

DEVELOPMENT OF REASONING MULTIPLE CHOICE QUESTIONS (RMCQS) TO ANALYZE HIGHER ORDER THINKING SKILLS

Diah Ayu Saputri¹, Mohammad Masykuri^{2**}, Suryadi Budi Utomo^{3***}

^{1,2,3} Program Studi Pendidikan Kimia, Universitas Sebelas Maret, Surakarta
Jl. Ir. Sutami 36A Surakarta

E-mail: **masykuri@staff.uns.ac.id

Received: 04 Juli 2021. Accepted: 19 November 2022. Published: 31 Desember 2022

DOI: 10.30870/educhemia.v7i1.11791

Abstract. The development of reasoning multiple choice questions (RMCQs) to analyze higher-order thinking skills is based on the need for appropriate assessment instruments to determine high-order thinking skills on learning requirements in the 21st century. The development used the model proposed by Thiagarajan, 4D-models with four process stages: (1) define; (2) design; (3) develop; (4) disseminate. The results were analyzed using the winstep application to obtain the value of validity, reliability, discriminatory power, distractor index, and difficulty level. The instrument has a good reliability coefficient, shown by the item reliability in the final test of 0.98, with the ability to distinguish three groups of samples, low, moderate, and high. 75% of the validity index and 80% of the distractor index function well, so the assessment instrument developed can be used to analyze high-order thinking skills. The assessment instrument is also applicative if used as an evaluation instrument in chemistry learning.

Keywords: assessment instrument, high order thinking, acid-base solution

Abstrak. Pengembangan instrumen asesmen pilihan ganda beralasan untuk mengukur keterampilan berpikir tingkat tinggi pada materi larutan asam basa dilatarbelakangi oleh minimnya ketersediaan instrumen asesmen yang layak untuk mengetahui tingkat keterampilan berpikir tinggi pada tuntutan pembelajaran abad 21. Pengembangan instrumen menggunakan model dari Thiagarajan yaitu 4D (*four-D models*) dengan menjadi 4 tahapan proses, yaitu: (1) *define*; (2) *design*; (3) *develop*; dan (4) *disseminate*. Hasil uji dianalisis menggunakan aplikasi *winstep* untuk mendapatkan nilai validitas, reliabilitas, daya pembeda, indeks pengecoh, dan tingkat kesukaran. Instrumen memiliki koefisien reliabilitas yang sangat baik ditunjukkan dengan nilai *item reliability* pada uji tahap akhir sebesar 0,98 dengan kemampuan membedakan 3 kelompok sampel yaitu sampel dengan abilitas rendah, sedang, dan tinggi. 75% indeks validitas dan 80% indeks pengecoh berfungsi dengan baik, sehingga dapat disimpulkan bahwa instrumen asesmen yang dikembangkan dapat digunakan untuk menganalisis keterampilan berpikir tingkat tinggi. Instrumen asesmen juga aplikatif apabila digunakan sebagai instrumen evaluasi pada pembelajaran Kimia.

Kata Kunci: instrumen asesmen; berpikir tingkat tinggi, larutan asam basa

PENDAHULUAN

Perkembangan dalam bidang sains dan teknologi pada abad 21 mengalami eskalasi yang sangat pesat. Di sisi lain, manusia sebagai pelaku peradaban juga dituntut untuk memiliki sumber daya yang berkualitas (Wijaya *et al.*, 2016). Untuk menciptakan sumber daya manusia yang berkualitas diperlukan suatu pendidikan yang baik. Salah satu ciri keberhasilan pendidikan dalam mencetak sumber daya manusia yang berkualitas adalah dimilikinya keterampilan untuk bertahan hidup (*life skills*). Abad 21 menuntut perubahan pandangan yang sangat besar dalam berbagai sendi kehidupan, salah satunya pendidikan nasional. Pembelajaran abad 21 menurut Kemendikbud (Baroya, 2018) menekankan pada keterampilan yang dimiliki siswa, antara lain keterampilan mencari tahu, menganalisis masalah, berpikir analitis, kolaborasi dalam menyelesaikan masalah atau dengan kata lain disebut sebagai berpikir tingkat tinggi (*higher order thinking*).

Dalam sains, khususnya ilmu kimia, diperlukan keterampilan bernalar dan berpikir tingkat tinggi untuk menganalisis peristiwa yang terjadi di alam. Menurut (Nurwahidah *et al.*, 2020) belajar sains memerlukan latihan dan pengalaman nyata dalam kehidupan. Dalam konteks

tersebut, kimia merupakan ilmu yang sangat relevan untuk mengasah keterampilan berpikir tingkat tinggi (Holbrook, 1974). Pembelajaran kimia mempelajari ilmu kimia untuk mendeskripsikan fakta dalam kehidupan sehari-hari yang bersifat abstrak (Munandar and Jofrisha, 2016). Karakteristik ilmu kimia yaitu: 1) abstrak; 2) penyederhanaan kenyataan; 3) konsektif dan berjenjang (Erlina, 2011). Karakteristik tersebut membentuk perspektif siswa bahwa kimia adalah ilmu yang sulit dipelajari. Strategi yang tepat dan pelatihan soal tingkat tinggi sangat diperlukan untuk meningkatkan keterampilan berpikir agar siswa dengan mudah memahami konsep kimia. Dalam implementasinya, siswa menganggap bahwa kompetensi dasar Larutan Asam Basa adalah materi sulit (Sahlan *et al.*, 2016). Hal ini disebabkan oleh materi yang bersifat abstrak, konseptual. Materi tersebut juga merupakan materi esensial bersifat submikroskopik dan simbolis (Mahfuzah *et al.*, 2018).

Berpikir tingkat tinggi dikelompokkan menjadi 5 kategori, yaitu a) analisis, evaluasi, dan kreasi b) *logic and reasoning* dan 3) *judgement* 4) *problem solving* 5) *creativity and creative thinking* (Brookhart, 2010). Menurut (Shidiq *et al.*, 2015) keterampilan berpikir tingkat tinggi

memiliki beberapa ciri salah satunya yaitu banyaknya variabel dalam berpikir. Sehingga dapat diartikan bahwa keterampilan berpikir tingkat tinggi merupakan aplikasi proses berpikir manusia untuk kondisi kompleks. Dalam penelitian tersebut juga disebutkan pentingnya mendorong siswa agar memiliki keterampilan berpikir tingkat tinggi, dengan cara melakukan pembelajaran yang melatih keterampilan berpikir. Keterampilan berpikir tingkat tinggi (*higher order thinking skills*) dapat diamati kemudian dinilai secara optimal menggunakan instrumen asesmen yang tepat. Dasar suatu penilaian pendidikan di Indonesia dapat dilihat dalam Peraturan Pemerintah Nomor 23 Tahun 2016 bahwa penilaian merupakan suatu proses untuk mendapatkan dan menganalisis data yang sebagai pengukur pencapaian siswa. Kontras dengan hal ini, pengembangan instrumen asesmen keterampilan berpikir tingkat tinggi masih sangat minim sehingga berdampak pada rendahnya kualitas hasil evaluasi.

Asesmen untuk mengukur hasil belajar dan keterampilan berpikir tingkat tinggi dapat berbentuk pilihan ganda, uraian, atau *two tier multiple-choice* yang dikenal dengan pilihan ganda beralasan atau *reasoning multiple choice questions*

(RMCQs). Menurut (Dufresne *et al.*, 2002) soal pilihan ganda atau *multiple-choice questions* (MCQs) merupakan salah satu bentuk soal yang mudah diterapkan dan dapat menghemat waktu dalam penilaiannya. Namun, kelemahannya adalah siswa dapat melakukan penembakan jawaban. Hal ini juga disampaikan oleh (Bush, 2001) bahwa dalam soal pilihan ganda, siswa dapat mencapai nilai tinggi karena menembak jawaban. Dengan empat atau lima pilihan jawaban pada setiap nomor soal, siswa dengan keterampilan rendah masih bisa berharap untuk mendapat skor 25% atau 20% (rata-rata). Salah satu bentuk soal yang dapat meminimalisir adanya penembakan jawaban oleh siswa adalah pilihan ganda beralasan atau *reasoning multiple choice questions (RMCQs)*. Bentuk soal tersebut dapat melatih siswa berpikir kritis, menemukan permasalahan dan solusi disertai alasan yang berdasar serta logis. Disamping itu, soal pilihan ganda beralasan juga merupakan penilaian yang cenderung objektif, tidak subjektif seperti soal uraian. Penggunaan soal pilihan ganda beralasan dinilai efektif sebagai asesmen untuk mengetahui tingkat keterampilan berpikir siswa.

Belum ada penelitian dan pengembangan instrumen asesmen untuk

mengukur keterampilan berpikir tingkat tinggi pada materi Larutan Asam Basa dengan menggunakan bentuk soal pilihan ganda beralasan atau *reasoning multiple choice questions* (RMCQs). Pada penelitian sebelumnya, pengembangan instrumen asesmen untuk mengukur keterampilan berpikir tingkat tinggi pada materi Larutan Asam Basa hanya menggunakan soal pilihan ganda atau *multiple-choice questions* (MCQs) yang menyebabkan instrumen tersebut kurang valid karena masih banyak siswa yang menjawab soal dengan cara menembak.

Dalam penelitian ini, dikaji pengembangan soal pilihan ganda beralasan atau *reasoning multiple choice questions* (RMCQs) pada materi Larutan Asam Basa. Dari sisi konten, Kompetensi Dasar Larutan Asam Basa mempelajari beberapa submateri yang terdiri dari teori asam basa; menentukan sifat larutan; menghitung pH larutan. Bagi siswa Kompetensi Dasar Larutan Asam Basa tergolong sulit untuk dipahami (Juandi, 2019). Hal ini menunjukkan betapa pentingnya peran keterampilan berpikir siswa dalam keberhasilan belajar. Dengan dimilikinya keterampilan berpikir tingkat tinggi yang baik tentu akan memudahkan siswa dalam mempelajari suatu materi yang tergolong sulit. Namun, dalam praktiknya siswa seringkali tidak

dilibatkan dalam membentuk keterampilan berpikir tingkat tinggi, salah satunya tidak adanya suatu latihan untuk mengukur serta mengevaluasi keterampilan berpikir tingkat tinggi. Oleh sebab itu diperlukan instrumen asesmen yang layak untuk menghasilkan penilaian hasil belajar yang objektif serta dapat mengukur keterampilan berpikir tingkat tinggi siswa khususnya mata pelajaran kimia materi Larutan Asam Basa.

METODE

Rancangan Penelitian

Penelitian ini merupakan jenis penelitian *Research and Development* (R&D) dengan model 4D (*four-D models*) yang dicetuskan oleh Thiagarajan dan memiliki 4 tahapan proses, yaitu : (1) *define*; (2) *design*; (3) *develop*; dan (4) *disseminate* (Thiagarajan *et al.*, 1976). Tahap *define* dilakukan kajian literatur dan kajian luas. Tahap *design* ditentukan bentuk instrumen, penyusunan kisi-kisi, serta perancangan instrumen. Proses *develop* terdiri dari validasi oleh praktisi, uji skala kecil, dan menengah. Tahap *disseminate* dilakukan uji skala luas dan analisis menggunakan *Rasch model*.

Sumber Data

Data yang didapat dari penelitian pengembangan asesmen ini terdiri dari data kualitatif dan kuantitatif. Data

kualitatif didapatkan dari validasi oleh ahli, angket yang diberikan pada *testee* dan wawancara guru kimia SMA N 2 Sukoharjo dan SMA Nur Hidayah. Data kuantitatif didapatkan dari hasil uji coba. Uji coba skala kecil 7 responden (5 siswa kelas XII MIPA dan satu guru SMA N 2 Sukoharjo serta 1 guru SMA IT Nur Hidayah). Pada skala menengah *testee* terdiri dari 63 siswa kelas XII dan 2 guru SMA N 2 Sukoharjo. Uji skala luas *testee* terdiri dari 143 siswa kelas XI MIPA SMA N 2 Sukoharjo. Pemilihan sampel dilakukan secara acak atau *simple random sampling*.

Teknik Pengumpulan

Teknik pengumpulan data dalam penelitian ini yaitu: (1) melakukan kajian literatur dan kajian luas; (2) menyusun instrumen (indikator pengembangan instrumen asesmen, kisi-kisi dan lembar soal, lembar validasi, lembar wawancara, angket kebutuhan dan respon siswa, rubrik penilaian); (3) melakukan validasi materi, penilaian, bahasa, dan guru sejawat dari ahli pada bidangnya; (4) revisi instrumen asesmen sesuai masukan ahli; (5) uji coba instrumen asesmen (skala kecil, menengah, luas); (6) analisis ketercapaian instrumen asesmen (validitas, reliabilitas, tingkat kesukatan,

indeks pengecoh, dan daya pembeda butir soal); (7) revisi instrumen asesmen dari hasil analisis uji.

Analisis Data

Analisis Kualitatif

Analisis kualitatif diterapkan saat mengkaji hasil dari validasi ahli (ahli materi, ahli penilaian, ahli bahasa, guru sejawat), angket kebutuhan, respon *testee*, analisis keterpenuhan indikator HOTS pada soal PAS, nilai akhir PTS siswa kelas XI, dan wawancara guru. Data hasil validasi ahli terdiri dari kategori relevan dan tidak relevan dari ahli yang dianalisis dengan menggunakan formula Gregory untuk menghitung *content validity coefficient*.

Analisis Kuantitatif

Analisis kuantitatif dilakukan saat mengkaji hasil uji coba tahap kecil, menengah, dan luas. Aplikasi yang digunakan untuk melakukan analisis secara kuantitatif adalah *Winstep 3.7*. Dalam penelitian ini, analisis kuantitatif dilakukan untuk mengetahui tingkat ketercapaian instrumen asesmen dalam mengukur keterampilan berpikir berdasarkan beberapa indikator yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan indeks pengecoh.

HASIL DAN PEMBAHASAN

Hasil akhir dari penelitian ini yaitu terciptanya instrumen asesmen pilihan ganda beralasan untuk mengukur keterampilan berpikir tingkat tinggi pada materi Larutan Asam Basa. Validasi ahli bertujuan untuk mendapatkan saran dan penilaian dari ahli materi, penilaian, bahasa, dan guru sejawat. Analisis penilaian butir soal HOTS menggunakan formula Gregory untuk mengetahui validitas konstruk.

Tabel 1. *Content Validity* pada Hasil Akhir dari Instrumen Asesmen

Jumlah	Panelis I		Panelis II	
	R	TR	R	TR
40 (20 <i>first tier</i> , 20 <i>second tier</i>)	32	6	31	9
	<i>Vi sebelum revisi</i> = 0,675 (analisis dapat dilanjutkan dengan perbaikan)			
	40	0	40	0
	<i>Vi setelah revisi</i> = 1 (>0,700, analisis dapat dilanjutkan)			

Keterangan: R: Relevan; TR: Tidak Relevan; Vi: Validitas konstruk

Berdasarkan hasil analisis validasi ahli pada Tabel 1, instrumen asesmen yang terdiri atas 20 butir soal *first tier* dan 20 soal *second tier* pada tahap akhir setelah revisi dinyatakan layak digunakan.

Dalam uji tahap kecil didapatkan lama waktu untuk mengerjakan instrumen asesmen, yaitu 100 menit. Informasi kedua yang didapatkan dari analisis butir soal pada uji tahap kecil, menengah, dan luas yaitu kualitas butir soal berdasarkan

indikator instrumen asesmen yang baik (validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan indeks pengecoh).

Validitas Instrumen

Tabel 2. *Item Fitness* Instrumen Asesmen

Tahap Uji	Item tidak sesuai kriteria			Item tidak fit
	ZS TD	MNSQ	Pt-Measure Corr	
Tahap kecil	3	32	17	3
Tahap menengah	4	10	22	4
Tahap luas	12	17	25	10

Berdasarkan Tabel 2, pada setiap tahap pengujian terdapat item yang tidak *fit* ditandai dengan nilai ZSTD, MNSQ, dan Pt-Measure Corr tidak *fit* secara serentak. Menurut Boone et al. (2013), item tidak perlu diganti jika hanya satu atau dua ketentuan yang tidak memenuhi, setidaknya masih ada satu kriteria yang terpenuhi. Berdasarkan data yang didapatkan pada setiap tahap pengujian didapatkan juga informasi bahwa beberapa *testee* tidak *memenuhi fit order*. Hasil perhitungan terhadap *person measure* terdapat dalam Tabel 3.

Tabel 3 *Person Fitness* Instrumen Asesmen

Tahap Uji	ZSTD	MNSQ	Pt-Measure Corr	Person Tidak Fit
Tahap kecil	1	5	2	1
Tahap menengah	9	40	15	7
Tahap luas	12	103	12	10

Berdasarkan Tabel 3, *testee* yang tidak *memenuhi fit order* tidak dihilangkan meskipun dapat memberikan hasil yang kurang baik. Hal ini terjadi karena *testee* tersebut bukan bagian dari instrumen. Selain itu tidak adanya eliminasi *testee* juga akan memberikan respon yang beragam (tidak sesuai dengan model ideal) akibat banyak faktor, sehingga memudahkan dalam perbaikan instrumen.

Reliabilitas Instrumen

Tabel 4 Koefisien Reliabilitas Instrumen Asesmen

Tahap Uji	<i>Item Reliability</i>	<i>Person Reliability</i>	Cronbach Alpha
Tahap kecil	0,52	0,92	0,93
Tahap menengah	0,95	0,85	0,83
Tahap luas	0,98	0,77	0,76

Koefisien reliabilitas hasil uji coba skala kecil yang terdapat dalam Tabel 4 menjelaskan bahwa instrumen asesmen mempunyai kualitas reprodutifitas instrumen yang rendah dan konsistensi *person* dalam menjawab butir-butir soal di dalamnya sangat baik dilihat dari nilai *person reliability* yang tinggi. Interaksi antara instrumen dengan *person* juga terjalin baik dilihat dari tingginya nilai cronbach alpha. Buruknya reprodutifitas instrumen mengharuskan peneliti untuk memperbaiki instrumen asesmen tersebut

dan kembali melakukan uji untuk skala menengah.

Hasil uji coba skala menengah terhadap instrumen asesmen yang diperbaiki setelah uji coba tahap kecil menunjukkan bahwa instrumen mempunyai kualitas reprodutifitas item yang sangat baik dilihat dari nilai *item reliability* >0,90 dengan konsistensi *person* dalam menjawab butir-butir soal di dalamnya tergolong baik, sehingga memberikan interaksi yang baik antara instrumen dengan *person*.

Pada uji skala luas menunjukkan hasil bahwa instrumen asesmen yang diperbaiki memiliki kualitas reprodutifitas yang tinggi diimbangi dengan kemampuan *person* dalam menjawab butir soal cukup konsisten. Hal ini menghasilkan interaksi cukup baik antara instrumen dengan *person*.

Dalam penelitian ini, reliabilitas instrumen juga dilihat dari nilai *separation index*. Berdasarkan hasil setiap tahap uji coba didapatkan nilai *separation index* yang terdapat dalam Tabel 5.

Tabel 5 *Separation Index* Instrumen Asesmen

Pelaksanaan Pengujian	<i>Item Separation</i>	H item	<i>Person Separation</i>	H person
Tahap kecil	1,03	1,71	3,43	4,91
Tahap menengah	4,26	6,01	2,37	3,49
Tahap luas	7,10	9,80	1,85	2,80

Berdasarkan Tabel 5 nilai *item separation* menunjukkan besarnya sampel yang tersebar sepanjang skala interval linear. Menurut (Widiyawati *et al.*, 2021) nilai indeks separasi yang masih rendah merupakan indikasi bahwa siswa belum terbiasa dengan jenis instrumen atau item yang terlalu sulit. Indeks *item separation* yang telah melebihi 2.00 merupakan indikasi bahwa siswa atau responden memiliki kemampuan yang beragam. *Person separation* pada Tabel 5 menunjukkan kemampuan instrumen dalam mengelompokkan sampel sesuai dengan kemampuannya.

Berdasarkan Tabel 5, terlihat bahwa pada uji tahap kecil instrumen asesemen yang disusun memiliki kemampuan untuk mengelompokkan sampel menjadi 5 kelompok dilihat dari nilai *H person* (sangat rendah, rendah, sedang/menengah, tinggi, dan sangat tinggi). Uji tahap menengah digunakan untuk menguji instrumen asesemen yang telah diperbaiki dari uji tahap sebelumnya, yaitu uji tahap kecil. Hasil uji tahap menengah dijelaskan bahwa instrumen memiliki kemampuan untuk mengelompokkan sampel menjadi 3 kelompok dilihat dari nilai *H person* (abilitas rendah, sedang/menengah, dan tinggi). Uji tahap luas bertujuan untuk menguji instrumen asesmen yang telah

direvisi dari uji sebelumnya, yaitu pada uji tahap menengah. Uji tahap luas menunjukkan hasil bahwa instrumen asesemen memiliki kemampuan untuk mengelompokkan sampel menjadi 3 kelompok dilihat dari nilai *H person* (abilitas rendah, sedang/menengah, dan tinggi).

Tabel 6 *Average Ability* Berdasarkan Hasil Pengujian Instrumen

Tahap Uji	Nomor item yang tidak memenuhi	Jumlah Item tidak memenuhi
Tahap kecil	5,6,8,21,24, 31, 33, 34, 40	9
Tahap menengah	4,7, 13, 14, 20, 26, 27	7
Tahap luas	3, 4, 17, 19, 20,22, 30,39	8

Dalam Tabel 6 terdapat beberapa item dengan *average ability* yang tidak memenuhi. Hal ini berarti distraktor pada nomor yang tidak memenuhi *average ability* kurang berfungsi dengan baik. Pada uji tahap akhir, terdapat 80% item dengan distraktor yang berfungsi dengan baik.

Tabel 7 Presentase Tingkat Kesukaran

Kategori	Presentase Tingkat Kesukaran (%)		
	Uji Skala Kecil	Uji Skala Besar	Uji Skala Luas
Sangat Mudah	35	30	15
Mudah	17,5	17,5	35
Sulit	15	17,5	20
Sangat Sulit	32,5	35	30

Tingkat kesukaran butir instrumen dapat dilihat dari respon *testee*. Dalam

penelitian ini respon *testee* dibagi menjadi 4 kelompok yaitu sangat mudah, mudah, sulit, dan sangat sulit. Hasil presentase pada setiap kelompok *testee* dihitung pada uji skala luas yaitu 15% untuk *testee* yang menganggap bahwa instrumen asesmen yang dikembangkan sangat mudah, 35% menganggap instrumen asesmen yang dikembangkan mudah, 20% *testee* menganggap bahwa instrumen asesmen yang dikembangkan sulit, dan 30% *testee* menganggap bahwa keseluruhan butir instrumen asesmen yang dikembangkan sangat sulit.

Pada setiap hasil uji coba yang dilakukan pada tabel 4, presentase *testee* yang menganggap bahwa butir instrumen sangat sulit untuk diselesaikan selalu mendominasi. Hal tersebut menunjukkan bahwa masih rendahnya keterampilan berpikir tingkat tinggi yang dimiliki oleh peserta didik.

KESIMPULAN

Pengembangan instrumen asesmen pilihan ganda beralasan atau *reasoning multiple choice questions (RMCQs)*

DAFTAR RUJUKAN

Baroya, E. H. (2018) "Strategi Pembelajaran Abad 21," *Jurnal Lembaga Penjaminan Mutu*

dilakukan menggunakan model Thiagarajan yaitu 4D (*four-D models*) dengan tahapan : (1) *define*; (2) *design*; (3) *develop*; dan (4) *disseminate*. Hasil uji dianalisis menggunakan aplikasi *winstep* untuk mendapatkan nilai validitas, reliabilitas, daya pembeda, indeks pengecoh, dan tingkat kesukaran.

Ketercapaian indeks pengecoh mencapai 80%. Validitas instrumen pada uji skala tahap luas mencapai 75%. Reliabilitas instrumen sangat baik ditunjukkan dengan nilai *item reliability* 0,98 dengan kemampuan membedakan 3 kelompok sampel (rendah, sedang, dan tinggi). Pada setiap tahap uji coba, tingkat kesukaran instrumen selalu tinggi ditandai dengan rata-rata respon *testee* yang memilih sangat sulit pada setiap uji mencapai 32,5%. Hasil uji pengembangan instrumen asesmen menunjukkan bahwa instrumen asesmen layak digunakan untuk menganalisis keterampilan berpikir tingkat tinggi dan aplikatif apabila digunakan sebagai evaluasi dalam materi Larutan Asam Basa.

Pendidikan Prov. DIYogyakarta, I(01), hal. 101–115.

Boone, W. J., Yale, M. S. dan Staver, J. R. (2014) *Rasch analysis in the human*

- sciences, Rasch Analysis in the Human Sciences*. doi: 10.1007/978-94-007-6857-4.
- Brookhart, S. M. (2010) *How to Assess Higher-Order Thinking Skills in Your Classroom, Journal of Education*. doi: 10.1177/002205741808801819.
- Bush, M. (2001) 'A Multiple Choice Test that Rewards Partial Knowledge', *Journal of Further and Higher Education*, 25(2), pp. 157–163.
- Dufresne, R. J., Leonard, W. J. and Gerace, W. J. (2002) 'Making Sense of Students' Answers Questions', *The Physics Teacher*, 40(March), pp. 174–180.
- Erlina (2011) "Deskripsi Kemampuan Berpikir Formal Mahasiswa Pendidikan Kimia Universitas Tanjungpura," *Jurnal Visi Ilmu Pendidikan*, 6(3), hal. 631–640.
- Holbrook, J. (1974) "Making chemistry relevant," *Journal of Chemical Education*, 51(1), hal. 45. doi: 10.1021/ed051p45.
- Juandi, A. (2019) "Salinan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016 tentang Standar Penilaian Pendidikan," 12, hal. 2–4. doi: 10.31227/osf.io/munp2.
- Mahfuzah, B. A., Utomo, Y. dan Munzil (2018) "Efektivitas GDL (Guided Discovery Learning) dan Problem Solving terhadap KBK (Keterampilan Berpikir Kritis) dan HOTS (Higher Order Thinking Skills)," *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 3(6), hal. 739–744.
- Munandar, H. dan Jofrisha (2016) "Analisis Pelaksanaan Pembelajaran Kimia di Kelas Homogen," *Lantanida Journal*, 4(1), hal. 98–110.
- Nurwahidah, I. *et al.* (2020) 'Development of Science Test to Measure HOTS and Digital Literacy of Junior High School Students on The Topic of City Noise', 12(2), pp. 203–213.
- Sahlan, F., Haetami, A. dan Marhadi, M. A. (2016) "Identifikasi Kesulitan Belajar Siswa Pada Materi," *Jurnal Pendidikan Kimia FKIP Universitas Halu Oleo*, 5(7), hal. 58–69.
- Shidiq, A. S., Masykuri, M. and Susanti, E. (2015) 'Analisis Higher Order Thinking Skills (HOTS) Menggunakan Instrumen Two-Tier Multiple Choice Pada Materi Kelarutan Dan Hasil Kali Kelarutan Untuk Siswa Kelas Xi Sma N 1 Surakarta', *Prosiding Seminar Nasional Pendidikan Sains*, (November), pp. 159–166.
- Thiagarajan, S., Dorothy S, S. dan Semmel, M. I. (1976) "Instructional development for training teachers of

- exceptional children: A sourcebook,” *Journal of School Psychology*, 14(1), hal. 75. doi: 10.1016/0022-4405(76)90066-2.
- Widiyawati, Y. *et al.* (2021) ‘The 21 st century science learning: HOTS and digital literacy among junior high school students in Semarang, Indonesia’, *Journal of Physics: Conference Series*, 1842(1). doi: 10.1088/1742-6596/1842/1/012081.
- Wijaya;, E. Y., Sudjimat;, D. A. dan Nyoto, A. (2016) “Transformasi Pendidikan Abad 21 Sebagai Tuntutan,” *Jurnal pendidikan*, 1, hal. 263–278. Tersedia pada: <http://repository.unikama.ac.id/840/32/263-278> Transformasi Pendidikan Abad 21 Sebagai Tuntutan Pengembangan Sumber Daya Manusia di Era Global .pdf.