

ANALISIS KUALITAS SOAL TES FORMATIF UNTUK MENGUKUR *HIGH ORDER THINKING SKILLS* PADA MATERI KESETIMBANGAN KIMIA

Nabila Nur Indah Dunggio^{1*}, Mangara Sihalo², Netty Ino Ischak³, Nurhayati Bialangi⁴, Lukman Abdul Rauf Laliyo⁵

Program Studi Pendidikan Kimia, Universitas Negeri Gorontalo, Indonesia

*Corresponding author email: nabilanurindahdunggio@gmail.com

Article History

Received: 28 May 2026

Revised: 6 August 2026

Published: 22 August 2026

ABSTRACT

This study aimed to develop a Higher Order Thinking Skills (HOTS)-based formative test instrument on chemical equilibrium material and to analyze the quality of the instrument based on validity, reliability, difficulty level, discrimination power, and distractor effectiveness using the Rasch Model. This study employed a Research and Development method involving 564 students as research subjects. The developed instrument consisted of HOTS multiple-choice questions at cognitive levels C4, C5, and C6. Instrument validity was analyzed through expert validation, while empirical analysis was conducted using Winsteps and Microsoft Excel. The results showed that the instrument obtained a validity percentage of 88.95%, categorized as very valid. Rasch Model analysis indicated that all items met the item fit criteria with a Cronbach Alpha reliability value of 0.78, categorized as good. The item difficulty levels consisted of easy, moderate, and difficult categories, with moderate items being dominant. All items demonstrated good discrimination power as indicated by positive Pt Measure Corr values. In addition, most distractors functioned effectively because they were selected by students. The distribution of HOTS items consisted of 44.44% C4 level, 33.33% C5 level, and 22.22% C6 level. Therefore, the HOTS-based formative test instrument on chemical equilibrium material was considered feasible for measuring students' higher-order thinking skills.

Keywords: HOTS, Formative Test, Chemical Equilibrium, Rasch Model, Instrument Quality.

Copyright © 2026, The Author(s).

How to cite: Dunggio, N.N.I., Sihalo, M., Ischak, N.I., Bialangi, N. & Laliyo, L.A.F. (2026). ANALISIS KUALITAS SOAL TES FORMATIF UNTUK MENGUKUR *HIGH ORDER THINKING SKILLS* PADA MATERI KESETIMBANGAN KIMIA. NUSRA: Jurnal Penelitian Dan Ilmu Pendidikan, 7 (3), 1950–1972. <https://doi.org/10.55681/nusra.v7i3.6525>



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

LATAR BELAKANG

Pendidikan abad ke-21 menuntut peserta didik tidak hanya menguasai pengetahuan konseptual, tetapi juga memiliki kemampuan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS), seperti kemampuan menganalisis, mengevaluasi, dan menciptakan solusi terhadap suatu permasalahan. Dalam pembelajaran kimia, kemampuan HOTS sangat diperlukan karena sebagian besar konsep kimia bersifat abstrak, kompleks, serta membutuhkan penalaran ilmiah yang mendalam. Namun, implementasi penilaian berbasis HOTS di sekolah masih menghadapi berbagai kendala, khususnya pada penyusunan instrumen evaluasi yang digunakan guru dalam proses pembelajaran.

Dalam berbagai penelitian menunjukkan bahwa instrumen penilaian yang digunakan di sekolah masih didominasi oleh soal berbasis *Lower Order Thinking Skills* (LOTS). Meskipun guru telah mengenal konsep HOTS, kenyataannya sebagian besar soal yang digunakan masih berada pada level mengingat, memahami, dan menerapkan. Yessi *et al.*, (2022) menyatakan bahwa hanya sekitar 1% soal yang benar-benar mengandung unsur HOTS, sedangkan 99% lainnya masih berada pada kategori LOTS. Kondisi tersebut menunjukkan adanya kesenjangan antara tuntutan kurikulum dengan praktik evaluasi pembelajaran yang diterapkan di sekolah.

Permasalahan ini menjadi semakin kompleks pada materi kimia yang bersifat abstrak, salah satunya materi kesetimbangan kimia. Materi kesetimbangan kimia menuntut peserta didik memahami konsep reaksi reversibel, kesetimbangan dinamis, tetapan kesetimbangan, serta prinsip Le Chatelier yang memerlukan kemampuan

analisis dan penalaran konseptual yang tinggi. Basyiroh *et al.*, (2022) menyatakan bahwa tingkat kesulitan belajar siswa pada materi kesetimbangan kimia berada pada kategori sedang hingga tinggi, terutama pada subkonsep kesetimbangan dinamis dan tetapan kesetimbangan. Rendahnya pemahaman peserta didik terhadap materi tersebut menunjukkan bahwa pembelajaran dan evaluasi yang dilakukan belum sepenuhnya mampu melatih kemampuan berpikir tingkat tinggi siswa.

Selain itu, hasil penelitian Husada *et al.*, (2021) menunjukkan bahwa sebagian besar paket soal kimia SMA kelas XI belum memenuhi karakteristik HOTS. Dari enam paket soal yang dianalisis, hanya satu paket soal yang memuat item berbasis HOTS dengan persentase sebesar 12%, sedangkan soal lainnya masih didominasi oleh LOTS. Hal ini menunjukkan bahwa kualitas instrumen evaluasi pembelajaran kimia masih perlu ditingkatkan agar mampu mengukur kemampuan berpikir kritis, analitis, dan kreatif peserta didik sesuai dengan tuntutan pembelajaran abad ke-21.

Secara teoritis, HOTS mengacu pada kemampuan berpikir tingkat tinggi yang meliputi keterampilan menganalisis (C4), mengevaluasi (C5), dan mencipta (C6) berdasarkan Taksonomi Bloom revisi Anderson dan Krathwohl. Kemampuan ini tidak hanya menuntut peserta didik menghafal konsep, tetapi juga menghubungkan informasi, menilai suatu fenomena, serta menghasilkan solusi baru terhadap permasalahan yang dihadapi (Listiani & Rachmawati, 2022). Oleh karena itu, instrumen penilaian yang digunakan dalam pembelajaran kimia harus mampu mengakomodasi indikator HOTS agar peserta didik terbiasa berpikir kritis dan

reflektif dalam menyelesaikan masalah kimia.

Salah satu bentuk evaluasi yang dapat digunakan untuk mendukung pengukuran HOTS adalah tes formatif. Tes formatif merupakan penilaian yang dilakukan selama proses pembelajaran untuk memperoleh informasi mengenai tingkat penguasaan konsep peserta didik dan kesulitan belajar yang mereka alami. Asriadi & Hadi (2021) menjelaskan bahwa tes formatif berfungsi sebagai alat diagnostik yang membantu guru memperbaiki proses pembelajaran melalui umpan balik yang diperoleh dari hasil tes siswa. Dengan adanya tes formatif yang berkualitas, guru dapat mengetahui sejauh mana kemampuan berpikir tingkat tinggi siswa berkembang selama proses pembelajaran berlangsung.

Instrumen tes formatif yang baik harus memenuhi beberapa aspek kualitas, di antaranya validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Instrumen yang valid dan reliabel akan menghasilkan data penilaian yang akurat dan konsisten dalam mengukur kemampuan peserta didik. Selain itu, analisis kualitas butir soal juga penting dilakukan untuk memastikan bahwa setiap item soal benar-benar mampu membedakan siswa yang memahami konsep dengan baik dan yang belum memahami konsep secara optimal.

Beberapa analisis yang dilakukan penelitian sebelumnya telah mengembangkan dan menganalisis instrumen HOTS pada pembelajaran kimia. Aulia & Rahmawati (2025) mengembangkan soal HOTS berbentuk pilihan ganda piktorial pada materi kesetimbangan kimia dan memperoleh hasil validitas serta reliabilitas yang tinggi. Shavira *et al.*, (2022) juga menemukan bahwa kualitas soal kesetimbangan kimia

dapat dianalisis melalui validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Namun demikian, penelitian yang secara khusus menganalisis kualitas soal tes formatif berbasis HOTS pada materi kesetimbangan kimia masih terbatas, terutama pada konteks pembelajaran kimia di SMA.

Berdasarkan uraian tersebut, diperlukan penelitian mengenai analisis kualitas soal tes formatif untuk mengukur HOTS pada materi kesetimbangan kimia. Penelitian ini penting dilakukan untuk mengetahui sejauh mana kualitas instrumen tes formatif yang digunakan mampu mengukur kemampuan berpikir tingkat tinggi peserta didik berdasarkan aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh, serta kesesuaian level kognitif HOTS. Dengan demikian, hasil penelitian diharapkan dapat menjadi acuan dalam pengembangan instrumen evaluasi pembelajaran kimia yang lebih berkualitas, relevan, dan sesuai dengan tuntutan pembelajaran abad ke-21.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan jenis penelitian deskriptif. Pendekatan kualitatif digunakan karena penelitian bertujuan untuk mendeskripsikan secara mendalam kualitas butir soal tes formatif dalam mengukur *Higher Order Thinking Skills* (HOTS) pada materi kesetimbangan kimia. Penelitian deskriptif berfokus pada penggambaran karakteristik suatu objek secara sistematis tanpa melakukan manipulasi variabel penelitian. Pendekatan ini dipilih karena sesuai untuk menganalisis kualitas instrumen tes berdasarkan data empiris hasil jawaban peserta didik serta hasil validasi ahli. Wulandari *et al.*, (2024) menyatakan

bahwa penelitian kualitatif bertujuan memahami fenomena secara alamiah melalui proses interpretasi yang mendalam, sedangkan penelitian deskriptif digunakan untuk menjelaskan kondisi variabel secara sistematis tanpa membandingkan ataupun menghubungkannya dengan variabel lain (Asri & Julisman, 2022).

Penelitian dilaksanakan pada empat sekolah menengah atas di Provinsi Gorontalo, yaitu SMAN 2 Limboto, SMAN 1 Kabila, SMAN 1 Gorontalo, dan SMAN 3 Gorontalo. Pemilihan lokasi penelitian dilakukan secara purposive dengan mempertimbangkan keterwakilan karakteristik pelaksanaan pembelajaran kimia dan evaluasi pembelajaran di tingkat SMA. Penelitian dilaksanakan pada bulan Desember 2025 sampai Januari 2026.

Subjek penelitian terdiri atas peserta didik kelas XI dan XII peminatan kimia yang telah mempelajari materi kesetimbangan kimia. Selain peserta didik, penelitian juga melibatkan validator ahli yang terdiri dari dosen pendidikan kimia dan guru kimia SMA untuk menilai kelayakan instrumen yang digunakan. Pemilihan validator dilakukan berdasarkan kompetensi dan pengalaman dalam bidang evaluasi pembelajaran kimia.

Instrumen penelitian berupa tes formatif berbentuk pilihan ganda sebanyak 18 butir soal yang disusun berdasarkan indikator HOTS pada level kognitif C4 (menganalisis), C5 (mengevaluasi), dan C6 (mencipta) sesuai Taksonomi Bloom revisi Anderson dan Krathwohl. Penyusunan instrumen dilakukan dengan mengacu pada kompetensi dasar dan indikator pencapaian kompetensi materi kesetimbangan kimia. Setiap soal dilengkapi dengan lima alternatif jawaban yang terdiri atas satu jawaban benar dan empat pengecoh (*distractor*). Selain instrumen tes, penelitian juga menggunakan

lembar validasi ahli untuk menilai aspek isi, konstruksi, dan bahasa soal.

Tahapan penelitian dimulai dari observasi awal di sekolah untuk memperoleh informasi mengenai pelaksanaan evaluasi pembelajaran kimia dan penggunaan soal berbasis HOTS oleh guru. Observasi dilakukan melalui wawancara informal dan studi dokumentasi terhadap instrumen evaluasi yang digunakan di sekolah. Hasil observasi dijadikan dasar dalam penyusunan instrumen tes formatif berbasis HOTS.

Tahap berikutnya adalah penyusunan instrumen tes formatif berbasis HOTS pada materi kesetimbangan kimia. Penyusunan soal memperhatikan karakteristik HOTS yang meliputi kemampuan menganalisis, mengevaluasi, dan menciptakan. Kisi-kisi instrumen dikembangkan berdasarkan indikator materi dan indikator soal sehingga setiap item memiliki keterkaitan yang jelas dengan kompetensi yang diukur. Setelah instrumen selesai disusun, dilakukan validasi oleh ahli untuk menilai kesesuaian konten, konstruksi, dan penggunaan bahasa pada setiap butir soal. Saran dan masukan dari validator digunakan sebagai dasar untuk merevisi instrumen sebelum dilakukan uji coba kepada peserta didik.

Pengumpulan data dilakukan melalui pemberian tes formatif kepada peserta didik secara tertulis. Jawaban peserta didik kemudian dianalisis untuk mengetahui kualitas butir soal berdasarkan beberapa aspek, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, efektivitas pengecoh, dan kesesuaian level kognitif HOTS. Validitas butir soal digunakan untuk mengetahui sejauh mana soal mampu mengukur kompetensi yang seharusnya diukur, sedangkan reliabilitas digunakan untuk mengetahui tingkat konsistensi instrumen. Tingkat kesukaran dianalisis untuk mengetahui kategori soal mudah,

sedang, atau sukar, sementara daya pembeda digunakan untuk mengetahui kemampuan soal dalam membedakan peserta didik berkemampuan tinggi dan rendah. Efektivitas pengecoh dianalisis untuk mengetahui fungsi alternatif jawaban pada soal pilihan ganda.

Analisis data dilakukan secara deskriptif kualitatif dengan mendeskripsikan hasil analisis kualitas butir soal berdasarkan data empiris yang diperoleh dari hasil tes peserta didik. Analisis juga didukung dengan pendekatan model Rasch untuk melihat kesesuaian item, reliabilitas, tingkat kesukaran, dan karakteristik butir soal secara lebih mendalam. Penggunaan model Rasch dalam penelitian evaluasi pendidikan dinilai mampu memberikan informasi yang lebih akurat mengenai kualitas instrumen dibandingkan analisis klasik karena dapat menunjukkan interaksi antara kemampuan peserta didik dan tingkat kesulitan soal secara objektif. Hasil analisis kemudian disajikan dalam bentuk naratif deskriptif agar dapat memberikan gambaran menyeluruh mengenai kualitas soal tes formatif dalam mengukur HOTS pada materi kesetimbangan kimia.

HASIL DAN PEMBAHASAN

Validitas Soal Tes Formatif untuk Mengukur HOTS pada Materi Kesetimbangan Kimia

Hasil Validasi Ahli

Tabel 1 Hasil Validasi Ahli

Nomor Soal	Rata-Rata	Persentase (%)	Kriteria
1	3,67	91,66	SV
2	3,58	89,58	SV
3	3,41	85,41	SV
4	3,50	87,50	SV

5	3,67	91,66	SV
6	3,58	89,58	SV
7	3,50	87,50	SV
8	3,67	91,66	SV
9	3,33	83,33	SV
10	3,58	89,58	SV
11	3,67	91,66	SV
12	3,58	89,58	SV
13	3,67	91,66	SV
14	3,58	89,58	SV
15	3,33	83,33	SV
16	3,58	89,58	SV
17	3,58	89,58	SV
18	3,58	89,58	SV
19	3,75	93,75	SV
20	3,41	85,41	SV
Total	3,56	88,85	SV

Sumber: Data diolah Winsteps, 2026

Berdasarkan Tabel 1, hasil validasi ahli menunjukkan bahwa seluruh 20 butir soal memperoleh kategori sangat valid dengan persentase penilaian berada pada rentang 81%–100%. Kategori tersebut mengacu pada kriteria interpretasi menurut Himmah & Sulaikho (2022), yang menyatakan bahwa persentase 81%–100% termasuk kategori sangat valid. Secara keseluruhan, persentase validasi instrumen mencapai 88,85%, sehingga instrumen dinyatakan layak secara teoritis dan dapat digunakan pada tahap uji coba.

Setelah dinyatakan valid secara teoritis, instrumen diujicobakan kepada peserta didik untuk mengevaluasi kualitas setiap butir soal sebelum digunakan pada penelitian utama. Hasil analisis menggunakan Model Rasch menunjukkan bahwa dari 20 butir soal yang diujikan, sebanyak 18 soal memenuhi kriteria item fit berdasarkan nilai Outfit MNSQ, Outfit ZSTD, dan *Pt-Measure Corr*, sedangkan 2

soal lainnya tidak memenuhi kriteria sehingga tidak digunakan dalam penelitian utama.

Kesesuaian Item Butir Soal

Tabel 2 Hasil Analisis Item Fit

Item	MNSQ	ZSTD	Pt-Measure Corr	Keterangan
S1	1,21	6,0	0,42	Fit
S2	1,21	1,1	0,48	Fit
S3	1,03	0,2	0,47	Fit
S4	0,97	-0,4	0,48	Fit
S5	0,98	-0,2	0,47	Fit
S6	0,83	-1,7	0,55	Fit
S7	1,07	0,9	0,42	Fit
S8	0,81	-2,0	0,52	Fit
S9	0,83	-1,4	0,52	Fit
S10	0,71	-2,4	0,56	Fit
S11	1,32	4,3	0,40	Fit
S12	0,76	-2,4	0,57	Fit
S13	1,06	4,1	0,45	Fit
S14	0,76	-2,7	0,59	Fit
S15	0,98	-0,2	0,49	Fit
S16	0,85	-1,3	0,55	Fit
S17	1,03	0,3	0,45	Fit
S18	0,90	-0,9	0,53	Fit

Sumber: Data diolah Winsteps, 2026

Berdasarkan Tabel 2, seluruh butir soal yang dianalisis dinyatakan fit. Nilai MNSQ berada pada rentang 0,71 sampai 1,32 dan nilai *Pt-Measure Corr* berada pada rentang 0,40 sampai 0,59. Hal ini menunjukkan bahwa butir soal telah sesuai dengan pola respons model Rasch dan dapat digunakan untuk mengukur konstruk yang dituju.

Beberapa butir memiliki nilai ZSTD di luar rentang ideal. Namun demikian, nilai

ZSTD sangat sensitif terhadap ukuran sampel yang besar. Karena jumlah peserta didik dalam penelitian ini mencapai 564 orang, maka penafsiran item fit lebih ditekankan pada konsistensi nilai MNSQ dan *Pt-Measure Corr*. Dengan demikian, hasil item fit tetap menunjukkan bahwa 18 butir soal yang dianalisis layak digunakan pada tahap pengukuran empiris.

Reliabilitas Soal Tes Formatif untuk Mengukur HOTS pada Materi Keseimbangan Kimia

Tabel 3 Hasil Uji Reliabilitas

Indikator	Nilai	Kategori	Interpretasi
Cronbach's Alpha (KR-20)	0,78	Bagus	Instrumen memiliki konsistensi internal yang memadai dalam mengukur HOTS.
Item reliability	0,99	Istimewa	Urutan tingkat kesukaran butir soal sangat stabil dan konsisten.
Separation	9,76	Sangat tinggi	Instrumen mampu membedakan butir soal ke dalam beberapa strata tingkat kesukaran.

Sumber: Data diolah Winsteps, 2026

Berdasarkan Tabel 3, nilai *Cronbach's Alpha* sebesar 0,78 berada pada kategori bagus sehingga menunjukkan bahwa instrumen memiliki konsistensi internal yang baik dalam mengukur kemampuan HOTS. Nilai item *reliability* sebesar 0,99 menunjukkan bahwa urutan tingkat kesulitan item sangat stabil. Sementara itu, nilai separation sebesar 9,76 menunjukkan

bahwa instrumen memiliki kemampuan yang sangat baik dalam membedakan tingkat kesulitan butir soal.

Tingkat Kesukaran Soal Tes Formatif untuk Mengukur HOTS pada Materi Keseimbangan Kimia

Tabel 4 Tingkat Kesukaran Butir Soal

Nilai Measur (logit)	Interpretasi	Jumlah Butir Soal	Nomor Soal	Persentase (%)	Kategori
Measur logit < -1,12	Item mudah	2 soal	2, 3	11,11	4 soal
-1,12 ≤ Measur logit < +1,12	Item sedang	13 soal	4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18	72,22	10 soal
Measur logit > +1,12	Item sukar	3 soal	1, 11, 13	16,67	4 soal

Sumber: Data diolah Winsteps, 2026

Hasil analisis tingkat kesukaran menggunakan Rasch Model pada Tabel 4 menunjukkan nilai rata-rata *item measure* sebesar 0,00 dengan standar deviasi 1,12. Berdasarkan nilai tersebut, kategori tingkat kesukaran ditentukan menjadi tiga, yaitu kategori mudah untuk nilai logit < -1,12, kategori sedang untuk nilai logit antara -1,12 sampai +1,12, dan kategori sulit untuk nilai logit > +1,12.

Dari 18 butir soal yang dianalisis, diperoleh 2 butir soal (11,11%) berkategori mudah, 13 butir soal (72,22%) berkategori sedang, dan 3 butir soal (16,67%) berkategori sulit. Hasil ini menunjukkan bahwa sebagian besar soal berada pada kategori sedang. Jika dibandingkan dengan proporsi ideal tingkat kesukaran menurut

Bano *et al.*, (2022), yaitu 25% soal mudah, 50% soal sedang, dan 25% soal sulit, distribusi soal pada instrumen ini belum sepenuhnya proporsional karena kategori sedang lebih dominan. Meskipun demikian, dominasi soal kategori sedang menunjukkan bahwa instrumen memiliki tingkat kesukaran yang cukup baik dan layak digunakan dalam penelitian.

Untuk mendukung hasil analisis tersebut, dilakukan pula perhitungan persentase jawaban benar dan salah pada setiap butir soal berdasarkan respons 564 peserta didik. Penyajian data ini bertujuan untuk memperjelas interpretasi tingkat kesukaran soal melalui proporsi peserta didik yang mampu menjawab benar pada setiap butir soal.

Table 5 Persentase Jawaban Benar dan Salah Setiap Butir Soal

Nomor Soal	Benar (f)	Benar (%)	Salah (f)	Salah (%)	Kategori Kesukaran
1	217	38,48	347	61,52	Sukar
2	494	87,59	70	12,41	Mudah
3	479	84,93	85	15,07	Mudah
4	327	57,98	237	42,02	Sedang
5	378	67,02	186	32,98	Sedang
6	412	73,05	152	26,95	Sedang
7	367	65,07	197	34,93	Sedang
8	415	73,58	149	26,42	Sedang
9	439	77,84	125	22,16	Sedang
10	451	79,96	113	20,04	Sedang
11	174	30,85	390	69,15	Sukar
12	423	75,00	141	25,00	Sedang
13	88	15,60	476	84,40	Sukar
14	403	71,45	161	28,55	Sedang
15	353	62,59	211	37,41	Sedang
16	432	76,60	132	23,40	Sedang
17	407	72,16	157	27,84	Sedang
18	421	74,65	143	25,35	Sedang

Sumber: Data diolah Winsteps, 2026

Tabel 5 menunjukkan bahwa butir nomor 13 memiliki persentase jawaban benar paling rendah, yaitu 15,60%, sehingga dapat dipahami sebagai butir yang paling menantang bagi peserta didik. Sebaliknya, butir nomor 2 memiliki persentase jawaban benar paling tinggi, yaitu 87,59%, sehingga tergolong paling mudah dikerjakan. Kesesuaian antara kategori Rasch dan persentase jawaban memperkuat deskripsi bahwa tingkat kesukaran soal dipengaruhi oleh kompleksitas konsep, banyaknya variabel yang harus dipertimbangkan, dan kedekatan konteks soal dengan pengalaman belajar peserta didik.

Daya Pembeda Soal Tes Formatif untuk Mengukur HOTS pada Materi Keseimbangan Kimia

Tabel 6 Kualitas Daya Pembeda Berdasarkan Tingkat Kesukaran

Tingkat Kesukaran	Jumlah Butir Soal	Persentase	Rentang Pt Mean Corr
Mudah	2 soal	11,11%	0,47 – 0,48
Sedang	13 soal	72,22%	0,42 – 0,59
Sukar	3 soal	16,67%	0,40 – 0,45

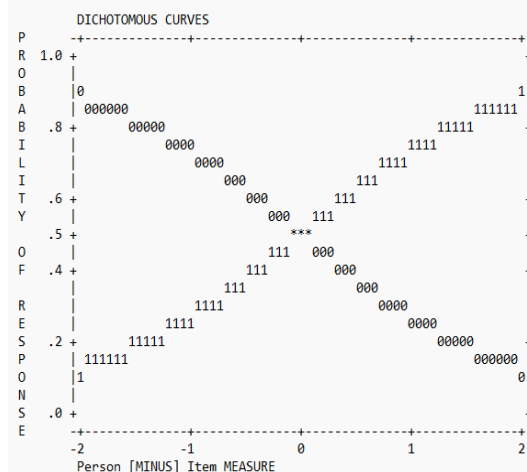
Sumber: Data diolah Winsteps, 2026

Berdasarkan hasil analisis Rasch, daya pembeda butir soal dianalisis melalui nilai *Pt Mean Corr* yang berdasarkan pada tingkat kesukaran soal. Hasil analisis menunjukkan bahwa soal kategori mudah memiliki rentang nilai *Pt Mean Corr* sebesar 0,47-0,48, soal kategori sedang rentang nilainya sebesar 0,42-0,59, sedangkan soal kategori sukar berada di rentang 0,40-0,45. Seluruh butir soal memiliki nilai *Pt Mean Corr* positif yang berarti berada pada kategori

baik, sehingga seluruh soal mampu membedakan peserta didik yang memiliki kemampuan tinggi dan rendah.

Efektivitas Pengecoh (Distraktor) Soal Tes Formatif untuk Mengukur HOTS pada Materi Keseimbangan Kimia

TABLE 21.1 DATA HASIL PENELITIAN ZOU278WS.TXT Mar 2 19:46 2026
INPUT: 564 Person 18 Item REPORTED: 564 Person 18 Item 2 CATS WINSTEPS 3.73



Gambar 1 Grafik Option Probability Curve

Karena data yang dianalisis bersifat dikotomis (benar-salah), grafik *Option Probability Curve* hanya menampilkan dua kategori jawaban, yaitu kategori 0 untuk jawaban salah dan kategori 1 untuk jawaban benar. Pada grafik tersebut, sumbu X (*horizontal*) menunjukkan kemampuan peserta didik dalam satuan logit dari rendah hingga tinggi, sedangkan sumbu Y (*vertical*) menunjukkan probabilitas peserta didik dalam memilih kategori jawaban tertentu. Simbol 0 menunjukkan probabilitas peserta didik menjawab salah, sedangkan simbol 1 menunjukkan probabilitas peserta didik menjawab benar.

Berdasarkan grafik, pada kemampuan rendah (sekitar logit -2 sampai -1), kategori 0 lebih dominan, yang berarti peserta didik dengan kemampuan rendah cenderung menjawab salah. Sebaliknya, pada kemampuan tinggi (sekitar logit 1 sampai 2),

kategori 1 lebih dominan, sehingga peserta didik dengan kemampuan tinggi lebih berpeluang menjawab benar. Pada titik sekitar logit 0, kedua kategori memiliki probabilitas yang hampir sama, ditandai dengan simbol (**), yang menunjukkan bahwa peluang menjawab benar dan salah berada pada kondisi seimbang. Pola tersebut menunjukkan hubungan yang logis dan sesuai dengan karakteristik Model Rasch.

Untuk memperoleh hasil analisis yang lebih rinci, dilakukan analisis lanjutan menggunakan rekapitulasi jawaban peserta didik pada Microsoft Excel. Pada analisis ini, pilihan jawaban yang salah dihitung sebagai pengecoh (*distractor*), sedangkan kunci jawaban tidak termasuk dalam perhitungan efektivitas pengecoh.

Tabel 7 Hasil Analisis Efektivitas Pengecoh Berdasarkan Persentase

Nom or Soal	Kunci Jawa ban	Pilihan Jawaban (%)				
		A	B	C	D	E
1	D	8,16 ^(B))	38,30^(C))	6,56 ^(B))	38,48^(C))	8,51 ^(C))
2	C	4,43^(B))	3,72 ^(B))	87,59^(C))	2,30^(R))	1,95^(C))
3	C	4,26 ^(B))	6,21^(B))	84,93^(C))	3,19 ^(B))	1,42^(C))
4	D	4,96 ^(B))	10,64 ^(C))	9,57 ^(B))	57,98^(C))	16,84^(B))
5	A	67,02^(C))	17,02^(C))	6,91 ^(B))	5,67 ^(B))	3,37 ^(C))
6	B	8,16 ^(B))	73,05^(C))	10,46^(C))	6,74 ^(B))	1,60^(C))
7	B	14,36^(C))	65,07^(C))	9,57 ^(B))	5,32 ^(B))	5,67 ^(C))
8	B	4,61 ^(B))	73,58^(C))	6,03 ^(B))	8,33^(B))	7,45 ^(C))
9	D	8,51^(B))	7,27 ^(B))	4,08 ^(B))	77,84^(C))	2,30^(C))
10	D	3,90 ^(B))	7,98^(B))	6,03 ^(B))	79,96^(C))	2,13^(C))
11	E	8,16 ^(B))	5,50 ^(B))	44,68^(C))	10,82 ^(C))	30,85^(C))

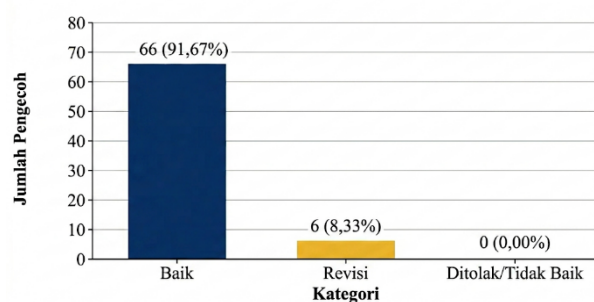
Nom or Soal	Kunci Jawa ban	Pilihan Jawaban (%)				
		A	B	C	D	E
12	D	5,50 ^(B))	5,50 ^(B))	10,11^(C))	75,00^(C))	3,90 ^(C))
13	E	51,95^(C))	17,02 ^(C))	9,40 ^(B))	6,03 ^(B))	15,60^(C))
14	D	7,09 ^(B))	7,62^(B))	6,74 ^(B))	71,45^(C))	7,09 ^(C))
15	B	4,96 ^(B))	62,59^(C))	15,25^(C))	8,16 ^(B))	9,04 ^(C))
16	A	76,60^(C))	6,91 ^(B))	4,79 ^(B))	7,98^(B))	3,72 ^(C))
17	C	7,62 ^(B))	8,51^(B))	72,16^(C))	5,50 ^(B))	6,21 ^(C))
18	C	5,50 ^(B))	3,90 ^(B))	74,65^(C))	8,51^(B))	7,45 ^(C))

Sumber: Data diolah, 2026

Berdasarkan Tabel 7, sebagian besar alternatif jawaban pengecoh pada setiap butir soal telah berfungsi dengan baik. Namun, masih terdapat beberapa pengecoh yang perlu direvisi, yaitu pada soal nomor 2 pilihan D dan E, soal nomor 3 pilihan E, soal nomor 6 pilihan E, soal nomor 9 pilihan E, dan soal nomor 10 pilihan E. Pengecoh tersebut termasuk kategori revisi berdasarkan kriteria efektivitas pengecoh menurut Bano *et al.*, (2022) karena memiliki nilai pengecoh di bawah 0,025 atau dipilih oleh kurang dari 2,5% peserta didik. Kondisi ini menunjukkan bahwa pengecoh tersebut kurang menarik perhatian peserta didik, sehingga perlu diperbaiki agar lebih logis, konsisten, dan seimbang dengan alternatif jawaban lainnya.

Jumlah keseluruhan pengecoh dihitung berdasarkan banyaknya soal dan jumlah pengecoh pada setiap butir soal. Setiap soal terdiri atas lima pilihan jawaban, yaitu satu kunci jawaban dan empat pengecoh. Dengan jumlah soal sebanyak 18 butir, maka total pengecoh yang dianalisis berjumlah 72 pengecoh. Rekapitulasi hasil

analisis efektivitas pengecoh kemudian disajikan pada gambar berikut.



Gambar 2 Rekapitulasi Hasil Efektivitas Pengecoh

Berdasarkan hasil tersebut, dari 72 pengecoh yang dianalisis, sebanyak 66 pengecoh atau 91,67% termasuk dalam kategori baik, 6 pengecoh diklasifikasikan ke dalam kategori revisi, dan tidak terdapat pengecoh yang tergolong dalam kriteria tolak atau tidak baik. Dengan demikian secara keseluruhan, hasil analisis membuktikan bahwa sebagian besar opsi pengecoh telah berfungsi secara optimal, tetapi beberapa pilihan pengecoh masih perlu diperbaiki.

Kualitas Soal Tes Formatif dalam Mengukur HOTS Berdasarkan Level Kognitif

Table 8 Pengelompokkan Soal Berdasarkan Level Kognitif HOTS

Level Kognitif	Mudah	Sedang	Sukar	Jumlah Soal	Persentase (%)
C4	2	5	1	8 soal	44,44
C5	0	4	2	6 soal	33,33
C6	0	4	0	4 soal	22,23
Total	2	13	3	18 soal	100

Sumber: Data diolah, 2026

Berdasarkan Tabel 8, level kognitif C4 memiliki variasi yang beragam karena tersebar pada kategori mudah, sedang, dan sukar dengan persentase soal paling tinggi

yaitu 44,44%. Sementara itu, soal level C5 berada pada kategori sedang dan sukar dengan persentase 33,33%, sedangkan soal level C6 berada pada kategori sedang dengan persentase paling rendah yaitu 22,22%. Hasil tersebut menunjukkan bahwa soal tes formatif lebih dominan dalam mengukur kemampuan menganalisis dibandingkan kemampuan mengevaluasi dan mencipta pada materi kesetimbangan kimia.

PEMBAHASAN

Kualitas Validitas Soal Tes Formatif untuk Mengukur HOTS pada Materi Kesetimbangan Kimia

Berdasarkan hasil validasi ahli, instrumen memperoleh persentase sebesar 88,95% dengan kategori sangat valid. Hasil tersebut menunjukkan bahwa butir soal telah memenuhi aspek kesesuaian indikator, kelayakan materi, penggunaan bahasa, dan kesesuaian dengan level HOTS pada materi kesetimbangan kimia. Sebagian besar soal memperoleh persentase validitas antara 81,25%–100%, meskipun beberapa soal masih memerlukan revisi pada aspek redaksi dan kejelasan konteks soal agar lebih mudah dipahami peserta didik. Tingginya validitas ini dipengaruhi oleh penyusunan soal yang mengacu pada indikator pembelajaran serta level kognitif HOTS C4, C5, dan C6.

Temuan ini sejalan dengan penelitian Nugrahnastiti & Kamaludin (2024) yang menyatakan bahwa instrumen HOTS perlu melalui validasi ahli dan analisis empiris untuk memastikan kesesuaian isi dan kualitas pengukuran. Setelah validasi ahli, instrumen dianalisis menggunakan Rasch Model berbantuan Winsteps dan menunjukkan bahwa seluruh 18 item akhir memenuhi kriteria *item fit* berdasarkan nilai Outfit MNSQ, Outfit ZSTD, dan *Pt Measure*

Corr. Seluruh nilai *Pt Measure Corr* yang positif menunjukkan bahwa peserta didik dengan kemampuan tinggi cenderung menjawab soal dengan benar dibandingkan peserta didik berkemampuan rendah.

Hasil penelitian ini juga didukung oleh penelitian Oktaviani & Harjono (2025) yang menjelaskan bahwa validasi isi, validasi konstruk, dan analisis Rasch Model penting dilakukan untuk memastikan kesesuaian instrumen dengan indikator dan tujuan pengukuran. Secara keseluruhan, hasil penelitian menunjukkan bahwa soal tes formatif berbasis HOTS pada materi kesetimbangan kimia layak digunakan sebagai instrumen untuk mengukur kemampuan berpikir tingkat tinggi peserta didik.

Kualitas Reliabilitas Soal Tes Formatif untuk Mengukur HOTS pada Materi Kesetimbangan Kimia

Hasil analisis reliabilitas menunjukkan bahwa instrumen memperoleh koefisien *Cronbach Alpha* sebesar 0,78 yang termasuk kategori bagus. Nilai tersebut menunjukkan bahwa instrumen memiliki konsistensi internal yang baik dalam mengukur HOTS peserta didik pada materi kesetimbangan kimia. Hal ini sejalan dengan pendapat Hayat (2022) yang menyatakan bahwa instrumen dinyatakan reliabel apabila memiliki nilai *Cronbach Alpha* minimal 0,60. Dengan demikian, instrumen yang dikembangkan mampu memberikan hasil pengukuran yang stabil dan konsisten.

Selain itu, nilai *item reliability* sebesar 0,99 termasuk kategori istimewa, yang menunjukkan bahwa tingkat kesukaran antarbutir soal tersusun secara sangat konsisten. Nilai *separation* sebesar 9,76 juga menunjukkan bahwa instrumen memiliki kemampuan yang sangat baik dalam membedakan strata tingkat kesukaran soal,

sehingga penyebaran tingkat kesukaran item tergolong jelas dan proporsional. Tingginya reliabilitas instrumen dipengaruhi oleh kesesuaian indikator pembelajaran, level HOTS, serta konstruksi soal yang disusun secara sistematis, didukung oleh jumlah responden yang memadai dalam analisis Rasch.

Temuan ini sejalan dengan pendapat Sumintono & Widhiarso (2015) yang menyatakan bahwa reliabilitas tinggi menunjukkan konsistensi instrumen dalam mengukur kemampuan peserta didik. Penelitian Ariani *et al.*, (2026) juga menegaskan bahwa reliabilitas merupakan aspek penting dalam evaluasi kualitas instrumen bersama validitas empiris, tingkat kesukaran, dan daya beda. Oleh karena itu, hasil penelitian ini menunjukkan bahwa instrumen tidak hanya layak secara isi, tetapi juga memiliki kestabilan empiris yang baik untuk digunakan dalam tes formatif berbasis HOTS.

Kualitas Tingkat Kesukaran Soal Tes Formatif Berbasis HOTS pada Materi Kesetimbangan Kimia

Hasil analisis menunjukkan bahwa tingkat kesukaran soal tidak hanya dilihat berdasarkan nilai logit dari Rasch Model, tetapi juga dikaitkan dengan persentase jawaban benar-salah serta level kognitif HOTS pada setiap item. Hal ini karena instrumen dirancang untuk mengukur kemampuan berpikir tingkat tinggi pada materi kesetimbangan kimia. Dengan demikian, kategori mudah, sedang, dan sulit tidak hanya menunjukkan banyaknya peserta didik yang menjawab benar, tetapi juga menggambarkan bagaimana peserta didik merespons tuntutan berpikir tingkat tinggi pada setiap soal.

Dari 18 butir soal, diperoleh 2 soal (11,11%) berkategori mudah, 13 soal

(72,22%) berkategori sedang, dan 3 soal (16,67%) berkategori sukar. Dominasi kategori sedang menunjukkan bahwa sebagian besar soal cukup menantang namun masih dapat dijawab oleh peserta didik. Hasil ini juga menunjukkan bahwa soal HOTS tidak selalu berada pada kategori sukar, karena tingkat kesukaran dipengaruhi oleh kejelasan stimulus, kompleksitas konsep, penyajian soal, serta kemampuan peserta didik dalam menghubungkan informasi.

Dalam penelitian ini, soal HOTS dikembangkan pada level kognitif C4, C5, dan C6. Soal C4 menuntut kemampuan menganalisis hubungan konsep kesetimbangan, soal C5 menuntut kemampuan mengevaluasi pengaruh suatu kondisi terhadap sistem kesetimbangan, sedangkan soal C6 menuntut kemampuan merancang strategi atau solusi berdasarkan konsep kesetimbangan kimia. Oleh karena itu, tingkat kesukaran soal juga dipengaruhi oleh tuntutan kognitif yang terdapat pada masing-masing item.

Meskipun distribusi tingkat kesukaran belum sepenuhnya sesuai dengan proporsi ideal, yaitu 25% soal mudah, 50% soal sedang, dan 25% soal sukar, dominasi soal kategori sedang masih dinilai relevan dengan tujuan pengukuran HOTS. Hal ini disebabkan karena materi kesetimbangan kimia bersifat abstrak dan menuntut peserta didik menganalisis hubungan antarvariabel, seperti konsentrasi, suhu, tekanan, arah pergeseran kesetimbangan, dan penerapan prinsip Le Chatelier. Oleh sebab itu, sebagian besar soal berada pada kategori sedang karena peserta didik tidak hanya menghafal konsep, tetapi juga melakukan analisis dan evaluasi dalam menentukan jawaban.

Temuan ini sejalan dengan penelitian Fadhilah & Wisudawati (2026) yang

menekankan bahwa instrumen berkualitas harus memiliki distribusi tingkat kesukaran yang cukup merata agar mampu mengukur kemampuan peserta didik secara komprehensif. Dalam konteks HOTS, soal yang tergolong mudah tetap dapat dikategorikan sebagai HOTS apabila menuntut kemampuan menganalisis, mengevaluasi, atau mencipta. Secara keseluruhan, hasil penelitian menunjukkan bahwa instrumen tes formatif HOTS pada materi kesetimbangan kimia telah memiliki variasi tingkat kesukaran yang cukup baik dan efektif dalam memberikan tantangan kognitif kepada peserta didik, meskipun beberapa item masih perlu disempurnakan agar distribusi tingkat kesukaran menjadi lebih proporsional.

Kualitas Daya Pembeda Tes Formatif Berbasis HOTS pada Materi Kesetimbangan Kimia

Hasil analisis menunjukkan bahwa seluruh butir soal memiliki nilai *Pt Measure Corr* positif dengan rentang 0,40–0,59. Nilai tersebut menunjukkan bahwa setiap item memiliki kemampuan yang cukup baik dalam membedakan peserta didik berkemampuan tinggi dan rendah. Korelasi positif menunjukkan adanya hubungan searah antara tingkat kemampuan peserta didik dan peluang menjawab soal dengan benar, sehingga tidak terdapat item yang bekerja berlawanan dengan konstruk yang diukur. Dengan demikian, seluruh item dapat dinyatakan memiliki daya pembeda yang baik.

Berdasarkan tingkat kesukarannya, soal kategori mudah memiliki rentang *Pt Measure Corr* sebesar 0,47–0,48, soal kategori sedang sebesar 0,42–0,59, dan soal kategori sukar sebesar 0,40–0,45. Hasil tersebut menunjukkan bahwa soal kategori sedang cenderung memiliki daya pembeda

yang lebih optimal dibandingkan soal yang terlalu mudah maupun terlalu sukar. Soal mudah umumnya dapat dijawab oleh hampir seluruh peserta didik sehingga kurang mampu menunjukkan perbedaan kemampuan, sedangkan soal sukar cenderung sulit dijawab oleh sebagian besar peserta didik sehingga kemampuan diskriminasinya menurun. Meskipun demikian, seluruh soal tetap memiliki nilai korelasi positif yang menunjukkan bahwa item masih berfungsi dengan baik dalam proses pengukuran.

Daya pembeda yang baik dipengaruhi oleh kualitas penyusunan soal yang selaras dengan indikator HOTS, terutama pada kemampuan analisis, evaluasi, dan mencipta. Temuan ini sejalan dengan penelitian Laliyo *et al.*, (2026) yang menyatakan bahwa korelasi positif pada Rasch Model menunjukkan adanya hubungan searah antara kemampuan peserta didik dan peluang menjawab soal dengan benar. Penelitian tersebut juga menegaskan bahwa item dengan daya pembeda yang baik merupakan syarat penting dalam pengukuran kemampuan berpikir tingkat tinggi karena mampu merefleksikan variasi kemampuan peserta didik secara akurat. Oleh karena itu, hasil penelitian ini menunjukkan bahwa instrumen tes formatif HOTS pada materi kesetimbangan kimia telah mampu membedakan tingkat kemampuan peserta didik secara memadai.

Kualitas Efektivitas Pengecoh (Distraktor) Tes Formatif Berbasis HOTS pada Materi Kesetimbangan Kimia

Hasil analisis efektivitas pengecoh dilakukan melalui Option Probability Curve pada Rasch Model dan rekapitulasi persentase jawaban peserta didik menggunakan Microsoft Excel. Kedua analisis tersebut digunakan untuk

memperoleh gambaran yang lebih rinci mengenai pola pemilihan jawaban peserta didik pada setiap butir soal. Berdasarkan hasil *Option Probability Curve*, peserta didik dengan kemampuan rendah cenderung berada pada kelompok jawaban salah, sedangkan peserta didik dengan kemampuan tinggi lebih banyak berada pada kelompok jawaban benar. Pola tersebut menunjukkan hubungan yang logis antara tingkat kemampuan peserta didik dan peluang menjawab soal secara tepat.

Analisis efektivitas pengecoh dilakukan berdasarkan nilai proportion endorsing, yaitu persentase peserta didik yang memilih setiap alternatif jawaban. Pengecoh dinyatakan berfungsi dengan baik apabila masih dipilih oleh sebagian peserta didik karena menunjukkan bahwa opsi tersebut cukup logis dan mampu mengecoh peserta didik yang belum memahami konsep secara optimal (Panjaitan *et al.*, 2024). Hasil analisis menunjukkan bahwa dari 72 pengecoh, sebanyak 66 pengecoh termasuk kategori baik dan 6 pengecoh termasuk kategori revisi. Pengecoh kategori baik menunjukkan bahwa alternatif jawaban berfungsi efektif dalam membedakan peserta didik yang memahami materi dan yang belum memahami materi, sedangkan pengecoh kategori revisi perlu diperbaiki pada aspek redaksi, konteks, maupun kemiripan dengan kunci jawaban.

Pada beberapa butir soal, persentase pemilihan kunci jawaban lebih tinggi dibandingkan pengecoh, yang menunjukkan bahwa sebagian besar peserta didik mampu menentukan jawaban yang benar. Namun, pengecoh pada soal tersebut tetap dikategorikan baik karena masih dipilih oleh sebagian peserta didik. Sebagai contoh, pada soal nomor 1, kunci jawaban opsi D dipilih sebesar 38,48%, sedangkan pengecoh opsi C dipilih sebesar 38,30%, sehingga seluruh

pengecoh pada soal tersebut tergolong efektif. Sebaliknya, pada soal nomor 2, pengecoh opsi D dan E termasuk kategori revisi karena dipilih kurang dari 2,5% peserta didik. Pada soal nomor 13, sebagian besar peserta didik memilih pengecoh opsi A sebesar 51,95%, sedangkan kunci jawaban opsi E hanya dipilih sebesar 15,60%, yang menunjukkan masih adanya kesulitan peserta didik dalam memahami konsep yang diujikan.

Secara keseluruhan, hasil analisis menunjukkan bahwa pengecoh pada soal HOTS materi kesetimbangan kimia telah berfungsi dengan baik. Pengecoh yang efektif tidak hanya berperan sebagai jawaban salah, tetapi juga mampu merepresentasikan miskonsepsi atau kesalahan berpikir peserta didik. Temuan ini sejalan dengan penelitian Rezigalla *et al.*, (2024) yang menyatakan bahwa efektivitas pengecoh berhubungan langsung dengan tingkat kesukaran dan daya pembeda soal. Penelitian Priyani & Sugiharto (2024) juga menegaskan bahwa pengecoh yang berfungsi optimal dapat meningkatkan ketepatan instrumen dalam mengukur kemampuan peserta didik. Oleh karena itu, hasil penelitian menunjukkan bahwa soal tes formatif berbasis HOTS pada materi kesetimbangan kimia telah memiliki kualitas pengecoh yang baik, meskipun beberapa opsi masih perlu disempurnakan agar lebih optimal dalam mendukung pengukuran HOTS.

Kualitas Soal Tes Formatif dalam Mengukur HOTS Berdasarkan Level Kognitif

a. Kualitas Soal Level Kognitif C4 dalam Mengukur HOTS

Berdasarkan hasil analisis, terdapat 8 butir soal pada level kognitif C4 (menganalisis), yaitu soal nomor 1–8.

Dari keseluruhan soal tersebut, 2 soal termasuk kategori mudah (nomor 2 dan 3), 5 soal kategori sedang (nomor 4, 5, 6, 7, dan 8), serta 1 soal kategori sukar (nomor 1). Variasi tingkat kesukaran ini menunjukkan bahwa meskipun berada pada level kognitif yang sama, kompleksitas proses analisis yang diperlukan peserta didik berbeda-beda. Tingkat kesukaran dipengaruhi oleh karakteristik stimulus, jumlah konsep yang dianalisis, keterkaitan antarvariabel, dan kedalaman penalaran yang diperlukan dalam menyelesaikan soal.

Soal nomor 2 dan 3 tergolong mudah karena stimulus yang digunakan membantu peserta didik memahami informasi secara lebih jelas. Pada soal nomor 2, peserta didik menganalisis grafik perubahan konsentrasi menuju keadaan setimbang sehingga proses analisis lebih bersifat interpretatif. Sementara itu, soal nomor 3 hanya menuntut peserta didik menghubungkan kondisi reaksi dengan konsep dasar kesetimbangan dinamis tanpa melibatkan integrasi beberapa konsep secara bersamaan.

Berbeda dengan kedua soal tersebut, soal nomor 1 termasuk kategori sukar karena menuntut peserta didik menganalisis hubungan antara fotosintesis dan respirasi dalam konteks kesetimbangan kimia. Soal ini memerlukan integrasi konsep lintas disiplin, yaitu kimia dan biologi, serta pemahaman hubungan reaksi dua arah dalam kehidupan sehari-hari. Kompleksitas proses analisis tersebut menyebabkan peserta didik memerlukan penalaran yang lebih mendalam.

Sementara itu, soal nomor 4–8 termasuk kategori sedang karena memerlukan proses analisis yang lebih

kompleks, tetapi masih berada dalam konteks yang cukup dipahami peserta didik. Soal nomor 4 menuntut analisis jenis kesetimbangan berdasarkan fase zat, soal nomor 5 mengintegrasikan konsep stoikiometri dan tetapan kesetimbangan (K_c), soal nomor 6 menganalisis hubungan suhu dengan sifat reaksi endoterm dan eksoterm, soal nomor 7 melibatkan hubungan derajat disosiasi dengan perhitungan jumlah mol gas, sedangkan soal nomor 8 menuntut analisis pengaruh perubahan volume terhadap arah pergeseran kesetimbangan berdasarkan asas Le Chatelier.

Berdasarkan uraian tersebut, dapat dipahami bahwa tingkat kesukaran soal HOTS pada level kognitif C4 tidak hanya dipengaruhi oleh level kognitifnya, tetapi juga oleh kompleksitas stimulus, jumlah konsep yang harus dihubungkan, serta tahapan penalaran yang diperlukan peserta didik dalam menyelesaikan soal. Soal dengan stimulus sederhana dan hanya melibatkan satu konsep utama cenderung lebih mudah, sedangkan soal yang membutuhkan integrasi beberapa konsep dan penalaran bertahap cenderung memiliki tingkat kesukaran lebih tinggi.

b. Kualitas Soal Level Kognitif C5 dalam Mengukur HOTS

Berdasarkan hasil analisis, terdapat 6 butir soal pada level kognitif C5 (mengevaluasi), yaitu soal nomor 9, 10, 11, 12, 13, dan 14. Soal-soal tersebut terdiri atas 4 soal kategori sedang (nomor 9, 10, 12, dan 14) serta 2 soal kategori sukar (nomor 11 dan 13). Variasi tingkat kesukaran tersebut menunjukkan bahwa meskipun berada pada level kognitif yang sama, kompleksitas proses evaluasi yang diperlukan peserta didik dalam menyelesaikan soal berbeda-beda. Pada level C5, peserta didik tidak hanya

dituntut memahami dan menganalisis konsep, tetapi juga melakukan penilaian, memberikan pertimbangan, dan menentukan keputusan berdasarkan konsep kesetimbangan kimia.

Soal nomor 9, 10, 12, dan 14 tergolong kategori sedang karena proses evaluasi yang diperlukan masih berada pada konteks yang cukup familiar bagi peserta didik. Pada soal-soal tersebut, peserta didik diminta mengevaluasi pengaruh perubahan kondisi reaksi terhadap sistem kesetimbangan, hubungan perubahan kondisi dengan nilai tetapan kesetimbangan, serta pengaruh tekanan dan volume terhadap arah pergeseran kesetimbangan berdasarkan asas Le Chatelier. Tingkat kesukarannya tergolong sedang karena informasi yang disajikan masih cukup terarah sehingga peserta didik dapat melakukan evaluasi tanpa melalui tahapan penalaran yang terlalu kompleks.

Sebaliknya, soal nomor 11 dan 13 termasuk kategori sukar karena menuntut proses evaluasi yang lebih mendalam. Pada kedua soal tersebut, peserta didik harus menelaah beberapa informasi atau kondisi sekaligus sebelum menentukan jawaban yang paling tepat. Soal nomor 11 menuntut peserta didik mengevaluasi beberapa kondisi reaksi secara bersamaan dan mempertimbangkan setiap alternatif jawaban secara menyeluruh. Sementara itu, soal nomor 13 mengharuskan peserta didik menafsirkan informasi terkait faktor-faktor yang memengaruhi kesetimbangan kimia serta menghubungkan beberapa konsep secara simultan. Kompleksitas proses evaluasi tersebut menyebabkan kedua soal memerlukan penalaran yang lebih mendalam dibandingkan soal kategori sedang.

Berdasarkan uraian tersebut, tingkat kesukaran soal pada level kognitif C5 dipengaruhi oleh kompleksitas proses evaluasi yang diperlukan peserta didik. Soal dengan informasi yang lebih terarah dan hanya menuntut evaluasi terhadap satu kondisi utama cenderung memiliki tingkat kesukaran lebih rendah. Sebaliknya, soal yang mengharuskan peserta didik mengevaluasi beberapa kondisi, membandingkan alternatif jawaban, dan melakukan penalaran bertahap cenderung memiliki tingkat kesukaran lebih tinggi. Dengan demikian, seluruh soal pada level C5 dalam penelitian ini tetap termasuk kategori HOTS karena menuntut kemampuan mengevaluasi, memberikan pertimbangan logis, dan menentukan keputusan berdasarkan konsep kesetimbangan kimia.

c. Kualitas Soal Level Kognitif C6 dalam Mengukur HOTS

Berdasarkan hasil analisis, terdapat 4 butir soal pada level kognitif C6 (mencipta), yaitu soal nomor 15, 16, 17, dan 18. Seluruh soal tersebut berada pada kategori sedang. Hasil ini menunjukkan bahwa soal level C6 telah mampu mengukur kemampuan berpikir tingkat tinggi peserta didik, khususnya dalam menyusun strategi penyelesaian, menentukan keputusan, dan mengintegrasikan beberapa konsep kesetimbangan kimia dalam penyelesaian soal. Meskipun seluruh soal berada pada kategori sedang, masing-masing item tetap memiliki karakteristik proses berpikir yang berbeda sesuai dengan kompleksitas konsep dan penalaran yang diperlukan.

Soal nomor 15 dan 16 tergolong kategori sedang karena peserta didik dituntut menentukan solusi atau

keputusan berdasarkan kondisi kesetimbangan yang disajikan. Pada kedua soal tersebut, peserta didik perlu menghubungkan beberapa informasi dan menyusun langkah penyelesaian berdasarkan konsep yang telah dipelajari. Namun, stimulus soal masih cukup terarah sehingga strategi penyelesaian dapat disusun tanpa memerlukan penalaran yang terlalu kompleks.

Seluruh soal level C6 berada pada kategori sedang karena stimulus yang digunakan masih berada pada konteks yang familiar bagi peserta didik dan informasi pada soal disajikan secara cukup jelas. Peserta didik juga telah diarahkan pada konsep-konsep yang relevan sehingga proses penyusunan solusi dapat dilakukan secara bertahap. Selain itu, soal-soal pada level ini belum menuntut peserta didik menghasilkan solusi yang sepenuhnya baru, tetapi lebih menekankan kemampuan menyusun strategi penyelesaian berdasarkan konsep yang telah dipelajari sebelumnya. Kondisi tersebut menyebabkan tingkat kesukaran soal tidak mencapai kategori sukar.

Berdasarkan uraian tersebut, dapat dipahami bahwa soal pada level kognitif C6 dalam penelitian ini telah mampu mengukur kemampuan berpikir kreatif peserta didik dalam materi kesetimbangan kimia. Soal-soal tersebut menuntut peserta didik menyusun strategi, menentukan keputusan, dan mengintegrasikan beberapa konsep secara simultan, sehingga seluruh item pada level C6 tetap termasuk kategori HOTS.

d. Karakteristik Tingkat Kesukaran pada Level Kognitif HOTS

Berdasarkan hasil analisis tingkat kesukaran pada setiap level kognitif

HOTS, diperoleh distribusi tingkat kesukaran yang bervariasi pada ranah C4, C5, dan C6. Pada level kognitif C4 ditemukan soal dengan kategori mudah, sedang, dan sukar. Pada level kognitif C5 terdapat soal kategori sedang dan sukar, sedangkan seluruh soal pada level kognitif C6 berada pada kategori sedang. Variasi tersebut menunjukkan bahwa tingkat kesukaran soal HOTS tidak hanya dipengaruhi oleh level kognitif, tetapi juga oleh kompleksitas stimulus, jumlah konsep yang dianalisis, dan proses penalaran yang diperlukan dalam menyelesaikan soal.

Pada level kognitif C4 (menganalisis), soal memiliki variasi tingkat kesukaran yang lebih beragam karena proses analisis dapat memiliki kompleksitas yang berbeda-beda. Soal yang hanya menuntut interpretasi informasi sederhana atau analisis satu konsep utama cenderung berada pada kategori mudah, sedangkan soal yang memerlukan integrasi beberapa konsep, interpretasi stimulus yang lebih mendalam, dan penalaran bertahap cenderung memiliki tingkat kesukaran lebih tinggi. Oleh karena itu, pada level C4 ditemukan distribusi tingkat kesukaran mulai dari mudah hingga sukar.

Pada level kognitif C5 (mengevaluasi), soal didominasi kategori sedang dan sukar. Hal ini menunjukkan bahwa kemampuan mengevaluasi memerlukan proses berpikir yang lebih kompleks dibandingkan kemampuan menganalisis. Peserta didik tidak hanya memahami dan menganalisis informasi, tetapi juga harus memberikan penilaian, mempertimbangkan beberapa alternatif jawaban, dan menentukan keputusan yang paling tepat berdasarkan konsep

ilmiah. Soal yang melibatkan evaluasi terhadap beberapa kondisi secara simultan cenderung memiliki tingkat kesukaran lebih tinggi.

Sementara itu, seluruh soal pada level kognitif C6 (mencipta) berada pada kategori sedang. Meskipun soal C6 menuntut kemampuan menyusun strategi penyelesaian dan menentukan solusi, stimulus soal masih berada pada konteks yang cukup familiar dan informasi yang tersedia cukup terarah. Selain itu, soal lebih menekankan kemampuan menyusun solusi berdasarkan konsep yang telah dipelajari sebelumnya, bukan menciptakan solusi yang sepenuhnya baru. Oleh karena itu, tingkat kesukaran soal level C6 cenderung berada pada kategori sedang.

Hasil penelitian ini menunjukkan bahwa soal HOTS tidak selalu memiliki tingkat kesukaran tinggi. Soal HOTS dapat tergolong mudah, sedang, maupun sukar bergantung pada kompleksitas proses berpikir, karakteristik stimulus, banyaknya konsep yang dihubungkan, dan kedalaman penalaran yang diperlukan peserta didik. Dengan demikian, level kognitif HOTS dan tingkat kesukaran soal saling berkaitan, tetapi tetap merupakan aspek yang berbeda dalam analisis kualitas instrumen.

e. Persentase Distribusi Soal HOTS Berdasarkan Level Kognitif

Distribusi soal HOTS pada penelitian ini terdiri atas 44,44% level C4, 33,33% level C5, dan 22,22% level C6. Hasil tersebut menunjukkan bahwa soal level C4 memiliki persentase paling tinggi dibandingkan level C5 dan C6. Distribusi ini masih dapat dianggap proporsional karena kemampuan analisis merupakan dasar dalam pengembangan

kemampuan evaluasi dan mencipta pada berpikir tingkat tinggi. Temuan ini sejalan dengan penelitian Efendi *et al.*, (2025) yang menyatakan bahwa soal level C4 cenderung lebih dominan karena kemampuan analisis menjadi landasan sebelum peserta didik mampu melakukan evaluasi dan menciptakan solusi.

Selain itu, proporsi ideal soal HOTS umumnya terdiri atas 40% level C4, 30% level C5, dan 30% level C6 agar kemampuan berpikir kritis, evaluatif, dan kreatif peserta didik dapat berkembang secara optimal (Masrifah *et al.*, 2025). Berdasarkan hal tersebut, distribusi soal HOTS dalam penelitian ini sudah tergolong cukup proporsional karena seluruh level HOTS telah terwakili dengan persentase yang relatif seimbang, meskipun persentase soal level C6 masih lebih rendah. Kondisi ini disebabkan karena soal level C6 cenderung lebih kompleks untuk dikembangkan, terutama dalam bentuk pilihan ganda, karena menuntut kemampuan peserta didik dalam menyusun strategi dan menentukan solusi secara lebih mendalam.

f. Kualitas Soal Tes Formatif dalam Mengukur HOTS Berdasarkan Level Kognitif

Berdasarkan hasil penelitian, semua soal yang berkaitan dengan materi kesetimbangan kimia masih bisa diklasifikasikan sebagai soal Higher Order Thinking Skills (HOTS) karena semua soal berada di tingkat kognitif C4 (analisis), C5 (evaluasi), dan C6 (kreasi). Hasil ini mengindikasikan bahwa kemampuan berpikir yang diperlukan untuk menyelesaikan soal sudah sejalan dengan ciri-ciri kemampuan berpikir tingkat tinggi yang berdasarkan taksonomi Bloom revisi.

Hasil dari penelitian ini memperlihatkan bahwa soal HOTS tidak selalu masuk dalam kategori sulit. Dalam penelitian ini, soal HOTS terbagi menjadi kategori mudah, sedang, dan sukar. Ini menunjukkan bahwa tingkat kesulitan soal tidak bisa dijadikan satu-satunya patokan untuk menilai kualitas instrumen HOTS. Soal yang tergolong dalam kategori mudah masih bisa dianggap sebagai soal HOTS jika tetap memerlukan kemampuan untuk menganalisis, mengevaluasi, atau menciptakan dalam proses penyelesaiannya. Di sisi lain, soal yang dianggap sulit belum tentu tergolong sebagai soal HOTS jika hanya memerlukan kemampuan untuk mengingat atau memahami konsep.

Selain didasarkan pada level kognitif, kualitas soal HOTS dalam penelitian ini juga didukung oleh analisis validitas, reliabilitas, kemampuan daya pembeda, dan efektivitas pengecoh. Analisis validitas menunjukkan bahwa semua soal dapat mengukur indikator pembelajaran sesuai dengan sasaran pengukuran HOTS. Analisis reliabilitas menunjukkan bahwa alat ukur tersebut memiliki konsistensi yang baik dalam menilai kemampuan berpikir tingkat tinggi siswa. Selain itu, sebagian besar soal memiliki kemampuan membedakan yang baik, sehingga bisa mengidentifikasi siswa dengan kemampuan tinggi dan rendah. Pengecoh di setiap soal juga berfungsi dengan optimal, karena setiap pilihan jawaban dipilih oleh siswa, yang menunjukkan bahwa mereka berpikir untuk menentukan jawaban yang benar.

Dengan demikian, kualitas soal HOTS dalam penelitian ini bukan hanya ditentukan oleh tingkat kesulitan soal,

tapi juga dipengaruhi oleh keselarasan level kognitif, karakteristik stimulus, kemampuan soal dalam membedakan potensi peserta didik, serta efektivitas pilihan salah dalam memandu proses berpikir peserta didik. Oleh sebab itu, seluruh butir soal pada penelitian ini dapat dikatakan memiliki kualitas yang baik dan layak digunakan dalam menilai kemampuan berpikir tingkat tinggi peserta didik mengenai materi kesetimbangan kimia.

KESIMPULAN

Berdasarkan hasil penelitian, instrumen tes formatif berbasis *Higher Order Thinking Skills* (HOTS) pada materi kesetimbangan kimia yang dikembangkan telah memenuhi kriteria kualitas instrumen yang baik ditinjau dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil validasi ahli menunjukkan bahwa instrumen berada pada kategori sangat valid sehingga layak digunakan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik. Hasil analisis menggunakan Rasch Model juga menunjukkan bahwa seluruh item akhir memenuhi kriteria *item fit* sehingga mampu mengukur konstruk HOTS secara konsisten.

Instrumen yang dikembangkan memiliki reliabilitas yang baik dengan nilai *Cronbach Alpha* sebesar 0,78 serta nilai *item reliability* yang tergolong istimewa. Hal tersebut menunjukkan bahwa instrumen mampu memberikan hasil pengukuran yang stabil dan konsisten. Selain itu, hasil analisis tingkat kesukaran menunjukkan adanya variasi tingkat kesukaran pada setiap level kognitif HOTS. Pada level C4 ditemukan

soal kategori mudah, sedang, dan sukar, pada level C5 terdapat soal kategori sedang dan sukar, sedangkan seluruh soal level C6 berada pada kategori sedang. Variasi tersebut menunjukkan bahwa tingkat kesukaran soal tidak hanya dipengaruhi oleh level kognitif, tetapi juga oleh kompleksitas stimulus, jumlah konsep yang dianalisis, dan proses penalaran yang diperlukan dalam menyelesaikan soal.

Hasil analisis daya pembeda menunjukkan bahwa seluruh item memiliki nilai *Pt Measure Corr* positif sehingga mampu membedakan peserta didik berkemampuan tinggi dan rendah dengan baik. Selain itu, hasil analisis efektivitas pengecoh menunjukkan bahwa sebagian besar pengecoh telah berfungsi secara optimal karena masih dipilih oleh peserta didik, meskipun beberapa pengecoh masih memerlukan revisi agar lebih efektif. Distribusi soal HOTS dalam penelitian ini juga tergolong cukup proporsional karena seluruh level HOTS, yaitu C4, C5, dan C6, telah terwakili dalam instrumen.

Secara keseluruhan, instrumen tes formatif berbasis HOTS pada materi kesetimbangan kimia yang dikembangkan dalam penelitian ini layak digunakan sebagai alat evaluasi untuk mengukur kemampuan berpikir tingkat tinggi peserta didik. Instrumen ini diharapkan dapat membantu guru dalam mengevaluasi kemampuan analisis, evaluasi, dan kreasi peserta didik secara lebih komprehensif pada pembelajaran kimia, khususnya materi kesetimbangan kimia.

DAFTAR PUSTAKA

- Ariani, S. N., Yerimadesi, Gazali, F., & Aini, F. Q. (2026). Pengembangan Instrumen Tes Literasi Kimia pada Materi Sistem Periodik Unsur SMA/MA. *Journal of Chemistry Education and Integration*, 5(1), 68–81.
- Asri, S. K., & Julisman, I. (2022). Pengaruh Citra Merek dan Kualitas Produk Philips Terhadap Minat Beli Konsumen pada Yogya Grand Majalengka. *Jurnal Impresi Indonesia (JII)*, 1(3), 282–287. <https://doi.org/10.36418/jii.v1i3.40>.Siti
- Asriadi, M., & Hadi, S. (2021). *Analysis of the Quality of the Formative Test Items for Physics Learning Using the Rasch Model in the 21 st Century Learning*. *Jipf (Jurnal Ilmu Pendidikan Fisika)*, 6(2), 158–166. <https://doi.org/dx.doi.org/10.26737/jipf.v6i2.2030>
- Aulia, A. F., & Rahmawati, S. (2025). *Development of Higher-Order Thinking Skills Oriented to Illustrated Multiple Choice Questions Using A Contextual Approach to Chemical Equilibrium Materials*. *Jurnal Pendidikan Sains Dan Matematika*, 13(130), 45–62. <https://doi.org/doi.org/10.23971/11ewke66>
- Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis Kriteria Butir Soal Ujian Sekolah Mata Pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, 8(1), 145. <https://doi.org/10.32884/ideas.v8i1.660>
- Basyiroh, U., Mulyani, B., & Dwi, R. (2022). Analisis Kesulitan Belajar pada Materi Kestimbangan Kimia dengan Tes Diagnostik Three-Tier Multiple Choice pada Siswa Kelas XI SMA Negeri 1 Surakarta Tahun Pelajaran 2020/2021. *Jurnal Pendidikan Kimia*, 11(1), 51–59. <https://doi.org/doi.org/10.20961/jp.kim.v11i1.49899>
- Efendi, F. R., Isnaeni, H. N., & Saputra, R. (2025). Analisis Soal HOTS dalam Latihan Buku Teks Cerdas Cergas Berbahasa dan Bersastra Indonesia Kelas 12. *Geram: Jurnal Pendidikan, Bahasa Dan Sastra*, 13(2), 269–282. <https://doi.org/10.25299/geram.2025.22882>
- Fadhilah, S. N., & Wisudawati, A. W. (2026). *The HOTS-Characteristic Diagnostic Tests Using the System Thinking Approach to Detect Acid-Base Misconceptions in High School Chemistry*. *Jurnal Kimia Dan Pendidikan Kimia*, 11(1), 109–129. <https://doi.org/doi.org/10.20961/jkpk.v11i1.111442>
- Hayat, B. (2022). *Adjustment for Guessing In a Basic Statistics Test for Indonesian Undergraduate Psychology Students Using the Rasch Model*. *Cogent Education*, 9(1). <https://doi.org/10.1080/2331186X.2022.2059044>
- Himmah, F., & Sulaikho, S. (2022). Pengembangan Media Pembelajaran Berbasis Android Dengan Pemanfaatan. *Journal of Education and Management Studies*, 5(4), 38–47.

- <https://creativecommons.org/licenses/by-nc/4.0/>
- Husada, V. M. S., Yamtinah, S., & Indriyanti, N. Y. (2021). Analisis Konten High Order Thinking Skill pada Soal PAS dan PAT Kimia Kelas XI Tahun Pelajaran 2019/2020 di Kota Surakarta. *Jurnal Pendidikan Kimia*, 10(2), 215–223. <https://doi.org/doi.org/10.20961/jp.kim.v10i2.48679>
- Laliyo, L. A. R., Katili, Y., Lukum, A., Kilo, A. La, & Pikoli, M. (2026). *Rasch PCM Diagnostic Analysis of Critical Thinking Item Responses in Indonesian Atomic Structure Learning*. *Jurnal Kimia Dan Pendidikan Kimia*, 11(1), 130–146. <https://doi.org/doi.org/10.20961/jkpk.v11i1.112310>
- Listiani, W., & Rachmawati. (2022). Transformasi Taksonomi Bloom dalam Evaluasi Pembelajaran Berbasis HOTS. *Jurnal Jendela Pendidikan*, 2(03), 397–402. <https://doi.org/doi.org/10.57008/jjp.v2i03.266>
- Masrifah, K., Ahsanuddin, M., Mahliatussikah, H., & Ismail, Z. (2025). *Evaluating Higher-Order Thinking Skills Representation in the 2022 Quadra Arabic Language Textbooks for Indonesian Madrasah Aliyah*. *Al-Lisan: Jurnal Bahasa*, 10(2), 255–269. <https://doi.org/10.30603/al.v10i2.6703>
- Nugrahnastiti, N., & Kamaludin, A. (2024). *Development of a Higher-Order Thinking Skills (HOTS) Test Instrument on Electrochemical Material for High School Students*. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, 10(2), 451–466. <https://doi.org/doi.org/10.33394/jk.v10i2.11370>
- Oktaviani, S. Z., & Harjono. (2025). *Development of Two-Tier Multiple-Choice Diagnostic Test Instrument Based on Science Literacy on Chemical Equilibrium Material*. *Chemistry in Education*, 14(1), 1–14. <https://doi.org/10.15294/chemined.v14i1.21813>
- Panjaitan, B. R., Ningrum, E., & Waluya, B. (2024). *Analysis of Multiple Choice HOTS Test Questions on the Final Semester Assessment*. *Jurnal Geografi*, 16(2), 170–180. <https://doi.org/10.24114/jg.v16i2.55504>
- Priyani, T., & Sugiharto, B. (2024). *Analysis of Biology Midterm Exam Items Using a Comparison of The Classical Theory Test and The Rasch model*. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 10(3), 939–958. <https://doi.org/10.22219/jpbi.v10i3.34345>
- Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaiji, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). *Item Analysis: The Impact of Distractor Efficiency on The Difficulty Index and Discrimination Power of Multiple-Choice Items*. *BMC Medical Education*, 24(1), 1–7. <https://doi.org/10.1186/s12909-024-05433-y>
- Shavira, A., Fazlia, R., Rahmayani, I., Soal, A. B., & Kimia, K. (2022). Analisis

- Tingkat Kesulitan dan Kualitas Soal Tes Pada Materi Keseimbangan Kimia. *Jurnal Ilmiah Mahasiswa Jurusan Pendidikan Kimia*, 9(2), 44–52. <https://doi.org/10.82415/jimpk.v9i2.31324>
- Sumintono, B., & Widhiarso, W. (2015a). Aplikasi Pemodelan RASCH pada Assesment Pendidikan. Penerbit Trim Komunikata.
- Wulandari, T., Sari, D. P., & Nasution, A. R. (2024). Deskripsi Mendalam untuk Memastikan Keteralihan Temuan Penelitian Kualitatif. *Jurnal Literasiologi*, 11(2), 124–131. <https://doi.org/doi.org/10.47783/literasiologi.v9i4>
- Yessi, Driana, E., & Ernawati. (2022). Tingkat Pemahaman dan Instrumen HOTS Buatan Guru IPA di SMP Tangerang Selatan Tahun Pelajaran 2019-2020. *Jurnal Penelitian Dan Penilaian Pendidikan (JPPP)*, 7654(June), 15–28. <https://doi.org/doi.org/10.22236/jppp.v3i1>