

Pengembangan Instrumen Tes 4TMC CBT Pada Materi Sistem Pernapasan untuk Mengukur Berpikir Kritis Siswa SMP Kelas VIII

Bahraeni Tajri Zahrah Via Faradisa^{1)*}, Septi Kurniasih¹⁾, Liska Berlian¹⁾

¹⁾Pendidikan IPA, FKIP, Universitas Sultan Ageng Tirtayasa

*Corresponding Author: bahraeniv@gmail.com

ABSTRAK

Permasalahan yang dihadapi dalam pembelajaran IPA salah satunya adalah kurangnya berpikir kritis siswa, sedangkan berpikir kritis merupakan keterampilan yang sangat penting dalam perkembangan keterampilan abad ke-21 (21st Century Skill). Berpikir kritis dapat diukur menggunakan tes diagnostik yang berbentuk pilihan ganda empat tingkat, yang dikenal sebagai Four Tier Multiple Choice (4TMC) berbasis CBT (Computer Based Test). Penelitian ini bertujuan untuk menghasilkan instrumen tes yang valid dan mengetahui efisiensi instrumen tes berbasis 4TMC CBT pada materi sistem pernapasan yang dirancang untuk mengukur berpikir kritis siswa. Penelitian ini menggunakan metode Research and Development (R&D), dengan tahapan Define, Design, Develop, and Disseminate. Sampel penelitian terdiri dari 45 siswa kelas VIII SMP. Instrumen yang digunakan meliputi angket validasi dan angket uji keterbacaan siswa. Hasil validasi instrumen tes oleh validator menunjukkan tingkat kevalidan, yaitu 91,60% kategori (sangat valid) oleh validator evaluasi, 93,63% kategori (sangat valid) oleh validator materi, dan 95,02% kategori (sangat valid) oleh guru IPA. Sementara itu, hasil angket uji keterbacaan mendapatkan nilai rata-rata 89,20% kategori (sangat baik). Instrumen ini juga menunjukkan reliabilitas tinggi sebesar 0,766 dan tingkat kesukaran soal 16 soal kategori (sedang) serta daya pembeda soal sebanyak 3 soal kategori (sangat baik) dan 19 soal kategori (baik). Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa instrumen tes berbasis 4TMC CBT pada materi sistem pernapasan dapat digunakan sebagai alat evaluasi untuk mengukur berpikir kritis siswa kelas VIII.

Kata Kunci: Instrumen Tes; Berpikir kritis; Four Tier Multiple Choice (4TMC); Sistem Pernapasan

Received: 7 Okt 2024; Revised: 12 Okt 2024; Accepted: 12 Okt 2024; Available Online: 14 Okt 2024

This is an open access article under the CC - BY license.



PENDAHULUAN

Abad 21 perkembangan era revolusi industri 4.0 mendorong peningkatan kualitas sumber daya manusia agar mampu bersaing dalam kemajuan ilmu pengetahuan dan teknologi (IPTEK), situasi ini menuntut hasirnya inovasi kreatif untuk menjadikan proses pembelajaran menjadi lebih efektif (Megahantara, 2017). Pembelajaran yang efektif mencakup dua aspek utama, yaitu metode belajar siswa dan strategi mengajar guru yang bertujuan memusatkan perhatian siswa agar pembelajaran menjadi lebih efisien (Fajeri et al., 2022). Dalam menghadapi tantangan abad 21, diperlukan keseimbangan antara pengetahuan dengan keterampilan, dengan penekanan pada penguasaan keterampilan berpikir kritis sebagai hal yang esensial (Nengsih et al., 2022). Keterampilan berpikir kritis membantu siswa memecahkan masalah secara logis, reflektif dan produktif dalam mengambil keputusan. Pengembangan keterampilan berpikir kritis terjadi melalui pembelajaran dengan permasalahan di kehidupan sehari-hari yang biasa ditemui dalam pembelajaran IPA (Utami et al., 2022).

Keterampilan berpikir kritis siswa dapat diukur melalui alat evaluasi berbasis *Computer Based Test* (CBT) yang berisi soal-soal HOTS (*Higher Order Thinking Skill*) yang dirancang untuk mengembangkan kemampuan berpikir tingkat tinggi pada siswa. Pada soal HOTS biasanya lebih kompleks, menawarkan beberapa solusi solusi (*Open Ended Approach*) dan menuntut pemikiran kritis dalam proses menyelesaikannya. Melalui bentuk soal ini mendorong siswa untuk menganalisis, mengevaluasi atau menciptakan solusi terhadap masalah yang dihadapi (Pursitasari et al., 2023). Selain itu, soal HOTS memungkinkan siswa untuk mengurai informasi, memberikan kritik serta menghasilkan solusi atas suatu masalah (Fathia et al., 2023). Penggunaan soal HOTS dalam evaluasi

berbasis *Computer Based Test* (CBT) menjawab kebutuhan abad 21, dengan menawarkan inovasi alat evaluasi yang efisien, mengurangi penggunaan kertas, meminimalkan kesalahan teknis dalam pengoreksian, serta meningkatkan akurasi skor hasil ujian (Pamungkas & Hakim, 2019).

Berdasarkan hasil wawancara guru IPA di 3 Sekolah Menengah Pertama (SMP) Kota Serang, diketahui bahwa banyak siswa dinilai memiliki keterampilan berpikir kritis rendah. Hal ini terlihat dari belum tercapainya Kriteria Ketuntasan Minimal (KKM) pada mata pelajaran IPA, khususnya materi sistem pernapasan. Selain itu, terdapat faktor lain yaitu kurangnya kemampuan guru dalam mengembangkan instrumen evaluasi yang tepat untuk mengukur keterampilan berpikir kritis siswa. Ujian yang diberikan masih berbasis *Paper Based Test* (PBT) dengan soal berbentuk PG (pilihan ganda) dan uraian. Namun, guru berpendapat bahwa soal PG (pilihan ganda) tidak mampu mengukur keterampilan berpikir kritis siswa dengan baik, karena mudah dijawab dengan menebak atau menduga-duga, sehingga sulit bagi guru untuk menilai seberapa dalam tingkat pemahaman siswa mengenai materi yang sudah dipelajari. Selain itu, soal PG (pilihan ganda) dianggap tidak efisien karena memerlukan banyak kertas dan biaya.

Berdasarkan permasalahan tersebut, penelitian ini menawarkan solusi melalui pengembangan instrumen tes 4TMC yang menggunakan indikator berpikir kritis yang berbasis *Computer Based Test* (CBT). Instrumen tes 4TMC berbasis *Computer Based Test* (CBT) ini dirancang dengan mengadopsi level kognitif analisis (C4), evaluasi (C5) dan mencipta (C6) (Anwar, 2022). Instrumen tes 4TMC (*Four-tier multiple choice*) merupakan bentuk soal pilihan ganda yang terdiri dari 4 tingkat. Tingkat ke-1 berisi soal dengan beberapa opsi pilihan jawaban, tingkat ke-2 mengukur keyakinan terhadap jawaban yang dipilih (yakin atau tidaknya), tingkat ke-3 meminta alasan atas jawaban yang telah dipilih, tingkat ke-4 mengukur keyakinan siswa dalam menjelaskan alasan. 4TMC (*Four Tier Multiple Choice*) ini efektif dalam mendeteksi adanya miskonsepsi atau tidak paham konsep pada siswa (Yasthopi & Ritongga, 2019). Instrumen tes 4TMC berbasis CBT membantu guru dalam proses evaluasi, menghemat waktu, dan menghasilkan penilaian yang lebih objektif (Rasyidin et al., 2023).

Penting bagi guru untuk mengukur keterampilan berpikir kritis siswa pada materi sistem pernapasan, karena materi ini relevan dengan kondisi yang sering dialami remaja, terutama ditingkat sekolah menengah pertama (SMP). Salah satu masalah serius yang sering ditemui adalah perilaku merokok, baik di dalam maupun di luar lingkungan sekolah. Berdasarkan data dari Kementerian Kesehatan yang disampaikan pada saat kegiatan *Indonesian Conference on Tobacco or Health*, sekitar 20% perokok remaja yang berusia 13-15 tahun, dan jumlah perokok remaja meningkat terus hingga mencapai 58,8% per tahun, yang dapat menimbulkan resiko kesehatan (Kurniati et al., 2019).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan instrumen tes yang valid serta mengetahui efisiensi instrumen tes berbasis (4TMC) *Computer Based Test* (CBT) pada materi sistem pernapasan yang dirancang untuk mengukur keterampilan berpikir kritis siswa. Instrumen tes ini akan diuji validitas, reliabilitas tingkat kesukaran dan daya beda, serta dilakukan uji keterbacaan oleh siswa. Hal ini bertujuan agar produk yang dihasilkan dapat berfungsi sebagai referensi bagi guru dan sebagai alat evaluasi efektif dalam mengukur keterampilan berpikir kritis.

METODE

Penelitian ini menerapkan metode *Research and Development* (R&D) dengan model yang dikemukakan oleh Thiagarajan et al., (1974), yaitu model 4-D yang terdiri dari tahapan: *define* (pendefinisian), *design* (perancangan), *develop* (pengembangan), dan *disseminate* (penyebaran). Namun penelitian ini dibatasi hanya sampai tahap *develop* (pengembangan) yang dilakukan melalui uji coba terbatas.

Tahap Define adalah fase untuk menentukan persyaratan yang diperlukan dalam mengembangkan produk yang sesuai dengan kebutuhan pembelajaran siswa. Tahap ini meliputi 5 langkah utama, yaitu: a) Analisis ujung depan (*Front-end Analysis*) yang bertujuan untuk mengidentifikasi masalah yang muncul dalam pembelajaran IPA, khususnya terkait alat evaluasi yang digunakan melalui observasi lapangan dan wawancara dengan guru IPA kelas VIII di 3 sekolah SMP Kota Serang. Ke-3 sekolah tersebut telah menerapkan kurikulum Merdeka, namun menghadapi masalah yang sama, yaitu instrumen tes yang belum mampu mengukur keterampilan berpikir kritis siswa, serta guru IPA SMP yang belum pernah mengembangkan instrumen tes 4TMC. b) Analisis siswa (*Learner Analysis*) yang merupakan langkah penting dalam awal perencanaan. Tujuannya

adalah untuk mengamati karakteristik, kemampuan dan pengalaman siswa. Berdasarkan wawancara dengan guru mengenai kemampuan siswa saat mengerjakan instrumen tes, diketahui bahwa keterampilan berpikir kritis siswa masih rendah, terlihat dari nilai yang berada dibawah KKM. Hal ini mengindikasikan bahwa banyak siswa yang belum mencapai tingkat keterampilan berpikir kritis yang diharapkan. c) Analisis tugas (*Taks Analysis*) menentukan capaian pembelajaran (CP) pada Fase D untuk dapat menetapkan indikator assessmen, berdasarkan hasil wawancara, diketahui bahwa materi sistem pernapasan menjadi permasalahan yang dihadapi siswa kelas VIII dari ketiga sekolah. Analisis tugas ini akan menentukan konsep materi yang akan dikembangkan untuk merumuskan solusi atas permasalahan tersebut. d) Analisis konsep (*Concept Analysis*) capaian pembelajaran (CP), tujuan pembelajaran (TP), dan konsep materi yang dijadikan dasar dalam menyusun kisi-kisi soal yang dirancang menggunakan level kognitif yang sesuai dengan tuntutan dari TP untuk mencapai tingkat keterampilan berpikir kritis siswa kelas VIII. e) Spesifikasi tujuan pembelajaran (*Specifying Instructional Objectives*) proses menentukan perubahan perilaku siswa yang diinginkan setelah menggunakan instrumen tes 4TMC CBT pada materi sistem pernapasan diharapkan dapat tercapainya keterampilan berpikir kritis siswa kelas VIII.

Tahap Design merupakan tahap perencanaan untuk merancang kebutuhan produk yang dikembangkan meliputi: a) melakukan analisis terhadap capaian pembelajaran (CP) yang akan digunakan. b) menentukan jumlah proporsi butir soal. d) menyusun rubrik penilaian. e) menyusun rancangan kisi-kisi soal yang mencakup hal-hal berikut: 1) Tujuan pembelajaran (TP) yang menurunkan indikator ketercapaian tujuan pembelajaran (IKTP); 2) Dalam menyusun kisi-kisi, ada 5 aspek yang harus diperhatikan: 1) Tujuan Pembelajaran (TP), 2) Indikator ketercapaian tujuan pembelajaran (IKTP), 3) Instrumen Assesmen, 4) Level kognitif yang relevan dengan HOTS, dan 5) Penomoran soal; 3) Bentuk, Jumlah dan Alokasi Waktu Pengerjaan Soal.

Tahap Develop merupakan pengembangan instrumen tes berbasis 4TMC CBT yang telah divalidasi dari validator evaluasi, materi dan guru IPA. Setelah menerima saran dan masukan dari validator maka peneliti akan melakukan revisi produk sesuai saran dan masukan agar instrumen valid dan siap untuk digunakan. Selanjutnya dilakukan uji coba terbatas dan uji keterbacaan oleh siswa untuk mengetahui efisiensi instrumen tes dalam mengukur ketetapan berpikir kritis. Subjek penelitian ini terdiri dari 45 siswa kelas VIII dari 3 sekolah SMPN Kota Serang, yang dipilih menggunakan teknik *Purposive Sampling* dengan mengambil sampel sebesar 5% dari populasi. Instrumen pengumpulan data penelitian ini menggunakan lembar validasi dan lembar uji keterbacaan dengan penilaian berdasarkan *Skala Likert* yang memiliki aturan pembobotan yang ditunjukkan pada Tabel 1.

Tabel 1. Skala Likert (Yusuf et al., 2021).

Kriteria	Skor
Sangat Baik (SB)	5
Baik (B)	4
Cukup Baik (CB)	3
Kurang Baik (KB)	2
Tidak Baik (TB)	1

Data dalam penelitian ini, selanjutnya dianalisis dan dihitung untuk mendapatkan persentase kevalidan soal dengan menggunakan rumus pada persamaan 1. (Handayani et al., 2018).

$$p = \frac{f}{n} \times 100 \quad (1)$$

Dimana p= Persentase, f= Skor yang dipeoleh, n= Skor maksimum yang dihitung. Hasil dari persentase yang diperoleh, kemudian diinterpretasikan ke dalam bentuk pernyataan dengan kriteria penilaian produk yang ditunjukkan pada Tabel 2.

Tabel 2. Interpretasi Kriteria Tingkat Validasi Instrumen Tes (Riduwan, 2015).

Persentase (%)	Katagori Validitas
0-20	Tidak Valid
21-40	Kurang Valid
41-60	Cukup Valid
61-80	Valid
81-100	Sangat Valid

Penilaian uji keterbacaan siswa digunakan sebagai data untuk mengukur efisiensi produk instrumen tes, menggunakan persamaan 2 (Sarip et al., 2022).

$$Pk = \frac{\text{Jumlah skor hasil pengumpulan data}}{\text{Skor kriteria}} \times 100 \% \quad (2)$$

Dimana PK= Persentase keterbacaan (%), skor kriteria= Total skor maksimal keterbacaan. Hasil persentase yang diperoleh selanjutnya akan diinterpretasikan ke dalam bentuk pernyataan dengan kriteria penilaian dapat dilihat pada Tabel 3.

Tabel 3. Interpretasi Kriteria Keterbacaan Instrumen Tes (Sarip et al., 2022).

Skor (%)	Katagori Keterbacaan
80,1% -100%	Sangat Baik
60,1% - 80%	Baik
40,1% - 60%	Cukup Baik
20,1% - 40%	Kurang Baik
0,0% - 20%	Tidak Baik

HASIL DAN PEMBAHASAN

Hasil Penilaian Validator Evaluasi, Materi, Guru IPA

Validasi ahli evaluasi yang mencakup 4 aspek yaitu: isi/materi, penyajian, kebahasaan dan kegrafikan memperoleh persentase sebesar 91,60% dengan kategori sangat valid yang ditunjukkan pada Tabel 4.

Tabel 4. Hasil Validasi Ahli Evaluasi

No	Aspek	Persentase	Kategori
1	Isi/Materi	88,87%	Sangat Valid
2	Penyajian	98,21%	Sangat Valid
3	Kebahasaan	89%	Sangat Valid
4	Kegrafikan	90,33%	Sangat Valid
Rata-rata persentase		91,60%	Sangat Valid

Hasil rata-rata dari validator ahli evaluasi pada aspek isi/materi mencapai 88,87% yang termasuk dalam kategori sangat valid menunjukkan bahwa Tujuan pembelajaran (TP) yang digunakan sudah sesuai dan merinci materi pembelajaran yang relevan dengan indikator ketercapaian tujuan pembelajaran (IKTP) yang diturunkan. Selaras dengan pernyataan (Yusuf, 2018) bahwa IKTP yang tepat sesuai dengan tujuan pembelajaran dapat mencerminkan pencapaian hasil belajar dalam proses pembelajaran. Masalah yang disajikan dalam soal sudah sesuai dengan level HOTS yang mengaitkan indikator berpikir kritis siswa SMP. Soal-soal HOTS berfokus pada pemahaman konsep, pemrosesan dan penerapan informasi berdasarkan permasalahan yang dihadapi, sehingga mendorong siswa untuk berpikir kritis melalui evaluasi atau menciptakan solusi untuk memecahkan masalah (Ahmad et al., 2020). Hasil persentase pada aspek isi/materi belum maksimal karena terdapat beberapa soal yang kurang memberikan penjelasan mendalam sesuai dengan materi permasalahan yang diberikan, sehingga diperlukan perbaikan.

Penilaian aspek penyajian sebesar 99% dengan kategori sangat valid, menunjukkan bahwa penyajian produk telah disusun secara sistematis meliputi cover, petunjuk pengerjaan soal, kartu soal 4TMC, kunci jawaban soal dan kunci jawaban alasan. Soal-soal yang disajikan juga jelas dan tidak menggunakan pernyataan "semua jawaban di atas salah/benar" sesuai dengan hasil penilaian dari 2 validator ahli evaluasi. Selaras dengan pernyataan Harsono et al., (2020) bahwa dalam pembuatan soal harus komprehensif dengan pilihan jawaban secara jelas, singkat, padat dan tegas, penyajian ilustrasi seperti gambar, tabel, dan grafik juga memberikan kejelasan informasi dan mempermudah dalam penyampaian pesan kepada pembaca (Khairunnisa et al., 2023).

Hasil penilaian aspek kebahasaan memperoleh persentase 89% kategori sangat valid, yang menunjukkan aspek bahasa belum mencapai nilai maksimal, karena pada soal nomor 40 terdapat penggunaan kalimat yang kurang komunikatif, serta terdapat penulisan *typing error* dan penggunaan tanda baca yang perlu diperbaiki oleh peneliti. Saran perbaikan dari validator ahli agar kaidah bahasa yang digunakan lebih komunikatif dan sesuai kaidah bahasa Indonesia serta ejaan PUEBI agar mudah dipahami. Hal ini selaras dengan pernyataan Winata,

(2020) yang menyatakan bahwa penggunaan bahasa yang sesuai dengan kaidah PUEBI akan mempermudah penyampaian pesan secara jelas dan mudah dipahami. Penelitian [Sagala & Naibaho, \(2023\)](#) juga menekankan pentingnya bahasa komunikatif dalam penulisan, yang mencakup penggunaan kata-kata dan kalimat yang jelas, ringkas dan mudah dipahami agar pesan dapat tersampaikan dengan efektif.

Aspek kegrafikan penulisan yang memperoleh persentase 90,33% dengan kategori sangat valid, berdasarkan hasil penilaian validator. Aspek kegrafikan memiliki 1 indikator yang mendeskripsikan hasil penilaian ahli evaluasi kesesuaian grafik penulisan baik ukuran, huruf yang sesuai, jarak spasi dan font tulisan. Selaras dengan pernyataan [Ningtyas & Rahmawati, \(2022\)](#) bahwa penggunaan grafik sangat penting untuk menciptakan desain menarik dengan tata letak dan penataan yang rapih serta harus mempertimbangkan ukuran, jenis huruf dan jarak penulisan agar memudahkan pembaca.

Validasi ahli materi yang terdiri dari 4 aspek yaitu: isi/materi, penyajian, kebahasaan dan kegrafikan memperoleh persentase sebesar 93,63% dengan kategori sangat valid yang ditunjukkan pada Tabel 5.

Tabel 5. Hasil Validasi Ahli Materi

No	Aspek	Persentase	Kategori
1	Isi/Materi	87,31%	Sangat Valid
2	Penyajian	96,80%	Sangat Valid
3	Kebahasaan	90,75%	Sangat Valid
4	Kegrafikan	99,66%	Sangat Valid
Rata-rata persentase		93,66%	Sangat Valid

Penilaian pada aspek Isi/Materi memperoleh persentase 87,31% dengan kategori sangat valid, berdasarkan penilaian dari 2 validator ahli materi. Hal ini membutuhkan kesesuaian Tujuan Pembelajaran (TP) dengan indikator ketercapaian tujuan pembelajaran (IKTP) yang diturunkan tepat dalam membentuk indikator asesmen pada butir soal. Soal-soal yang disajikan menggunakan level kognitif HOTS dengan kata kerja operasional (KKO) yang dapat mengukur berpikir kritis siswa. Selaras dengan pernyataan [Dewi et al., \(2020\)](#) bahwa pemilihan kata kerja operasional (KKO) yang tepat diperlukan untuk merinci indikator tujuan pembelajaran agar isi atau materi lebih terukur pada capaian pembelajaran. Persentase aspek isi/materi belum maksimal karena beberapa soal kurang tepat dalam penggunaan kata kerja operasional (KKO) yang sesuai dengan indikator asesmen pada butir soal, sehingga perlu perbaikan.

Aspek penyajian mendapatkan persentase penilaian dari validator sebesar 96,80% dengan kategori sangat valid, yang menunjukkan bahwa konsistensi penyajian telah disusun secara sistematis yang sesuai dengan soal berbasis 4TMC. Penyajian soal tidak memiliki pilihan jawaban yang bermakna sama yang membingungkan siswa dalam menjawab soal. Selaras dengan pernyataan [Kadir et al., \(2015\)](#) yang mengidentifikasi beberapa hal yang perlu diperhatikan dalam penyusunan soal, yaitu: 1) butir soal yang sesuai dengan indikator, 2) pokok soal pilihan ganda harus jelas, singkat, padat dan tegas, 3) soal tidak boleh memberikan petunjuk ke jawaban yang benar, 4) soal dan pilihan jawaban tidak ada pernyataan yang bersifat sama/ganda, 5) pilihan jawaban tidak menggunakan bunyi “semua pilihan jawaban diatas salah/benar”.

Penilaian aspek kebahasaan memperoleh persentase 90,75% dengan kategori sangat valid, yang menunjukkan ketepatan tata bahasa yang komunikatif kaidah bahasa Indonesia yang benar, meskipun persentase tinggi aspek bahasa belum mencapai nilai maksimal. Para ahli memberikan saran bahwa terdapat satu soal dengan redaksi yang kurang tepat dan ketepatan ejaan PUEBI yang masih belum dipahami oleh siswa yaitu, kesalahan dalam penggunaan PUEBI mencakup pemakaian huruf, penggunaan kata serta penggunaan tanda baca atau penulisan kata serapan. Berdasarkan hal itu, penting untuk menggunakan kalimat yang teratur, lengkap dan ejaan cermat, karena keteraturan dan kelengkapan kalimat dengan ejaan yang jelas akan memudahkan pembaca dalam memahami tulisan dan menyampaikan gagasan dengan lebih efektif ([Leksono, 2019](#)).

Aspek kegrafikan penulisan yang memperoleh hasil sebesar 99,66% dengan kategori sangat valid dari ahli materi mengenai kesesuaian grafik penulisan pada ukuran huruf, jarak spasi dan font tulisan yang sesuai sehingga mempermudah siswa dalam mengerjakannya soal. Selaras dengan pernyataan [Nisrokha, \(2019\)](#) kerapihan grafik termasuk ukuran dan spasi sangat penting untuk keterbacaan suatu tulisan dan memudahkan pembaca saat

membaca teks, pemilihan jenis, ukuran huruf dan spasi sebaiknya tidak terlalu rapat maupun terlalu renggang agar tidak menyulitkan pembaca.

Validasi guru IPA menilai mengenai aspek isi/materi, penyajian, kebahasaan dan kegrafikan yang memperoleh persentase sebesar 95,02% ditunjukkan pada Tabel 6.

Tabel 6. Hasil Validasi Guru IPA

No	Aspek	Persentase	Kategori
1	Isi/Materi	94,27%	Sangat Valid
2	Penyajian	93,99%	Sangat Valid
3	Kebahasaan	94,08%	Sangat Valid
4	Kegrafikan	97,77%	Sangat Valid
Rata-rata persentase		95,02%	Sangat Valid

Penilaian pada aspek Isi/Materi memperoleh persentase 94,27% dengan kategori sangat valid. Hal ini menunjukkan bahwa kesesuaian tujuan penelitian (TP) yang memerinci materi pembelajaran untuk membentuk indikator asesmen pada butir soal yang tepat. Kedalaman materi yang disajikan dalam berbentuk soal HOTS serta berkaitan dengan fakta-fakta dalam kehidupan sehari-hari. Selaras dengan pernyataan [Ananda, \(2023\)](#) yang menyatakan bahwa soal HOTS memerlukan pemikiran kritis dan kreatif untuk memecahkan masalah yang menghubungkan pengetahuan dengan konteks kehidupan sehari-hari, sehingga menghasilkan solusi atau ide baru. Dengan demikian, alat evaluasi berupa instrumen tes 4TMC terbukti efektif dalam mendiagnosa pemahaman siswa dan sejalan dengan model pembelajaran yang diterapkan guru. Pernyataan ini juga selaras dengan pendapat [Islami et al., \(2018\)](#) yang menyebutkan bahwa instrumen 4TMC adalah alat evaluasi yang sangat efektif untuk mengidentifikasi miskonsepsi, karena dapat menentukan sifat dan kekuatan miskonsepsi siswa pada konsep yang diamati serta mampu mengukur tingkat kepercayaan diri siswa dalam memberikan jawaban yang yakin serta tingkat keyakinan alasan yang tepat.

Penilaian aspek penyajian memperoleh persentase 93,99% dengan kategori sangat valid, dapat diartikan bahwa konsistensi penyajian yang tersusun secara sistematis yang mencakup petunjuk pengerjaan soal yang jelas dan mudah dipahami oleh siswa, pedoman penilaian instrumen tes 4TMC yang berbentuk penskoran terhadap jawaban mudah dimengerti oleh guru, kesesuaian penyajian jawaban yang disajikan tidak memiliki pilihan jawaban yang bermakna sama. Selaras dengan pernyataan [Gilbert & Sutandi, \(2021\)](#) bahwa dalam pemilihan kata dan kalimat pada pilihan jawaban harus tepat agar menciptakan suasana yang kondusif dan mencegah kesalahpahaman dalam berkomunikasi maupun makna yang sama. Kesesuaian dan ketepatan ilustrasi yang disajikan berupa gambar, tabel dan grafik pada soal sesuai dengan materi dan *website* yang memfasilitasi bentuk soal 4TMC serta alokasi waktu pengerjaan yang sudah sesuai dengan 20 butir soal tergolong HOTS pada tipe soal 4TMC. Selaras dengan pernyataan [Novialdi et al., \(2020\)](#) bahwa *website* yang menyertakan gambar, tabel atau grafik dapat memberikan pengalaman menarik dan berarti bagi siswa, sehingga informasi mudah dipahami, diingat dan memungkinkan siswa untuk mengungkapkan alasan balik jawaban dengan lebih mudah.

Aspek kebahasaan memperoleh penilaian dari 3 guru IPA sebesar 94,08% dalam kategori sangat valid, yang menunjukkan bahwa bahasa yang digunakan komunikatif dan mudah dipahami. Setiap soal memiliki ketepatan ejaan PUEBI (Pedoman Umum Ejaan Bahasa Indonesia) dan mudah dimengerti oleh siswa. Hasil persentase pada kebahasaan belum maksimal dikarenakan terdapat beberapa soal yang kurang tepat dalam penyajian soal, pada frasa terlalu panjang dan menggunakan terlalu banyak kata untuk menyampaikan pesan sehingga tidak komunikatif sehingga pesan yang disampaikan tidak mudah di cerna oleh siswa. Oleh karena itu peneliti memperbaiki saran guru IPA. Selaras dengan pernyataan [Khaerudin, \(2017\)](#) bahwa bahasa yang digunakan dalam butir soal harus memenuhi kalimat yang komunikatif dan tidak ambigu serta mengikuti aturan bahasa indonesia yang baik dan benar.

Hasil penilaian aspek kegrafikan sebesar 97,77% dengan kategori sangat valid. dibuktikan bahwa aspek kegrafikan berdasarkan nilai dari guru IPA mengungkapkan grafik penulisan baik ukuran, huruf yang sesuai, jarak spasi dan font tulisan telah sesuai sehingga memudahkan siswa dan guru. Selaras dengan pernyataan [Kartika, \(2015\)](#) bahwa penulisan suatu pesan harus memperhatikan aspek verbal seperti, pemilihan jenis huruf (*typeface*), spasi yang tidak terlalu rapat agar mempermudah pembaca selain itu, memilih ukuran huruf dan huruf (*font*) dengan menggunakan *bolds* dan *italic* agar juga penting untuk menegaskan atau membedakan kata.

Setelah melakukan perbaikan berdasarkan hasil validasi dari validator, langkah selanjutnya adalah melakukan uji coba terbatas pada 45 siswa kelas VIII di 3 SMPN Kota Serang yang menjadi subjek penelitian. Tujuan dari tahap ini adalah untuk menilai kualitas butir soal yang telah dikembangkan secara empiris. Sebaran jumlah siswa yang terlibat dalam uji coba produk ditampilkan pada Tabel 7.

Tabel 7. Jumlah Siswa Pada Uji Coba Produk

Subjek Penelitian	Jumlah Siswa Dalam Penelitian
SMPN A Kota Serang	15
SMPN B Kota Serang	15
SMPN C Kota Serang	15
Total	45 Siswa

Pada pengukuran kualitas soal dilakukan pada penelitian ini dengan menggunakan uji validitas, reliabilitas, tingkat kesukaran soal dan daya pembeda, untuk memudahkan pengujian instrumen ini peneliti menggunakan alat bantu berupa software Microsoft Excel.

Uji validitas

Uji validitas instrumen tes dilakukan menggunakan *software Microsoft Excel*. Berdasarkan hasil penghitungan indeks validitas pada 40 butir soal, soal nomor 1 termasuk soal yang valid dengan indeks validitas valid sebesar (0,677), sedangkan butir soal nomor 35 memiliki indeks validitas tidak valid sebesar (-0,426). Hal ini selaras dengan pernyataan [Fadli et al., \(2023\)](#) yang mengungkapkan bahwa instrumen yang dianggap valid dapat digunakan dalam penelitian mendatang untuk meningkatkan kualitas penelitian tersebut. Persentase hasil indeks validitas instrumen tes ditunjukkan pada Tabel 8.

Tabel 8. Persentase Hasil Indeks Validitas Instrumen Tes

Keterangan	Σ	Butir soal	(%)
Valid	20	1,5,8,9,11,12,17,18,19,21,22,23,25,26,27,32,33,34,37,38	50%
Tidak valid	20	2,3,4,6,7,10,13,14,15,16,20,24,28,29,30,31,35,36,39,40	50%

Reliabilitas

Uji Reliabilitas dilakukan menggunakan *software Microsoft Excel* dengan koefisien korelasi *Alpha-Cronbach* > 0,61 dianggap sebagai reliabel. Hasil uji reliabilitas tes di dapatkan nilai sebesar 0,766 menunjukkan “Reliabilitas Tinggi” pada Tabel 9.

Tabel 9. Reliabilitas Instrumen Tes

Nilai Acuan	Cronbach's Alpha	Kategori
0,61	0.766	Reliabilitas Tinggi

Tingkat kesukaran

Butir soal dikategorikan sukar jika indeksnya < 0,30, sedang jika berada pada rentang 0,31-0,70 dan kategori mudah indeks > 0,70. Uji tingkat kesukaran dilakukan menggunakan *software Microsoft Excel*. Hasil penghitungan indeks kesukaran pada 40 butir yang dilihat pada Tabel 10.

Tabel 10. Persentase Hasil Indeks Tingkat Kesukaran Instrumen Tes

Kategori	Σ	Nomor soal	(%)
Sedang	16	1,5,9,12,17,18,22,23,25,26,27,32,33,34,35,38	40%
Mudah	24	2,3,4,6,7,8,10,11,13,14,15,16,19,20,21,24,28,29,30,31,36,37,39,40	60%

Daya pembeda

Uji daya pembeda dilakukan menggunakan *software Microsoft Excel*. Dari 40 butir soal, diketahui bahwa soal nomor 23 termasuk dalam kategori “Sangat Baik” dengan indeks sebesar (0,75) sedangkan soal nomor 39 tergolong dalam kategori “Sangat Buruk” dengan indeks sebesar (-0,472), yang diartikan bahwa soal tersebut tidak diterima atau tidak dapat digunakan, dan tidak mampu membedakan antara siswa yang menguasai konsep dengan yang belum. Hasil penghitungan indeks daya pembeda dilihat pada Tabel 11.

Tabel 11. Persentase Hasil Indeks Daya Pembeda Instrumen Tes

Kategori	\sum	Butir soal	(%)
Sangat Baik	3	23,33,38	7,5%
Baik	19	1,5,8,11,12,16,17,18,19,20,21,22,27,28,29,30,34,37,40	47,5%
Cukup	4	6,25,26,32	10%
Buruk	6	7,9,13,14,15,36	15%
Sangat Buruk	8	2,3,4,10,24,31,35,39	20%

Berdasarkan perhitungan instrumen tes yang telah melalui seleksi dengan keseluruhan uji validitas, reliabilitas, tingkat kesukaran dan daya pembeda sehingga peneliti dapat mengidentifikasi kategori soal yang digunakan, yang akan dibuang atau direvisi. Dari rekapitulasi soal terdapat 20 soal yang berkualitas baik yang berorientasi pada HOTS (*High Order Thinking Skill*). Hal ini selaras dengan pernyataan Saputri et al., (2023) yang menyatakan bahwa kualitas instrumen tes yang baik melalui tahap validitas, reliabilitas, tingkat kesukaran dan daya pembeda, sehingga dapat disimpulkan bahwa kualitas instrumen sangatlah penting untuk memastikan keakuratan serta keandalan hasil pengukuran.

Hasil Uji Keterbacaan Siswa

Setelah dilakukan uji coba instrumen tes, selanjutnya siswa diberikan lembar angket di akhir pembelajaran untuk mengetahui efisiensi instrumen tes. Persentase hasil untuk setiap aspek ditunjukkan pada Tabel 12.

Tabel 12. Persentase Hasil Penilaian Keterbacaan Instrumen Tes

Aspek	Peresentase	Kategori
Kebahasaan	88,32%	Sangat Baik
Kegrafikan	86,66%	Sangat Baik
Penyajian	88,66%	Sangat Baik
Isi/materi	93,18%	Sangat Baik
Rata-rata	89,20%	Sangat Baik

Pada tabel 12 menunjukkan hasil uji keterbacaan siswa menggunakan angket keterbacaan untuk mengetahui keefisienan instrumen tes. Angket ini mencakup 4 aspek pernyataan yaitu: kebahasaan, kegrafikan, penyajian dan isi/materi. Hasil keterbacaan instrumen tes menunjukkan persentase sebesar 89,20% dengan kategori sangat baik diartikan bahwa instrumen tes dinyatakan efisien dalam mengukur keterampilan berpikir kritis siswa pada materi sistem pernapasan. Selain itu, pada uji keterbacaan produk yang dilakukan tidak ada masukan atau saran yang diberikan, hanya komentar positif tentang instrumen tes yang diuji.

SIMPULAN

Berdasarkan hasil pembahasan, dapat disimpulkan bahwa instrumen tes 4TMC CBT pada materi sistem pernapasan telah diuji kelayakan dan dinyatakan sangat valid oleh para ahli dengan presentase sebesar 93,41%. Rincian hasil validasi ahli menunjukkan bahwa ahli evaluasi memperoleh persentase sebesar 91,60% (sangat valid), validasi ahli materi memperoleh persentase sebesar 93,63% (sangat valid) dan Guru IPA memperoleh persentase sebesar 95,02% (sangat valid). instrumen tes ini memiliki reliabilitas tinggi sebesar 0,766 dan tingkat kesukaran soal dengan kategori (sedang) sebanyak 16 soal serta daya pembeda soal kategori (sangat baik) sebanyak 3 soal dan kategori (baik) 19 soal, dan memiliki keterbacaan produk dengan kategori (sangat baik) sebesar 89,20%, Hal ini membuktikan bahwa produk instrumen tes berbasis 4TMC CBT pada materi sistem pernapasan dapat menjadi alat evaluasi untuk mengukur berpikir kritis siswa kelas VIII.

Daftar Pustaka

- Ahmad, D. N., Setyowati, L., Pujaning, A., & Suhendri, H. (2020). Analisis Sistem Penilaian HOTS (*Higher Order Thinking Skill*) Dalam mengukur Kemampuan Berpikir Kritis dan Kreatif. *Jurnal Biotik*, 8(1), 11-23.
- Ananda, A., Sulasteri, S., Angriani, A. D., Kusumayanti, A., & Halimah, A. (2023). Pengembangan Instrumen Tes Tipe *Higher Order Thinking Skills* (HOTS) Pokok Bahasan Bangun Ruang Sisi Datar. *Al Asma: Journal of Islamic Education*, 5(2), 114-122.

- Anwar, A. (2022). Media Sosial Sebagai Inovasi Pada Model PJBL Dalam Implementasi Kurikulum Merdeka. *Jurnal UPI*, 19(2), 237-249.
- Dewi, N. K., Suparwa, I., & Putra, A. P. (2020). Penerapan Taksonomi Bloom Pada Pembelajaran Bahasa di Kelas 2 SD Bali Kiddy. *Jurnal Linguistika*, 27(2), 183-195.
- Fadli, R., Hidayati, S., Cholifah, M., Siroji, R. A., & Afgani, M. W. (2023). Validitas dan Reliabilitas Pada Penelitian Motivasi Belajar Pendidikan Agama Islam Menggunakan *Product Moment*. *JIIIP (Jurnal Ilmu Pendidikan)*, 6(3), 1734-1739.
- Fajeri, L., Berlian, L., & Biru, L.T (2023). Pengaruh Model Pembelajaran Tipe *Talking Stick* Berbantuan Media *Mystery Box* Terhadap Hasil Belajar Siswa. *Jurnal Pendidikan IPA (PENDIPA)*, 7(2), 150-157.
- Fathia, A., Berlian, L., & El Islami, R. A. Z (2023). Pengembangan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Tema Energi Ramah Lingkungan Pada Siswa Kelas IX. *Jurnal Pendidikan IPA (PENDIPA)*, 7(2), 232-240.
- Gilbert, J. M., & Sutandi, S. (2021). Perbandingan Kata Zongsuan, Suanshi, Zhongyu, Jieguo Serta Padanan Katanya Dalam Bahasa Indonesia. *Jurnal Bahasa dan Sastra*, 9(3), 268-279.
- Hardayani, N. D., Astutik, S., & Lesmono, A. D. (2018). Identifikasi Miskonsepsi Siswa Menggunakan *Four-Tier Diagnostic Test* Pada Materi Hukum Termodinamika Di SMA Bondowosa. *Jurnal Pembelajaran Fisika*, 7(2), 189-195.
- Harsono., Hafsi, A. R., & Budiman, H (2020). Pengembangan Profesionalisme Guru SDI Al Munawwarah Pemekasan Melalui Pelatihan Penyusunan Soal. *Jurnal Masyarakat Mandiri (JMM)*, 4(2), 281-290.
- Islami, D., Suryaningsih, S., & Bahriah, E. S. (2018). Identifikasi Miskonsepsi Siswa Pada Konsep Ikatan Kimia Menggunakan Tes *Four-Tier Multiple-Choice (4TMC)*. *Jurnal Riset Pendidikan Kimia*, 9(1), 21-29.
- Kadir, A. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Jurnal Al-Ta'dib*, 8(2), 70-81.
- Kartika, R. (2015). Memilih dan Memanfaatkan Tipografi. *Jurnal Humaniora*, 6(3), 312-318.
- Khaerudin. (2017). Administrasi, Analisis Butir, dan Kaidah Penulisan Tes. *Jurnal Madaniyah*, 1(12), 97-128.
- Khairunnisa, Fitri, A., & Husin, M. (2023). Penerapan Model Pembelajaran *Project Based Learning (PJBL)* Terhadap Keterampilan Menggambar Ilustrasi di Kelas V SD Negeri 36 Banda Aceh. *Elementary Education Research*, 8(4), 409-419.
- Kurniati, G., Widiatutik, O., & Suwarni, L. (2020). Efektivitas Media Video Terhadap Peningkatan Pengetahuan dan Sikap Tentang Bahaya Merokok Pada Anak Sekolah Menengah Pertama. *Jurnal Endurance*, 5(2), 251-258.
- Leksono, M. L. (2019). Analisis Kesalahan Penggunaan Pendoman Ejaan Bahasa Indonesia (PUEBI) Pada Tugas Makalah dan Laporan Mahasiswa IT Telkom Purwokerto. *Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 4(2), 116-120.
- Megahantara, G. S. (2017). Pengaruh Teknologi Terhadap Pendidikan Di Abad 21. Yogyakarta: Universitas Negeri Yogyakarta.
- Nengsih, C. O., Zulyusri., & Lufri. (2022). Studi Literatur: Penggunaan Alat Evaluasi Pembelajaran IPA Yang Mendukung Keterampilan Abad 21. *Jurnal ESABI (Edukasi dan Sains Biologi)*, 4(1), 10-20.
- Ningtyas, H. A., & Rahmawati, L. E. (2023). Kelayakan Isi, Penyajian, Kebahasaan, dan Kegrafikan Bahan Ajar Teks Deskripsi di SMP Kelas VII. *Imajeri: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 6(1), 52-71.
- Nisrokha. (2019). Desain Teknologi Cetak, *Jurnal Madaniyah*, 9(1), 79-99.
- Novialdi., Amir, Z., & Thahir, M. (2020). Pengembangan Media Pembelajaran Berbasis *Website* Untuk Memfasilitasi Pemahaman Konsep Siswa SMK Negeri 5 Pekanbaru. *Milenial: Journal For Teachers And Learning*, 1(1), 25-33.
- Pamungkas, N. T., & Hakim, L. (2019). Pengembangan Alat Evaluasi Berbasis *Computer Based Test (CBT)* Pada

- Materi Jurnal Penyesuaian Perusahaan Dagang di SMA Negeri 1 Puri Mojokerto. *Jurnal Pendidikan Akuntansi*, 7(1), 90-95.
- Pursitasari, I. D., Permatasari, A., Rubini, B., Didit., Heliawati., Nullhakim L., Kurniasih. S., Taufik, A. N. (2023). Pelatihan Penyusunan Desain Praktikum dan Penggunaan KIT Praktikum IPA Bagi Guru IPA SMP di Kabupaten Serang. *Jurnal ABDINUS: Jurnal Pengabdian Nusantara*, 7(2), 516-530.
- Rasyidin, E. Y., Sjaifuddin., & Berlian L. (2023). Validitas Asesmen Formatif Berbasis *Wizer Me* Pada Tema Revolusi Pangan Dalam Mengukur Kemampuan Literasi Sains Peserta Didik SMP. *Eduproxima: Jurnal Ilmiah Pendidikan IPA*, 5(2), 92-104.
- Riduwan. (2015). *Skala Pengukuran Variabel-Variabel Penelitian*, Bandung: Alfabeta.
- Sagala, D., & Naibaho, D. (2023). Mampu Menggunakan Bahasa Yang Komunikatif. *Pediaqu: Jurnal Pendidikan Sosial dan Humaniora*, 2(4), 12724-12735.
- Saputri, H.P., Larasati, N. J., & Shaleh. (2023). Analisis instrumen assesmen : validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Jurnal ilmiah PGSD FKIP Universitas Mandiri*, 9(5), 2986-2995.
- Sarip, M., Amintarti, S., & Utami, N. H. (2022). Validitas Dan Keterbacaan Media Media Ajar *E-Booklet* Untuk Siswa SMA/MA Materi Keanekaragaman Hayati. *JUPIES: Jurnal Pendidikan dan Ilmu Sosial*, 1(1), 43-59.
- Thiagarajan. S., Semmel. D., & Semmel. M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Washington: Leadership Training Institute, University of Minnesota.
- Utami, T. P., & Sjaifuddin, Berlian, L. (2022). Pengembangan Soal Uraian Berbasis Indikator Kemampuan Berpikir Tingkat Tinggi Pada Konsep Sistem Pencernaan Pada Manusia Untuk Siswa Kelas VIII SMP/MTs. *Jurnal Pendidikan IPA (PENDIPA)*, 6(1), 129-135.
- Winata, N. T. (2021). Pembinaan Bahasa Indonesia Yang Baik dan Benar Dikalangan Mahasiswa di Era Milenial Melalui Media Sosial. *Jurnal Penilitia Bahasan dan Sastra Indonesia*, 6(2), 267-275.
- Yasthophi, A., & Ritonga, P. S. (2019). Pengembangan Instrumen Tes Diagnostik *Multiple Choice Four Tier* Pada Materi Ikatan Kimia. *Jurnal Konfigurasi*, 3(1), 24-31.
- Yusuf, A., Suardana, I. N., & Selamat, K. (2021). Pengembangan Media Pembelajaran *Flashcard* IPA SMP Materi Tata Surya. *Jurnal Pendidikan Dan Pembelajaran Sains (JPPSI)*, 4(1), 69-80.
- Yusuf, B. B. (2018). Konsep dan Indikator Pembelajaran Efektif, *Jurnal Kajian Pembelajaran Dan Keilmuan*, 1(2), 13-20.