

## **Pengembangan Instrumen Tes *Two-Tier Higher Order Thinking Skill* (HOTS) Berbasis IBT pada Materi Kesetimbangan Kimia untuk Siswa SMA/MA**

### ***Development of a Two- Tier Higher Order Thinking Skill (HOTS) Test Instrument Based on IBT on Chemical Equilibrium Material for SMA/MA Students***

Zikra Effendi<sup>1</sup> and Andromeda Andromeda<sup>1\*</sup>

<sup>1</sup>Program Studi Pendidikan Kimia, Fakultas Matematika dan IPA, Universitas Negeri Padang, Sumatera Barat, Indonesia.

\* Email: [andromeda@fmipa.unp.ac.id](mailto:andromeda@fmipa.unp.ac.id)

#### **ABSTRACT**

HOTS is an ability to transfer, connect one material with another, possess information and solve problems derived from the information in a creative and critical manner. Higher Order Thinking Skill (HOTS) are abilities that must be possessed by students in facing global competition in the 21<sup>st</sup> century industrial revolution 4.0 in accordance with the 2013 curriculum. This study aims to produce a test instrument for higher order thinking skill (HOTS) in chemical equilibrium and determine the level of validity and its reliability. This type of research is research and development (R&D) and uses the research model that has been developed by Treagust (1988). The results showed that the value of the validity of the content instrument test was 1 and the validity of the construct instrument was 95% in the material aspect, 94% in the presentation aspect, 96% in the language aspect, and 08% on the additional rules. The instrument reliability tes coefficient is 0,85 for the question (firsrt tier) and 0,89 for the reason (second tier).

*Keywords:* HOTS, Two-Tier, Test Instrument , Chemical Equilibrium.

#### **ABSTRAK**

HOTS merupakan suatu kemampuan yang dapat mentransfer, menghubungkan suatu materi dengan materi yang lainnya, memproses informasi serta menyelesaikan masalah yang didapat dari informasi tersebut dengan kreatif dan kritis. Higher order thinking skill (HOTS) merupakan kemampuan yang harus dimiliki oleh siswa dalam menghadapi persaingan global abad ke 21 revolusi industri 4.0 sesuai dengan tuntutan kurikulum 2013. Penelitian ini bertujuan untuk menghasilkan instrument tes two-tier higher order thinking skill (HOTS) pada materi kesetimbangan kimia serta menentukan tingkat validitas dan reliabilitasnya. Jenis penelitian ini adalah penelitian dan pengembangan atau research and development (R&D) dan menggunakan model penelitian yang telah dikembangkan oleh Treagust (1988). Hasil penelitian menunjukkan bahwa nilai validitas konten instrumen tes adalah 1 dan

validitas konstruk instrumen adalah 95% pada aspek materi, 94% pada aspek penyajian, 96% pada aspek bahasa, dan 98% aturan tambahan. Koefisien reliabilitas instrumen tes adalah 0,85 untuk butir pertanyaan soal (*first tier*) dan 0,89 untuk butir alasan soal (*second tier*).

**Kata Kunci:** HOTS, *Two-Tier*, Instrumen Tes, Keseimbangan Kimia.

## PENDAHULUAN

Abad 21 dikenal dengan abad pendidikan, dimana abad ini menjadi landasan utama berbagai aspek kehidupan. Abad 21 menitikberatkan pada kemampuan berpikir serta komunikasi peserta didik dalam pembelajaran sains. Tujuan suatu pendidikan dapat dicapai melalui kurikulum. Kurikulum yang digunakan saat sekarang ini ialah kurikulum 2013 yang menilai pada tiga aspek, yaitu: penilaian keterampilan, sikap, dan pengetahuan. Kurikulum 2013 menjadikan siswa lebih aktif dan guru menjadi lebih kreatif (Sada, 2019).

Pembelajaran yang baik dapat dilihat dari kualitas penilainya (Maryanto, 2015). Penilaian yang dilakukan harus sesuai dengan pencapaian kompetensi dasar (KD) yang ditetapkan. Menurut (Conklin, 2011) terdapat dua kategori keterampilan berpikir yaitu LOTS dan HOTS. *Lower-Order Thinking Skill* (LOTS) mencakup kemampuan mengingat (C1), memahami (C2), dan mengaplikasikan (C3). Sedangkan *Higher Order Thinking Skill* (HOTS) meliputi kemampuan menganalisis (C4), mengevaluasi (C5), dan mengkreasikan (C6).

Kemampuan berpikir tingkat tinggi atau yang dikenal dengan *Higher Order Thinking Skill* (HOTS) menjadi salah satu kemampuan yang harus dimiliki oleh siswa dalam menghadapi perkembangan zaman dan persaingan global saat ini. Kemampuan berpikir tingkat tinggi merupakan suatu kemampuan yang dapat menuntun siswa mentransfer dan menghubungkan suatu materi dengan materi yang lainnya. Salah

satu cara yang dapat meningkatkan kemampuan berpikir tingkat tinggi yaitu dengan latihan atau mengerjakan soal-soal yang bersifat kritis (Kartimi dkk., 2012). Soal HOTS merupakan instrumen penilaian yang dapat digunakan untuk mengukur kemampuan berpikir tingkat tinggi siswa serta dapat melatih siswa berpikir kritis dan kreatif. Oleh sebab itu, soal HOTS sangat penting untuk *study* kelayakan kerja di masa depan (Habiddin & Page, 2020).

Wiwik (2019) mengatakan bahwa terdapat 3 prinsip penilaian berpikir tingkat tinggi, yaitu: 1) Menyajikan stimulus, 2) Membedakan tingkat kesulitan soal dengan level kognitif, dan 3) Menggunakan permasalahan-permasalahan baru serta tidak menggunakan soal yang melibatkan proses mengingat.

Dalam menyusun soal HOTS sama dengan soal pada umumnya. Akan tetapi pada soal HOTS terdapat stimulus soal yang digunakan sebagai rangsangan dalam menuntun siswa untuk berpikir lebih kritis dan menggali konsep-konsep yang telah dikuasainya. Adapun langkah-langkah dalam menyusun soal HOTS menurut (Fanani, 2018) ialah sebagai berikut: 1) Menganalisis kompetensi dasar, 2) Membuat kisi-kisi soal, 3) Menggunakan permasalahan yang kontekstual (yang ada dalam kehidupan sehari-hari), 4) Menyusun butir soal, dan 5) Membuat penskoran.

Tidak semua kompetensi dasar yang ada dapat dibuatkan soal HOTS. Hanya saja kompetensi dasar yang memiliki KKO (kata kerja operasional) pada ranah C4, C5, atau C6 (Mujib, 2019). Adapun materi yang

diambil dalam pengembangan soal HOTS ini ialah materi kesetimbangan kimia. Materi kesetimbangan kimia merupakan salah satu materi yang terdapat dalam kurikulum 2013 untuk siswa SMA kelas XI. Materi ini membutuhkan kemampuan berpikir tingkat tinggi siswa karena berada pada level kognitif C4. Kesetimbangan kimia merupakan materi yang kompleks sehingga dibutuhkan proses berpikir yang lebih dan tidak hanya sekedar menghafal konsep-konsep.

*Two tier multiple choice question* merupakan bentuk pertanyaan pilihan ganda yang berbeda dari pertanyaan pilihan ganda lainnya. Instrumen tes *two tier* memiliki 2 bagian, dimana bagian pertama merupakan jawaban dari pertanyaan soal sedangkan bagian kedua merupakan alasan yang mendasari pilihan jawaban pada bagaian pertama (Bayrak, 2013). Instrumen tes *two tier* dapat mengidentifikasi konsep-konsep yang ada pada peserta didik. Dibandingkan dengan penggunaan pilihan ganda tradisional, *two tier* item berpotensi dalam memunculkan cara peserta didik beralasan (Kahveci, 2013).

Instrumen penilaian dapat digunakan oleh guru dalam melaksanakan penilaian sebagai bentuk evaluasi dari suatu pembelajaran. Salah satu metode yang dapat digunakan dalam evaluasi pembelajaran yaitu dengan memanfaatkan teknologi saat ini khususnya internet. Evaluasi pembelajaran berbasis internet atau yang lebih dikenal dengan IBT (*Internet Based Test*) merupakan suatu sistem yang yang membutuhkan jaringan untuk mengaksesnya dan juga menjadikan proses assesmen lebih mudah dan efektif (Amarulloh dkk., 2019). Salah satu platform pembelajaran yang memerlukan internet dalam mengaksesnya ialah *google classroom*. *Google classroom* merupakan layanan yang disediakan oleh google sebagai suatu sistem *e-learning* berbasis

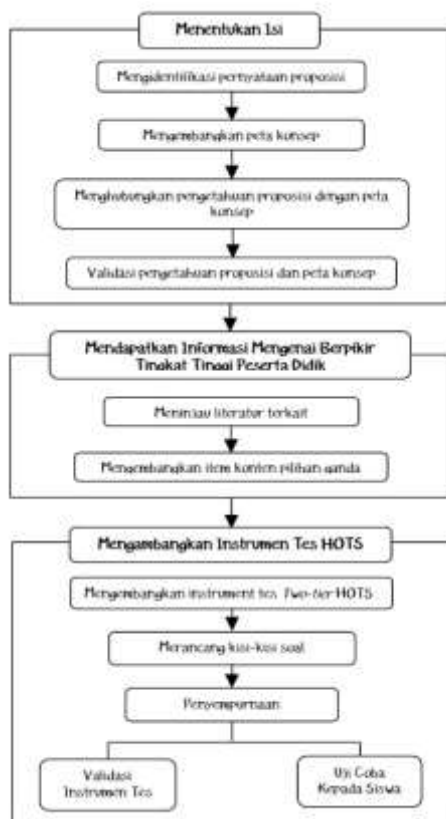
*virtual class* yang membutuhkan jaringan (internet) dalam mengaksesnya (Barir, 2016).

Berdasarkan kepada latar belakang yang dipaparkan, maka perlu dikembangkan instrumen tes *two tier higher order thinking skill* (HOTS) untuk meningkatkan kemampuan berpikir tingkat tinggi siswa sehingga menghasilkan instrument tes yang berkualitas. Adapun yang menjadi tujuan dalam penelitian ini adalah menentukan tingkat validitas dan reliabilitas instrumen Tes *Two Tier Higher Order Thinking Skill* (HOTS) pada materi kesetimbangan kimia untuk siswa SMA/MA serta menganalisis butir soal pada bagian daya beda dan indek kesukaran.

## METODE

Penelitian ini merupakan penelitian pengembangan instrumen tes *Two Tier Higher Order Thinking Skill* (HOTS) pada materi kesetimbangan kimia untuk siswa SMA/MA. Jenis penelitian ini adalah penelitian dan pengembangan atau *research and development* (R&D). Penelitian pengembangan menggunakan model yang telah dikembangkan oleh (Treagust, 1988). Terdapat tiga tahapan dasar dalam pengembangan ini, yaitu: 1) Menentukan isi, 2) Mendapatkan informasi mengenai berpikir tingkat tinggi peserta didik, 3) Mengembangkan instrumen tes HOTS. Tahap penentuan isi meliputi identifikasi pernyataan proposisi, mengembangkan peta konsep, menghubungkan pengetahuan proposisi dengan peta konsep, dan validasi pernyataan pengetahuan proposisi dan peta konsep. Informasi mengenai keterampilan berpikir tingkat tinggi peserta didik dapat diperoleh melalui data hasil observasi di sekolah. Tahapan pengembangan instrumen tes HOTS meliputi pengembangan instrumen tes *two tier* HOTS, merancang kisi-kisi dan penyempurnaan. Pada tahapan penyempurnaan dilakukan validasi

instrumen tes dan uji coba kepada siswa. Validasi dilakukan oleh 5 orang validator, 3 diantaranya merupakan dosen kimia dan 2 orang lagi ialah guru SMA.



Gambar 1. Skema prosedur penelitian Treagust (1988)

Penelitian ini dilakukan di SMAN 7 PADANG, dan yang menjadi subjek dalam penelitian ini ialah dosen jurusan kimia FMIPA UNP dan guru kimia SMA di kota Padang. Objek penelitian ini adalah kumpulan instrumen tes two-tier higher order thinking skill (HOTS) pada materi kesetimbangan kimia untuk siswa SMA/MA. Jenis data yang digunakan dalam penelitian ini adalah data primer yaitu data yang diambil langsung dari guru melalui angket *google form* dan lembar validasi.

Instrumen yang digunakan untuk mengumpulkan data dalam penelitian ini ialah lembar angket wawancara dan angket validasi. Validasi instrumen tes dilakukan pada dua aspek yaitu validasi konten dan konstruk. Data validitas konten diolah

dengan menggunakan CVR yang dikembangkan oleh (Lawshe, 1975) dengan pilihan jawaban yang terdiri dari dua kategori yaitu: iya dan tidak. Untuk kategori iya mendapatkan nilai 1 dan kategori tidak mendapatkan nilai 0.

Skor CVR berkisar antara -1 sampai 1. Skor penilaian pada CVR dapat ditentukan berdasarkan jumlah validator. Item dinyatakan valid apabila angka CVR melebihi batas minimum. Jika angka CVR berada dibawah batas minimum maka item dinyatakan tidak valid. Batas minimum untuk 5 orang validator adalah 0,736 (Wilson & Holmvall, 2013).

Analisis validitas konstruk diukur pada 4 aspek penyajian item tes yang terdiri dari aspek materi, aspek penyajian, aspek penggunaan bahasa, dan aturan tambahan. Analisis data validasi konstruk diolah dengan menggunakan *Aiken's V* yang disusun menggunakan empat tingkatan terdiri dari: Tidak Relevan (TR), Kurang Relevan (KR), Cukup Relevan (CR), dan Relevan (R) yang kemudian diolah sesuai dengan skala yang terdapat pada *Aiken's V*. Nilai skala *Aiken's V* untuk 5 orang validator ialah 0,93.

Setelah validasi, instrumen tes diujicobakan kepada 20 orang siswa. Uji coba ini bertujuan untuk menentukan reliabilitas soal dan menganalisis butir soal pada bagian daya pembeda, tingkat kesukaran dan fungsi distraktor (pengecoh). Hasil uji coba dianalisis dengan menggunakan aplikasi anates yang dikembangkan oleh Mr. Drs. Karno To, M.Pd., dan Mr. Yudi Wibisono.

## HASIL DAN DISKUSI

### Hasil

Hasil penelitian dari pengembangan instrumen tes *two tier* HOTS ini adalah sebagai berikut:

Hasil pengembangan produk awal

### ***Menentukan Isi***

Tahapan pertama yang dilakukan ialah mengidentifikasi pernyataan pengetahuan proposisi. Kegiatan ini bertujuan untuk menganalisis dan menguraikan Kompetensi Dasar (KD) menjadi Indikator Pencapaian Kompetensi (IPK), yaitu KD 3.9 kelas XI SMA, yang berbunyi “Menganalisis faktor-faktor yang mempengaruhi pergeseran arah kesetimbangan kimia dan penerapannya dalam industri”. KD diturunkan menjadi 2 indikator. Berdasarkan indikator yang diturunkan maka dikembangkanlah konsep-konsep dasar (proposisi) pada materi kesetimbangan kimia. Selanjutnya mengembangkan peta konsep berdasarkan pernyataan proposisi. Pernyataan proposisi dan peta konsep harus memiliki keterkaitan dan saling berhubungan. Validasi konten pernyataan proposisi dan peta konsep divalidasi oleh 4 orang validator, 2 diantaranya dosen kimia dan 2 orang lagi guru SMA. Hasil validasi di analisis menggunakan analisa formula CVI yang dapat menghitung *content validity coefficient*. Item dinyatakan valid apabila memiliki I-CVI sebesar 0,80 (Polit & Beck, 2006). Hasil validasi konten pernyataan proposisi dan peta konsep dari 4 orang validator memiliki nilai I-CVI adalah 1, hal ini menandakan valid.

### ***Mendapatkan informasi mengenai keterampilan berpikir tingkat tinggi peserta didi***

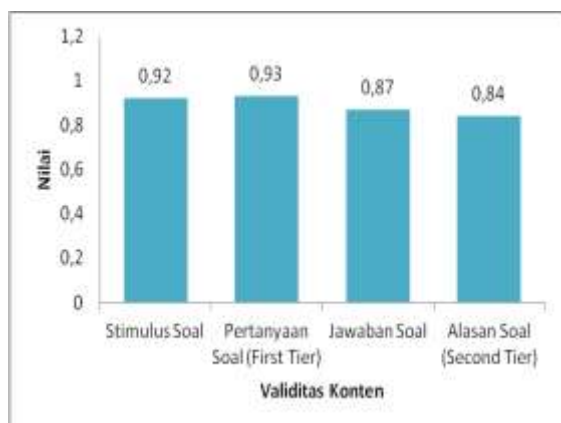
Mendapatkan informasi mengenai berpikir tingkat tinggi peserta didik dilakukan dengan peninjauan literatur terkait. Hasil tinjauan literatur didapatkan informasi mengenai pentingnya keterampilan berpikir tingkat tinggi yang harus dimiliki oleh siswa sesuai dengan tuntutan kurikulum saat ini. Hal ini dapat dijadikan acuan dalam mengembangkan instrumen tes HOTS. Pengembangan item konten pilihan ganda berdasarkan konsep-

konsep pada kompetensi dasar yang telah ditetapkan.

### ***Mengembangkan instrumen tes HOTS***

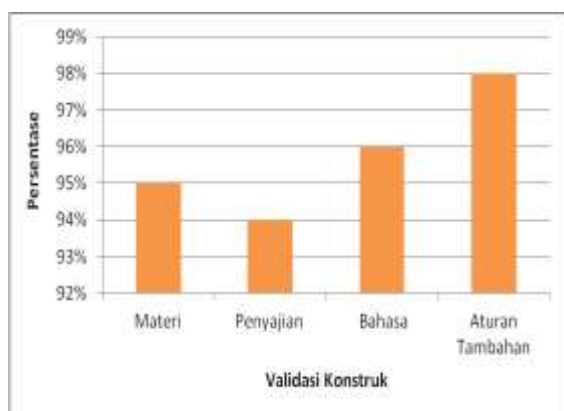
Instrumen tes *two tier* HOTS dikembangkan berdasarkan pada konten pilihan ganda yang telah dikembangkan sebelumnya. Tes *two tier* merupakan tes dua tingkat yang terdiri dari pertanyaan soal (*first tier*) dan alasan soal (*second tier*). Kisi-kisi soal memuat butir-butir soal yang telah dikembangkan sebelumnya. Rancangan kisi-kisi didasarkan pada indikator soal dan IPK yang sesuai dengan KD. Setelah rancangan kisi-kisi terbentuk maka dilakukan penyempurnaan terhadap soal.

Pada tahap penyempurnaan dilakukan validasi dan uji coba ke siswa. Hasil uji coba akan digunakan dalam menentukan reliabilitas soal, daya beda, indek kesukaran dan fungsi distraktor. 2 tahapan validasi yang dilakukan yaitu validasi konten dan konstruk. Pada validasi konten terdapat 4 aspek yang akan dinilai diantaranya; stimulus soal, pertanyaan soal, jawaban soal, dan alasan soal. Sedangkan pada validasi konstruk yang menjadi aspek penilaiannya yaitu aspek materi, penyajian, bahasa dan aturan tambahan. Untuk hasil validasi konten di analisis menggunakan metode CVR. Validasi instrumen tes divalidasi terhadap 25 orang validator dengan jumlah soal yaitu 25 butir soal. Hasil validasi pertama terdapat 13 soal valid dan 12 soalnya lagi belum valid. Dan setelah dilakukan revisi didapatkan 25 soal valid dengan angka CVR yang diperoleh adalah 1 untuk 5 orang validator. Berikut grafik rata-rata validitas konten sebelum revisi.



Gambar 2. Grafik rata-rata validitas konten sebelum revisi

Hasil validasi konstruk di analisis menggunakan *Aiken's V*. persentase hasil analisis data validitas konstruk ialah 95% untuk aspek materi, 94% aspek penyajian, 96% aspek bahasa, dan 98% aturan tambahan. Dari hasil persentase tersebut dapat diketahui validitas konstruk yang diperoleh sangat tinggi. Berikut grafik persentase validitas konstruk:



Gambar 3. Grafik persentase validitas konstruk

Setelah semua soal dinyatakan valid tahapan selanjutnya yaitu uji coba ke siswa. Soal di uji cobakan kepada 20 orang siswa secara online melalui *google form*. Dari hasil uji coba didapatkanlah nilai reliabilitas, daya beda, indek kesukaran soal yang di analisis menggunakan aplikasi anates.

Reliabilitas soal dengan menggunakan aplikasi anates dibagi menjadi dua bagian. Bagian pertama

analisis butir pada pertanyaan soal (*first tier*) dan bagian kedua analisis butir alasan soal (*second tier*). Hasil analisis menunjukkan bahwa pada bagian pertama nilai reliabilitas butir soal yaitu 0,85 sedangkan pada bagian kedua nilai reliabilitas butir soal yaitu 0,89. Hal ini menunjukkan bahwa soal memiliki nilai reliabilitas yang tinggi sehingga soal dianggap konsisten dalam meningkatkan keterampilan berpikir tingkat tinggi siswa. Nilai koefisien reliabilitas soal antara 0 sampai 1 (Khumaedi, 2012).

Hasil analisis tingkat kesukaran soal menggunakan aplikasi anates dibagi menjadi dua bagian. Hasil analisis tingkat kesukaran soal dikategorikan menjadi 3 yaitu: sedang, sukar, dan sangat sukar. Pada bagian pertama analisis tingkat kesukaran jawaban soal (*first tier*) terdapat 18 soal berada pada level sedang, 4 soal berada pada level sukar, dan 3 soal berada pada level sangat sukar. Sedangkan untuk hasil analisis tingkat kesukaran pada bagian alasan soal (*second tier*) terdapat 24 soal berada pada level sedang, 7 soal berada pada level sukar, dan 4 soal berada pada level sangat sukar.

Daya pembeda soal dianalisis menggunakan aplikasi anates dibagi menjadi dua bagian. Hasil analisis menunjukkan bahwa daya beda soal dikategorikan menjadi 4 bagian, yaitu jelek, sedang, baik, dan sangat baik dengan persentasenya untuk bagian pertama yaitu jawaban soal (*first tier*) memiliki daya beda berturut-turut adalah 28%, 16%, dan 12% sedangkan untuk bagian kedua yaitu alasan soal (*second tier*) memiliki daya beda berturut-turut adalah 36%, 28%, dan 12%. Nilai daya beda berkisar anatar 0% hingga 100%.

### Pembahasan

Model penelitian yang telah dikembangkan oleh (Treagust, 1988) terdiri atas 3 tahapan dasar yaitu; 1) Menentukan isi, 2)

Mendapatkan informasi terkait keterampilan berpikir tingkat tinggi peserta didik, 3) Mengembangkan instrumen tes. Dalam penelitian ini dilakukan 3 validasi, yaitu validasi konten pernyataan proposisi dan peta konsep, validasi konten instrumen tes, dan validasi konstruk instrumen tes. Setelah selesai dilakukan validasi dan soal dinyatakan valid tahap selanjutnya yaitu pengujian instrument tes kepada siswa yang bertujuan untuk menganalisis butir soal.

Validasi konten pernyataan proposisi dan peta konsep divalidasi oleh 5 orang validator dan analisis data dengan menggunakan analisa formula CVI. Pernyataan proposisi dan peta konsep dinyatakan valid apabila memiliki nilai I-CVI sebesar 0,80. Hasil validasi konten pernyataan proposisi dan peta konsep materi kesetimbangan kimia memiliki nilai I-CVI sebesar 1 yang menandakan bahwa pernyataan proposisi dan peta konsep yang dikembangkan ialah valid. Dan semua validator setuju dengan konsep-konsep yang telah dikembangkan.

Validitas konten instrumen tes divalidasi oleh 5 orang validator dengan 4 aspek penilaian, yaitu; stimulus soal, jawaban soal (*first tier*), pertanyaan soal, dan alasan soal (*second tier*). Data hasil validitas konten dianalisis menggunakan CVR dengan batas minimum CVI untuk 5 orang validator ialah 0,736. Validasi konten instrumen tes dilakukan sebanyak 2 kali validasi. Hasil validasi pertama terdapat 13 soal valid dan 12 soalnya lagi belum valid. Hal ini disebabkan oleh beberapa faktor seperti stimulus soal yang tidak kontekstual (tidak sesuai dengan kenyataan dalam kehidupan sehari-hari), jawaban soal benar lebih dari satu, serta alasan soal belum tepat. 4 soal dinyatakan tidak valid secara kontekstual yaitu soal no 2, 5, 6 dan 20 sehingga soal harus diganti secara keseluruhan. Soal no 8, 15, 21, dan 22 memiliki *option* jawaban benar lebih dari

satu sehingga perlu direvisi. Alasan soal yang untuk soal no 11, 13, dan 25 belum tepat sehingga perlu diperbaiki. Berdasarkan hasil validasi pertama didapatkan nilai CVI sebesar 0,6, artinya nilai CVI belum memenuhi batas minimum untuk 5 orang validator sehingga soal dinyatakan belum valid dan harus diperbaiki. Hasil analisis data validitas konten instrumen tes setelah dilakukan revisi nilai CVI yang didapatkan adalah 1, artinya soal sudah valid secara konten.

Hasil validitas konstruk instrumen tes yang divalidasi oleh 5 orang validator diukur pada 4 aspek, yaitu: aspek materi, penyajian, bahasa, dan aturan tambahan. Analisis data validitas menggunakan metode *Aiken's V* dan skala yang digunakan ada 4 yaitu; tidak relevan (TR), kurang relevan (KR), cukup relevan (CR), dan relevan (R). Hasil analisis data menunjukkan nilai validitas konstruk instrumen tes sangat tinggi dengan persentase 95% pada aspek materi, 94% pada aspek penyajian, 96% pada aspek bahasa, dan 98% pada aspek aturan tambahan. Setelah soal dinyatakan valid secara konten dan konstruk tahap selanjutnya yaitu uji coba soal kepada siswa.

Uji coba kepada siswa dilakukan secara online dengan menggunakan *google form*. Pengerjaan soal dilakukan secara serentak dengan waktu pengerjaan selama 90 menit. Hasil uji coba soal digunakan dalam menentukan nilai reliabilitas dan analisis butir soal pada bagian tingkat kesukaran dan daya pembeda. Analisis reliabilitas soal menggunakan aplikasi anates terbagi atas 2 bagian, yaitu: bagian pertama jawaban soal (*first tier*). Hasil analisis bagian pertama didapatkan nilai reliabilitas soal 0,85. Sedangkan untuk bagian kedua jawaban soal (*second tier*). Hasil analisis bagian kedua didapatkan nilai reliabilitas 0,89. Hasil uji reliabilitas pertanyaan soal dan jawaban soal yang

didapatkan sangatlah tinggi. Suatu instrument dikatakan reliabel apabila memiliki nilai reliabilitas besar dari 0,7.

Analisis tingkat kesukaran butir soal dianalisis menggunakan aplikasi anates dan dibagi menjadi 2 bagian. Analisis tingkat kesukaran instrumen tes dikategorikan menjadi 3 bagian, yaitu sedang, sukar, dan sangat sukar. Persentase hasil analisis bagian pertama pertanyaan soal (*first tier*) berturut-turut ialah 72%, 16%, dan 12%. Sedangkan persentase hasil analisis data bagian kedua alasan soal (*second tier*) berturut-turut adalah 56%, 28%, dan 16%.

Analisis daya beda butir soal dilakukan dengan menggunakan aplikasi anates dibagi menjadi dua bagian. Daya beda soal mampu membedakan siswa yang berada pada kelompok bawah dengan kelompok atas. Daya beda dikatakan baik jika memiliki nilai positif. Berdasarkan hasil penelitian yang telah dilakukan, daya beda dikategorikan menjadi 4 bagian, yaitu jelek, sedang, baik, dan baik sekali. Hasil persentase daya beda pada bagian pertama pertanyaan soal (*first tier*) berturut-turut adalah 28%, 16%, 12%, dan 44%. Sedangkan untuk persentase daya beda pada bagian alasan soal (*second tier*) berturut-turut adalah 36%, 28%, 12%, dan 16%.

Berdasarkan pada hasil penelitian yang telah dipaparkan. Pengembangan instrumen tes *two-tier higher order thinking skill* (HOTS) pada materi kesetimbangan kimia layak untuk dikembangkan. Hal ini dikarenakan instrument tes telah dinyatakan valid secara konten dan konstruk sehingga bisa di ujicobakan. Hasil uji coba digunakan dalam menentukan nilai reliabilitas dan analisis butir soal pada tingkat kesukaran dan daya pembeda.

### KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, instrumen tes *two-tier higher*

*order thinking skill* (HOTS) pada materi kesetimbangan kimia dinyatakan valid dan reliabel. Nilai validitas konstruk yang sangat tinggi, yaitu 95% pada aspek materi, 94% pada aspek penyajian, 96% pada aspek bahasa, dan 98% aturan tambahan. Untuk nilai reliabilitas yang didapatkan ialah 0,85 pada pertanyaan soal (*first tier*) dan 0,89 pada alasan soal (*second tier*) artinya instrumen tes memiliki koefisien reliabilitas yang sangat tinggi.

### KETERBATASAN DAN IMPLIKASI UNTUK PENELITIAN LAIN

Adapun yang menjadi keterbatasan dalam penelitian ini adalah metoda yang belum terlaksana secara tuntas dan sistematis.

Diharapkan peneliti selanjutnya dapat melaksanakan semua tahapan metode secara sistematis.

### REFERENSI

- Amarulloh, A., Surahman, E., & Meylani, V. (2019). Refleksi Peserta Didik Terhadap Pembelajaran Berbasis Digital. *Metaedukasi*, 1(1), 13–23.
- Barir, H., A. (2016). Efektivitas Penggunaan ELearning Moodle, Google Classroom Dan Edmodo. *Jurnal I-Statement*, Vol.02, 2.
- Bayrak, B. K. (2013). Using Two-Tier Test to Identify Primary Students' Conceptual Understanding and Alternative Conceptions in Acid Base. *Mevlana International Journal of Education*, 3(2), 19–26. <https://doi.org/10.13054/mije.13.21.3.2>
- Conkil, W. (2011). *Higher Order Thinking Skills to Develop 21<sup>st</sup> Century Learners*. Shell Education.
- Fanani, M. Z. (2018). Strategi Pengembangan Soal Hots Pada Kurikulum 2013. *Edudeena*, 2(1), 57–76. <https://doi.org/10.30762/ed.v2i1.582>
- Habiddin, H., & Page, E. M. (2020). Probing Students' Higher Order



- Thinking Skills Using Pictorial Style Questions. *Macedonian Journal of Chemistry and Chemical Engineering*, 39(2), 251–263. <https://doi.org/10.20450/mjce.2020.2133>
- Kahveci, A. (2013). *Diagnostic Assessment of Student Understanding of the Particulate Nature of Matter: Decades of Research*. 249–278. [https://doi.org/10.1007/978-94-007-5914-5\\_12](https://doi.org/10.1007/978-94-007-5914-5_12)
- Kartimi, Liliyasi, & Permanasari, A. (2012). Pengembangan Alat Ukur Berpikir Kritis pada Konsep Senyawa Hidrokarbon untuk Siswa SMA di Kabupaten Kuningan. *Jurnal Pendidikan MIPA*, 13(1), 19–25.
- Khumaedi, M. (2012). Reliabilitas Instrumen Penelitian Pendidikan. In *Jurnal Pendidikan Teknik Mesin Unnes* (Vol. 12, Issue 1, p. 129541).
- Lawshe, C. H. (1975). a Quantitative Approach To Content Validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Maryanto, W. (2015). Model Penilaian untuk Pembelajaran Abad 21 (Sebuah Kajian untuk Mempersiapkan SDM Kritis dan Kreatif). *Prosiding Pendidikan Teknik Boga Busana*, 21(1), 10.
- Mujib, M. F. R. (2019). *Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills)*. Direktorat Pembinaan Sekolah Menengah Atas Direktorat Jenderal Pendidikan Dasar Dan Menengah Kementerian Pendidikan Dan Kebudayaan.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur>
- Sada, C. (2019). Exploring Teaching Learning Process in Developing Higher Order Thinking Skill (HOTS) to Higher Secondary School (SMA) Students in Pontianak. *JETL (Journal Of Education, Teaching and Learning)*, 4(1), 228. <https://doi.org/10.26737/jetl.v4i1.1020>
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169. <https://doi.org/10.1080/0950069880100204>.
- Wilson, N. L., & Holmvall, C. M. (2013). The development and validation of the Incivility from Customers Scale. *Journal of Occupational Health Psychology*, 18(3), 310–326.
- Wiwik, S. (2019). Penilaian Berorientasi Higher Order Thinking Skills. *Direktorat Jenderal Guru Dan Tenaga Kependidikan Kementerian Pendidikan Dan Kebudayaan*.