



Validitas dan Realibilitas Instrumen Tes *Two Tier* Berbasis *Higher Order Thinking Skills* pada Materi Sifat Koligatif Larutan untuk Siswa SMA/MA

Syaza Syahana^{1✉}, Andromeda²

Universitas Negeri Padang, Indonesia^{1,2}

E-mail : syazasyahana1@gmail.com¹, andromedasaidir@yahoo.com²

Abstrak

Kompetensi abad ke-21 menuntut siswa untuk dapat meningkatkan kemampuan berpikir tingkat tinggi dengan adanya ketersediaan instrumen tes berbasis HOTS. Penelitian ini bertujuan untuk menghasilkan instrumen tes berbasis *Higher Order Thinking Skills* (HOTS) dalam bentuk *two tier* yang valid dan reliabel pada materi sifat koligatif larutan. Jenis penelitian yang digunakan adalah *research and development* (R&D) dengan model pengembangan plomp yang terdiri dari tiga tahapan yaitu *preliminary research*, *prototyping stage* dan *assessment phase* serta teknik analisis data pada validitas konten menggunakan metode CVR dan validitas konstruk menggunakan Aiken V. Berdasarkan hasil penelitian nilai CVR dari validitas konten yang terdiri dari stimulus soal, pertanyaan soal (*first tier*), jawaban soal dan alasan soal (*second tier*) yaitu 1 dan validitas konstruk dari aspek materi, penyajian, bahasa dan aturan tambahan berurut-turut 95.17%, 94,11%, 98.67% dan 99.78%. Instrumen tes *two tier* memiliki tingkat realibilitas pada *first tier* 0.82 dan *second tier* 0.91. Data yang diperoleh dari instrumen tes *two tier* berbasis *Higher Order Thinking Skills* (HOTS) pada materi sifat koligatif larutan sudah valid dan memiliki realibilitas item soal yang sangat tinggi.

Kata Kunci: instrumen tes, *higher order thinking skills*, *two tier*, sifat koligatif larutan, plomp.

Abstract

The demands of the 21st century enhance the thinking ability of the students to have basic competence in higher order thinking through the availability of HOTS test instrument bases. The purpose of the research was develop the test instrument of higher order thinking skills (HOTS) in two tier test which have the validity and reliability in colligative properties matter. The type of the research is research and development (R&D) using plomp model. In the plomp model, there are three phase namely preliminary research, prototyping phase and assessment phase. The analysis technique of content validity and construct validity with CVR and Aiken V methode. Based on the research the value of the content validity is 1 and the construsct validity in the matter, presentation, language, and additional rules are 95.17%, 94.11%, 98.67% and 99.78%. The reliability of the test instrument in first tier is 0.82 and the second tier is 0.91. The data obtained from instrument test two tier based higher order thinking skills (HOTS) in colligative properties is valid and verh high reliability.

Keywords: test instrument, *higher order thinking skills*, *two tier*, colligative properties, plomp.

PENDAHULUAN

Tantangan terbesar pendidikan pada abad ke-21 yaitu memperbaiki kemampuan siswa dalam meningkatkan kemampuan berpikir. Siswa bukan hanya dituntut untuk dapat meningkatkan kemampuan akademik, berpengetahuan luas dan mandiri tetapi siswa juga dituntut untuk kreatif, inovatif, efektif dan mempunyai kemampuan dalam menyelesaikan suatu masalah (Wiyarsi et al., 2019). Kurikulum yang digunakan dalam membekali kompetensi dasar siswa untuk mampu berpikir tingkat tinggi adalah kurikulum 2013.

Kurikulum 2013 membekali siswa pada empat kompetensi yaitu: *critical thinking*, *creativity*, *collaboration* dan *communicatin* (Putriani, 2021). HOTS (*Higher Order Thinking Skills*) merupakan kemampuan berpikir tingkat tinggi dalam menyelesaikan suatu masalah untuk dapat berpikir secara analisis, kritis dan kreatif serta mendukung untuk membekali siswa pada empat kompetensi yang harus dimiliki pada kurikulum 2013 (Fitriyana et al., 2019).

Dalam *Taxonomy Bloom* yang terdapat dalam (Anderson et al., 2001) dimensi proses berpikir terdiri atas kemampuan mengingat (*remembering-C1*), memahami (*understanding-C2*), menerapkan (*applying-C3*), menganalisis (*analysing-4*), mengevaluasi (*evaluating-C5*), dan mencipta (*creating-C6*). Untuk mengukur kemampuan berpikir tingkat tinggi (HOTS) dimensi proses berpikir terdiri atas kemampuan menganalisis (*analysing-4*), mengevaluasi (*evaluating-C5*), dan mencipta (*creating-C6*) (Mujib & Rasyid, 2019).

Lemahnya kemampuan siswa Indonesia untuk memecahkan masalah pada soal level tinggi karena siswa Indonesia terbiasa dengan soal level rendah. Hal ini dibuktikan dari laporan hasil PISA pada tahun 2015 dan 2018 yang menyatakan terjadinya penurunan kemampuan matematika dari 386 menjadi 379 dan kemampuan kinerja sains dari 403 menjadi 380 (Suhady & Roza, 2020). Sehingga diperlukan adanya upaya pembaharuan dan inovasi untuk dapat meningkatkan kemampuan siswa dalam meningkatkan kemampuan berpikir tingkat tinggi (*High Order Thinking Skill*) dengan mengembangkan instrumen tes dalam bentuk *two tier*.

Berdasarkan hasil angket yang disebarakan terhadap 8 (delapan) orang guru kimia di beberapa kabupaten Sumatera Barat yaitu Pasaman Barat, Payakumbuh, Lubuk Alung dan Kota Batam, diperoleh hasil bahwa lebih dari 50% guru menggunakan alat evaluasi berbasis LOTS (*Lower Order Thinking Skill*). Hal ini disebabkan karena dalam instrumen evaluasi yang digunakan oleh guru bersumber dari internet, buku cetak, kumpulan soal maupun LKS. Berdasarkan hasil angket tersebut sebanyak 87.5% guru memerlukan instrumen tes berbasis HOTS salah satunya pada materi sifat koligatif larutan.

Instrumen tes *two tier* berupa soal pilihan ganda (*multiple choice*) yang terdiri dari dua tingkat yaitu pertanyaan soal (*first tier*) dan alasan soal (*second tier*). Instrumen tes *two tier* dapat mengukur sejauh mana kemampuan siswa dalam berpikir tingkat tinggi melalui alasan yang sesuai dengan jawaban yang diberikan. Penelitian mengenai pengembangan instrumen tes berbasis *Higher Order Thinking Skills* (HOTS) sudah banyak dikembangkan, seperti pada penelitian (Hutapea & Sudrajat, 2019) mengenai pengembangan instrumen penilaian berbasis HOTS pada pembelajaran sifat koligatif larutan dalam bentuk pilihan ganda. Selanjutnya penelitian (Sarah et al., 2021) mengenai *The Development Higher Order Thinking Skills (HOTS) as a Question in Chemistry* dalam bentuk *multiple choice*, dan pada penelitian (Hartono & Pahlevi, 2020) mengenai *Development Of Assessment Instruments Based on Hots At Surabaya Vocational High School* dalam bentuk *multiple choice*. Pada penelitian ini pengembangan instrumen tes dalam bentuk *two tier multiple choice* berbasis *Higher Order Thinking Skills* (HOTS) yang memiliki kriteria pengembangan soal berdasarkan analisis KD yang berada pada level C4 atau menganalisis.

Tujuan dari penelitian ini adalah untuk menghasilkan instrumen tes berbasis *Higher Order Thinking Skills* (HOTS) pada materi sifat koligatif larutan yang valid dan reliabel agar ketersediaan instrumen tes pada materi terkait dapat memenuhi kompetensi dasar.

METODE

Jenis penelitian yang digunakan adalah penelitian dan pengembangan atau *Research and Development* (R&D). Sugiyono (2015) menyatakan bahwa penelitian ini digunakan untuk menghasilkan suatu produk tertentu dan menguji keefektifan dari suatu produk. Model pengembangan yang digunakan dalam penelitian ini adalah model Plomp. Model plomp disebut juga dengan *design research*. Plomp (2007) menyatakan bahwa tahapan dalam penelitian ini terdiri dari tiga tahapan yaitu: *preliminary research*, *prototyping stage*, dan *assessment phase*.

Penelitian dilakukan di SMA N 1 Lembah Melintang, Pasaman Barat pada semester genap Januari - Juni tahun pelajaran 2021/2022. Subjek penelitian ini adalah dosen Jurusan Kimia FMIPA UNP dan guru Kimia SMA di Pasaman Barat dan objek penelitian ini adalah kumpulan instrumen tes *Higher Order Thinking Skills* (HOTS) pada materi pokok Sifat koligatif larutan untuk siswa SMA/MA.

Dalam penelitian pengembangan soal HOTS pada materi sifat koligatif larutan digunakan instrumen untuk mengumpulkan data pada penelitian ini yaitu : lembar angket wawancara, angket validasi dan tes. Pada lembar angket validasi isi digunakan metode *content validity ratio* (CVR) dengan pilihan jawaban terdiri dari 2 kategori yaitu: iya dan tidak. Informasi ini dapat dilihat pada tabel 1.

Tabel 1. Nilai Angket Validitas Isi

Nilai	Kategori
1	Iya
0	Tidak

Nilai dari CVR berkisar antara -1 sampai +1, jika nilai validasi konten menunjukkan negatif maka item dari soal tersebut tidak esensial, sedangkan jika nilai validasi konten menunjukkan positif maka item soal tersebut dinyatakan esensial. Nilai dari CVR juga berdasarkan pada jumlah validator dan batas kritis atau minimal CVR untuk dinyatakan bahwa item tersebut dinyatakan valid, namun jika nilai CVR berada di bawah batas kritis nilai CVR maka item soal tersebut dinyatakan tidak valid. Pada penentuan validitas isi instrumen ini, menggunakan 5 orang validator sehingga batas minimum dari nilai CVR yaitu 0.736 (Nengsih et al., 2019). Formula yang diajukan oleh Lawshe untuk menganalisis validitas konten yaitu :

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}}$$

Keterangan :

CVR = *Content Validity Ratio*

ne = Jumlah ahli yang setuju

N = Jumlah ahli yang memvalidasi

(Lawshe, 1985)

Analisis validitas konstruk diukur dengan melakukan analisis terhadap penyajian item tes yang terdiri dari aspek materi, penyajian, aspek penggunaan bahasa dan aturan tambahan. Analisis validitas konstruk dari data yang dihasilkan menggunakan Aiken V, dengan rumus sebagai berikut :

$$V = \frac{\sum s}{n(c - 1)}$$
$$s = r - l_o$$

Keterangan :

V = Indeks kesepakatan rater.

l_o = Angka penilaian validitas terendah
(dalam hal ini = 1)

c = Angka penilaian validitas tertinggi

(dalam hal ini = 4)

Angket uji validitas konstruk menggunakan indeks validitas Aiken disusun menggunakan empat tingkatan yang terdiri dari : Tidak Relevan (TR), kurang relevan (KR), cukup relevan (CR) dan relevan (R) yang kemudian diolah dan diinterpretasikan sesuai dengan skala yang terdapat pada Aiken V. Informasi ini dapat dilihat pada tabel 2.

Tabel 2. Interpretasi Nilai Sesuai Dengan Skala Aiken V

Interval	Kesepakatan	Kategori
0.0 – 0.2	Sangat Rendah	Sangat Tidak Valid
0.21 – 0.4	Rendah	Tidak Valid
0.41 – 0.6	Cukup Tinggi	Kurang Valid
0.61 – 0.8	Tinggi	Valid
0.81 – 0.1	Sangat Tinggi	Sangat Valid

(Marlisyah et al., 2018)

Setelah dilakukan validasi selanjutnya instrumen tes diujikan dalam kelompok kecil (*small group*) kepada 20 orang siswa SMA N 1 Lembah Melintang. Kemudian hasil dari uji coba yang dilakukan secara *online* melalui *google form*, selanjutnya dilakukan analisis butir soal yang terdiri dari realibilitas soal, daya beda, dan indeks kesukaran. Analisis butir soal menggunakan *software* ANATES 4.0.9. Adapun keunggulan dari *Software* anates yaitu menggunakan bahasa indonesia, mampu menampilkan berbagai fitur dan analisis perhitungan yang terdiri dari realibilitas data, skor data dibobot, kelompok atas dan kelompok bawah. Tingkat kesukran butir soal, daya beda butir soal dan indeks korelasi dari skor item dengan skor total (Arif et al., 2014).

HASIL DAN PEMBAHASAN

Pengembangan instrumen tes *two tier* berbasis *Higher Order Thinking Skills* (HOTS) pada materi sifat koligatif larutan terdapat tiga tahapan yaitu *preliminary research*, *prototyping phase* and *assessment phase*. Adapun langkah *preliminary research* dibagi menjadi 4 bagian yaitu:

1) Analisis Kebutuhan

Pada analisis kebutuhan dilakukan penyebaran angket menggunakan *google form* dengan tujuh orang guru mengenai pengembangan instrumen tes berbasis *Higher Order Thinking Skills* (HOTS) di beberapa SMA di Sumatera Barat dan satu SMA di kota Batam. Berdasarkan angket yang disebarakan mengungkapkan bahwa sekolah tersebut secara keseluruhan telah menerapkan kurikulum 2013, namun dalam proses pembelajaran dan jenis instrumen tes yang digunakan masih berjenis *Lower Order Thinking Skills* (LOTS). Berdasarkan hasil angket yang disebarakan, salah satu penyebab jenis soal yang digunakan dalam bentuk LOTS yaitu kurangnya ketersediaan instrumen tes berjenis HOTS salah satunya pada materi sifat koligatif larutan.

Analisis Konteks

Proses analisis konteks dilakukan dengan cara melihat kurikulum yang digunakan di sekolah yaitu kurikulum 2013. Setelah itu dilakukan analisis terhadap kurikulum 2013 seperti yang terdapat pada Permen Nomor 37 tahun 2018 mengenai materi yang memiliki persyaratan dalam mengembangkan instrumen tes *Higher Order Thinking Skills* (HOTS), salah satunya pada materi sifat koligatif larutan yang berada pada tingkat menganalisis atau C4. Adapun kompetensi dasar dan indikator pencapaian kompetensi dari sifat koligatif larutan adalah sebagai berikut :1. Menganalisis fenomena sifat koligatif larutan (penurunan tekanan uap larutan, kenaikan titik didih larutan, penurunan titik beku dan tekanan osmotik larutan).

Studi Literatur

Untuk mendukung proses pengembangan instrumen tes *two tier* berbasis *Higher Order Thinking Skills* (HOTS) maka dilakukan studi literatur terkait untuk mengembangkan instrumen tes berupa buku-buku, jurnal terkait dan sumber lain seperti internet.

Pengembangan Kerangka Konseptual

Pada tahap ini dilakukan rancangan awal mengenai pengembangan instrumen tes *two tier multiple choice* yang digunakan setelah melalui tahap studi literatur. Langkah awal dari pengembangan kerangka konseptual adalah melalui penurunan indikator pencapaian kompetensi menjadi indikator soal yang diturunkan dari kompetensi dasar dari sifat koligatif larutan. Setelah dilakukan penyusunan indikator soal, maka dilakukan penyusunan kisi-kisi soal. Pada instrumen tes yang dikembangkan terdapat 30 soal yang dikembangkan.

Langkah kedua dalam pengembangan instrumen tes *two tier* yaitu *prototyping phase* yang dibagi menjadi 4 prototipe dalam pembentukan instrumen tes yaitu:

Prototipe 1

Pengembangan prototipe 1 dimulai dengan merancang soal berdasarkan kisi-kisi soal yang telah diturunkan dari indikator pencapaian kompetensi dan kompetensi dasar sampai kepada tahap pembuatan desain dari cover soal. Adapun langkah dari pembentukan prototipe 1 yaitu: Pertama, penyusunan stimulus soal pada materi sifat koligatif larutan berupa fenomena-fenomena dalam kehidupan sehari-hari yang berkaitan dengan sifat koligatif larutan, grafik, tabel, data, dan juga gambar yang dapat membantu siswa dalam memahami pertanyaan soal yang diberikan. Kedua, penyusunan pertanyaan soal (*first tier*) dan alasan soal (*second tier*) berupa pertanyaan *multiple choice* yang terdiri dari 5 *option*. Ketiga, pembuatan kunci jawaban soal yang dilakukan dengan membuat langkah berpikir dari C1 yaitu mengingat hingga mencapai ke tahap C4 yaitu menganalisis. Keempat, pembuatan identitas soal memuat kompetensi dasar, indikator pencapaian kompetensi, materi, indikator soal dan level kognitif yang terdapat pada setiap item soal. Kelima, mendesain cover soal menggunakan aplikasi *canva* terdiri dari judul dari materi instrumen tes yang akan dikembangkan, logo UNP, tahun penulisan, nama penulis dan juga nama pembimbing.

Prototipe 2

Pada tahap pembuatan prototipe 2 dilakukan *self evaluation* terkait pengembangan instrumen tes. Hal yang dilakukan pada *self evaluation* terdiri dari penggunaan bahasa dari instrumen tes yang dapat menimbulkan makna ganda sehingga menyebabkan kurangnya pemahaman siswa dalam memahami soal, kesesuaian antara indikator soal dengan indikator pencapaian kompetensi, kelengkapan dari cover dari instrumen tes, identitas pemilik, kata pengantar, daftar isi, petunjuk penggunaan instrumen tes, level kognitif yang terdapat pada kartu soal, stimulus pada soal, dan kunci jawaban dari soal.

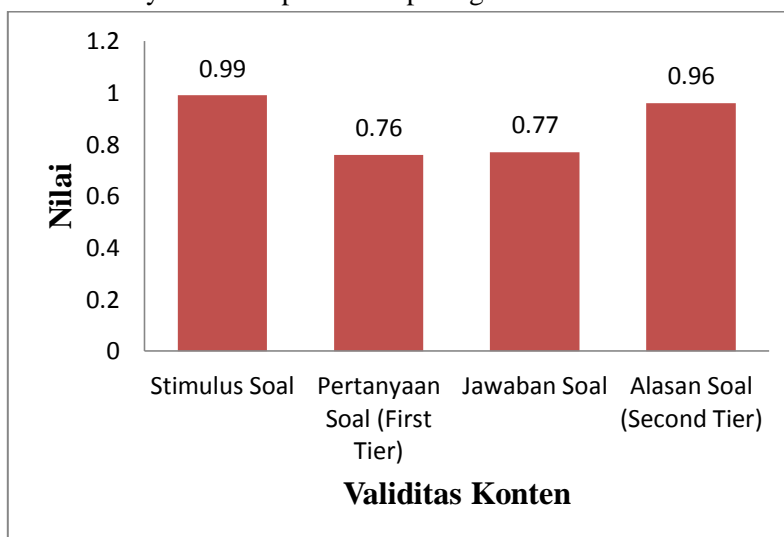
Prototipe 3

Pada tahap pembentukan prototipe 3 dilakukan 2 tahapan yaitu *one to one evaluation* dan *expert review*. Pada tahap *one to one evaluation* dilakukan uji coba kepada 3 orang siswa terkait instrumen tes *two tier multiple choice* berbasis *higher order thinking skills* (HOTS) pada materi sifat koligatif larutan. Setelah dilakukannya uji coba terhadap 3 orang siswa maka diberikan angket secara langsung untuk menanyakan hal terkait instrumen tes.

Berdasarkan hasil uji *one to one evaluation* terhadap 3 orang siswa SMA N 1 Lembah Melintang dari aspek bahasa memiliki bahasa yang mudah dimengerti, petunjuk penggunaan soal yang mudah dipahami, waktu pengerjaan pada instrumen tes yang cukup sebanyak 90 menit. Kemudian data, tabel, gambar serta grafik yang mudah dipahami dan membantu dalam pengerjaan soal. Kendala yang dihadapi siswa yaitu soal yang memiliki kalimat yang panjang, hal ini disebabkan karena berdasarkan KD dan IPK dari soal yang menjelaskan mengenai fenomena-fenomena, sehingga soal disusun dan dirancang berdasarkan fenomena yang terjadi.

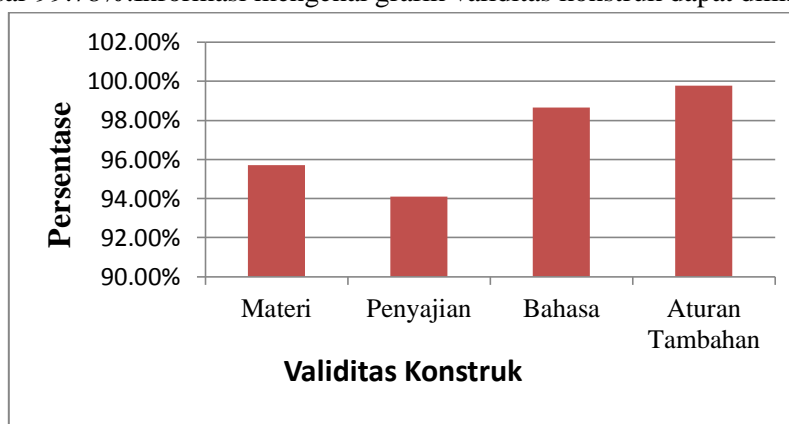
Pada tahap *expert review* dilakukan uji validitas konten dan validitas konstruk terhadap instrumen tes yang dikembangkan. Analisis dari instrumen tes menggunakan angket, pada analisis data validitas konten menggunakan metode CVR sedangkan pada analisis konstruk menggunakan Aiken V. Jumlah validator dalam analisis pengembangan instrumen tes terdiri dari 5 orang validator yaitu 3 orang dosen FMIPA UNP Universitas Negeri Padang dan 2 orang guru SMA N 1 Lembah Melintang.

Pada validitas konten terdapat 4 bentuk penilaian yaitu stimulus soal, pertanyaan soal (*first tier*), jawaban soal, dan alasan soal (*second tier*). Berdasarkan hasil validasi pertama instrumen tes terhadap 5 orang validator dari 30 butir soal terdapat 9 soal yang valid yaitu nomor 1, 3, 10, 17, 19, 24, 25, 27, dan 28 serta 21 soal belum valid. Hal ini disebabkan karena soal berada pada level C3 yaitu menghitung dan kurangnya stimulus yang diberikan pada pertanyaan soal sehingga perlu dilakukan revisi. Berdasarkan rata-rata analisis validitas konten sebelum dilakukannya revisi dapat dilihat pada grafik di bawah ini:



Grafik 1. Rata- Rata Validitas Konten Sebelum Revisi

Setelah dilakukan revisi pada validitas konten maka didapatkan rata- rata nilai CVR dari instrumen tes yaitu 1 yang terdiri dari stimulus soal, pertanyaan soal (*first tier*), jawaban soal, dan alasan soal (*second tier*). Pada validitas konstruk dengan menggunakan Aiken V terdapat 4 aspek yang dilakukan penilaian yaitu materi, penyajian, bahasa, dan aturan tambahan dari instrumen tes yang dikembangkan. Berdasarkan kepada nilai dari validitas konstruk materi memiliki persentase 95.71%, penyajian sebesar 94.11%, bahasa 98.67% dan aturan tambahan sebesar 99.78%. Informasi mengenai grafik validitas konstruk dapat dilihat pada grafik 1.



Grafik 2. Hasil Validitas Konstruk menggunakan Aiken V

Pada tahap pembentukan prototipe 4 dilakukan uji coba skala kecil (*small group*) terhadap 20 siswa untuk menganalisis realibilitas, daya beda, dan tingkat kesukaran dari instrumen tes *two tier multiple choice* pada materi sifat koligatif larutan. Analisis butir soal dianalisis menggunakan aplikasi *anates* pada setiap butir soal yang dibagi menjadi dua bentuk analisis yaitu pertanyaan soal (*first tier*) dan alasan soal (*second tier*).

Berdasarkan analisis butir soal dengan menggunakan aplikasi *anates* dalam menganalisis realibilitas butir soal dibagi menjadi dua bagian yaitu: analisis butir soal pada pertanyaan soal (*first tier*) dan alasan soal (*second tier*). Adapun hasil analisis pada pertanyaan soal (*first tier*) menunjukkan nilai realibilitas adalah 0.82 sedangkan pada alasan soal (*second tier*) adalah 0.91. Pada pertanyaan soal (*first tier*) dan alasan soal (*second tier*) memiliki realibilitas sangat tinggi. Suatu instrumen dapat dikatakan reliabel bila memiliki nilai realibilitas di atas 0.7 yang artinya instrumen tes ini memiliki konsistensi yang sangat tinggi.

Tingkat kesukaran pada pertanyaan soal (*first tier*) berada pada kategori sedang berjumlah 6 soal, soal yang berada pada kategori sukar berjumlah 15 soal, dan soal yang berada pada kategori sangat sukar berjumlah 9 soal. Hasil analisis tingkat kesukaran dari butir soal pada bagian alasan soal (*second tier*) menggunakan aplikasi *anates* menunjukkan bahwa jumlah soal yang berada pada kategori sedang berjumlah 10 soal, soal yang berada pada kategori sukar berjumlah 16 soal, dan butir soal yang berada pada kategori sangat sukar berjumlah 4 soal, sehingga butir soal pada instrumen tes *two tier* menunjukkan bahwa tingkat kesukaran soal dapat berfungsi dengan baik.

Analisis pada daya beda bertujuan untuk mengukur sejauhmana butir soal tersebut dapat membedakan antara siswa yang memiliki penguasaan terhadap suatu materi dengan siswa yang belum atau kurang menguasai suatu materi dengan kriteria tertentu. Berdasarkan koefisien dari daya beda terhadap item soal menyatakan bahwa semakin tinggi daya beda terhadap item soal, maka semakin tinggi item soal tersebut dapat membedakan penguasaan kemampuan siswa terhadap suatu materi (Arifin, 2013). Butir soal dapat memiliki daya beda yang baik apabila daya beda tersebut bernilai positif, namun apabila daya beda dari instrumen tes bernilai negatif berarti kelompok bawah lebih banyak menjawab benar daripada kelompok atas. Berdasarkan hasil analisis butir soal pada *first tier* dan *second tier* instrumen tes memiliki daya beda soal yang efektif.

Analisis 30 butir daya beda soal dianalisis menggunakan aplikasi *anates*. Pada analisis butir soal dibagi menjadi dua bagian yaitu pertanyaan soal (*first tier*) dan alasan soal (*second tier*). Informasi mengenai daya beda soal dapat dilihat pada tabel 3.

Tabel 3. Analisis Daya Beda Soal

Instrumen Tes	Persentase			
	Jelek	Sedang	Baik	Baik Sekali
(<i>First Tier</i>)	13.33%	66.67%	13.33%	6.67%
(<i>Second Tier</i>)	16.67%	53.33%	27.67%	3.33%

Butir soal memiliki daya beda soal yang efektif apabila butir soal tersebut memiliki daya beda sedang, baik dan baik sekali (Astuti, 2020). Berdasarkan data analisis daya beda di atas menunjukkan bahwa soal yang memiliki daya beda jelek pada pertanyaan soal (*first tier*) dan alasan soal (*second tier*) hanya 13.33% dan 16.67%.

Berdasarkan uji validiitas, realibilitas dan analisis butir soal pada indeks kesukaran dan daya beda soal, instrumen tes *two tier* layak untuk dikembangkan. Hal ini disebabkan karena instrumen tes memiliki nilai validitas isi yang terdiri dari 4 kategori yaitu stimulus soal, pertanyaan soal (*first tier*), jawaban soal dan alasan soal (*second tier*) dengan nilai CVR yaitu 1. Butir soal dapat valid apabila memiliki CVR di atas atau sama dengan nilai minimum terhadap jumlah validator yaitu 0.736 (Siswaningsih et al., 2017). Selain itu instrumen tes ini juga memiliki validitas konstruk dari aspek materi, penyajian, bahasa dan aturan tambahan

yang sangat tinggi, sehingga instrumen tes ini dapat diuji cobakan untuk melihat realibilitas soal, indeks kesukaran dan daya pembeda soal.

Keterbatasan pada penelitian ini adalah jumlah sampel yang digunakan 20 siswa yang sampai pada tahap ujicoba skala kecil karena berdasarkan kepada tujuan utama penelitian ini adalah untuk mengembangkan suatu produk instrumen tes *two tier* berbasis *Higher Order Thinking Skills* (HOTS) pada Materi Sifat Koligatif Larutan untuk siswa SMA/MA. Pada uji coba skala kecil (*small group*) instrumen tes memiliki realibilitas yang sangat tinggi pada pertanyaan soal (*first tier*) dan alasan soal (*second tier*), serta indeks kesukaran yang dapat berfungsi dengan baik dan daya beda soal yang dapat membedakan siswa dengan kemampuan yang tinggi dan rendah.

KESIMPULAN

Berdasarkan pada penelitian yang telah dilakukan dapat disimpulkan bahwa instrumen tes *two tier* berbasis *Higher Order Thinking Skills* (HOTS) pada materi sifat koligatif larutan untuk siswa SMA/ MA memiliki validitas konten dan konstruk dan realibilitas soal yang sangat tinggi serta menghasilkan analisis butir soal pada indeks kesukaran dan daya pembeda yang efektif dan baik untuk digunakan.

DAFTAR PUSTAKA

- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. 41(4), 352.
- Arif, M., Madura, U. T., & Wibisono, Y. (2014). *Penerapan aplikasi anates bentuk soal pilihan ganda*. 1(1), 1–9.
- Arifin. (2013). *Evaluasi Pembelajaran*. PT. Remaja Rosdakarya.
- Astuti, A. (2020). Analisis Soal Ujian Akhir Semester Genap Mata Pelajaran Matematika Siswa Sd Negeri 005 Bunuang. *Jurnal Pendiidikan Tambusai*, 4, 867–880.
- Fitriyana, N., Marfuatun, M., & Priyambodo, E. (2019). The Profile of Students' Analytical Thinking Skills on Chemistry Systemic Learning Approach. *Scientiae Educatia*, 8(2), 207.
- Hartono, R., & Pahlevi, T. (2020). Development of Assessment Instruments Based on Hots At Surabaya Vocational High School. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 6(2), 210–218.
- Hutapea, N., & Sudrajat, A. (2019). *Pengembangan Instrumen penilaian berbasis kemampuan berpikir tingkat tinggi (HOTS) untuk pembelajaran sifat koligatif larutan pada siswa kelas XII IPA SMA Negeri 12 Medan*. 1, 498–502.
- Lawshe. (1985). *A Quantitative Approach to Content Validity*. Personel Psychology.
- Marlisyah, S., Anom, K. W., & Sukaryawan, M. (2018). *Pengembangan Modul Kimia Muatan Lokal Valid dan Praktis Tentang Pakan Ampas Tahu Terfermentasi untuk Pertumbuhan Ikan Patin Berbasis STEM-PBL*. 2(November).
- Mujib, & Rasyid, M. F. (2019). *modul penyusunan keterampilan berpikir tingkat tinggi (higher order thinking skills) kimia*. Direktorat Pembinaan Sekolah Menengah Atas.
- Nengsih, N. R., Yusmaita, E., & Gazali, F. (2019). *Evaluasi validitas konten dan konstruk bahan ajar asam basa berbasis REACT*. 1–10.
- Plomp, T., & Nieveen, N. (2007). *An Introduction to Educational Design Research*.
- Putriani, J. D. (2021). *EDUKATIF : JURNAL ILMU PENDIDIKAN Penerapan Pendidikan Indonesia di Era*

- 1058 *Validitas dan Realibilitas Instrumen Tes Two Tier Berbasis Higher Order Thinking Skills pada Materi Sifat Koligatif Larutan untuk Siswa SMA/MA – Syaza Syahana, Andromeda*
DOI: <https://doi.org/10.31004/edukatif.v3i3.505>
- Revolusi Industri 4 . 0.* 3(3), 831–838.
- Sarah, F., Khaldun, I., & Gani, A. (2021). *The Development Higher Order Thinking Skill (Hots) As Questions In Chemistry Study (Solubility And Solubility Product Constant)*. 9(1), 51–60.
- Siswaningsih, W., Firman, H., & Indonesia, U. P. (2017). *Pada Materi Larutan Elektrolit Dan Nonelektrolit*. 144–149.
- Sugiyono. (2015). *Metode Penelitian Pendidikan*. Alfabeta.
- Suhady, W., & Roza, Y. (2020). *Pengembangan Soal untuk Mengukur Higher Order Thinking Skill (HOTS) Siswa*. 2, 143–150.
- Wiyarsi, A., Fachriyah, A. R., Supriadi, D., & Bin Muhamad Damanhuri, M. I. (2019). A test of analytical thinking and chemical representation ability on ‘rate of reaction’ topic. *Cakrawala Pendidikan*, 38(2), 228–242.