir soal esai HOTS, penyusunan kunci jawaban, serta penyusunan rubrik penskoran. Instrumen dikembangkan berdasarkan level kognitif menganalisis (C4), mengevaluasi (C5), dan mencipta (C6) sesuai dengan Taksonomi Bloom revisi. Tahap *develop* meliputi validasi oleh ahli, revisi instrumen berdasarkan masukan validator, dan uji coba instrumen. Selanjutnya, tahap *disseminate* dilakukan secara terbatas melalui penyebarluasan instrumen penilaian HOTS yang telah memenuhi kriteria kualitas instrumen, meliputi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, kepada guru kimia di sekolah penelitian.

Subjek penelitian terdiri atas 29 peserta didik kelas XI IPA SMAS Adhyaksa 1 Kota Jambi Tahun Ajaran 2025/2026. Instrumen penelitian yang digunakan meliputi lembar validasi ahli dan instrumen penilaian HOTS berbentuk soal esai yang dikembangkan pada level kognitif menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Data penelitian diperoleh melalui validasi ahli dan uji coba instrumen. Validitas isi dianalisis menggunakan koefisien Aiken's V, sedangkan validitas butir dianalisis menggunakan korelasi *Pearson Product Moment*. Reliabilitas instrumen dihitung menggunakan koefisien Alpha Cronbach dengan bantuan perangkat lunak IBM SPSS Statistics 23. Selain itu, karakteristik butir soal dianalisis berdasarkan tingkat kesukaran dan daya beda untuk menentukan kualitas instrumen yang dikembangkan.

HASIL DAN PEMBAHASAN

Tahap *Define*

Tahap *define* dilakukan melalui lima analisis utama, yaitu analisis awal-akhir (*front-end analysis*), analisis karakteristik peserta didik (*learner analysis*), analisis tugas (*task analysis*), analisis konsep (*concept analysis*), dan analisis tujuan pembelajaran (*learning objectives analysis*). Hasil wawancara dengan guru kimia di SMAS Adhyaksa 1 Kota Jambi menunjukkan bahwa instrumen penilaian yang digunakan meliputi pertanyaan pemantik sebagai asesmen diagnostik, lembar observasi aktivitas sebagai asesmen formatif, serta tes tertulis berbentuk pilihan ganda sebagai asesmen sumatif.

Hasil analisis terhadap soal sumatif menunjukkan bahwa distribusi tingkat kognitif belum merata. Soal pada kategori *Lower Order Thinking Skills* (LOTS) (C2–C3) mendominasi butir soal nomor 1–4, kategori *Middle Order Thinking Skills* (MOTS) (C4) terdapat pada nomor 5–6, sedangkan kategori *Higher Order Thinking Skills* (HOTS) (C5) hanya terdapat pada nomor 7–10. Selain itu, belum terdapat soal yang mengukur kemampuan mencipta (C6), dan seluruh soal masih berbentuk pilihan ganda sehingga belum mampu mengukur kemampuan peserta didik dalam mengonstruksi jawaban secara mandiri.

Analisis tugas mengacu pada Capaian Pembelajaran (CP) Kimia Fase F Kurikulum Merdeka. Instrumen yang dikembangkan mencakup tiga level HOTS, yaitu menganalisis (C4) distribusi pasangan elektron dan bentuk molekul, mengevaluasi (C5) ketepatan geometri molekul, serta mencipta (C6) prediksi bentuk molekul berdasarkan teori VSEPR. Hasil analisis konsep menghasilkan tujuh konsep utama, yaitu (1) pasangan elektron valensi, (2) teori VSEPR, (3) notasi AX_mE_n , (4) geometri domain elektron, (5) bentuk molekul, (6) sudut ikatan, dan (7) kepolaran molekul. Hasil analisis tersebut menjadi dasar dalam penyusunan instrumen penilaian HOTS.

Tahap *Design*

Tahap *design* menghasilkan rancangan instrumen penilaian HOTS berupa kisi-kisi soal yang terdiri atas lima soal utama dengan total lima belas subsoal pada level kognitif menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Instrumen disusun dalam bentuk tes esai (*paper-based test*) dengan skor maksimum 100.

Pemilihan bentuk soal esai didasarkan pada pertimbangan bahwa format ini lebih mampu mengukur kemampuan peserta didik dalam mengonstruksi, menjelaskan, dan memberikan alasan terhadap jawabannya secara mandiri. Dibandingkan dengan soal pilihan ganda, soal esai dinilai lebih efektif untuk

mengukur kemampuan berpikir tingkat tinggi, khususnya pada aspek analisis, evaluasi, dan kreasi yang menjadi karakteristik utama HOTS.

Tabel 1. Kisi-kisi Instrumen Penilaian HOTS Materi Bentuk Molekul Berdasarkan Teori VSEPR

No.	Indikator HOTS	Materi	Level Kognitif	Nomor Soal	Nomor Sub Soal	Total
1	Menganalisis distribusi pasangan elektron bebas (PEB), pasangan elektron ikatan (PEI), notasi AX_nE_m , dan bentuk molekul	Bentuk Molekul Berdasarkan Teori VSEPR	C4	1, 2	6	40
2	Mengevaluasi ketepatan geometri molekul serta pengaruh pasangan elektron bebas terhadap sudut ikatan	Bentuk Molekul Berdasarkan Teori VSEPR	C5	3, 4	6	40
3	Menciptakan prediksi bentuk dan kepolaran molekul baru beserta alasan ilmiahnya	Bentuk Molekul Berdasarkan Teori VSEPR	C6	5	3	20
Total				5 soal	15 sub soal	100

Tahap Develop

Validasi Isi (Aiken's V)

Validasi isi dilakukan melalui dua tahap, yaitu *expert review* (tahap I) dan penilaian kuantitatif menggunakan rumus Aiken's V (tahap II). Pada tahap I, validator memberikan masukan terkait redaksi soal, penggunaan kata kerja operasional, serta konstruksi butir soal. Berdasarkan saran tersebut, dilakukan revisi berupa perbaikan kata kerja operasional, penyempurnaan konstruksi soal, serta penambahan istilah yang lebih relevan agar sesuai dengan indikator HOTS dan materi bentuk molekul berdasarkan teori VSEPR.

Setelah dilakukan revisi, validasi tahap II menggunakan rumus Aiken's V menghasilkan nilai rata-rata sebesar **0,88** yang termasuk dalam kategori **valid**. Hasil tersebut menunjukkan bahwa seluruh butir soal memiliki tingkat validitas isi yang tinggi dan telah memenuhi kesesuaian dengan indikator HOTS pada materi bentuk molekul berdasarkan teori VSEPR.

Tabel 2. Hasil Analisis Validitas Isi Instrumen Menggunakan Koefisien Aiken's V

Item	Ahli		S1	S2	ΣS	n(c-1)	V	Keterangan
	1	2						
Butir 1A	4	4	3	3	6	6	1,00	Valid
Butir 1B	4	3	3	2	5	6	0,83	Valid
Butir 1C	4	3	3	2	5	6	0,83	Valid
Butir 2A	3	4	2	3	5	6	0,83	Valid

Butir 2B	3	4	2	3	5	6	0,83	Valid
Butir 2C	3	4	2	3	5	6	0,83	Valid
Butir 3A	4	4	3	3	6	6	1,00	Valid
Butir 3B	4	3	3	2	5	6	0,83	Valid
Butir 3C	4	3	3	2	5	6	0,83	Valid
Butir 3A	3	4	2	3	5	6	0,83	Valid
Butir 4B	4	4	3	3	6	6	1,00	Valid
Butir 4C	4	4	3	3	6	6	1,00	Valid
Butir 4D	4	3	3	2	5	6	0,83	Valid
Butir 5A	4	4	3	3	6	6	1,00	Valid
Butir 5B	4	3	3	2	5	6	0,83	Valid
Rata-rata V							0,88	Valid

Validitas Butir Soal

Uji validitas butir soal dilakukan menggunakan korelasi *Pearson Product Moment* dengan jumlah responden sebanyak 29 orang dan nilai r tabel sebesar 0,367 pada taraf signifikansi 5% ($\alpha = 0,05$). Hasil analisis secara manual maupun menggunakan bantuan perangkat lunak IBM SPSS Statistics 23 menunjukkan hasil yang konsisten. Dari 15 butir soal yang diujikan, terdapat 9 butir soal yang dinyatakan valid, yaitu butir 1A, 1B, 2C, 3A, 3B, 3C, 4A, 4B, dan 5A. Sebanyak 4 butir soal dinyatakan tidak valid, yaitu butir 1C, 2A, 2B, dan 5B. Selain itu, terdapat 2 butir soal, yaitu 4C dan 4D, yang gugur karena seluruh peserta didik memperoleh skor nol sehingga varians butir sama dengan nol dan koefisien validitas tidak dapat dihitung.

Tabel 3. Hasil Uji Validitas Butir Soal

Item	r Manual	r SPSS	Perbedaan	Perhitungan Manual	SPSS	Keterangan
1A	0,618	0,610	0,008	Valid	Valid	Konsisten
1B	0,684	0,683	0,001	Valid	Valid	Konsisten
1C	0,337	0,317	0,020	Tidak Valid	Tidak Valid	Konsisten
2A	0,423	0,348	0,075	Tidak Valid	Tidak Valid	Konsisten
2B	0,208	0,327	0,119	Tidak Valid	Tidak Valid	Konsisten
2C	0,457	0,374	0,083	Valid	Valid	Konsisten
3A	0,696	0,710	0,014	Valid	Valid	Konsisten
3B	0,523	0,589	0,066	Valid	Valid	Konsisten
3C	0,668	0,654	0,014	Valid	Valid	Konsisten
4A	0,693	0,672	0,021	Valid	Valid	Konsisten
4B	0,476	0,503	0,027	Valid	Valid	Konsisten
4C	-	-	-	Gugur	Gugur	Konsisten
4D	-	-	-	Gugur	Gugur	Konsisten
5A	0,410	0,429	0,019	Valid	Valid	Konsisten
5B	0,160	0,166	0,006	Tidak Valid	Tidak Valid	Konsisten

Reliabilitas

Uji reliabilitas dilakukan terhadap 13 butir soal setelah dua butir soal (4C dan 4D) dikeluarkan karena memiliki varians nol. Analisis reliabilitas menggunakan rumus Alpha Cronbach yang dihitung secara manual dan dengan bantuan perangkat lunak IBM SPSS Statistics 23. Hasil perhitungan manual

memperoleh koefisien reliabilitas sebesar 0,722, sedangkan hasil analisis menggunakan IBM SPSS Statistics 23 memperoleh koefisien sebesar 0,735. Perbedaan nilai sebesar 0,013 disebabkan oleh perbedaan tingkat ketelitian pembulatan dalam proses perhitungan. Kedua nilai koefisien reliabilitas tersebut berada di atas batas minimum 0,70, sehingga instrumen yang dikembangkan termasuk dalam kategori reliabilitas tinggi.

Tabel 4. Perbandingan Hasil Uji Reliabilitas Menggunakan Perhitungan Manual dan IBM SPSS Statistics 23

Aspek	Perhitungan Manual	IBM SPSS Statistics 23	Keterangan
Nomor Item	13	13	—
Cronbach's Alpha	0,722	0,735	Perbedaan = 0.013
Kategori	Reliable (tinggi)	Reliable (tinggi)	—

Tingkat Kesukaran

Analisis tingkat kesukaran dilakukan terhadap 15 butir soal untuk mengetahui tingkat kesulitan setiap butir dalam mengukur kemampuan peserta didik. Hasil analisis menunjukkan bahwa terdapat 1 butir soal berkategori mudah, yaitu butir 1A, 5 butir soal berkategori sedang, yaitu butir 1B, 1C, 2A, 2B, dan 3A, serta 9 butir soal berkategori sukar, yaitu butir 2C, 3B, 3C, 4A, 4B, 4C, 4D, 5A, dan 5B.

Tabel 5. Hasil Analisis Tingkat Kesukaran Butir Soal

Item	TK	Kategori	Item	TK	Kategori
1A	0,823	Mudah	3C	0,082	Sulit
1B	0,448	Sedang	4A	0,181	Sulit
1C	0,443	Sedang	4B	0,230	Sulit
2A	0,460	Sedang	4C	0	Sulit
2B	0,362	Sedang	4D	0	Sulit
2C	0,091	Sulit	5A	0,145	Sulit
3A	0,339	Sedang	5B	0,021	Sulit
3B	0,103	Sulit	-	-	-

Daya Beda

Analisis daya beda dilakukan menggunakan indeks daya beda (D) dengan membandingkan kelompok atas (27%) dan kelompok bawah (27%) dari total 29 peserta didik. Hasil analisis menunjukkan bahwa terdapat 2 butir soal berkategori baik, yaitu butir 1A dan 1B, 9 butir soal berkategori cukup, yaitu butir 1C, 2A, 2B, 3A, 3B, 3C, 4A, 4B, dan 5A, serta 4 butir soal berkategori kurang, yaitu butir 2C, 4C, 4D, dan 5B.

Tabel 6. Hasil Analisis Daya Beda Butir Soal

Item	Upper Class	Lower Class	Skor Maksimal	DP	Category	Decision
1A	7,875	3,125	8	0,594	Good	Used
1B	4,250	0,875	6	0,563	Good	Used
1C	3,000	1,500	6	0,250	Enough	Revised
2A	3,750	1,875	6	0,313	Enough	Revised
2B	2,500	1,250	6	0,208	Enough	Revised
2C	1,125	0,375	8	0,094	Not enough	Revised
3A	3,375	1,000	6	0,396	Enough	Used
3B	1,625	0,000	6	0,271	Enough	Used
3C	1,750	0,000	8	0,219	Enough	Used
4A	3,500	0,625	8	0,359	Enough	Used
4B	2,750	1,250	6	0,250	Enough	Used
4C	0,000	0,000	6	0,000	Not enough	Thrown away
4D	0,000	0,000	10	0,000	Not enough	Thrown away
5A	2,000	0,125	5	0,375	Enough	Used
5B	0,250	0,125	5	0,025	Not enough	Revised

Rekapitulasi Hasil Analisis Butir Soal

Berdasarkan rekapitulasi hasil analisis validitas, reliabilitas, tingkat kesukaran, dan daya beda, diperoleh 8 butir soal yang memenuhi kriteria dan layak digunakan, yaitu butir 1A, 1B, 3A, 3B, 3C, 4A, 4B, dan 5A. Sebanyak 5 butir soal memerlukan revisi, yaitu butir 1C, 2A, 2B, 2C, dan 5B, sedangkan 2 butir soal dinyatakan gugur, yaitu butir 4C dan 4D.

Pembahasan

Penelitian ini berhasil mengembangkan instrumen penilaian *Higher Order Thinking Skills* (HOTS) pada materi bentuk molekul berdasarkan teori VSEPR menggunakan model pengembangan 4D yang terdiri atas tahap *define*, *design*, *develop*, dan *disseminate*. Model ini dipilih karena memiliki tahapan yang sistematis dalam menghasilkan produk yang sesuai dengan kebutuhan pembelajaran (Indaryanti et al., 2025). Berdasarkan analisis kebutuhan, instrumen penilaian yang digunakan di SMAS Adhyaksa 1 Kota Jambi masih didominasi soal pada level LOTS dan MOTS, sehingga belum mampu mengukur kemampuan berpikir tingkat tinggi peserta didik secara optimal. Temuan ini sejalan dengan penelitian Kriswantoro & Lucya Wulandari (2024) yang menyatakan bahwa instrumen evaluasi kimia di SMA belum sepenuhnya mengakomodasi kemampuan HOTS, terutama pada level mencipta (C6). Kondisi tersebut juga didukung oleh hasil PISA 2022 yang menunjukkan bahwa kemampuan literasi sains peserta didik Indonesia masih berada di bawah rata-rata OECD (Naura Ayu et al., 2025).

Instrumen yang dikembangkan berupa soal esai pada level kognitif C4, C5, dan C6. Pemilihan bentuk soal esai didasarkan pada kemampuannya dalam mengungkap proses berpikir peserta didik secara lebih mendalam dibandingkan

soal pilihan ganda. Menurut Viniyasi et al. (2022), soal esai lebih efektif untuk mengukur kemampuan analisis, evaluasi, dan kreasi karena peserta didik dituntut mengonstruksi jawaban berdasarkan penalarannya sendiri. Oleh karena itu, format ini dinilai sesuai untuk mengukur kemampuan HOTS pada materi bentuk molekul berdasarkan teori VSEPR yang bersifat abstrak dan membutuhkan penalaran konseptual.

Hasil validasi isi menunjukkan nilai rata-rata Aiken's V sebesar 0,88 yang termasuk kategori valid. Nilai tersebut menunjukkan bahwa butir soal telah sesuai dengan indikator pembelajaran, materi, dan karakteristik HOTS. Hasil ini sejalan dengan pendapat Utami et al. (2024) yang menyatakan bahwa instrumen dengan nilai Aiken's V di atas batas minimum memiliki representasi isi yang baik. Selain itu, proses validasi yang dilakukan melalui *expert review* dan penilaian kuantitatif memungkinkan perbaikan terhadap redaksi, penggunaan kata kerja operasional, dan konstruksi soal sehingga kualitas instrumen menjadi lebih baik (Sirya & Rahayu, 2024)

Uji validitas butir menunjukkan bahwa dari 15 butir soal yang dikembangkan terdapat 9 butir soal yang valid, sedangkan dua butir soal dinyatakan gugur karena seluruh peserta didik memperoleh skor nol sehingga tidak memiliki varians. Kondisi tersebut mengindikasikan bahwa tingkat kesulitan kedua butir soal terlalu tinggi sehingga belum mampu membedakan kemampuan peserta didik. Hasil uji reliabilitas memperoleh koefisien Alpha Cronbach sebesar 0,735 yang menunjukkan bahwa instrumen memiliki konsistensi internal yang tinggi. Hasil ini sejalan dengan penelitian Azmi et al. (2021) yang menyatakan bahwa instrumen HOTS yang dikembangkan melalui proses validasi yang sistematis mampu menghasilkan reliabilitas yang baik.

Analisis tingkat kesukaran menunjukkan bahwa sebagian besar butir soal berada pada kategori sukar, sedangkan analisis daya beda menunjukkan sebagian besar butir soal memiliki kategori cukup hingga baik. Dominasi soal berkategori sukar merupakan karakteristik yang umum pada instrumen HOTS karena dirancang untuk mengukur kemampuan berpikir tingkat tinggi pada level analisis, evaluasi, dan kreasi Widana (2017). Sementara itu, daya beda yang didominasi kategori cukup menunjukkan bahwa sebagian besar butir soal telah mampu membedakan peserta didik yang memiliki kemampuan tinggi dan rendah. Berdasarkan rekapitulasi seluruh analisis, diperoleh delapan butir soal yang memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, dan daya beda sehingga layak digunakan sebagai instrumen penilaian HOTS pada materi bentuk molekul berdasarkan teori *VSEPR*.

KESIMPULAN

Penelitian ini berhasil mengembangkan instrumen penilaian *Higher Order Thinking Skills* (HOTS) pada materi bentuk molekul berdasarkan teori VSEPR menggunakan model pengembangan 4D yang meliputi tahap *define*, *design*, *develop*, dan *disseminate*. Instrumen yang dikembangkan berupa soal esai yang mengukur kemampuan berpikir tingkat tinggi pada level kognitif menganalisis (C4), mengevaluasi (C5), dan mencipta (C6). Hasil validasi isi menunjukkan nilai rata-rata Aiken's V sebesar 0,88 dengan kategori valid, sedangkan uji reliabilitas memperoleh koefisien Alpha Cronbach sebesar 0,735 yang menunjukkan reliabilitas tinggi. Berdasarkan analisis validitas butir, tingkat kesukaran, dan daya beda, diperoleh delapan butir soal yang memenuhi seluruh kriteria kualitas instrumen. Dengan demikian, instrumen yang dikembangkan dinyatakan layak digunakan sebagai alternatif instrumen penilaian untuk mengukur kemampuan *Higher Order Thinking Skills* (HOTS) peserta didik pada materi bentuk molekul berdasarkan teori VSEPR.

DAFTAR PUSTAKA

- Afriani, Setyarini, M., & Efkar, T. (2018). Pengembangan Instrumen Asesmen Pengetahuan Berbasis HOTS pada Materi Elektrolit dan Non Elektrolit. *Jurnal Pendidikan Dan Pembelajaran Kimia*. <https://doi.org/http://dx.doi.org/10.23960/jppk>
- Anderson, L. W. & Krathwohl, D. R.(2001). *A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives*. Longman.
- Azizah, I. N., & Silfianah, I. (2024). Pengembangan Instrumen Penilaian Kognitif Berbasis Higher Order Thinking Skills (HOTS) Berbantu Aplikasi Quizizz Pada Materi Ikatan Kimia Dan Gaya Antarmolekul. *UNESA Journal of Chemical Education*, 13(2), 159–170. <https://doi.org/10.26740/ujced.v13n2.p159-170>
- Azmi, N. L., Nurhayati, S., Priatmoko, S., & Wardani, S. (2021). Pengembangan Instrumen Tes Untuk Mengukur Hots Peserta Didik Pada Materi Laju Reaksi. *Journal of Chemistry In Education*, 10(1). <https://doi.org/https://doi.org/10.15294/chemined.v10i1.41123>
- Effendy. (2016). *Kimia Dasar untuk Perguruan Tinggi*. Bayumedia Publishing.
- Kriswantoro, K., & Lucy Wulandari. (2024). Karakteristik Instrumen Tes Higher Order Thinking Skills (HOTS) Kimia SMA. *Jurnal Pendidikan MIPA*, 14(4), 1060–1066. <https://doi.org/10.37630/jpm.v14i4.2063>
- Lestari, M., Musta, R., & Rahmanpiu. (2023). S A I N S Analisis Kualitas Soal Ujian Akhir Semester Mata Pelajaran Kimia SMA Negeri 1 Napaballano.

- Muchlis, I. P., & Andromeda, A. (2020). Pengembangan Instrumen Tes Berbasis Higher Order Thinking Skill Pada Materi Hidrolisis Garam Untuk Siswa SMA/MA. *Jurnal Eksakta Pendidikan (JEP)*, 4(2), 218. <https://doi.org/10.24036/jep/vol4-iss2/520>
- Naura Ayu, G., Andini Putri, C., Rifky Riyanto, A., & Koto, I. (2025). The Scientific Literacy Competence of Students in Indonesia and Mexico Based on PISA 2022: An International Comparative Study. *The Future of Education Journal*, 4, Page. <https://journal.tofedu.or.id/index.php/journal/index>
- Petrucci, R. H., F. Geoffrey Herring, Jeffry D. Madura, & Carey Bissonnette. (2011). *Kimia Dasar: Prinsip dan Terapan Modern*. Erlangga.
- Sinaga, Y. E., & Yusuf, M. (2023). PENGEMBANGAN INSTRUMEN TES BERBASIS HIGHER ORDER THINKING SKILLS (HOTS) PADA MATERI ASAM BASA. *Jurnal Pendidikan Sosial Dan Humaniora*, 2(3). <https://publisherqu.com/index.php/pediaqu>
- Sirya, Z. A., & Rahayu, Y. S. (2024). Validity of Misconception Test Instrument Using Four-Tier True-False Test in Plant Tissue Structure and Function Material. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, 13(1). <https://doi.org/https://doi.org/10.26740/bioedu.v13n1.p22-29>
- Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. In *Journal of Intelligence* (Vol. 11, Number 3). MDPI. <https://doi.org/10.3390/jintelligence11030054>
- Utami, L., Festiyed, Dian Purnama Ilahi, Arista Ratih, Elvi yenti, & Lazulva. (2024). Analisis Indeks Aiken Untuk Mengetahui Validitas Isi Instrumen Scinetific Habbits Of Mind. *Journal of Research and Education Chemistry*, 6(1), 59. [https://doi.org/10.25299/jrec.2024.vol6\(1\).17430](https://doi.org/10.25299/jrec.2024.vol6(1).17430)
- Viniasari, H., Susilowati, E., & Mulyani, D. B. (2022). *Implementasi Penilaian Higher Order Thinking Skills (HOTS) Dalam Pembelajaran Kimia Di Sma Negeri 1 Magelang*. <https://doi.org/10.20961/jpkim.v11i2.58710>

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>