

## Desain Pengembangan Instrumen dan Penilaian Kemampuan Berpikir Kritis dalam Mengidentifikasi Zat Aditif

Ipat Pitriani Saripah<sup>a\*</sup>, Ade Yeti Nuryantini<sup>b</sup>, Ida Farida<sup>c</sup>

<sup>a,b,c</sup> Tadris IPA Pascasarjana, UIN Sunan Gunung Djati, Bandung, Indonesia.

\* Corresponding author: [ipatpitriani87@email.com](mailto:ipatpitriani87@email.com)

### Informasi Artikel

#### Histori Artikel

Submission: 6/9/2024

Accepted: 23/12/2024

Published: 30/12/2024

#### Kata Kunci

Desain Instrumen;

Berpikir Kritis;

Evaluasi;

Zat Aditif

### Abstrak

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi instrumen tes kemampuan berpikir kritis dalam mata pelajaran IPA pada materi Zat Aditif. Instrumen yang dikembangkan berupa soal pilihan ganda yang dirancang berdasarkan lima indikator berpikir kritis. Sebanyak 10 butir soal diuji coba pada siswa untuk mengukur validitas dan reliabilitasnya. Hasil analisis menunjukkan bahwa dua soal telah dinyatakan valid dan layak digunakan, tujuh soal memerlukan revisi untuk meningkatkan kualitas pengecoh serta daya pembeda, sedangkan satu soal harus diganti karena memiliki validitas yang rendah. Analisis validitas, reliabilitas, dan daya pembeda menggunakan perangkat lunak Anates mengindikasikan perlunya perbaikan pada beberapa butir soal guna meningkatkan efektivitas instrumen dalam mengukur kemampuan berpikir kritis siswa. Penelitian ini berkontribusi pada pengembangan instrumen tes berpikir kritis yang lebih valid dan reliabel untuk materi Zat Aditif dalam pembelajaran IPA.

### Abstract

This study aims to develop and evaluate a critical thinking ability test instrument for the subject of Science (IPA) on the topic of Food Additives. The developed instrument consists of multiple-choice questions with five critical thinking indicators. This research involves 10 questions that were tested on students. The analysis results show that 2 questions are valid and usable, 7 questions need revisions to improve the quality of distractors and discrimination power, and 1 question must be replaced due to low validity. The validity, reliability, and discrimination power analyses using Anates software indicate the need for improvements on several items to achieve a more effective instrument in measuring students' critical thinking abilities. This study contributes to the development of a more valid and reliable critical thinking test instrument for the topic of Food Additives.

### Keywords

Instrument Design;

Critical Thinking;

Evaluation;

Additives

©2024 The Author's

This is an open-access article under the [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) license.



<https://doi.org/10.37058/metaedukasi.v6i2.12629>

## Pendahuluan

Pendidikan pada abad ke-21 menuntut siswa untuk tidak hanya menguasai pengetahuan faktual, tetapi juga mengembangkan keterampilan berpikir kritis, kemampuan analitis, dan *problem solving*. Dalam konteks ini, kemampuan berpikir kritis menjadi salah satu keterampilan yang sangat dibutuhkan oleh siswa (Triyanto et al., 2022), khususnya dalam mata pelajaran Ilmu Pengetahuan Alam (IPA). Pembelajaran IPA tidak hanya mengajarkan konsep-konsep ilmiah, tetapi juga melibatkan proses berpikir yang mendalam, seperti analisis, sintesis, dan evaluasi, yang semuanya merupakan komponen penting dalam berpikir kritis (Septiani et al., 2021) mengukur kemampuan berpikir kritis siswa dalam IPA haruslah valid, reliabel, dan mampu membedakan siswa berdasarkan tingkat kemampuan berpikir kritis mereka.

Kemampuan berpikir kritis di dalam IPA mencakup sejumlah kemampuan kognitif, seperti kemampuan mengidentifikasi masalah, merumuskan pertanyaan yang relevan, mengumpulkan dan mengevaluasi informasi, serta menarik kesimpulan yang logis berdasarkan data yang ada. Dalam hal ini, pengukuran kemampuan berpikir kritis pada siswa dapat dilakukan melalui berbagai instrumen penilaian, salah satunya adalah tes pilihan ganda (Ernanda et al., 2022). Tes pilihan ganda sering digunakan karena kemampuannya untuk mencakup berbagai aspek pengetahuan dan keterampilan dalam waktu yang relatif singkat. Namun, untuk tes pilihan ganda menjadi instrumen yang efektif dalam mengukur kemampuan berpikir kritis, soal-soal yang digunakan harus dirancang dengan hati-hati, terutama dalam hal kualitas pengecoh dan daya pembeda.

Seiring dengan perkembangan kurikulum dan kebutuhan pembelajaran di berbagai negara, semakin banyak upaya yang dilakukan untuk mengembangkan instrumen tes yang dapat mengukur kemampuan berpikir kritis siswa secara tepat. Banyak penelitian yang mengkaji bagaimana desain soal pilihan ganda dapat mencerminkan kemampuan berpikir kritis siswa. Namun, meskipun ada berbagai pendekatan yang telah diterapkan, masih banyak instrumen tes yang belum dapat mengukur kemampuan berpikir kritis secara optimal, terutama dalam konteks materi-materi yang lebih spesifik seperti Zat Aditif dalam IPA. Zat Aditif adalah bahan yang ditambahkan dalam produk makanan atau minuman dengan tujuan untuk mempertahankan kualitas, memperpanjang umur simpan, atau meningkatkan rasa. Mengingat kompleksitas dan pentingnya pemahaman terhadap zat aditif, materi ini menjadi penting untuk diuji dengan instrumen yang dapat mengukur kemampuan berpikir kritis siswa secara menyeluruh.

Salah satu masalah utama yang dihadapi dalam pengembangan instrumen tes untuk mengukur berpikir kritis adalah bagaimana soal-soal tersebut dapat dibedakan dengan baik antara siswa yang memiliki kemampuan tinggi dan rendah. Soal yang kurang tepat dalam hal kualitas pengecoh (distraktor) atau daya pembeda dapat menyebabkan penilaian yang tidak akurat, sehingga instrumen tes menjadi kurang efektif dalam mengukur kemampuan berpikir kritis siswa (Hardianti et al., 2023). Pengecoh yang tidak efektif atau tidak sesuai dengan tingkat pemahaman siswa dapat mengurangi kualitas soal dan membuatnya tidak dapat membedakan antara siswa yang memahami materi dengan baik dan mereka yang kurang memahami.

Di sisi lain, pentingnya instrumen tes yang reliabel dalam pengukuran berpikir kritis juga menjadi sorotan utama dalam pengembangan tes. Reliabilitas mengacu pada sejauh mana instrumen tes dapat menghasilkan hasil yang konsisten jika digunakan berulang kali (Ono, 2020). Oleh karena itu, instrumen tes yang digunakan untuk mengukur kemampuan berpikir kritis harus memiliki tingkat reliabilitas yang tinggi agar dapat memberikan hasil yang akurat dan konsisten dalam berbagai kondisi. Beberapa penelitian menunjukkan bahwa instrumen dengan reliabilitas rendah cenderung menghasilkan ketidakpastian dalam penilaian, yang dapat memengaruhi pengambilan keputusan dalam pendidikan (Maulana, 2022).

Berdasarkan latar belakang ini, penelitian ini bertujuan untuk mengembangkan instrumen tes berbentuk pilihan ganda yang dapat mengukur kemampuan berpikir kritis siswa dalam mata pelajaran IPA, khususnya pada materi Zat Aditif. Instrumen yang dikembangkan tidak hanya diharapkan memiliki kualitas soal yang baik, tetapi juga harus memiliki daya pembeda yang mampu membedakan kemampuan siswa yang tinggi dan rendah, serta kualitas pengecoh yang efektif. Melalui penelitian ini, diharapkan dapat diperoleh instrumen tes yang valid dan reliabel dalam mengukur kemampuan berpikir kritis siswa, serta memberikan kontribusi untuk pengembangan pembelajaran IPA yang lebih baik.

Meskipun penelitian mengenai pengembangan instrumen tes berpikir kritis telah banyak dilakukan, sebagian besar berfokus pada pengembangan tes untuk mata pelajaran lain atau belum memperhatikan kualitas pengecoh dan daya pembeda soal secara mendalam (Maulana, 2022). Belum banyak instrumen tes yang dapat diandalkan untuk mengukur berpikir kritis siswa dalam materi Zat Aditif pada pelajaran IPA, padahal materi ini memiliki potensi untuk melibatkan analisis dan evaluasi yang mendalam dari siswa terkait dampak penggunaan bahan-bahan aditif dalam kehidupan sehari-hari. Gap ini yang menjadi latar belakang bagi peneliti untuk mengembangkan instrumen tes yang lebih spesifik, dengan harapan instrumen ini dapat lebih efektif mengukur kemampuan berpikir kritis pada siswa.

Penelitian ini bertujuan untuk mengembangkan instrumen tes pilihan ganda yang dapat mengukur kemampuan berpikir kritis siswa pada mata pelajaran IPA dengan materi Zat Aditif. Instrumen tes yang dikembangkan akan diuji untuk validitas, reliabilitas, daya pembeda, serta efektivitas pengecoh. Dengan demikian, penelitian ini tidak hanya bertujuan untuk menghasilkan instrumen tes yang valid dan reliabel, tetapi juga untuk memberikan wawasan baru dalam pengembangan instrumen yang dapat mengukur kemampuan berpikir kritis siswa secara lebih tepat.

## **Metode**

Penelitian ini dilaksanakan di Madrasah Tanawiyah, Kabupaten Sumedang. Jenis penelitian yang digunakan adalah Penelitian dan Pengembangan (R & D), yang mengacu pada model yang dikemukakan oleh Sugiyono (2021), dengan penyesuaian sesuai kebutuhan penelitian ini. Penelitian ini mengacu pada taksonomi Bloom revisi yang mencakup tiga tingkatan utama, yaitu tingkat menengah dan tingkat yang lebih tinggi dari Bloom's (menganalisis, mengevaluasi, dan mencipta), yang memerlukan kemampuan kognitif tingkat tinggi. Taksonomi ini awalnya dirancang sebagai hirarki kumulatif, di mana penguasaan pada tingkat yang lebih tinggi menjadi prasyarat bagi kemampuan di tingkat yang lebih kompleks.

Pada tahap pertama pengembangan instrumen, dilakukan validasi oleh ahli, yang mencakup validitas isi dan konstruk dari instrumen penilaian, baik yang berbentuk tes maupun non-tes. Validitas ini sangat penting, karena instrumen yang digunakan untuk mengukur kemampuan berpikir kritis harus memiliki validitas konstruk dan isi yang baik sebelum digunakan dalam penelitian. Hasil validasi menunjukkan bahwa instrumen tersebut dinyatakan valid setelah dilakukan revisi terkait penulisan dan keterbacaan, kesesuaian antara indikator kemampuan berpikir kritis dengan butir soal, kesesuaian taksonomi kognitif revisi Bloom yang diterapkan pada setiap soal, keakuratan konsep, keakuratan gambar yang digunakan, serta penyederhanaan istilah-istilah ilmiah.

Selanjutnya, dilakukan perancangan draft instrumen berdasarkan hasil studi pendahuluan. Draft instrumen tersebut terdiri dari soal pilihan ganda (PG) yang bertujuan untuk mengukur kemampuan berpikir kritis siswa. Setelah disusun, draft instrumen tersebut divalidasi oleh para ahli, termasuk dosen dan praktisi (guru), untuk memastikan validitas isinya dan menyempurnakan draft instrumen tersebut. Upaya untuk meningkatkan validitas instrumen dilakukan melalui dua langkah: (1) menyusun kisi-kisi tes sebelum menulis atau memilih butir-butir soal yang akan diuji, dan (2) menyusun butir soal yang berpedoman pada kurikulum yang berlaku (Farida, 2019).

Setelah validasi oleh ahli, instrumen diuji coba kepada siswa. Proses ini bertujuan untuk melakukan validasi kuantitatif yang mencakup analisis reliabilitas, validitas, tingkat kesukaran, daya pembeda, dan kualitas pengecoh. Pengujian ini menggunakan program Anates V4, dengan tafsiran dan kriteria yang dijelaskan pada Tabel 1.

**Tabel 1.** Tafsiran harga koefisien korelasi, tingkat kesukaran, dan daya pembeda

Indeks Korelasi		Tingkat Kesukaran		Daya Pembeda	
Skor	Kriteria	Skor	Kriteria	Skor	Kriteria
0,00 – 0,20	Sangat Rendah	0,00 – 0,15	Sangat Sukar	0,00 – 0,19	Buruk
0,21 – 0,40	Rendah	0,16 – 0,30	Sukar	0,20 – 0,29	Sedang
0,41 – 0,60	Cukup	0,31 – 0,70	Sedang	0,30 – 0,39	Cukup
0,61 – 0,80	Tinggi	0,71 – 0,85	Mudah	0,40 – 0,70	Baik
0,81 – 1,00	Sangat Tinggi	0,86 – 1,00	Sangat Mudah	0,71 – 1,00	Baik Sekali

Validitas instrumen mengukur sejauh mana instrumen tersebut dapat memberikan informasi yang sesuai dan dapat digunakan untuk mencapai tujuan tertentu. Sementara itu, reliabilitas mengukur konsistensi instrumen, yakni sejauh mana instrumen memberikan hasil yang konsisten bila digunakan berkali-kali. Analisis reliabilitas dilakukan untuk seluruh butir soal, bukan per butir soal (Farida, 2019). Untuk menilai tingkat kesukaran soal, digunakan indeks kesukaran dengan kategori tertentu, di mana soal yang tidak terlalu mudah atau sulit dinilai baik. Indeks kesukaran yang seimbang menandakan soal yang efektif. Daya pembeda soal (D) mengukur kemampuan soal dalam membedakan antara siswa yang berkemampuan tinggi dan rendah. Rentang nilai indeks diskriminasi (D) antara -1 hingga 1, dengan kategori sebagai berikut: jika D berada dalam rentang 0,00 – 0,19, soal tersebut perlu diperbaiki. Jika D bernilai negatif, soal tersebut tidak dapat digunakan karena menunjukkan kualitas siswa yang terbalik (Farida, 2019).

Analisis Distraktor (pengecoh) dilakukan untuk menilai fungsi pilihan jawaban dalam soal pilihan ganda. Distraktor yang tidak dipilih oleh siswa menunjukkan bahwa pengecoh tersebut buruk, sementara sebuah pengecoh dikatakan berfungsi baik jika dipilih oleh setidaknya 5% siswa yang mengikuti tes (Farida, 2019). Setelah diperoleh data dari Anates, dilakukan revisi atau perbaikan instrumen berdasarkan saran dari para ahli dan praktisi. Instrumen yang telah diperbaiki kemudian siap digunakan untuk mengukur kemampuan berpikir kritis siswa.

Proses uji coba instrumen ini dilakukan kepada 32 siswa kelas VIII di MTs Plus Darul Hufadz. Data hasil uji coba selanjutnya dianalisis menggunakan Confirmatory Factor Analysis (CFA) melalui aplikasi IBM SPSS, untuk menguji sejauh mana data yang diperoleh sesuai dengan model teoretis yang telah dirumuskan sebelumnya. Pada tahap evaluasi, skor rata-rata dari data kuantitatif dihitung dan diinterpretasikan secara kualitatif dengan menggunakan skala penilaian yang dimodifikasi dari aturan yang dikembangkan oleh Sudijono, seperti yang ditunjukkan pada Tabel 2. Hasil analisis kualitatif digunakan sebagai dasar untuk menentukan apakah instrumen yang dikembangkan sudah layak digunakan.

**Tabel 2.** Kriteria penilaian instrumen

Skor Rata-Rata	Kriteria	Kesimpulan
>4,2	Sangat Baik	Dapat menjadi acuan
>3,4 – 4,2	Baik	Dapat digunakan tanpa revisi
>2,6 – 3,4	Cukup Baik	Dapat digunakan dengan sedikit revisi
>1,8 – 2,6	Kurang Baik	Dapat digunakan dengan beberapa revisi
≤ 1,8	Tidak Baik	Tidak dapat digunakan

## Hasil dan Pembahasan

### Hasil

Pada bagian berikut ini, disajikan hasil penelitian yang telah diperoleh dalam bentuk tabel untuk memudahkan pembaca dalam memahami serta menganalisis data yang telah dikumpulkan. Penyajian dalam bentuk tabel bertujuan untuk memberikan gambaran yang lebih jelas, sistematis, dan terstruktur mengenai temuan penelitian, sehingga informasi yang disampaikan dapat diinterpretasikan dengan lebih mudah dan akurat.

**Tabel 3.** Hasil Penelitian

Aspek	Hasil
Jumlah Soal yang Dikembangkan	10 soal
Soal Layak Digunakan	2 soal (nomor 3 dan 6)
Soal Perlu Direvisi	7 soal (nomor 1, 4, 5, 7, 8, 9, dan 10)
Soal Harus Diganti	1 soal (nomor 2)
Indikator Berpikir Kritis	5 indikator yang diukur dalam instrumen
Rekomendasi	Melakukan revisi pada soal yang perlu diperbaiki dan mengembangkan instrumen lebih lanjut
Kualitas Soal	Beberapa soal memiliki daya pembeda yang baik, sementara yang lain tidak efektif

### Pembahasan

Produk yang dikembangkan dalam penelitian ini adalah asesmen berbasis *Critical Thinking Skills* (CTS) pada materi zat aditif. Instrumen penilaian ini berupa soal tes pilihan ganda yang terdiri dari 10 butir soal. Soal-soal ini dirancang untuk mengukur kemampuan berpikir kritis siswa serta untuk menilai indikator-indikator pembelajaran yang telah disusun sesuai dengan Kompetensi Inti-Kompetensi Dasar (KI-KD). Berpikir kritis adalah pemikiran yang masuk akal dan refleksi yang berfokus untuk memutuskan apa yang harus dipercaya atau dilakukan. Hal ini menunjukkan bahwa dalam berpikir, individu menggunakan kemampuan berpikirnya secara intelektual untuk menghadirkan solusi atas masalah yang dihadapi, yang pada akhirnya membantu dalam pengambilan keputusan sebuah aspek penting dari berpikir kritis.

Menurut John Dewey (Farida, 2019), berpikir kritis adalah "pertimbangan yang aktif, terus-menerus, dan teliti mengenai sebuah keyakinan atau pengetahuan yang diterima begitu saja, dengan disertai alasan-alasan yang mendukung dan kesimpulan yang rasional." Penjelasan ini menekankan pentingnya alasan yang mendasari setiap keputusan yang diambil dalam berpikir kritis.

Sebaran soal yang disusun dalam instrumen ini sesuai dengan indikator-indikator berpikir kritis yang diukur dapat dilihat pada Tabel 4 berikut.

**Tabel 4.** Sebaran soal sesuai indikator berpikir kritis

No.	Indikator Berpikir Kritis	Kesimpulan
1	Menganalisis suatu kenyataan	1,2
2	Mengobservasi dan memberikan argumen	3,4
3	Membuat induksi dan meninjau hasil induksi	5,6
4	Mengonstruksi argumen atau memberikan klarifikasi lanjutan	7,8
5	Merumuskan alternatif	9,10

Untuk menghasilkan perangkat penilaian yang berkualitas, ada tiga hal utama yang perlu diperhatikan, yaitu: 1) analisis validitas, 2) analisis reliabilitas, dan 3) analisis butir soal (daya pembeda, tingkat kesukaran, kualitas pengecoh) (Farida, 2019).

Proses validasi dilakukan terhadap tiga ahli, yaitu 1 orang guru dan 2 orang dosen. Hasil validasi ahli dapat dilihat pada Tabel 5.

**Tabel 5.** Hasil validasi ahli

No.	Indikator Berpikir Kritis	Rata-Rata	Kriteria
1	Kejelasan petunjuk instrumen	4,97	Sangat Layak
2	Kelengkapan indikator instrumen	4,98	Sangat Layak
3	Kesesuaian indikator dengan pernyataan kuesioner	4,97	Sangat Layak
4	Keefektifan bahasa	4,81	Sangat Layak

Dari hasil pengolahan data yang telah dijabarkan sebelumnya, diperoleh informasi mengenai daya pembeda, tingkat kesukaran, korelasi, dan reliabilitas soal.

### Tingkat Kesukaran Soal

Tingkat kesukaran soal diukur berdasarkan persentase jumlah siswa yang menjawab soal dengan benar atau salah. Tingkat kesukaran soal ini lebih dilihat dari sudut pandang kemampuan siswa dalam menjawab soal, bukan dari sudut pandang guru yang membuat soal (Yadnyawati, 2019). Hal ini sejalan dengan (Fatimah & Alfath, 2019), yang menyatakan bahwa tingkat kesukaran soal adalah peluang siswa untuk menjawab soal dengan benar berdasarkan tingkat kemampuannya, yang digunakan untuk menentukan apakah soal tersebut mudah atau sulit. Tingkat kesukaran soal biasanya dinyatakan dalam bentuk indeks yang berkisar antara 0,00 hingga 1,00. Kriteria tingkat kesukaran dapat dijelaskan dalam Tabel 6.

**Tabel 6.** Kriteria Tingkat Kesukaran

Interval	Kriteria
$0,00 < P \leq 0,30$	Butir soal susah
$0,30 < P \leq 0,70$	Butir soal sedang
$0,70 < P \leq 1,00$	Butir soal mudah

Berdasarkan hasil analisis, terdapat 1 soal yang termasuk dalam kategori mudah dan 9 soal lainnya dalam kategori sedang. Hal ini menunjukkan bahwa tingkat kesukaran soal sudah sesuai dan seimbang, karena soal-soal yang tidak terlalu mudah maupun terlalu sulit lebih efektif digunakan dalam pengukuran. Indeks kesukaran soal yang baik adalah yang memiliki variasi antara mudah dan sedang. Berdasarkan rentang indeks kesukaran (P): 0,00-0,29 (sukar), 0,30-0,69 (sedang), dan 0,70-1,00 (mudah) (Farida, 2019). Hasil penelitian menunjukkan bahwa soal yang dikembangkan sudah memenuhi kriteria tingkat kesukaran yang baik. Dari 10 soal yang diuji, soal nomor 8 adalah satu-satunya soal yang tergolong mudah, sementara soal lainnya berada dalam kategori sedang. Tidak ada soal yang tergolong sulit, karena materi zat aditif umumnya dianggap standar bagi siswa.

## Daya Pembeda

Dalam soal pilihan ganda, daya pembeda merupakan kemampuan suatu instrumen soal untuk mengklasifikasikan siswa berdasarkan tingkat kemampuannya (Mustaqim & Sulisti, 2024). Pendapat ini sejalan dengan (Nurhalimah et al., 2022) yang menyatakan bahwa daya pembeda soal adalah kemampuan soal untuk membedakan peserta tes antara kelompok siswa dengan kemampuan tinggi dan kelompok siswa dengan kemampuan rendah. Daya pembeda soal mengindikasikan sejauh mana soal dapat memisahkan siswa yang benar-benar menguasai materi dari yang tidak menguasai materi, seperti pada Tabel 7.

**Tabel 7.** Kriteria Daya Pembeda Soal

Interval Daya Pembeda (D)	Kriteria
0,00-0,19	Buruk
0,20-0,29	Sedang
0,30-0,39	Cukup
0,40-0,70	Baik
0,70-1,00	Baik Sekali

Berdasarkan hasil analisis daya pembeda menggunakan aplikasi Anates versi 4.0.9, diperoleh hasil bahwa terdapat 2 soal yang termasuk dalam kategori baik sekali, yaitu soal nomor 3 dan 6. Ini menunjukkan bahwa soal-soal tersebut dapat dengan efektif membedakan siswa dengan kemampuan tinggi dan rendah. Selain itu, terdapat 7 soal lainnya yang memiliki daya pembeda baik, yaitu soal nomor 1, 4, 5, 7, 8, 9, dan 10. Hanya ada 1 soal yang memiliki daya pembeda buruk, yaitu soal nomor 2. Soal tersebut perlu diperbaiki untuk meningkatkan kualitasnya.

## Efektivitas Pengecoh

Pengecoh atau distraktor dalam soal pilihan ganda adalah pilihan jawaban yang dirancang untuk menjebak siswa agar memilih jawaban yang salah, meskipun mirip dengan jawaban yang benar (Mustaqim & Sulisti, 2024). Efektivitas pengecoh dapat dilihat dari seberapa sering siswa memilih jawaban pengecoh tersebut. Sebuah pengecoh dapat dikatakan efektif jika dipilih oleh minimal 5% peserta tes.

Dalam penelitian ini, jumlah pengecoh yang dikembangkan untuk mengukur kemampuan berpikir kritis adalah sebanyak 50 opsi dari 10 soal. Hasil analisis kualitas pengecoh dapat dilihat pada Tabel 8 berikut:

**Tabel 8.** Hasil analisis kualitas pengecoh

Kriteria	Nomor Soal	Jumlah	Persentase (%)
Di terima	3 dan 6	2	20
Revisi	1, 4,5,7,8,9,10	7	70
Diganti	2	1	10
Jumlah		10	100

Berdasarkan analisis pada tabel, ditemukan bahwa 2 soal (nomor 3 dan 6) diterima atau digunakan tanpa perlu revisi. Sementara itu, 7 soal lainnya perlu direvisi, dan soal nomor 2 harus diganti. Pengecoh yang efektif dapat meningkatkan kualitas soal, karena dapat menantang siswa dalam

membuat keputusan yang lebih cermat. Namun, pengecoh yang kurang efektif atau tidak dipilih oleh sebagian besar siswa perlu diperbaiki atau diganti agar lebih efektif dalam mengukur kemampuan berpikir kritis.

Efektivitas pengecoh dapat dianalisis berdasarkan hasil penyebaran jawaban siswa pada setiap butir soal. Berikut adalah kriteria efektivitas pengecoh yang diadaptasi dari Skala Likert: Sangat Baik: Jika terdapat tiga pengecoh yang dipilih oleh peserta tes. Baik: Jika terdapat dua pengecoh yang dipilih. Cukup Baik: Jika hanya ada satu pengecoh yang dipilih. Tidak Baik: Jika tidak ada pengecoh yang dipilih oleh peserta tes.

Berdasarkan analisis penyebaran jawaban siswa pada setiap soal, dapat disimpulkan bahwa soal nomor 3 dan 6 memiliki pengecoh yang efektif, sedangkan soal nomor 2 memiliki kualitas pengecoh yang buruk dan harus diganti. Selain itu, soal-soal lainnya menunjukkan bahwa beberapa pengecoh perlu direvisi agar lebih efektif dalam mengukur kemampuan berpikir kritis siswa.

Berdasarkan hasil analisis daya pembeda, tingkat kesukaran, dan kualitas pengecoh, disarankan untuk melakukan revisi terhadap soal-soal yang tidak memenuhi kriteria kualitas. Secara khusus, soal nomor 2 harus diganti karena daya pembedanya yang negatif dan kualitas pengecohnya yang buruk. Soal nomor 1, 4, 5, 7, 8, 9, dan 10 memerlukan perbaikan pada pengecohnya untuk meningkatkan efektivitasnya dalam mengukur kemampuan berpikir kritis siswa. Soal nomor 3 dan 6 dapat digunakan tanpa revisi karena sudah memenuhi semua kriteria kualitas yang diinginkan.

## **Validitas**

Validitas adalah ukuran untuk menilai kecermatan suatu instrumen dalam mengukur apa yang dimaksud (Maulana, 2022). Menurut Amelia & Erita (2024), validitas bertujuan untuk memastikan bahwa instrumen yang digunakan dapat mengukur dengan tepat apa yang ingin diukur, sehingga hasil pengukuran tersebut bisa dianggap sah. Validitas butir soal dalam penelitian ini dianalisis menggunakan program komputer Anates versi 4.0.9 for Windows.

Validitas (validity) berasal dari kata "valid" yang berarti sah atau tepat. Validitas mengacu pada sejauh mana instrumen dapat mengukur apa yang seharusnya diukur. Oleh karena itu, instrumen yang valid berarti alat ukur tersebut mampu memberikan hasil yang tepat sesuai dengan tujuan pengukurannya (Mustaqim & Sulisti, 2024).

Tingkat validitas soal nomor 3 dan 6 sudah dianggap baik, karena keduanya memiliki nilai korelasi  $\geq 0,215$  berdasarkan analisis point biserial. Nilai korelasi yang lebih besar dari 0,215 menunjukkan bahwa soal-soal tersebut dapat membedakan dengan baik antara siswa yang berada dalam kategori tinggi (siswa yang menguasai materi) dan siswa yang berada dalam kategori rendah (siswa yang tidak menguasai materi). Suatu instrumen penilaian yang baik harus memiliki validitas yang tinggi untuk memastikan bahwa instrumen tersebut benar-benar mengukur apa yang dimaksud (Farida, 2019). Berdasarkan hasil uji validitas, instrumen yang digunakan dalam penelitian ini memiliki validitas yang tinggi dan dapat digunakan untuk mengukur kemampuan berpikir kritis siswa.

## **Reliabilitas**

Reliabilitas mengukur sejauh mana suatu instrumen konsisten dalam memberikan hasil yang stabil ketika digunakan dalam kondisi yang sama (Mustaqim & Sulisti, 2024). Dalam penelitian ini, reliabilitas soal diuji menggunakan Anates. Hasil uji reliabilitas disajikan pada Tabel 9.

Tabel 9. Uji Reliabilitas Soal

No	Jumlah item	Reliabilitas	Keterangan
1	10	0,35	Rendah

Berdasarkan hasil perhitungan reliabilitas yang diperoleh sebesar 0,35, menunjukkan bahwa tingkat reliabilitas soal pilihan ganda yang diuji coba rendah karena nilai  $r < 0,35$ . Berdasarkan kriteria reliabilitas, jika nilai  $\alpha > 0,90$ , maka reliabilitasnya dianggap sempurna. Jika antara 0,70 hingga 0,90, reliabilitasnya dianggap tinggi, dan jika antara 0,50 hingga 0,70, reliabilitasnya dianggap moderat. Jika nilai  $\alpha < 0,50$ , reliabilitasnya dianggap rendah, yang mengindikasikan bahwa beberapa item soal mungkin tidak reliabel (Sanaky, 2021).

Dalam penelitian ini, reliabilitas yang rendah dapat disebabkan oleh beberapa item soal yang tidak konsisten dalam mengukur kemampuan berpikir kritis siswa. Oleh karena itu, perbaikan pada instrumen soal perlu dilakukan untuk meningkatkan konsistensi pengukuran dan memastikan soal dapat digunakan untuk mendapatkan keputusan yang valid dalam pengujian kemampuan berpikir kritis siswa.

## Kesimpulan

Berdasarkan hasil analisis yang dilakukan, dapat disimpulkan bahwa instrumen tes kemampuan berpikir kritis pada mata pelajaran IPA untuk materi Zat Aditif yang dikembangkan belum sepenuhnya dapat digunakan secara efektif. Instrumen tes yang terdiri dari soal pilihan ganda dengan lima indikator kemampuan berpikir kritis dan total 10 soal menunjukkan hasil evaluasi sebagai berikut: dua soal dinyatakan sudah layak digunakan tanpa revisi, tujuh soal perlu direvisi untuk meningkatkan kualitasnya, terutama dalam hal efektivitas pengecoh dan daya pembeda, dan satu soal harus diganti karena memiliki validitas negatif dan tidak dapat membedakan dengan baik antara siswa dengan kemampuan tinggi dan rendah. Oleh karena itu, setelah dilakukan revisi, peneliti diharapkan dapat mengembangkan instrumen tes ini lebih lanjut, agar lebih efektif dalam mengukur kemampuan berpikir kritis siswa dan dapat meningkatkan kualitas pembelajaran IPA pada materi Zat Aditif.

## Ucapan Terima Kasih

Ucapan terima kasih yang setinggi-tingginya kami sampaikan kepada pihak MTs Plus Darul Hufadz Kecamatan Jatiningor Kabupaten Sumedang atas ijin dan pendampingan yang diberikan pada penelitian ini dalam proses memperoleh data.

## Referensi

- Amelia, N., & Erita, S. (2024). Eksplorasi Validitas dan Reliabilitas Soal Pemahaman Konsep dalam Asesmen Pembelajaran. *Jurnal BIMA: Pusat Publikasi Ilmu Pendidikan Bahasa Dan Sastra*, 2(1), 222–232. <https://doi.org/10.61132/bima.v2i1.559>
- Ernanda, M., Suharsono, S., & Triyanto, S. A. (2022). The Effect of Implementing Problem Based Learning in Lesson Study on Students Critical Thinking Skills. *Bioedukasi: Jurnal Pendidikan Biologi*, 15(2), 112–125. <https://doi.org/10.20961/bioedukasi-uns.v15i2.61383>
- Farida, I. (2019). *Evaluasi Pembelajaran*. PT Remaja Rosdakarya.
- Fatimah, L. U., & Alfath, K. (2019). Analisis kesukaran soal, daya pembeda dan fungsi distraktor. *AL-MANAR: Jurnal Komunikasi Dan Pendidikan Islam*, 8(2), 37–64. <https://journal.stainsyk.ac.id/index.php/almanar/article/download/115/104>

- Hardianti, H., Liliawati, W., & Tayubi, Y. R. (2023). Karakteristik tes kemampuan berpikir kritis siswa SMA pada materi momentum dan impuls: Perbandingan classical theory test (CTT) dan model Rasch. *WaPfi (Wahana Pendidikan Fisika)*, 8(1), 21–28. <https://doi.org/10.17509/wapfi.v8i1.30958>
- Maulana, A. (2022). Analisis Validitas, Reliabilitas, dan Kelayakan Instrumen Penilaian Rasa Percaya Diri Siswa. *Jurnal Kualita Pendidikan*, 3(3), 133–139. <https://doi.org/10.51651/jkp.v3i3.331>
- Mustaqim, M., & Sulisti, H. (2024). Analisis Butir Soal Pas Matematika Peminatan: Daya Pembeda, Tingkat Kesukaran, Dan Kualitas Pengecoh. *Al-'Adad: Jurnal Tadris Matematika*, 3(1), 44–56. <https://doi.org/10.24260/add.v3i1.3011>
- Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan Antara Validitas Item Dengan Daya Pembeda Dan Tingkat Kesukaran Soal Pilihan Ganda Pas. *Natural Science Education Research*, 4(3), 249–257. <https://doi.org/10.21107/nser.v4i3.8682>
- Ono, S. (2020). Uji Validitas dan Reliabilitas Alat Ukur SG Posture Evaluation. *Jurnal Keterampilan Fisik*, 5(1), 55–61. <https://doi.org/10.37341/jkf.v5i1.167>
- Sanaky, M. M. (2021). Analisis Faktor-Faktor Keterlambatan Pada Proyek Pembangunan Gedung Asrama Man 1 Tulehu Maluku Tengah. *Jurnal Simetrik*, 11(1), 432–439. <https://doi.org/10.31959/js.v11i1.615>
- Septiani, T., Wuryastuti, S., & Suratno, T. (2021). Analisis Kemampuan Berpikir Kritis dalam Pembelajaran IPA Materi dan Sifatnya Siswa Kelas V SDN Kebon Dalem. *Didaktika*, 1(4), 695–704. <https://doi.org/10.17509/didaktika.v1i4.38138>
- Triyanto, S. A., Wahidin, W., Hartania, N., Solihat, A., & Sutrisno, S. (2022). Blended-problem based learning with integrated social media-based learning media in improving students' critical thinking skills. *Biosfer*, 15(2), 242–254. <https://doi.org/10.21009/biosferjpb.25792>
- Yadnyawati, I. A. G. (2019). *Evaluasi Pembelajaran* (I Ketut Suda (ed.); Kesatu). Undiksha Press. <http://repo.unhi.ac.id/bitstream/123456789/183/1/EVALUASI%20PEMBELAJARAN.pdf>