



Pengembangan Instrumen Tes *Four-Tier Multiple Choice (4TMC)* Berbasis *Higher Order Thinking Skill (HOTS)* Materi Unsur, Senyawa dan Campuran



Berliana Wanodya Diva Saputra, Lulu Tunjung Biru, R. Ahmad Zaky El Islami

Program Studi Pendidikan IPA, Universitas Sultan Ageng Tirtayasa

*Email: berlianawanodyadivasaputra@gmail.com

DOI: <https://doi.org/10.33369/pendipa.9.2.442-449>

ABSTRACT

It is known that some schools in Serang City are still not optimal in empowering students' Higher Order Thinking Skill (HOTS) habits and tend to lead to the C1 - C3 cognitive domain. The purpose of this study was to describe the level of validity and analyze students' responses to the Higher Order Thinking Skill (HOTS)-based Four-Tier Multiple Choice (4TMC) on Elements, Compounds, and Mixtures material in class VIII junior high school students (Phase D). Research and development was conducted using the R n D (Research and Development) method. The research was conducted using the ADDIE model which refers to the Analysis, Design, Development, Implementation, and Evaluation stages. The instruments used are interview instruments, validation questionnaires and learner response questionnaires. The subjects in this study were 3 expert lecturers of Science Education and 60 junior high school students in grade VIII. The validity level of the HOTS-based test instrument reached a value of 93.20% with the category "Very Feasible" and obtained the results of students' responses with the category "Good" at a percentage of 79.98%. Therefore, the HOTS-based 4TMC test instrument is very feasible to be used by teachers as a reference for conducting tests on science materials, especially elements, compounds, and mixtures.

Keywords: Test Instrument; 4TMC; HOTS; Elements; Compounds.

ABSTRAK

Diketahui bahwa beberapa diantara sekolah di Kota Serang masih belum optimal dalam memberdayakan kebiasaan berpikir *Higher Order Thinking Skill (HOTS)* pada peserta didik dan cenderung mengarah kepada ranah kognitif C1 – C3. Tujuan dari penelitian ini adalah untuk menguraikan tingkat kevalidan serta menganalisis respon peserta didik terhadap Instrumen Tes *Four – Tier Multiple Choice (4TMC)* Berbasis *Higher Order Thinking Skill (HOTS)* Materi Unsur, Senyawa dan Campuran pada siswa kelas VIII jenjang SMP (Fase D). Penelitian dan pengembangan dilakukan dengan menggunakan metode R n D (*Research and Development*). Penelitian dilakukan dengan menggunakan model ADDIE mengacu pada tahap *Analyze, Design, Development, Implementation and Evaluation*. Instrumen yang digunakan berupa instrumen wawancara, angket validasi dan angket respon peserta didik. Subjek dalam penelitian ini adalah 3 dosen ahli Pendidikan IPA dan 60 peserta didik SMP kelas VIII. Tingkat kevalidan dari instrumen tes berbasis HOTS mencapai nilai 93,20% dengan kategori "Sangat Layak" dan mendapat hasil respon peserta didik dengan kategori "Baik" pada presentase sebesar 79,98%. Oleh karena itu, instrumen tes 4TMC berbasis HOTS sangat layak digunakan oleh guru sebagai referensi pelaksanaan tes pada materi IPA khususnya unsur, senyawa dan campuran.

Kata kunci: Instrumen tes; 4TMC; HOTS; Unsur; Senyawa.

PENDAHULUAN

Pendidikan abad ke – 21 juga dikenal dengan era *Society 5.0* yang dimana memuat tantangan esensial untuk menyamaratakan

kualitas pembelajaran secara menyeluruh diseluruh dunia, termasuk negara Indonesia. Kualitas pendidikan yang unggul dapat dimulai melalui perencanaan kurikulum yang matang dan

pelaksanaan optimal sesuai harapan. Sehingga, kurikulum merdeka dihadirkan sebagai respon agar kualitas pendidikan di Indonesia lebih berkembang. Kurikulum merdeka memberikan kebebasan kepada peserta didik untuk menaruh minat lebih terhadap sesuatu yang mereka gemari, termasuk gaya belajar (Ainia, 2020).

Sistem pembelajaran saat ini mengharuskan pendidik dan peserta didik memiliki kemampuan dan keterampilan 4C. Kemampuan 4C didalamnya memuat cakupan berupa *collaboration, communication, creativity and innovation* serta *critical thinking and problem solving*. Kemampuan 4C dapat dioptimalkan melalui pembelajaran berbasis HOTS atau *Higher Order Thinking Skill*. Pembelajaran HOTS memuat mekanisme berpikir peserta didik dengan level kognitif tinggi. Hal tersebut dibangun atas berbagai konsep, pengetahuan dan takson pembelajaran seperti taksonomi bloom.

Kemampuan berpikir tingkat tinggi didalamnya mencakup beberapa kemampuan spesifik lain, seperti: (1) kemampuan berpikir kritis; (2) kemampuan berpikir kreatif; (3) kemampuan pemecahan masalah (Dinni, 2018). Kecakapan tersebut kemudian diolah agar peserta didik mampu menemukan solusi atas situasi yang sedang berlangsung, menghasilkan ide orisinal dan inovatif, serta memiliki kecakapan dalam mengambil tindakan berdasarkan hasil analisa dan pertimbangan yang matang.

Umumnya level kognitif HOTS mengacu ranah Taksonomi Bloom revisi C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta). Namun, level kognitif yang diberikan kepada peserta didik masih didominasi oleh level C1 sampai C3. Sehingga, kualitas tes hasil belajar masih memerlukan perbaikan konten dan konstruk instrumen yang digunakan. Berdasar pada hasil studi internasional *Programme for International Student Assessment (PISA)* menyatakan bahwa capaian literasi membaca, matematika dan sains yang dicapai peserta didik Indonesia masih rendah (Zamroni, 2018).

Selaras dengan data *konkrit* hasil observasi kepada guru IPA di dua SMP Negeri Kota Serang memaparkan bahwa kemampuan berpikir tingkat tinggi peserta didik cenderung rendah. Peserta didik masih mengalami kesulitan dalam membedakan konsep nyata dengan abstrak. Peserta didik juga masih kesulitan dalam

mengembangkan ide gagasan kreatif yang dimiliki, serta merumuskan solusi atas masalah dalam pembelajaran IPA khususnya pada materi kimia seperti unsur, senyawa campuran. Materi dengan fokus kimia maupun fisika cenderung mengandung konsep simbolis, istilah khusus maupun rumus yang perlu dipelajari secara menyeluruh dan berulang, sehingga dapat mengasah kemampuan berpikir tingkat tinggi yang sebenarnya dimiliki peserta didik.

Pengembangan instrumen tes dengan butir HOTS sangat disarankan dalam proses pembelajaran. Karena, keterampilan berpikir dengan level kognitif tinggi tidak hanya memuat seberapa sulit pembelajaran atau soal yang dibuat, namun seberapa besar pengaruhnya dalam merangsang peserta didik untuk berpikir dengan kritis, kreatif, kolaboratif serta komunikatif. Soal HOTS acap kali dirancang lebih sistemik dan menawarkan beberapa solusi, sehingga menuntut peserta didik untuk menggunakan logisme berpikir mereka dalam menganalisis, mengevaluasi dan mencipta solusi (Faradisa *et al*, 2024).

Instrumen tes tipe *Four-Tier Multiple Choice (4TMC)* adalah salah satu instrumen yang selaras dengan kemampuan HOTS karena dapat digunakan untuk menaksir pendapat kritis peserta didik. Instrumen 4TMC dapat meminimalisir adanya miskonsepsi dalam pembelajaran, serta mampu meningkatkan kemampuan peserta didik pada level tinggi melalui pengolahan asumsi, gagasan serta menyelesaikan permasalahan dengan pendapat lugas terhadap hal kompleks (Kristanto & Setiawan, 2020).

Berdasarkan temuan observasi sebelumnya, peneliti memiliki gagasan yang diharapkan menjadi sebuah solusi melalui pengembangan instrumen tes *Four-Tier Multiple Choice (4TMC)* berbasis *Higher Order Thinking Skill (HOTS)* pada materi IPA fase D khususnya unsur, senyawa dan campuran. Tujuan penelitian ini adalah untuk menguraikan tingkat kevalidan serta menganalisis respon peserta didik terhadap instrumen tes *Four-Tier Multiple Choice (4TMC)* berbasis *Higher Order Thinking Skill (HOTS)* materi unsur, senyawa dan campuran. Pengembangan penelitian ini diharapkan dapat membantu pendidik dalam mendiagnosa dan mengasah kemampuan berpikir tingkat tinggi

pada peserta didik terhadap materi yang diajarkan.

METODE PENELITIAN

Penelitian pengembangan mengacu pada metode *Research and Developmet* (R&D). Metode ini bersifat deskriptif serta menggambarkan metode secara menyeluruh dalam mencapai produk, sehingga dikategorikan ke dalam metode prosedural. Riset pengembangan menerapkan model ADDIE melalui lima tahapan pengembangan berupa *analyze, design, development, implementation and evaluation* (Sugiyono, 2016).

Penelitian pengembangan ini terdiri atas beberapa subjek validitas dan uji coba terbatas. Subjek validitas mencakup 3 validator ahli Pendidikan IPA yang dipilih berdasarkan kriteria akademis dan berkompeten dalam bidangnya. Selain melibatkan ahli, subjek penelitian melibatkan peserta didik kelas VIII di dua SMP Negeri Kota Serang sebanyak 60 jiwa yang dipilih melalui teknik *random sampling*. Data yang diperoleh berupa data kualitatif dan kuantitatif menggunakan angket observasi pra penelitian, angket validasi ahli dan respon peserta didik. Perolehan data dalam penelitian mengacu pada analisis perhitungan data penelitian sebagai berikut.

Perolehan data angket validasi kemudian dihitung menggunakan rumus:

$$Me = \frac{\sum Xi}{n} \quad (1) \quad p = \frac{f}{n} \times 100\% \quad (2)$$

Keterangan:

- Me = Nilai rata – rata per indikator
- Σ = Kalkulasi nilai
- Xi = Nilai yang muncul
- n = Kalkulasi skor maksimum
- p = Besarnya presentase hasil
- f = Skor yang diperoleh
- % = Presentase tetap

Kemudian, dilakukan interpretasi data hasil validasi teoritis terhadap kriteria berikut:

Tabel 1. Kriteria Validitas Teoritis Instrumen

Presentase	Kriteria
0-20	Sangat lemah
21-40	Lemah
41-60	Cukup
61-80	Layak
81-100	Sangat Layak

(Riduwan, 2012)

Setelah uji validitas ahli secara teoritis, kemudian dilakukan uji validitas empiris, uji reliabilitas, tingkat kesukaran dan daya pembeda instrumen tes yang dihitung berdasarkan rumusan berikut:

a) Uji Validitas Empiris

Uji validitas empiris dihitung menggunakan *software* IBM SPSS versi 20, dengan nilai signifikan diantara $<0,05 = \text{valid}$, sedangkan $>0,05 = \text{tidak valid}$.

Tabel 2. Interval Validitas Instrumen Tes

Nilai	Kriteria	Interpretasi Kriteria
$0,80 > 1,00$	Sangat Tinggi	Valid
$0,60 > 0,79$	Tinggi	Valid
$0,40 > 0,59$	Cukup	Valid
$0,20 > 0,39$	Rendah	Tidak Valid
$>0,19$	Sangat Rendah	Tidak Valid

(Puspitasari & Febrinita, 2021)

b) Uji Reliabilitas

Uji reliabilitas dibuktikan menggunakan *software* IBM SPSS versi 20 dengan pengujian *Cronbach's Alpha*, jika nilai koefisien $>0,60$ maka tidak reliabel (Wola *et al*, 2016).

Tabel 3. Kriteria Koefisien Reliabilitas Tes

Koefisien R	Kriteria R
$0,80 \leq r \leq 1,00$	Sangat Tinggi
$0,60 \leq r \leq 0,79$	Tinggi
$0,40 \leq r \leq 0,59$	Cukup Tinggi
$0,20 \leq r \leq 0,39$	Rendah
$0,00 \leq r \leq 0,19$	Sangat Rendah

(Arikunto, 2012)

c) Tingkat Kesukaran

Analisis tingkat kesukaran instrumen tes dihitung menggunakan *software* Anates versi 4.0.9 dan diinterpretasikan pada tabel berikut:

Tabel 4. Indeks Tingkat Kesukaran Tes

Rentang TK	Kategori TK
0,00 – 0,29	Sukar
0,30 – 0,69	Sedang
0,70 – 1,00	Mudah

(Arikunto, 2012)

d) Daya Pembeda

Daya pembeda dihitung menggunakan *software* Anates versi 4.0.9 guna membedakan proporsi butir soal dalam suatu kelompok. Hal

tersebut berarti pembeda dalam soal ditujukan untuk membedakan peserta didik dalam penguasaan materi. Jika kedua kelompok memiliki proporsi sama besar, maka tes tidak memiliki DP. Berikut interpretasinya:

Tabel 5. Kriteria Daya Pembeda

Indeks Diskriminan	Kriteria DP
$0,71 \geq 1,00$	Sangat Baik
$0,41 \geq 0,70$	Baik
$0,21 \geq 0,40$	Cukup
$0,00 \geq 0,20$	Buruk
Bertanda negatif	Buruk

sekali/Diganti/Dihapus

(Agustiawan & Fuady, 2023)

Setelah pengujian data kuantitatif, dilakukan uji coba terbatas kepada peserta didik sehingga diperoleh hasil analisa kemampuan berpikir tingkat tinggi dan respon peserta didik terhadap instrumen. Analisa perolehan jawaban peserta didik menggunakan pedoman penskoran 0 untuk jawaban salah, dan 1 jika benar sesuai dengan empat tingkatannya. Kemudian, dihitung menggunakan rumus sebagai berikut:

$$p = \frac{f}{n} \times 100\%$$

Keterangan:

- p = Besarnya presentase hasil
- f = Skor yang diperoleh
- n = Kalkulasi skor maksimum
- % = Presentase tetap

Hasil analisa diinterpretasikan dalam sajian tabel berikut.

Tabel 6. Interpretasi Level HOTS

Kriteria HOTS	Kategori
$76 \leq X \leq 100$	Sangat Tinggi
$51 \leq X \leq 75$	Tinggi
$26 \leq X \leq 50$	Rendah
$1 \leq X \leq 25$	Sangat Rendah

(Azmi *et al*, 2021)

Kemudian, dilakukan pengujian data terhadap respon peserta didik menggunakan rumus:

$$P = \frac{TS_e}{TS_h} \times 100\% \quad (1) \quad \bar{x} = \frac{F}{N} \times 100\% \quad (2)$$

Keterangan:

- P = Presentase peserta didik
- TS_e = Total skor empirik
- TS_h = Total skor maksimum

- \bar{x} = Rerata keseluruhan respon
- F = Jumlah presentase skor
- N = Jumlah keseluruhan indikator

Setelah itu, diinterpretasikan dalam kategori respon sebagai berikut:

Tabel 7. Kriteria Presentase Tingkat Respon

Presentase	Kriteria
$81,25\% < x \leq 100\%$	Sangat Baik
$62,50\% < x \leq 81,25\%$	Baik
$43,75\% < x \leq 62,50\%$	Cukup Baik
$25,00\% < x \leq 43,75\%$	Kurang Baik

(Arikunto, 2012)

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menguraikan tingkat kevalidan dan menganalisis respon peserta didik terhadap produk instrumen yang dikembangkan. Setelah melalui beberapa tahapan penelitian, diperoleh hasil validasi teoritik oleh ahli Pendidikan IPA, uji validitas empiris, reliabilitas, tingkat kesukaran dan daya pembeda pada instrumen tes, serta hasil uji coba terbatas terhadap respon dan hasil analisa peserta didik pada level HOTS.

Hasil Penilaian Validator Ahli Pendidikan IPA

Berdasarkan uji validitas teoritik yang dilakukan oleh 3 ahli Pendidikan IPA diperoleh hasil yang dipaparkan dalam tabel berikut.

Tabel 8. Hasil Penilaian Valid Ahli

No	Indikator	Presentase	Kriteria
1	Kelayakan Isi/Materi	91,66%	Sangat Layak
2	Konstrutivitas Instrumen Tes	93,75%	Sangat Layak
3	Kebahasaan	93,20%	Sangat Layak
Rata – Rata		93,20%	Sangat Layak

Selaras dengan tabel tersebut, diketahui bahwa instrumen tes sangat layak digunakan dibuktikan dengan indeks presentase sebesar 93,20%. Presentase tersebut menjelaskan bahwa pada setiap indikator memiliki kriteria kelayakan tersendiri. Indikator kelayakan isi menunjukkan bahwa butir soal sesuai dengan capaian dan

tujuan pembelajaran. Ditinjau dari sisi konstruk, butir soal disusun dengan jelas, padat dan bebas dari ambiguitas dan didukung oleh aspek kebahasaan pada butir soal menggunakan kaidah EYD dan istilah ilmiah yang tepat. Indikator tersebut, secara menyeluruh menegaskan bahwa instrumen tes 4TMC memenuhi kriteria penilaian yang baik dan layak digunakan dalam pembelajaran.

Hasil Uji Validitas Empiris, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda

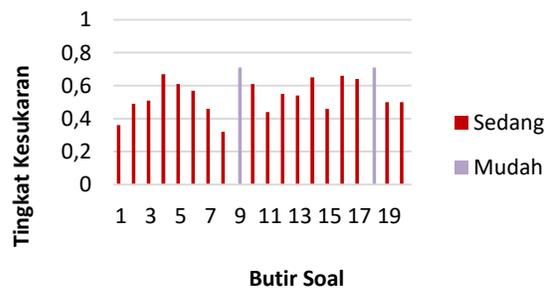
Uji validitas empiris dilakukan dengan memberikan instrumen tes kepada peserta didik sesuai dengan subjek yang telah ditentukan. Kemudian diperoleh hasil bahwa terdapat 3 butir soal dengan kategori tidak valid. Butir soal tidak valid terdapat pada nomor soal 7 dengan indeks signifikan sebesar 0,079, soal nomor 8 dengan indeks 0,206 dan soal nomor 19 dengan indeks 0,062. Nilai signifikansi pada ketiga butir soal tersebut >0,05 sehingga dinyatakan tidak valid empiris karena memiliki propabilitas kesalahan yang terlalu tinggi.

Terdapat beberapa asumsi yang menjawab pertanyaan mengapa item soal tidak valid, salah satunya adalah penggunaan kata – kata yang sukar dan dikemas dengan jelek, sehingga peserta didik kesulitan dalam mencerna bahasa yang digunakan. Hal ini dijadikan pertimbangan bahwa sangat penting dalam menentukan pemilihan diksi yang tepat, penyusunan kalimat yang efektif, serta penyesuaian bahasa soal dengan tingkat pemahaman siswa agar instrumen dapat digunakan seluruhnya untuk mengukur kompetensi yang diharapkan. Soal dengan indeks validitas yang >0,05 harus dipertimbangkan untuk diubah atau tidak digunakan kembali (Yusup, 2018).

Berdasarkan hasil uji validitas empiris, dihasilkan analisa reliabilitas instrumen tes menunjukkan angka 0,712 yang berarti indeks reliabilitas tinggi >0,60. Rentang tersebut dipercaya mampu mengasah kemampuan berpikir tingkat tinggi pada peserta didik. Reliabilitas menjadi sebuah ketetapan bahwa instrumen dapat digunakan secara berkala dengan hasil yang konsisten, sehingga hasil yang diperoleh dapat diandalkan. Artinya, apabila tes dilakukan secara berulang dalam kurun waktu

berbeda, maka hasilnya tidak jauh beda (Fatimah & Alfath, 2019).

Selaras dengan validitas dan reliabilitas, tingkat kesukaran suatu instrumen tes menjadi indikator esensial dalam menentukan kualitas sebuah butir soal. Tingkat kesukaran tidak ditinjau berdasarkan asumsi penyusun, melainkan berdasar kepada kesanggupan peserta didik dalam menjawab pertanyaan. Tingkat kesukaran digunakan sebagai bentuk penilaian interpretasi instrumen dalam bentuk sulit atau mudah. Tingkat kesukaran ditujukan supaya kualitas tes stabil, sehingga materi yang tersaji didalamnya tidak sukar atau terlalu mudah, sehingga dikatakan seimbang. Dalam penelitian ini, dihasilkan data tingkat kesukaran dalam sajian grafik sebagai berikut:



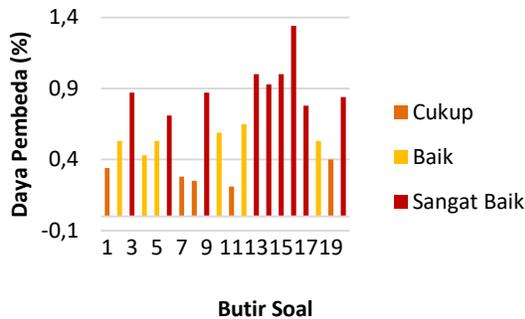
Grafik 1. Tingkat Kesukaran Instrumen

Meninjau data yang dihasilkan, terdapat hasil analisa tingkat kesukaran 20 butir soal menjelaskan bahwa sebesar 90% atau 18 butir soal berada dalam kategori sedang, sedangkan 10% lainnya terindikasi mudah. Artinya, butir soal yang disusun memiliki proporsi tingkat kesukaran yang baik dalam mengukur kemampuan peserta didik. Selaras dengan pernyataan Rajagukguk & Naibabo (2023) berpendapat, jika tingkat kesukaran ditujukan supaya memastikan suatu butir soal tidak terlalu mudah atau sulit, sehingga sesuai dengan tingkat pemahaman peserta didik.

Daya pembeda bertujuan untuk mengklasifikasikan peserta didik dengan kelompok tinggi dan peserta didik dengan kelompok rendah.

Perolehan hasil analisa terhadap daya pembeda, dinyatakan terdapat 9 butir soal yang memperoleh indeks diskriminan sangat baik ($DP \geq 71\%$). Kemudian, 6 butir lainnya dalam

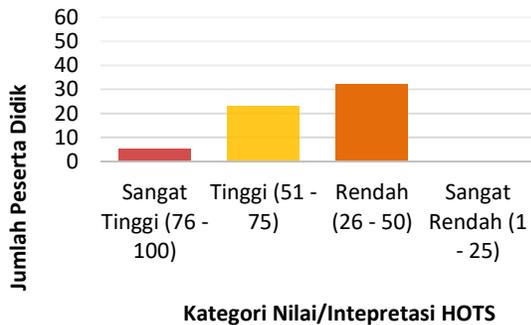
kategori baik dan 5 diantaranya berada pada kategori cukup. Data tersebut tidak menunjukkan adanya soal dengan nilai $DP \leq 20\%$ atau bernilai negatif. Dengan demikian, dapat dikatakan bahwa sebagian besar butir soal memiliki kualitas daya pembeda yang baik sehingga layak digunakan sebagai upaya pengasahan kemampuan berpikir tingkat tinggi peserta didik pada materi unsur, senyawa dan campuran.



Grafik 2. Daya Pembeda Instrumen

Hasil Analisis Respon Peserta Didik

Instrumen yang sudah valid kemudian diuji coba terbatas selama 80 menit kepada 60 peserta didik. Setelah dilakukan uji coba terbatas, dilakukan pemberian angket respon peserta didik terhadap butir soal pada instrumen tes 4TMC.



Grafik 3. Kemampuan HOTS Peserta Didik

Data nilai peserta didik tersebut dianalisa berdasarkan skor akhir. Berdasarkan kriteria yang ditetapkan, klasifikasi tingkat kemampuan HOTS peserta didik, terdapat 5 peserta didik yang mencapai kategori sangat tinggi. Peserta didik dengan kategori tersebut, sudah mampu menunjukkan penguasaan materi yang sangat baik, paham konsep serta berkembang dalam tingkat kognitif kritis, analitis dan evaluatif.

Sebanyak 23 peserta didik lainnya mencapai kategori tinggi, dengan capaian bahwa sebagian besar dari populasi paham konsep dan mampu menjawab soal dengan baik. Sedangkan, 32 peserta lainnya masih berada dalam kategori rendah. Peserta didik yang termasuk ke dalam kelompok bawah cenderung kesulitan menjawab soal yang menuntut kemampuan berpikir tingkat tinggi. Mereka membutuhkan pembelajaran yang terarah disesuaikan dengan kebutuhan dan gaya belajar masing – masing, selain itu diperlukan bimbingan khusus, diskusi serta latihan soal berulang sehingga dapat mengasah kemampuan berpikir tingkat tinggi yang mereka miliki.

Tabel 10. Rerata Skor Setiap Soal

Nomor Soal	Rata – Rata Skor	Nomor Soal	Rata – Rata Skor
Soal 1	1,18	Soal 11	1,60
Soal 2	2,22	Soal 12	2,50
Soal 3	2,13	Soal 13	2,53
Soal 4	2,37	Soal 14	2,57
Soal 5	2,87	Soal 15	1,80
Soal 6	2,70	Soal 16	2,70
Soal 7	2,00	Soal 17	2,68
Soal 8	1,43	Soal 18	2,38
Soal 9	3,02	Soal 19	2,43
Soal 10	2,13	Soal 20	2,13

Berdasarkan hasil rata – rata skor pada setiap butir soal, peserta didik dapat menjawab dengan baik pada butir soal nomor 9 dengan indeks rerata sebesar 3,02. Sebagian besar peserta didik sudah memahami konsep unsur dan informasi yang disajikan dalam soal bersifat faktual, eksplisit serta mudah dipahami. Soal tersebut membahas Unsur Nihonium (Nh) disertai dengan informasi faktual lainnya, sehingga memudahkan peserta didik dalam menganalisis jawaban serta alasan yang tersedia.

Butir soal nomor 5 menempati urutan ke dua nomor soal yang dijawab benar oleh peserta didik. Banyak peserta didik yang menjawab benar pada butir soal tersebut dikareakan konteks bahasan yang nyata dan sesuai dengan kehidupan sehari – hari. Soal nomor 5 mengangkat situasi mengenai contoh senyawa yang ada di sekitar kita, yaitu proses pembuatan manisan mangga dengan timun, serta bertanya tentang zat yang

paling efektif dalam mengawetkan sayur – sayuran.

Berbeda dengan butir soal nomor 1 yang memiliki indeks rata – rata paling rendah diantara soal lainnya yaitu 1,18 disebabkan karena peserta didik tidak percaya diri dalam membedakan unsur, senyawa dan campuran. Kemudian, butir soal nomor 8 juga memiliki indeks rata – rata terendah kedua dengan nilai 1,43 yang dimana soal tersebut berisikan ilustrasi cerita mengenai jenis campuran. Peserta didik cenderung kesulitan dalam membedakan jenis homogen dan heterogen, sehingga pemahaman yang didapatkan tertukar.

Terlepas dari rasa tidak percaya diri pada diri peserta didik, beberapa lainnya mampu mengungkapkan alasan pemilihan jawaban dengan tepat. Salah satunya adalah, terdapat peserta didik yang mengungkapkan bahwa *“Senyawa dapat terbentuk karena adanya 2 unsur atau lebih yang berikatan secara kimia”*. Kemudian, pada butir soal nomor 5 *“cuka mengandung senyawa kimia, salah satunya asam karboksilat”* yang dimana asam asetat merupakan kelompok dari asam karboksilat.

Dengan demikian, beberapa diantara peserta didik tersebut mampu mengaitkan konsep materi unsur, senyawa dan campuran dengan fenomena yang ada di alam sekitar, melalui cara berpikir menganalisis, kreatif serta menyelesaikan persoalan yang dihadapinya dengan menjelaskan jawaban secara konseptual dan ilmiah.

Tabel 10. Hasil Analisis Respon Peserta Didik

No	Aspek Pertanyaan	Presentase	Kategori
1	Instrumen Tes 4TMC	83,80%	Sangat Baik
2	Kemampuan Berpikir Kritis	77,07%	Baik
3	Kemampuan Berpikir Kreatif	75,41%	Baik
4	Kemampuan Pemecahan Masalah	79,98%	Baik
Rata – Rata Keseluruhan Aspek		79,98%	Baik

Setelah semua tahapan dilaksanakan, kemudian dilakukan uji respon peserta didik

terhadap instrumen tes 4TMC. Uji respon dilakukan melalui pemberian angket kepada peserta didik, sehingga memperoleh presentase sebesar 79,98%.

Berdasarkan hasil analisis rerata keseluruhan aspek pada angket respon peserta didik, dinyatakan bahwa respon dalam kategori “baik” dengan presentase sebesar 79,98%. Pada aspek Instrumen 4TMC didalamnya meliputi beberapa indikator terkait seperti struktur penyajian tes, kelayakan isi dan materi serta kebahasaan memperoleh presentase sebesar 83,80%. Artinya, aspek tersebut sangat baik dalam menjelaskan kelengkapan panduan penggunaan soal, kematerian serta visualisasi pendukung yang disajikan dalam tes sebagai stimulus, sehingga dapat terpahami dengan baik oleh peserta didik dan meminimalisir terjadinya kesalahan dalam menemukan jawaban.

Aspek berikutnya merujuk pada ranah kognitif HOTS diantaranya berpikir kritis, kreatif dan pemecahan masalah. Kemampuan berpikir kritis peserta didik memperoleh presentase sebesar 77,07% yang dimana dikategorikan dalam interpretasi “Baik”. Peserta didik sudah mampu menganalisis perbedaan jenis unsur, senyawa dan campuran. Peserta didik juga sudah mampu mencetuskan gagasan dengan baik, walaupun beberapa diantaranya masih enggan menyampaikan gagasan yang dimilikinya. Hal tersebut didasari atas perasaan tidak percaya diri dan rasa “takut salah”, sehingga gagasan yang dimilikinya masih tersimpan rapih dalam pikiran dan belum tertuangkan secara lugas.

Aspek HOTS berikutnya memuat ranah indikator kemampuan berpikir kreatif yang memperoleh presentase sebesar 75,41% dengan kategori “Baik”. Peserta didik sudah mampu melihat sebuah permasalahan dengan sudut pandang yang berbeda. Beberapa diantaranya sudah mampu mengkombinasi materi dengan gagasan yang mereka miliki dan berani mencetuskan gagasan baru untuk memperkuat jawaban yang mereka miliki, contohnya adalah ketika peserta didik mampu melihat bahwa asam asetat termasuk ke dalam kelompok asam karboksilat.

Indikator berikutnya berupa kemampuan pemecahan masalah yang memperoleh presentase hasil sebesar 79,16% dengan kategori “Baik”. Sebagian besar peserta didik sudah

mampu menelaah lebih lanjut permasalahan yang mereka hadapi dalam butir soal. Peserta didik sudah mampu memikirkan langkah maupun metode yang harus diselesaikan. Sebagai contoh, peserta didik mampu merencanakan metode pemisahan campuran yang tepat untuk memisahkan serbuk kayu, gula dan air dalam satu wadah. Adapun kritik dan saran yang tertera pada lembar angket maupun secara langsung, peserta didik memberikan komentar positif dan baik terhadap produk yang dikembangkan.

KESIMPULAN

Hasil validasi instrumen tes oleh ahli Pendidikan IPA memperoleh nilai sebesar 93,20% dengan kategori “Sangat Layak” ditinjau dari aspek kelayakan materi, konstruktivitas instrumen serta bahasa yang digunakan dalam butir soal. Kemudian, hasil respon peserta didik memperoleh indeks sebesar 79,98% dengan kategori “Baik”. Validitas empirik, reliabelitas tes, tingkat kesukaran dan daya pembeda memiliki rentang nilai selaras antara satu sama lain. Hal tersebut membuktikan bahwa instrumen tes *Four-Tier Multiple Choice* (4TMC) dinilai sangat layak dan baik digunakan oleh guru maupun pendidik sebagai upaya pengasahan kemampuan berpikir tingkat tinggi peserta didik, khususnya materi unsur, senyawa dan campuran.

DAFTAR PUSTAKA

- Ainia, D.K. (2020). Merdeka belajar dalam pandangan Ki Hadjar Dewantara dan relevansinya bagi pengembangan pendidikan karakter. *Jurnal Filsafat Indonesia*, 3(3), 95–101.
- Agustiawan, E., Fuady, A. (2023). Analisis Butir Soal Tes Uraian Ujian Tengah Semester Mata Kuliah Statistik. *Educare: Jurnal Ilmu Pendidikan*, 2(1), 16 – 24.
- Arikunto, S. (2012). *Prosedur Penelitian Pendekatan Suatu Praktek*. Jakarta: Rineka Cipta.
- 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), 45-52.
- Dinni, H.N. (2018). HOTS (High Order Thinking Skills) dan Kaitannya dengan Kemampuan Literasi Matematika. *PRISMA, Prosiding Seminar Nasional Matematika*, 1, 170-176.
- Fatimah, L.U., Alfath, K. (2019). Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor. *Jurnal Komunikasi dan Pendidikan Islam*, 8(2), 37-64.
- Faradisa, B.T.Z.V., Kurniasih,S., Berlian, L. (2024). Pengembangan Instrumen Tes 4TMC CBT Pada Materi Sistem Pernapasan untuk Mengukur Berpikir Kritis Siswa SMP Kelas VIII. *Jurnal Pendidikan MIPA*, 14(4), 909-918.
- Puspitasari, W.D., Febrinita, F. (2021). Pengujian Validasi Isi (*Content Validity*) Angket Persepsi Mahasiswa Terhadap Pembelajaran Daring Matakuliah Matematika Komputasi. *Journal Focus Action of Reseacrh Mathematic (Factor M)*, 4(1), 77-90.
- Rajagukguk, M.J.K., Naibabo, D. 2023. Mampu Memilih Soal Berdasarkan Tingkat Kesukaran. *Jurnal Pendidikan Sosial dan Humaniora*. 2(4).
- Sugiyono. 2016. *Metode Penelitian Kuantitatif Kualitatif Dan R&D*. Bandung: Alfabeta.
- Wola, B.R., Ibrahim, M., Purnomo, T. (2020). Development of a Four-Tier Multipl Choice Test on The Concept of Transport Across Membranes. *Sciene Education Journal*, 4(2), 77-97.