



Analisis Soal HOTS Materi Asam Basa Untuk Mengukur Keterampilan Berpikir Kritis Dengan Rasch Model

Luluk Kamilia

Universitas Negeri Semarang, Indonesia

Korespondensi Penulis: lulukkamilia011@gmail.com*

Abstract. One of the competencies that must be mastered to face global competition in the 21st century world of work is individuals who have critical thinking skills. Critical thinking is one form of HOTS application. The purpose of this study was to analyze the HOTS test instrument to measure critical thinking skills. The method used is Research and Development which is reduced to 3D (define, design, development). The subjects in this study were students of class XI MIPA MAN 2 Wonosobo, with data collection techniques namely interviews, tests and questionnaires. The data analysis technique in this study was validation by experts and analysis with the Rasch model assisted by Ministep software. Analyzes with the Rasch model include summary statistics, item fit, item size, item correlation, and person size. Validation from experts gets an average score of 140 out of 160 with an average proportion of 87.7% which indicates a very valid instrument. Summary statistics for the testing phase show that the measurement of the person is 0.69 and 0.73 and the measurement of the item is 0.86 and 0.87, which means that the consistency of the student's answers is in the sufficient category, but the quality of the items is good. For the person measure implementation test stage it showed 0.69 and 0.74 meaning that the consistency of student answers was also in the sufficient category, while the quality of the items for the reliability aspect was in the special category with values of 0.96 and 0.97. Item Fit or the level of adding item items shows a 100% fit item from the testing stage to implementation.

Keywords: HOTS, Acid-Base, Critical thinking, Rasch Model

Abstrak. Salah satu kompetensi yang harus dikuasai untuk menghadapi persaingan global dalam dunia kerja abad 21 adalah individu dengan keterampilan berpikir kritis yang merupakan salah satu wujud penerapan HOTS. Tujuan penelitian ini adalah menganalisis instrumen tes HOTS untuk mengukur keterampilan berpikir kritis. Metode yang digunakan adalah *Research and Development* yang direduksi menjadi 3D (*define, design, development*). Subjek dalam penelitian ini siswa kelas XI MIPA MAN 2 Wonosobo, dengan teknik pengumpulan data yaitu wawancara, tes, dan angket. Teknik analisis data dalam penelitian ini adalah validasi pakar dan analisis dengan rasch model berbantuan *software* Ministep. Analisis dengan rasch model meliputi *summary statistic, item fit, item measure, item correlation, dan person measure*. Validasi dari ahli mendapatkan nilai rata-rata 140 dari skor 160 dengan persentase rata-rata 87,7% yang menunjukkan instrumen sangat valid. *Summary statistic* tahap uji coba menunjukkan *person measure* 0,69 dan 0,73 dan *item measure* 0,86 dan 0,87, artinya konsistensi jawaban siswa masuk kategori cukup, namun kualitas butir soalnya bagus. Untuk tahap uji implementasi *person measure* menunjukkan 0,69 dan 0,74 artinya konsistensi jawaban siswa juga masuk kategori cukup, sedangkan kualitas butir soal untuk aspek reliabilitasnya masuk kategori istimewa dengan nilai 0,96 dan 0,97. *Item Fit* atau tingkat kesesuaian butir soal menunjukkan soal 100% *fit*.

Kata kunci: HOTS, Asam-Basa, Berpikir Kritis, Rasch Model

1. PENDAHULUAN

Pendidikan memegang peranan yang sangat penting bagi generasi penerus kedepannya yang mana mempunyai tugas mengembangkan kemampuan untuk membentuk manusia dan peradaban yang layak untuk mencerdaskan kehidupan bangsa sesuai dengan ketentuan UU No. 20 Tahun 2003, Pasal 3 (Lubis *et al.*, 2021). Dalam undang-undang yang sama, pasal 1 ayat 1 menyatakan bahwa pendidikan adalah usaha terencana dan sadar guna mewujudkan

kan semangat belajar dan proses pembelajaran yang didalamnya peserta didik secara aktif mengembangkan kemampuannya. Potensi diri yang dimaksud adalah memiliki kekuatan spiritual keagamaan, pengendalian diri, budi pekerti, kecerdasan, akhlak mulia dan keterampilan (Nuryanti *et al.*, 2018). Dari pernyataan tersebut dapat kita lihat bahwa

pendidikan tidak hanya menambah pengetahuan tetapi juga menyiapkan keterampilan dan kemampuan yang akan dibutuhkan dalam kehidupan yang akan datang.

Kehidupan masa depan sering dikaitkan dengan Revolusi Industri 4.0. Revolusi Industri 4.0 membutuhkan sumber daya manusia yang berkualitas agar dapat bertahan di lingkungan yang semakin kompetitif (Pebriani *et al.*, 2021). Padahal, upaya penciptaan SDM yang dilakukan pemerintah akan diimplementasikan melalui kurikulum 2013. Silabus 2013 dirilis untuk mencerminkan era Revolusi Industri. Kurikulum K-13 menuntut siswa untuk lebih aktif, guru berperan sebagai fasilitator sedangkan siswa harus menggali pengetahuannya sendiri. Hal tersebut dapat membantu siswa dalam mengembangkan keterampilannya, yang meliputi kemampuan berpikir siswa.

Tujuan perancangan kurikulum 2013 adalah untuk menyiapkan warga negara yang berpendidikan karakter dan dapat mengintegrasikan keterampilan abad 21 (Yuliantaningrum, 2020). Keterampilan yang dibutuhkan untuk bersaing secara global di tempat kerja abad ke-21 meliputi kepribadian kreatif, pemikiran kritis, kemandirian, kerja sama tim, kreativitas, informasi, komunikasi, dan kemandirian belajar (Pratiwi *et al.*, 2019). Yang paling penting adalah wawasan tentang pemikiran logis dan kritis, karena ini mutlak diperlukan untuk pengembangan kemajuan belajar (Hallatu, 2017). Pembelajaran harus fokus pada keterampilan berpikir kritis untuk memungkinkan siswa berlatih dan mentransfer pemahaman mereka (Azizah *et al.*, 2018).

Kementerian Pendidikan dan Kebudayaan menerapkan standar internasional di bidang matematika, literasi, dan sains, yang membutuhkan pemikiran yang canggih. Keterampilan HOTS merupakan keterampilan penting dalam menghadapi dunia modern (Liana *et al.*, 2020). HOTS merupakan kemampuan berpikir yang tidak hanya diarahkan pada verbalisasi verbal, tetapi terkandung dalam suatu objek. Contoh interpretasi HOTS adalah kemampuan mendeskripsikan pemikiran integralistik dengan menggunakan analisis, sintesis, dan asosiasi untuk mencapai kesimpulan yang mengarah pada munculnya ide-ide kreatif dan inovatif (Fajarini *et al.*, 2019). HOTS merupakan keterampilan berpikir kognitif yang membutuhkan kemampuan lebih dari sekedar hafalan, atau repetisi. Keterampilan HOTS juga dapat dipahami sebagai proses berpikir untuk memecahkan suatu masalah dalam situasi yang kompleks. (Mari'a *et al.*, 2021).

Salah satu upaya memahami evolusi program Pendidikan di Indonesia dapat dilihat dari keikutsertaan Indonesia dalam *Program for International Student Assessments* (PISA). Nilai membaca PISA Indonesia 2018 adalah 371, di bawah rata-rata OECD 487. Nilai prestasi siswa Indonesia 379, di bawah rata-rata OECD 489. Sementara itu, nilai prestasi sains siswa

Indonesia 397, di bawah rata-rata Rata-rata OECD adalah 489 (Naufal *et al.*, 2021). Hasil tersebut menunjukkan bahwa penggunaan instrumen HOTS di kalangan siswa Indonesia masih rendah. Salah satu bentuk penerapan berpikir tingkat tinggi adalah keterampilan berpikir kritis.

Berdasarkan fakta di lapangan diketahui bahwa kemampuan berpikir kritis siswa SMA/MA masih cukup rendah yang tercermin dari hasil penelitian di SMAN 2 Sanggau. Hasil tes pelacakan berpikir kritis pertama yang dilakukan oleh Rasmawan (2015) pada 28 siswa kelas XI IPA dengan materi asam basa menunjukkan bahwa sebagian besar siswa termasuk dalam kategori keterampilan berpikir kritis rendah dan tidak terlatih (Tenti *et al.*, 2019). Hasil wawancara dengan guru kimia di MAN 2 Wonosobo diperoleh fakta bahwa guru masih jarang mengembangkan soal-soal HOTS secara mandiri untuk kegiatan penilaian dan pembelajaran. Sedangkan hasil wawancara dengan beberapa murid SMA di Kabupaten Wonosobo menyebutkan bahwa rata-rata guru masih melakukan penilaian dengan menggunakan instrumen tes yang diambil dari buku paket atau lembar kerja siswa (LKS) yang belum memenuhi standar HOTS salah satunya yaitu pada materi asam basa.

Materi asam basa merupakan salah satu contoh materi dalam kimia yang mendorong kemampuan berpikir kritis (Astuti *et al.*, 2022). Materi asam basa merupakan materi yang membutuhkan ketelitian dan keterampilan dalam perhitungan matematis (Supanti & Rusmawati, 2018). Disamping itu materi asam basa sangat penting dikuasai sebagai dasar pengetahuan untuk materi selanjutnya. Hasil evaluasi materi asam basa juga menjadi tolak ukur untuk pengetahuan materi selanjutnya.

Penilaian bukanlah hal yang baru di dunia pendidikan, namun dalam proses pembelajaran seringkali pendidik hanya berusaha meningkatkan kemampuan kognitif dengan mengabaikan beberapa aspek penilaian kinerja siswa. Akibatnya, kemampuan berpikir siswa masih tergolong lemah, karena guru hanya fokus pada apa yang diajarkan, tetapi instrumen yang dipakai kurang mengenal tujuan yang dinilai (Agustin *et al.*, 2018). Evaluasi guru tidak lepas dari penggunaan alat. Alat penilaian digunakan untuk menentukan seberapa baik pemahaman siswa setelah mempelajari suatu keterampilan. Pencapaian tujuan pembelajaran kimia memerlukan penggunaan alat asesmen yang melatih kemampuan berpikir, bukan hanya hafalan dan pemahaman (Amalia & Susilaningsih, 2014).

Penelitian ini bertujuan menganalisis soal HOTS materi asam basa yang dirancang untuk menilai kemampuan berpikir kritis siswa dengan menggunakan model Rasch. Diperlukan analisis untuk mengetahui kelayakan alat, profil kemampuan berpikir kritis siswa, serta respon guru dan siswa terhadap alat yang akan dikembangkan. Alat tes yang dikembangkan harus sesuai dan berkualitas, karena siswa membutuhkan alat yang tidak hanya

mengukur kemampuan kognitif hafalan dan pemahaman, tetapi juga dapat melatih kemampuan berpikir kritisnya.

2. METODE PENELITIAN

Penelitian ini menggunakan model 4-D yang direduksi menjadi 3-D yaitu: *define, design, develop*. Pada tahap pendefinisian dilakukan dengan mengidentifikasi kebutuhan dalam proses pembelajaran dan mengumpulkan berbagai informasi terkait instrumen tes HOTS yang akan dikembangkan. Tahap *design* atau desain bertujuan untuk merancang instrumen tes HOTS. Dimulai dengan menyusun kisi-kisi, memilih format dan media, kemudian membuat draf instrumen tes. Tahap terakhir adalah *development* atau pengembangan yaitu validasi soal oleh ahli dan analisis soal menggunakan *software* Ministep. Analisis dengan rasch model adalah analisis butir soal yang meliputi *summary statistic, item fit, item measure, item correlation*, dan *person measure*. Penelitian dilakukan pada bulan Maret 2023 yang berlokasi di MAN 2 Wonosobo. Populasi penelitian ini terdiri dari siswa kelas XI MIPA dengan sampel sebanyak 29 untuk uji coba dan 60 subjek implementasi. Instrumen dalam penelitian ini yaitu soal, angket dan lembar validasi. Teknik analisis data memakai dua jenis, yaitu analisa instrumen tes atau soal dan analisis instrumen nontes atau analisis angket siswa dan guru.

3. HASIL DAN PEMBAHASAN

Tahap Define

Pembuatan instrumen penilaian diawali dengan analisis kebutuhan dalam proses pembelajaran yang dilakukan di MA Negeri 2 Wonosobo, dimulai dengan analisis masalah, analisis siswa, analisis konsep dan analisis tujuan. Penelitian diawali dengan wawancara dengan subjek penelitian yaitu guru dan siswa. Hasil dari wawancara menunjukkan ada beberapa hal yang menjadi permasalahan: (1) Beberapa siswa merasa bahwa meskipun materi asam basa terlihat tidak begitu sulit tetapi pada faktanya mereka masih kesulitan untuk membedakan antara teori asam basa yang ada; (2) Guru belum maksimal dalam menggunakan soal-soal HOTS dalam pembelajaran, sehingga sebagian besar siswa masih kesulitan dalam menghadapi soal-soal HOTS. (3) Alat penilaian guru masih terbatas pada soal-soal yang terdapat dalam modul; (4) Standar soal yang disediakan dalam modul sebagian besar berada pada rentan C1 hingga C3 sehingga belum sesuai dengan level HOTS.

Permasalahan yang diperoleh dapat digunakan sebagai landasan dalam pembuatan instrumen dengan melakukan analisis konsep dan menentukan tujuan yang akan dicapai. Instrumen yang akan dikembangkan dirancang pada tahap selanjutnya yaitu tahap *design*.

Instrumen tersebut dirancang agar dapat digunakan untuk menyelesaikan masalah di lapangan. Tahap ini juga harus mendapatkan perhatian khusus karena akan menentukan hasil produk pada tahap selanjutnya. Tujuannya yaitu instrumen yang dihasilkan sungguh sesuai dengan standar dan relevan dengan tujuan penelitian.

Tahap *Design*

Dari hasil pengamatan dan interview yang didapat pada tahap *define*, instrumen tes HOTS mulai dikembangkan untuk mengukur kemampuan berpikir kritis siswa. Instrumen pertama yang dikembangkan adalah kisi-kisi. Kisi-kisi yang dibuat kemudian digunakan untuk mengembangkan soal dengan tujuan untuk mengukur keterampilan berpikir kritis. Bentuk soal yang dikembangkan dalam penelitian ini adalah pilihan ganda. Soal HOTS yang dikembangkan terdiri dari level soal C4 sampai C6 sesuai dengan level kognitif yang ditunjukkan oleh HOTS. Soal yang dikembangkan berjumlah 25 soal yang terintegrasi indikator berpikir kritis dengan materi Asam Basa. Soal terdiri dari 10 soal dengan jenjang soal C4, 12 soal dengan jenjang C5, dan 3 soal dengan jenjang C6. Masing-masing soal terintegrasi dengan indikator HOTS dan indikator keterampilan berpikir kritis. Dari uraian yang telah dijelaskan dapat disimpulkan bahwa setiap soal memiliki karakter yang berbeda. Media yang digunakan dalam pengembangan instrumen tes HOTS yaitu media cetak sehingga tes akan dilakukan secara *paper-based test* (PBT). Pemilihan format menentukan format yang sesuai untuk media yang digunakan. Pemilihan format dan media cetak diyakini lebih efektif karena menyesuaikan dengan kondisi di lapangan.

Tahap *Development*

Tahap pengembangan meliputi validasi produk dan analisis soal tahap uji coba dan implementasi.

Validasi Produk

Validasi instrumen dari materi, konstruksi dan bahasa dilakukan oleh tiga validator, dua validator dosen dan 1 validator guru kimia di MA Negeri 2 Wonosobo. Hasil validasi soal dapat dilihat pada Tabel 1. Hasil validasi angket jawaban guru dan siswa ditunjukkan pada Tabel 2. Berdasarkan hasil validasi dosen dan guru yang berpengalaman, dapat diketahui bahwa alat yang dibuat berada pada kategori sangat valid dan siap untuk diuji coba.

Tabel 1. Hasil Validasi Instrumen Tes HOTS oleh Validator Ahli

No	Validator	Skor Total	Kategori	Presentase
1.	Validator 1	137	Sangat valid	85,625%
2.	Validator 2	153	Sangat valid	95,625%
3.	Validator 3	131	Sangat valid	81,875%
Rata-rata		140	Sangat valid	87,7%

Tabel 2. Hasil Validasi Angket Respon Guru dan Siswa

No	Instrumen	Validator	Skor	Kategori	Presentase
1.	Angket respon guru	Validator 1	78	Sangat valid	97,5%
2.	Angket respon peserta didik	Validator 2	80	Sangat valid	100%

Analisis Butir Soal dengan Rasch Model

a. Reliabilitas Instrumen (*Summary Statistic*)

Summary statistic atau reliabilitas butir soal diperoleh menggunakan model rasch. Informasi mengenai ringkasan statistik dari produk yang dikembangkan dilihat di menu utama *Output Tables* pada Tabel 3.1 *Summary Statistic*. Ringkasan statistik yang diperoleh memberikan informasi secara keseluruhan meliputi *person reliability*, *item reliability*, dan *cronbach alpha reliability*. Hasil analisis secara keseluruhan menunjukkan bahwa instrumen yang dikembangkan memiliki kualitas yang sangat baik. Tabel 3 menunjukkan hasil dari masing-masing tahap uji.

Tabel 3. Reliabilitas Instrumen (*Summary Statistic*)

Tahap Uji	Analisis Rasch Model		
	<i>Person Reliability</i>	<i>Item Reliability</i>	<i>Cronbach Alpha Reliability</i>
Uji Coba	0,69 dan 0,73	0,86 dan 0,87	0,76
Uji Implementasi	0,69 dan 0,74	0,96 dan 0,97	0,85

Tingkat Kesesuaian Butir Soal (*Item Fit*)

Item fit dalam model rasch digunakan guna mengetahui instrumen yang dikembangkan dapat mengukur keterampilan berpikir kritis siswa dengan baik atau tidak. Pada setiap tahap uji, soal yang tidak sesuai (*misfit*) akan dievaluasi untuk menghasilkan soal yang sesuai (*fit*). Berikut ini hasil analisis kesesuaian butir soal diuraikan untuk setiap tahap uji pada Tabel 4.

Tabel 4. Tingkat Kesesuaian Butir Soal (*Item Fit*)

Tahap Uji	Kesesuaian Butir Soal	Nomor Soal	Jumlah
Uji Coba	Sesuai (<i>Fit</i>)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25	25 soal
	Tidak sesuai (<i>Misfit</i>)		0 soal
Uji Implementasi	Sesuai (<i>Fit</i>)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25	25 soal
	Tidak sesuai (<i>Misfit</i>)		0 soal

Tingkat Kesulitan Butir Soal (*Item Measure*)

Hasil analisa *item measure* menunjukkan soal yang dikembangkan lebih banyak pada kategori sangat sulit, sulit dan sedang dibandingkan dengan soal yang mudah. Hal ini sudah sesuai karena instrumen yang dikembangkan bertujuan untuk mengukur keterampilan berpikir kritis siswa. Di bawah ini dapat dilihat hasil pengelompokan soal berdasarkan tingkat kesulitannya pada Tabel 5.

Tabel 5. Tingkat Kesulitan Butir Soal (*Item Measure*)

Tahap Uji	Kriteria Indeks Kesukaran	Nomor Soal	Jumlah
Uji Coba	Sangat sulit	1, 2, 3, 14	4 soal
	Sulit	5, 7, 8, 10, 15, 16, 19, 20, 23, 24	10 soal
	Sedang	6, 9, 11, 12, 17, 18	6 soal
	Mudah	4, 13, 21, 22, 25	5 soal
Uji Implementasi	Sangat sulit	18, 2, 7, 22, 3, 23, 24, 10	8 soal
	Sulit		0 soal
	Sedang	1, 17, 25, 5, 13, 14, 16, 19, 4, 8, 11, 9, 12, 15, 20, 21	16 soal
	Mudah	6	1 soal

Daya Pembeda Butir Soal (*Item Corellation*)

Daya pembeda butir soal diperoleh melalui tabel *item correlation* dalam model rasch. Dari keseluruhan tahap uji, daya beda pada kategori cukup dan kategori baik lebih banyak dibandingkan dengan jumlah pada kategori lemah. Hasil analisis daya pembeda soal disajikan pada Tabel 6.

Tabel 6. Daya Pembeda Butir Soal (*Item Corellation*)

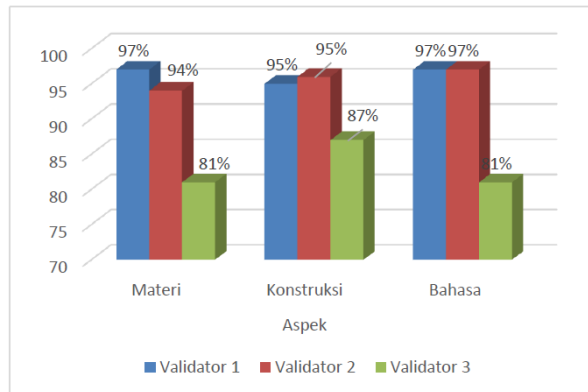
Tahap Uji	Kriteria Daya Pembeda	Nomor Soal	Jumlah
Uji Coba	Tidak ada daya pembeda	-	0 soal
	Lemah	4, 7, 21, 22	4 soal
	Cukup	6, 11, 12, 13, 18, 19	6 soal
	Baik	1, 2, 3, 5, 8, 9, 10, 14, 15, 16, 17, 20, 23, 24, 25	15 soal
	Baik sekali	-	0 soal
Uji Implementasi	Tidak ada daya pembeda	-	0 soal
	Lemah	2, 7, 18, 22	4 soal
	Cukup	3, 10, 24, 23, 6	5 soal
	Baik	1, 4, 5, 8, 9, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 25	16 soal
	Baik sekali	-	0 soal

Kelayakan Instrumen Test HOTS untuk Mengukur Keterampilan Berpikir Kritis Peserta Didik

a. Validasi Produk Instrumen Soal

Validasi instrumen produk melibatkan tiga aspek kelayakan: materia, konstruksi, dan bahasa. Hasil validasi menunjukkan bahwa instrumen HOTS yang dikembangkan untuk

mengukur kemampuan berpikir kritis siswa memiliki kategori sangat baik dengan rata-rata sebesar 87,7% seperti terlihat pada Grafik 1.



Grafik 1. Konsistensi Penilaian Instrumen oleh Validator Ahli

b. Validasi Angket

Aspek yang dinilai pada tahap validasi ini diantaranya kelengkapan komponen lembar angket, kelengkapan komponen pada kisi-kisi lembar angket, kesesuaian isi atau pernyataan yang terdapat pada lembar angket, kesesuaian penggunaan bahasa pada lembar angket, dan petunjuk pengisian angket. Hasil dari validasi angket respon guru menunjukkan kategori sangat valid dengan persentase skor yang diperoleh sebesar 97,5%. Sementara itu hasil angket respon siswa juga menunjukkan kategori sangat valid dengan skor 80 atau persentase yang diperoleh sebesar 100% menunjukkan angket sangat valid dan siap digunakan.

Hasil Analisis pada Tahap Uji Coba dan Uji Implementasi dengan Rasch Model

a. Reliabilitas Instrumen (Summary Statistic)

Reliabilitas instrumen dalam model Rasch ditentukan oleh kualitas pola respon siswa secara keseluruhan, dari instrumen yang dipakai, dan interaksi antara subjek dengan soal yang disajikan secara lengkap. Ringkasan statistik dari fase uji coba dan implementasi disajikan pada Gambar 1 dan Gambar 2.

SUMMARY OF 25 MEASURED PERSON									
TOTAL SCORE	COUNT	MEASURE	S.E.	MODEL	DIFF1	DIFF2	DIFF3	DIFF4	DIFF5
MEAN	12.0	25.0	-1.34	.00	.00	.00	.00	.00	.00
SD	5.0	.0	.00	.00	.00	.00	.00	.00	.00
P-SD	2.5	.0	1.50	.00	.00	.00	.00	.00	.00
N-SD	2.5	.0	1.50	.00	.00	.00	.00	.00	.00
REAL	18.0	25.0	2.00	1.50	2.00	1.50	1.50	1.50	1.50
PCN	8.0	25.0	-0.00	.00	.00	.00	.00	.00	.00
REAL	18.0	25.0	2.00	1.50	2.00	1.50	1.50	1.50	1.50
MODEL	18.0	25.0	2.00	1.50	2.00	1.50	1.50	1.50	1.50
S.E. OF PERSON MEAN	.41								
PERSON RELIABILITY	.97								
CRONBACH ALPHA (N=25) PERSON RAW SCORE	.97								
TEST RELIABILITY	.97								
SD	2.14								

SUMMARY OF 25 MEASURED ITEM									
TOTAL SCORE	COUNT	MEASURE	S.E.	MODEL	DIFF1	DIFF2	DIFF3	DIFF4	DIFF5
MEAN	18.0	25.0	-1.34	.00	.00	.00	.00	.00	.00
SD	5.0	.0	.00	.00	.00	.00	.00	.00	.00
P-SD	2.5	.0	1.50	.00	.00	.00	.00	.00	.00
N-SD	2.5	.0	1.50	.00	.00	.00	.00	.00	.00
REAL	18.0	25.0	2.00	1.50	2.00	1.50	1.50	1.50	1.50
PCN	8.0	25.0	-0.00	.00	.00	.00	.00	.00	.00
REAL	18.0	25.0	2.00	1.50	2.00	1.50	1.50	1.50	1.50
MODEL	18.0	25.0	2.00	1.50	2.00	1.50	1.50	1.50	1.50
S.E. OF ITEM MEAN	.41								
ITEM RELIABILITY	.96								
CRONBACH ALPHA (N=25) ITEM RAW SCORE	.96								
TEST RELIABILITY	.96								
SD	2.14								

Gambar 1. Hasil Output Tables 3.1 Summary Statistic Tahap Uji Coba

Gambar 2. Hasil Output Tables 3.1 Summary Statistic Tahap Implementasi

Hasil analisis pada tahap uji coba memberikan nilai reliabilitas person sebesar 0,69 dan 0,73 dan nilai reliabilitas item sebesar 0,86 dan 0,87. Hasil penelitian menunjukkan bahwa kesinambungan siswa dalam menjawab soal berada pada rentang “cukup” dan kualitas soal juga berada pada rentang “baik”. Hasil analisis untuk *text reliability* atau nilai *alpha Cronbach* tergolong pada kategori bagus, yang menandakan bahwa interaksi antara *person* dan butir-butir soal secara keseluruhan bagus. Hasil ketiga kategori pada *summary statistic* ini menunjukkan hasil bagus yang menandakan bahwa instrumen yang dikembangkan layak untuk digunakan.

Hasil analisis pada tahap uji implementasi menghasilkan nilai person reliability sebesar 0.69 dan 0.74, menunjukkan konsistensi siswa ketika menjawab soal tergolong pada kategori cukup. Nilai item reliability didapatkan sebesar 0.96 dan 0,97, yang artinya soal masuk pada kategori istimewa. *Alpha Cronbach* dengan nilai sebesar 0,83 menandakan soal masuk kategori bagus sekali. Hasil memenuhi standar pada setiap bagiannya yang menandakan bahwa instrumen yang dikembangkan layak untuk digunakan.

Tingkat Kesesuaian Butir Soal (*Item Fit*)

Item fit digunakan untuk menentukan kesesuaian soal terhadap model. Item fit menjelaskan item soal berfungsi sebagai suatu ukuran atau tidak. Berikut hasil analisis *item fit* pada tahap uji coba dan implementasi dapat dilihat pada Gambar 3 dan Gambar 4.

Item STATISTICS: MISFIT ORDER

ENTRY	TOTAL	TOTAL	DMLE	MODEL	INFIT	OUTFIT	[PTMEASURE-AL]	[EXACT MATCH]	
NUMBER	SCORE	COUNT	MEASURE	S.E.	INFOQ	ZSTD	CORR.	EXP.	Item
18	33	20	-1.20	.47	1.33	.91	1.44	1.27	36
6	32	20	-1.80	.45	1.15	.73	1.41	1.39	12
21	26	20	-2.13	.62	1.10	.26	1.35	.74	18
22	27	20	-2.59	.74	1.00	.30	1.35	.67	107
13	24	20	-1.52	.58	1.13	.55	1.21	.64	105
32	20	20	-.72	.41	1.15	.60	1.19	.92	102
4	25	20	-1.80	.55	1.07	.33	1.14	.43	102
31	32	20	-1.00	.40	1.00	.32	1.11	.40	111
10	10	20	.83	.40	1.00	.66	1.06	.37	112
20	12	20	.52	.39	1.07	.80	1.06	.47	115
25	24	20	-1.52	.58	1.03	.23	1.07	.39	113
14	6	20	1.57	.47	1.06	.31	.07	.40	116
10	6	20	1.00	.41	1.05	.37	1.03	.17	110
9	20	20	-.72	.41	1.01	.00	.00	.04	124
3	7	20	1.36	.45	1.00	.00	.00	.00	123
8	15	20	.87	.39	.90	.13	.99	.03	129
15	10	20	.83	.40	.99	.02	.94	.20	129
5	8	20	1.17	.43	.95	.17	.85	.50	136
1	6	20	1.57	.47	.92	.23	.88	.50	139
7	6	20	1.17	.43	.91	.00	.88	.42	140
23	9	20	1.00	.41	.86	.81	.81	.69	150
2	6	20	1.57	.47	.80	.73	1.03	.00	122
24	9	20	1.00	.41	.80	1.26	.73	1.00	140
16	10	20	.83	.40	.79	1.53	.74	1.00	165
17	16	20	-.80	.39	.79	1.53	.77	1.00	165
MEAN	15.0	20.0	.00	.40	1.00	-.00	1.00	.00	70.4
P.50	7.3	0	-1.20	.40	.12	.75	.21	.85	9.6

Gambar 3. Hasil Output Tables 10. *Item Fit* Order Tahap Uji Coba

Item STATISTICS: MISFIT ORDER

ENTRY	TOTAL	TOTAL	DMLE	MODEL	INFIT	OUTFIT	[PTMEASURE-AL]	[EXACT MATCH]	
NUMBER	SCORE	COUNT	MEASURE	S.E.	INFOQ	ZSTD	CORR.	EXP.	Item
3	6	60	3.12	.40	1.36	1.7	1.90	1.80	23
7	3	60	3.91	.60	1.00	.7	1.80	.16	95.0
22	3	60	3.91	.60	1.02	.7	1.80	.09	18
24	7	60	2.63	.42	1.11	.6	1.65	.00	25
2	3	60	3.91	.60	1.00	.7	1.22	1.33	10
10	9	60	2.61	.58	1.26	1.7	1.70	1.40	26
23	6	60	3.12	.40	.99	.6	1.60	.61	14
25	47	60	-1.27	.38	1.39	1.54	1.17	.55	36
30	2	60	4.35	.73	1.05	.31	1.22	.33	124
4	49	60	-1.58	.40	1.16	.69	1.13	.42	53
21	52	60	-1.11	.44	1.03	.20	.93	.04	108
1	30	60	-.22	.51	.98	.01	.97	-.07	111
6	57	60	-3.05	.64	.93	-.02	.20	-.59	143
9	51	60	-1.92	.41	.90	-.23	.65	.57	151
14	48	60	-1.42	.39	.89	-.41	.60	-.70	162
13	48	60	-1.42	.39	.87	-.40	.61	-.60	164
10	60	60	-1.42	.39	.85	-.55	1.17	.45	164
11	50	60	-1.74	.41	.81	-.69	.39	-.51	162
15	51	60	-1.92	.41	.78	-.80	.60	-.10	165
12	51	60	-1.92	.41	.76	-.90	.69	-.67	163
8	50	60	-1.74	.41	.71	-1.19	.50	-.86	168
5	48	60	-1.42	.39	.69	-1.19	.47	-.60	173
20	51	60	-1.92	.41	.69	-1.23	.42	1.22	168
17	45	60	-1.60	.36	.53	-2.46	.40	.90	181
16	48	60	-1.42	.39	.50	-2.46	.25	1.85	182
MEAN	14.8	60.0	.00	.45	.86	-.26	1.13	.14	89.1
P.50	20.0	0	2.46	.10	.22	.90	2.69	1.93	3.3

Gambar 4. Hasil Output Tables 10. *Item Fit* Order Tahap Uji Implementasi

Hasil di atas, pada bagian pertanyaan yang tidak mencapai kriteria MNSQ, ZSTD, dan PT MEASURE CORR disorot dengan warna merah di setiap tabel. Sekalipun item tersebut memiliki nilai MNSQ dan PT MEASURE CORR yang tidak memenuhi kriteria, belum tentu item tersebut tidak memenuhi syarat. Jika skor ZSTD sudah memenuhi kriteria, dan kategori lain tidak cocok, maka soal tetap dipertahankan. Hasil di atas menunjukkan bahwa soal hanya gagal memenuhi syarat salah satu kriteria, sehingga semua soal tetap digunakan. Berdasarkan hasil analisis, semua soal termasuk dalam kategori “fit”.

Tingkat Kesulitan Butir Soal (*Item Measure*)

Analisis untuk menentukan item measure ditunjukkan pada *Output Tables, Tabel 13. Item Measure*. Hasil analisis tingkat kesulitan butir soal ditunjukkan pada Gambar 5 dan Gambar 6.

Item STATISTICS: MEASURE ORDER										Item STATISTICS: MEASURE ORDER													
ENTRY	TOTAL	TOTAL	2MLE		INFIT	OUTFIT	PTMEASURE	AL-CORR.	EXACT	ENTRY	TOTAL	TOTAL	2MLE		INFIT	OUTFIT	PTMEASURE	AL-CORR.	EXACT				
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	INFIT	EXP.	OBSZ	NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	INFIT	EXP.	OBSZ				
1	6	25	1.57	.47	.92	-.21	.80	-.50	.39	22	79.3	18	2	60	4.35	-.73	1.66	-.31	1.21	.53	.85	.14	96.7
2	6	25	1.57	.47	.80	-.73	.66	-1.05	.68	22	79.3	2	3	60	3.91	.60	1.88	-.33	5.21	2.63	-.87	.16	95.0
14	6	25	1.57	.47	1.06	.31	.97	.02	.16	22	79.3	7	3	60	3.91	.60	1.88	-.33	7.09	3.38	-.10	.16	95.0
3	7	29	1.36	.45	1.00	.06	.96	-.06	.25	23	75.9	22	3	60	3.91	.60	1.88	-.33	5.98	3.18	-.83	.16	95.0
5	8	29	1.17	.43	.95	-.17	.85	-.55	.36	24	69.0	3	6	60	3.12	.45	1.36	1.14	9.00	5.68	-.51	.33	90.0
7	8	29	1.17	.43	.91	-.39	.88	-.42	.39	24	75.9	10	9	60	2.03	.42	.99	.07	2.03	1.64	.14	.23	96.0
10	9	29	1.00	.41	1.05	.37	1.02	.17	.18	25	55.6	24	7	60	2.03	.42	1.11	.47	6.85	3.58	-.82	.25	88.3
23	9	29	1.00	.41	.86	-.01	.81	-.90	.58	25	72.4	10	9	60	2.03	.42	1.11	.47	6.85	3.58	-.82	.25	88.3
24	9	29	1.00	.41	.80	-1.24	.73	-1.49	.62	25	72.4	1	38	60	-.21	.22	.98	-.05	.97	-.07	.55	.53	73.3
15	10	29	.83	.40	.99	-.02	.94	-.26	.29	26	60.0	17	45	60	-1.06	.30	.53	-2.47	.44	-2.18	.81	.55	95.0
16	10	29	.83	.40	.79	-1.33	.74	-1.59	.64	26	75.9	25	47	60	-1.27	.38	1.39	1.54	1.17	.55	.96	.54	88.0
19	10	29	.83	.40	1.09	.66	1.06	.37	.32	26	62.1	5	48	60	-1.42	.39	.87	-.48	.81	-.08	.64	.54	85.0
20	12	29	.52	.39	1.07	.65	1.06	.47	.15	26	48.3	13	48	60	-1.42	.39	.87	-.48	.81	-.08	.64	.54	85.0
8	15	29	.07	.39	.98	-.13	.99	-.03	.23	27	65.5	14	48	60	-1.42	.39	.89	-.41	.88	-.76	.62	.54	85.0
17	16	29	-.08	.39	.79	-2.15	.77	-2.00	.63	26	79.3	16	48	60	-1.42	.39	.59	-2.45	.39	2.35	.82	.54	91.7
9	20	29	-.71	.41	1.01	.09	.98	-.04	.24	25	72.4	19	20	29	-1.08	.45	1.86	.32	1.11	.48	.11	.23	75.9
12	20	29	-.71	.41	1.15	.94	1.19	.93	.03	25	65.5	20	20	29	-1.08	.45	1.86	.32	1.11	.48	.11	.23	75.9
6	22	29	-1.08	.45	1.15	.93	1.41	1.39	.12	23	75.9	11	50	60	-1.74	.41	.71	-1.10	.52	-1.06	.68	.52	90.0
11	22	29	-1.08	.45	1.06	.32	1.11	.48	.11	23	75.9	8	50	60	-1.74	.41	.69	-.71	.51	-.52	.52	.50	88.3
18	23	29	-1.29	.47	1.23	.91	1.44	1.27	.26	21	79.3	9	51	60	-1.92	.43	.92	-.23	.65	-.57	.57	.51	88.3
13	24	29	-1.52	.50	1.13	.51	1.21	.62	.06	20	82.8	12	51	60	-1.92	.43	.76	-.92	.61	-.67	.63	.51	91.7
25	24	29	-1.52	.50	1.03	.21	1.07	.30	.13	20	82.8	15	51	60	-1.92	.43	.78	-.80	.46	1.10	.65	.51	80.3
4	25	29	-1.80	.55	1.07	.31	1.14	.45	.02	18	86.2	20	51	60	-1.92	.43	.69	-1.23	.42	-1.22	.68	.51	91.7
21	26	29	-2.13	.62	1.10	.36	1.35	.74	-.10	16	89.7	21	52	60	-2.11	.44	1.03	.20	.91	.04	.48	.48	85.0
22	27	29	-2.59	.74	1.86	.30	1.35	.67	.07	13	93.1	6	57	60	-.46	.54	.93	-.02	.24	-.59	.43	.33	95.0
MEAN	15.0	29.0	.80	.46	1.00	-.02	1.02	-.04	.74.6			MEAN	34.8	60.0	-.08	.45	.94	-.26	2.13	.34			89.1
P.SD	7.3	.0	1.29	.08	.12	.75	.21	.84	9.6			P.SD	20.8	.0	2.46	.10	.22	.90	2.09	1.95			5.3

Gambar 5. Hasil *Output Tables 13. Item Measure* Tahap Uji Coba

Gambar 6. Hasil *Output Tables 13. Item Measure* Tahap Uji Implementasi

Pada *item measure* disajikan informasi nilai deviasi standar yang digunakan untuk melakukan pengelompokan item soal berdasarkan nilai rata-rata logit. Nilai untuk tahap uji coba adalah 1,29. Nilai logit atau *measure* berguna untuk mengetahui tingkat kesulitan sebuah soal. Semakin tinggi nilai logit soal, menandakan bahwa soal memiliki tingkat kesulitan yang semakin tinggi. Hasil analisis menunjukkan bahwa soal yang masuk pada kategori sangat sulit berjumlah 4 soal, 10 soal berada di kategori sulit, 6 soal pada kategori sedang, dan 5 soal berada di kategori mudah. Sesuai kategori kelompok kesulitan yang didapatkan berdasarkan rata-rata nilai logit pada tahap uji implementasi, dapat dilihat soal yang masuk pada kategori sangat sulit berjumlah 8 soal, 0 soal berada di kategori sulit, 16 soal masuk kategori sedang, dan 1 soal berada dikategori mudah. Soal dengan kategori sedang lebih banyak artinya soal layak dikembangkan karena soal tidak terlalu sulit dan tidak terlalu mudah.

Daya Pembeda Butir Soal (*Item Correlation*)

Daya pembeda ialah kemampuan suatu soal untuk membedakan antara siswa yang memiliki kecakapan tinggi dan rendah. Soal yang berkualitas baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar (Erfan et al., 2020). Kategori soal berdasarkan indeks daya pembeda item ditunjukkan pada Tabel 7. (Loka Son, 2019):

Tabel 7. Kategori Indeks Daya Pembeda Butir Soal

Daya Pembeda	Kategori
$D_p \leq 0,00$	Tidak ada daya pembeda
$0,00 < D_p \leq 0,20$	Lemah
$0,20 < D_p \leq 0,40$	Cukup
$0,40 < D_p \leq 0,70$	Baik
$0,70 < D_p \leq 1,00$	Baik sekali

Item Statistics: Correlation Index

Item	Total	Total Correct	Mean	SD	Item Mean	Item SD	Item Total Correlation	Item Total Correlation Squared	Item Total Correlation Squared Multiple	Item Total Correlation Squared Multiple Adjusted	Item Total Correlation Squared Multiple Adjusted Squared
1	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
2	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
3	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
4	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
5	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
6	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
7	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
8	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
9	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
10	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
11	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
12	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
13	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
14	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
15	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
16	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
17	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
18	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
19	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
20	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07

Gambar 7. Hasil Output Tables 26. Item Correlation Tahap Uji Coba

Item Statistics: Correlation Index

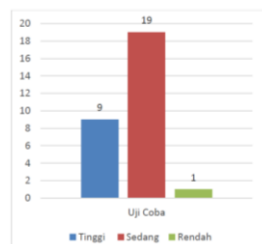
Item	Total	Total Correct	Mean	SD	Item Mean	Item SD	Item Total Correlation	Item Total Correlation Squared	Item Total Correlation Squared Multiple	Item Total Correlation Squared Multiple Adjusted	Item Total Correlation Squared Multiple Adjusted Squared
1	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
2	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
3	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
4	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
5	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
6	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
7	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
8	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
9	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
10	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
11	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
12	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
13	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
14	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
15	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
16	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
17	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
18	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
19	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07
20	20	12	0,60	0,49	0,60	0,49	0,27	0,07	0,07	0,07	0,07

Gambar 8. Hasil Output Tables 26. Item Correlation Tahap Uji Implementasi

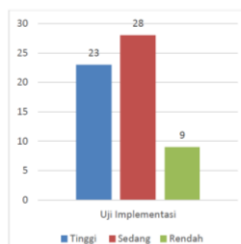
Daya pembeda untuk ditunjukkan dengan tanda kotak hitam. Hasil menunjukkan semua soal memiliki daya pembeda, soal dengan *item measure* pada kategori lemah berjumlah 4 soal, soal dengan daya pembeda kategori cukup terdapat 6 butir, dan 15 menunjukkan kategori baik. Analisis butir soal pada tahap implementasi didapatkan hasil untuk daya pembeda butir setiap soal sesuai dengan Gambar 4.12 yang ditandai dengan kotak hitam. Hasil yang diperoleh pada tahap implementasi menunjukkan 4 soal berada pada kategori lemah dan 5 soal berada pada kategori cukup, 16 soal memiliki daya baik dan tidak ada soal yang termasuk dalam kategori tidak memiliki daya pembeda dan kategori sangat baik.

Profil Keterampilan Berpikir Kritis Peserta Didik

Analisis diukur menggunakan model rasch. Analisis didapatkan melalui *Output Tables 17. Person Measure* yang memperlihatkan urutan *person measure* dari yang terbesar hingga terkecil (Sumintono & Widhiarso, 2015). Berdasarkan hasil analisis *person measure* pada tahap uji coba dan uji implementasi dengan rasch model, didapatkan sebaran profil keterampilan berpikir kritis ditunjukkan pada Grafik 2. Dan Grafik 3.



Grafik 2. Sebaran Kriteria Keterampilan Berpikir Kritis dengan Rasch Model Tahap Uji Coba



Grafik 3. Sebaran Kriteria Keterampilan Berpikir Kritis dengan Rasch Model Tahap Uji Implementasi

4. KESIMPULAN

Kelayakan instrumen tes HOTS untuk mengukur keterampilan berpikir kritis pada materi asam basa dinilai dari berbagai aspek. Penilaian instrumen soal dari validator ahli menunjukkan nilai rata-rata 87,7% menandakan instrumen sangat layak. Analisis butir soal menggunakan rasch model berbantuan *software* Ministep. Analisis pertama dilihat dari *Summary Statistic* pada kriteria *Cronbach alpha* didapatkan nilai 0,76 untuk tahap uji coba yang menunjukkan kategori bagus. Pada tahap uji implementasi didapatkan nilai 0,85 yang menunjukkan kategori bagus sekali. Aspek analisis butir soal *item fit* atau kesesuaian butir soal menunjukkan 100% soal valid berdasarkan kriteria nilai MNSQ, ZSTD, dan Pt Measure Corr. Profil keterampilan berpikir kritis peserta didik berdasarkan *person measure* yang didapat dengan rasch model didapatkan persentase sebesar 36% untuk kategori tinggi, kategori sedang mendapatkan 53%, dan 11% sisanya masuk dalam kategori rendah.

DAFTAR PUSTAKA

- Agustin, N., Susilogati, S., & Addiani, K. (2018). Desain instrumen tes bermuatan etnosains untuk mengukur kemampuan berpikir kritis siswa SMA. *Jurnal Inovasi Pendidikan Kimia*, 12(2), 2159–2169.
- Amalia, N. F., & Susilaningsih, E. (2014). Pengembangan instrumen penilaian keterampilan berpikir kritis siswa SMA pada materi asam basa. *Jurnal Inovasi Pendidikan Kimia*, 8(2), 1380–1389. <https://doi.org/10.24127/bioedukasi.v10i1.2004>
- Astuti, S. R. D. S., Puspita, A. R., & Amelia, R. N. K. (2022). Instrumen penilaian terintegrasi: Pengukuran kemampuan berpikir kritis dan keterampilan sains siswa pada materi larutan elektrolit dan nonelektrolit. *Jurnal Tarbiyah*, 29(2), 340–355.
- Azizah, M., Sulianto, J., & Cintang, N. (2018). Analysis of critical thinking skills of elementary school students in learning mathematics curriculum 2013. *Jurnal Penelitian Pendidikan*, 35(1), 61–70.
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes klasik dan model Rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19. <https://ejournal.undiksha.ac.id/index.php/IJERR/article/view/24080/pdf>
- Fajarini, I., Gafari, M. O. F., & Adisaputera, A. (2019). The development of the assessment's instrument based on higher order thinking skills to measure dimension of persuasive text skills on students of class VIII in Junior High School Muhammadiyah 7 Medan. *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, 2(2), 175–186. <https://doi.org/10.33258/birle.v2i2.289>
- Hallatu, Y. A. (2017). (Retracted) Pengaruh model problem based learning terhadap kompetensi pengetahuan dan keterampilan berpikir kritis siswa Madrasah Aliyah BPD IHA tentang konflik. *The Indonesian Journal of Social Studies*, 1(1), 11. <https://doi.org/10.26740/ijss.v1n1.p11-22>

- Liana, Y. R., Linuwih, S., & Sulhadi, S. (2020). The development of HOTS test instrument of thermodynamics law for senior high school. *Berkala Ilmiah Pendidikan Fisika*, 8(2), 103. <https://doi.org/10.20527/bipf.v8i2.8493>
- Loka Son, A. (2019). Instrumentasi kemampuan pemecahan masalah matematis: Analisis reliabilitas, validitas, tingkat kesukaran dan daya beda butir soal. *Gema Wiralodra*, 10(1), 41–52. <https://doi.org/10.31943/gemawiralodra.v10i1.8>
- Lubis, A. R., Sudrajad, A., & Nugraha, A. W. (2021). Pengembangan instrument assessment higher order thinking skill (HOTS) untuk mengukur kemampuan berpikir tingkat tinggi pada materi hidrolisis garam. *Prosiding Seminar Nasional Kimia Dan Kependidikan Kimia*, 173–181.
- Mari'a, H., & Ismono, I. (2021). Pengaruh implementasi model pembelajaran problem solving dipadukan dengan keterampilan HOTS terhadap hasil belajar peserta didik berdasarkan studi literatur. *UNESA Journal of Chemical Education*, 10(1), 10–19. <https://doi.org/10.26740/ujced.v10n1.p10-19>
- Naufal, L. A., Nurhayati, S., Priatmoko, S., & Wardani, S. (2021). Pengembangan instrumen tes untuk mengukur HOTS peserta didik pada materi laju reaksi. *Journal of Chemistry in Education*, 10(1), 45–52.
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224–233. <http://journal.uny.ac.id/index.php/jipi>
- Pebriani, N. P. I., Putrayasa, I. B., & Margunayasa, I. G. (2021). Pengembangan e-LKPD berbasis HOTS (kemampuan berpikir tingkat tinggi) dengan pendekatan saintifik pada pembelajaran IPA tema 8 kelas V SD. *PENDASI: Jurnal Pendidikan Dasar Indonesia*, 5(2), 301–311.
- Pratiwi, D. P., Risnani, L. Y., & Nofiana, M. (2019). Profil life and career skills siswa sekolah menengah atas (SMA) di Kabupaten Banyumas dalam konteks 21st century skills. *BIOEDUKASI (Jurnal Pendidikan Biologi)*, 10(2), 112. <https://doi.org/10.24127/bioedukasi.v10i2.2460>
- Sumintono, B., & Widhiarso, W. (2015). *Penilaian pendidikan dan ujian: Aplikasi pemodelan Rasch pada assessment pendidikan* (1st ed., pp. 1–4).
- Supanti, S., & Rusmawati, D. I. (2018). Peningkatan hasil belajar kimia materi larutan asam-basa dan keaktifan siswa dengan model pembelajaran IBL berpendekatan SETS. *Jurnal Inovasi Pendidikan Kimia*, 14(2), 2593. <https://journal.unnes.ac.id/nju/index.php/JIPK/article/view/20621>
- Tenti, M., Hairida, & Rahmat, R. (2019). Keterampilan berpikir kritis siswa menggunakan model pembelajaran inkuiri terbimbing pada materi asam basa. *Kemampuan Koneksi Matematis (Tinjauan Terhadap Pendekatan Pembelajaran SAVI)*, 53(9), 1689–1699.
- Yuliantaningrum, L., & Sunarti, T. (2020). Pengembangan instrumen soal HOTS untuk mengukur keterampilan berpikir kritis, berpikir kreatif, dan pemecahan masalah materi gerak lurus pada peserta didik SMA. *Inovasi Pendidikan Fisika*, 09(02), 76–82.