

## Pengembangan Instrumen Tes Berbasis *Higher Order Thinking Skill* Pada Materi Hidrolisis Garam Untuk Siswa SMA/MA

I P Muchlis<sup>1)</sup> Andromeda<sup>2)</sup>

<sup>1)</sup>Universitas Negeri Padang

<sup>2)</sup>Universitas Negeri Padang

[andromedasaidir@yahoo.com](mailto:andromedasaidir@yahoo.com)

### ABSTRACT

*The purpose of this study was to develop a higher order thinking skill (HOTS) based test instrument on salt hydrolysis for senior high school students. The type of research used is Research and Development (R&D) using the Plomp model which consists of three stages, namely Preliminary Research, Prototyping Stage and Assessment Phase. The test instrument was validated by 7 experts, namely 4 chemistry lecturers of FMIPA UNP, and 3 chemistry teachers using a validation questionnaire. The small group trial was carried out by 6 students and the field test was carried out by 26 students of Senior High School number 1 in Padang. The data analysis technique for content validity uses Lawshe's CVR and the construct validity uses the aiken's formula. The results showed that the instrument developed had good characteristics in terms of logical validity with a content validity index (CVI) of 0.841 and a construct validity of 0.82. The reliability coefficient of 0.89 is very high. The analysis of discriminating power showed 16 items had very good distinguishing quality, 4 items had good distinguishing quality, 4 items had moderate distinguishing quality. The difficulty index for 88% of the questions is categorized as medium. The correlation of the item score with a significant total score was 21 items. The dictator function on all items is effective. The data obtained shows that the test instrument based on higher order thinking skills (HOTS) on salt hydrolysis material for senior high school students is valid and has good item quality.*

**Keywords :** *Test Instrument, Higher Order Thinking Skill (HOTS), Salt Hydrolysis, Research and Development (R&D)*



This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2018 by author and Universitas Negeri Padang.

### PENDAHULUAN

Evaluasi diartikan sebagai proses memperoleh informasi yang digunakan sebagai dasar pengambilan keputusan terkait dengan kemampuan siswa dan daya serap materi pembelajaran (Sudjiono, 2009). Penilaian hasil belajar diharapkan dapat membantu siswa meningkatkan keterampilan berpikir tingkat tinggi atau keterampilan berpikir tingkat tinggi (HOTS), serta dapat mendorong siswa untuk berpikir secara luas dan mendalam tentang topik yang diajarkan oleh guru.

Kurikulum 2013 diperlukan untuk menumbuhkan berbagai kemampuan siswa di abad 21 yang beberapa di antaranya dirumuskan dalam 4C, yaitu (1) Berpikir kritis, sehingga siswa dapat menggunakan logika kritis dan rasional untuk menyelesaikan berbagai masalah kontekstual; (2) Kreativitas, yaitu mendorong siswa untuk berkreasi dalam menemukan berbagai solusi, merancang strategi baru dan menemukan metode baru; (3) Kolaborasi,

meningkatkan kemampuan siswa dalam bekerja dengan tim, toleransi, memahami perbedaan untuk mencapai tujuan dan (4) Komunikasi dapat mendorong siswa untuk berkomunikasi Kemampuan luas, kemampuan menangkap ide/informasi, kemampuan menafsirkan informasi, dan kemampuan berargumentasi dalam arti luas (Kemendikbud, 2019)

Kemampuan peserta didik di Indonesia khususnya di bidang kimia masih sangat rendah dibandingkan dengan negara lain. Hal ini dibuktikan dengan studi internasional yaitu tes PISA (*Program for International Student Assessment*) yang diselenggarakan oleh OECD (*Organization for Economic Co-operation and Development*). Hasil PISA 2018 menunjukkan bahwa Indonesia berada di peringkat 69 dari 76 negara (OECD, 2019). Hasil tersebut menunjukkan bahwa Indonesia masih jauh tertinggal dari beberapa Negara lainnya, dimana siswa Indonesia masih kurang sekali yang dapat menjawab soal yang mengukur keterampilan

berpikir tingkat tinggi dan siswa Indonesia hanya mampu menjawab soal yang tergolong dalam kategori rendah (Hanifah, 2019).

Hasil PISA yang lebih rendah tentunya disebabkan oleh banyak faktor. Faktor-faktor penyebabnya antara lain siswa Indonesia kurang terlatih dalam memecahkan masalah kontekstual, yang memerlukan penalaran, argumentasi, dan kreativitas dalam menyelesaikan masalah tersebut yang merupakan ciri khas soal PISA (Puspendik, 2011).

Ini dibuktikan dengan hasil analisis butir soal yang dilakukan Direktorat Pembinaan SMA pada tahun 2018/2019 pada 26 mata pelajaran yang ditujukan pada 136 SMA rujukan pada 34 provinsi di Indonesia. Sebanyak 1.779 soal yang dianalisis menunjukkan hasil bahwa sebagian soal berada pada tingkatan kognitif C<sub>1</sub> dan C<sub>2</sub>. Sebanyak 135 SMA rujukan, hanya 27% sekolah yang menyusun soal *HOTS* dari keseluruhan soal keseluruhan soal USBN yang dibuat. Hal ini tidak sesuai dengan tuntutan kurikulum 2013, yang semakin meningkatkan implementasi model evaluasi (Kemendikbud, 2019).

Salah satu kompetensi dasar yang dapat mengembangkan soal-soal keterampilan berpikir tingkat tinggi (*HOTS*) adalah KD 3.11 menganalisis kesetimbangan ion dalam larutan garam dan menentukan pH nya. Berdasarkan hal tersebut, agar siswa memahami dan menggunakan pengetahuan yang telah diketahuinya, siswa harus dilatih untuk memecahkan masalah, menemukan masalah, dan mengembangkan ide-idenya sendiri.

Hasil wawancara dengan 7 orang guru kimia pada tujuh sekolah di Kota Padang dan Kota Pariaman menunjukkan bahwa 71% guru memiliki keterbatasan dalam menemukan sumber soal *HOTS*, 80% dari 35 soal yang digunakan pada penilaian hasil belajar ulangan harian masih berada pada tingkatan C<sub>1</sub>-C<sub>3</sub>. Guru menggunakan soal-soal dibuku paket sebagai alat ukur kemampuan peserta didik. Hal ini dibuktikan dengan hasil analisis soal Ulangan Harian (UH) pada materi Hidrolisis Garam kelas XI semester genap tahun ajaran 2019/2020 di tujuh sekolah di Kota Padang dan Kota Pariaman menunjukkan tingkat kognitif mengingat (C1) dengan presentase 20%, memahami (C2) 40%, mengaplikasikan (C3) 20%, menganalisis (C4) 20%, mengevaluasi

(C5) 0%, dan mencipta (C6) 0%. Item soal kognitif mengevaluasi dan mencipta masih 0%.

Berdasarkan latar belakang yang telah dipaparkan perlu dikembangkan instrumen tes *Higher Order Thinking Skill (HOTS)* yang dapat digunakan guru sebagai standar penilaian dalam pembelajaran serta menjadi kumpulan soal yang berkualitas pada materi Hidrolisis Garam. Maka dari itu, dilakukanlah penelitian yang berjudul "Pengembangan Instrumen tes berbasis *Higher Order Thinking Skill (HOTS)* pada Materi Hidrolisis Garam untuk Siswa SMA/MA"

## METODE PENELITIAN

Penelitian ini merupakan penelitian pengembangan atau *Research and Development (R & D)*, yang memiliki tujuan untuk menghasilkan instrumen tes berbasis *HOTS* materi hidrolisis garam untuk siswa kelas XI SMA/MA. Penelitian dan pengembangan adalah proses atau langkah untuk mengembangkan produk baru atau menyempurnakan produk yang sudah ada, yang dapat dipertanggungjawabkan (Sugiyono, 2012).

Model pengembangan yang digunakan adalah Model plomp yang memiliki tiga tahapan yaitu penelitian awal (*preliminary research*), tahap pembentukan prototipe (*prototyping stage*) dan tahap penilaian (*assasment phase*) (Plomp, 2013).

Jenis data dalam penelitian ini termasuk kedalam data primer, karena data diperoleh secara langsung dari dosen, guru dan peserta didik.

Data dikumpulkan melalui angket validasi, angket wawancara dan tes. Angket uji validitas isi disusun menurut metode *content validity ratio (CVR)* dengan jawaban berikut ini:

Tabel 1. Skor Angket Uji Validitas Isi

| Pilihan | Kategori     |
|---------|--------------|
| 1       | Setuju       |
| 0       | Tidak setuju |

Data di analisis menggunakan metode CVR untuk validitas isi (konten). Metode lawshe menghasilkan nilai-nilai yang berkisar antara -1 sampai 1. Nilai positif menunjukkan bahwa lebih dari setengah ahli memberikan tanggapan valid kepada item. Semakin besar nilai CVR dari 0 maka semakin tinggi validitas isinya (Lawshe, 1985).

$$CVR = \frac{(n - \frac{N}{2})}{\frac{N}{2}}$$

Keterangan:

CVR = *content validity ratio*

n = Banyaknya ahli yang menjawab valid

N = Jumlah total ahli.

Butir soal dikatakan diterima apabila nilai CVR pada butir soal tersebut sama atau lebih besar dari nilai kritis, sedangkan butir soal akan ditolak apabila nilai CVR soal tersebut kurang dari nilai kritis dalam jumlah ahli. Nilai kritis CVR sebagai berikut:

Tabel 2. Nilai Kritis CVR (one-tailed, α = 0,5)

| Jumlah Ahli | Nilai Kritis CVR |
|-------------|------------------|
| 5           | 0,736            |
| 6           | 0,672            |
| 7           | 0,622            |
| 8           | 0,582            |

(Wilson, 2012)

Angket uji validitas konstruk dianalisis dengan menggunakan indeks validitas yang diusulkan oleh Aiken (Retnawati, 2016). Angket penilaian uji validitas konstruk yang digunakan disusun berdasarkan skala Likert seperti pada Tabel 1.

Tabel 3. Skala Likert

| Skala Likert | Penilaian           |
|--------------|---------------------|
| 1            | Sangat tidak setuju |
| 2            | Tidak setuju        |
| 3            | Netral              |
| 4            | Setuju              |
| 5            | Sangat setuju       |

(Retnawati, 2016)

Data hasil uji validitas konstruk yang diperoleh dianalisis dengan indeks validitas butir indeks *Aiken's V* (V) dirumuskan sebagai berikut :

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

$$s = r - l_0$$

Keterangan:

V = Indeks kesepakatan rater

l<sub>0</sub> = Angka penilaian validitas yang terendah (dalam hal ini = 1)

c = Angka penilaian validitas yang tertinggi (dalam hal ini = 5)

r = Angka yang diberikan oleh seorang penilai

n = Jumlah rater

Tabel 4. Keputusan Berdasarkan Indeks *Aiken's V* (V)

| Interval      | Kategori |
|---------------|----------|
| ≤ 0,4         | Kurang   |
| 0,4 < V ≤ 0,8 | Sedang   |
| 0,8 < V       | Valid    |

(Retnawati, 2016).

Angket validasi sesuai dengan ketentuan yang meliputi kelayakan isi, kelayakan konstruksi, kelayakan bahasa, dan kompetensi instrumen yang dihasilkan. Data Kualitas tes diperoleh dengan melakukan uji coba kepada 26 orang siswa kelas XII MIA di SMAN 1 Padang. Tes yang digunakan merupakan tes tertulis. Siswa menjawab soal dengan memilih satu pertanyaan yang dianggap benar terhadap lima pilihan jawaban lainnya.

Kualitas tes dianalisis dengan *software* ANATES 4.0.9. *Software* ANATES merupakan perangkat lunak yang dikembangkan oleh kartono dan wibisono. ANATES versi 4.0.9 dapat menganalisis validitas empiris, reliabilitas, indeks kesukaran soal, daya pembeda dan kualitas pengecoh butir soal pada instrumen tes berbasis *HOTS* pada materi hidrolisis penelitian ini menggunakan jenis penelitian pengembangan.

## HASIL DAN PEMBAHASAN

### Penelitian Awal (Preliminary Research)

#### *Analisis Kebutuhan*

Analisis kebutuhan bertujuan untuk mengetahui masalah dalam instrumen tes yang digunakan dalam proses evaluasi pada sekolah di Kota Padang dan Kota Pariaman. Berdasarkan hasil wawancara dengan 7 guru diperoleh informasi bahwa 1) 71% guru yang telah diwawancarai memiliki keterbatasan dalam menemukan sumber soal *HOTS*, 2) 85% dari 35 soal yang digunakan masih berada pada tingkatan C1 sampai dengan C3.

#### *Analisis Konteks*

Analisis konteks bertujuan untuk mengetahui hal-hal yang berhubungan dengan pembelajaran kimia, kurikulum, serta alat evaluasi yang digunakan dalam pembelajaran.

#### *Studi Literatur*

Studi literatur bertujuan sebagai sumber yang digunakan dalam mengembangkan instrumen tes berbasis *HOTS*.

### *Pengembangan Kerangka Konseptual*

Pengembangan kerangka konseptual dilakukan dengan penyusunan format kisi-kisi soal meliputi kompetensi dasar, indikator pencapaian kompetensi, materi pokok, level kognitif, indikator soal serta stimulus pada materi hidrolisis garam.

### **Tahap Pembentukan Prototipe**

#### *Prototipe I*

Instrumen tes yang dikembangkan berbasis *HOTS* yang pada setiap soal diberikan stimulus untuk merangsang keterampilan berpikir siswa dalam penyelesaian soal serta 5 pilihan jawaban yang sesuai dengan bentuk tes pilihan ganda *multiple choice*. Pemilihan format penulisan instrumen tes disesuaikan dengan komponen-komponen yang terdapat dalam instrumen diantaranya sebagai berikut:

#### 1. Cover

Bagian *cover* instrumen memuat judul instrumen, nama materi, nama penulis, nama instansi penulis, tahun, target sasaran dan nama pembimbing. Nama penulis dan instansi penulis berguna untuk memberikan informasi kepada pengguna tentang penulis instrumen tes dan instansi penulis. Target sasaran berguna untuk memberikan informasi kepada pengguna mengenai target keterpakaian instrumen yang ditulis oleh penulis. Nama pembimbing pada instrumen tes berguna untuk memberikan informasi kepada pengguna mengenai pembimbing penulis dalam penulisan instrumen tes.

#### 2. Kata Pengantar

Kata pengantar merupakan sebuah apresiasi yang berisi ucapan terimakasih kepada semua pihak yang telah membantu penulis dalam menyelesaikan instrumen tes berbasis *HOTS* (*higher order thinking skill*) pada materi hidrolisis garam untuk siswa SMA/ MA.

#### 3. Format Kisi-kisi Soal

Kisi-kisi soal berisikan Kompetensi Dasar (KD), Indikator Pencapaian Kompetensi (IPK), materi pokok, indikator soal, level kognitif, jumlah soal, nomor soal serta kunci jawaban pada materi hidrolisis garam berdasarkan kurikulum 2013 revisi 2018. Perumusan format kisi-kisi soal berguna untuk memudahkan dalam penulisan soal berbasis *HOTS*.

#### 4. Petunjuk Pengerjaan Soal

Komponen ini berisi petunjuk-petunjuk pengerjaan soal pada instrumen tes berbasis

*HOTS* sehingga proses evaluasi dapat berjalan secara efisien.

#### 5. Kartu Soal

Kartu soal dirumuskan sebanyak 30 item berbentuk pilihan ganda (*multiple choice*) yang dilengkapi dengan lima alternatif jawaban. Kartu soal berisikan identitas mata pelajaran, target sasaran serta kurikulum. Pada kartu soal juga dilengkapi dengan kompetensi dasar, indikator pencapaian kompetensi (IPK), indikator soal serta level kognitif. Soal-soal yang dirumuskan pada kartu soal memiliki stimulus yang merupakan ciri khas soal-soal berbasis *HOTS*.

#### 6. Kunci Jawaban/ Pedoman penskoran

Kunci jawaban berfungsi untuk mempermudah siswa dalam mengoreksi dan mengevaluasi jawabannya. Pada kunci jawaban juga dilengkapi dengan langkah berpikir dan penyelesaian tahap-tahap analisis soal yang membantu guru dalam mengoreksi jawaban yang diberikan siswa dalam menyelesaikan soal-soal berbasis *HOTS*.

#### *Prototipe II*

Pada protipe II dilakukan evaluasi formatif berupa evaluasi diri sendiri (*self evaluation*). *Self evaluation* dilakukan dengan menggunakan daftar cek (*check list*) pada angket yang digunakan berdasarkan karakteristik instrumen tes berbasis *HOTS* yang dikembangkan. Tahap ini difokuskan pada kesalahan dalam instrumen tes berbasis *HOTS*, seperti kelengkapan komponen instrumen tes.

#### *Prototipe III*

Pada tahap ini dilakukan uji coba satu-satu dan penilaian Ahli. Uji coba satu-satu dilakukan kepada tiga guru yang berasal dari SMAN 1 Padang, SMAN 7 Padang dan SMAN 8 Padang. Berdasarkan hasil wawancara diperoleh hasil bahwa instrumen tes yang dikembangkan telah memenuhi kriteria *HOTS*. Instrumen tes yang dikembangkan sudah berada pada level kognitif C4 dan Stimulus yang disajikan pada soal dapat merangsang keterampilan berpikir siswa karena mengimplementasikan permasalahan yang terjadi pada kehidupan sehari-hari. Bahasa yang digunakan mudah dipahami siswa.

Penilaian ahli dilakukan uji validitas logis sebagai berikut:

#### *Uji Validitas Logis*

Uji validitas logis terdiri dari validitas isi (*content validity*) dan validitas konstruk

(*construc validity*). Validitas isi mencangkup stimulus soal, pertanyaan soal dan jawaban soal, sedangkan hasil uji validitas konstruk digunakan untuk mengungkapkan kelayakan materi, komponen kebahasaan, komponen penyajian, dan aturan tambahan pada instrumen tes berbasis *HOTS* (*higher order thiking skill*) pada materi hidrolisis garam yang dikembangkan. Uji validitas logis dilakukan oleh 7 orang validator yang meiputi 4 orang dosen dan 3 orang guru.

Rata-rata CVR validitas konten dari masing-masing komponen yang dinilai dari ketujuh validator tersebut, maka didapatkan hasil sebagai berikut:

Tabel 5. Rekapitulasi Hasil Validitas Konten dengan CVR

| Nomor Soal | CVR   | CVR Minimum | Keterangan    |
|------------|-------|-------------|---------------|
| 1          | 0,810 | 0,622       | Soal diterima |
| 2          | 0,905 | 0,622       | Soal diterima |
| 3          | 0,905 | 0,622       | Soal diterima |
| 4          | 1     | 0,622       | Soal diterima |
| 5          | 0,810 | 0,622       | Soal diterima |
| 6          | 1     | 0,622       | Soal diterima |
| 7          | 0,619 | 0,622       | Soal ditolak  |
| 8          | 0,905 | 0,622       | Soal diterima |
| 9          | 0,714 | 0,622       | Soal diterima |
| 10         | 0,905 | 0,622       | Soal diterima |
| 11         | 0,810 | 0,622       | Soal diterima |
| 12         | 0,524 | 0,622       | Soal ditolak  |
| 13         | 0,810 | 0,622       | Soal diterima |
| 14         | 0,810 | 0,622       | Soal diterima |
| 15         | 0,905 | 0,622       | Soal diterima |
| 16         | 0,714 | 0,622       | Soal diterima |
| 17         | 0,619 | 0,622       | Soal ditolak  |
| 18         | 1     | 0,622       | Soal diterima |
| 19         | 0,905 | 0,622       | Soal diterima |
| 20         | 0,810 | 0,622       | Soal diterima |
| 21         | 0,429 | 0,622       | Soal ditolak  |
| 22         | 0,810 | 0,622       | Soal diterima |
| 23         | 0,714 | 0,622       | Soal diterima |
| 24         | 0,905 | 0,622       | Soal diterima |
| 25         | 0,905 | 0,622       | Soal diterima |
| 26         | 0,905 | 0,622       | Soal diterima |
| 27         | 0,905 | 0,622       | Soal diterima |
| 28         | 0,810 | 0,622       | Soal diterima |
| 29         | 0,714 | 0,622       | Soal diterima |
| 30         | 0,714 | 0,622       | Soal diterima |

Berdasarkan data rekapitulasi diperoleh bahwa 4 butir soal ditolak karena memiliki nilai CVR kecil dari nilai minimum yang diberikan.

Soal yang ditolak adalah butir soal nomor 7, 12, 17 dan 21. Nilai CVI yang didapatkan adalah sebesar 0,841 yang menandakan soal yang terdapat pada instrumen tes berbasis *HOTS* pada materi hidrolisis garam sudah valid secara konten. Nilai CVI suatu instrumen tidak boleh lebih rendah dari 0,78 (Lynn, 1986).

Validitas konstruk dianalisis dengan Indeks *Aiken's V* yang menunjukkan hasil sebagai berikut:

Tabel 6. Hasil Validitas Konstruk dengan *Aiken's V* (V)

| Nomor Soal | V    | Kategori |
|------------|------|----------|
| 1          | 0,87 | Tinggi   |
| 2          | 0,80 | Sedang   |
| 3          | 0,85 | Tinggi   |
| 4          | 0,86 | Tinggi   |
| 5          | 0,88 | Tinggi   |
| 6          | 0,86 | Tinggi   |
| 7          | 0,81 | Tinggi   |
| 8          | 0,88 | Tinggi   |
| 9          | 0,80 | Sedang   |
| 10         | 0,87 | Tinggi   |
| 11         | 0,80 | Sedang   |
| 12         | 0,69 | Sedang   |
| 13         | 0,84 | Tinggi   |
| 14         | 0,85 | Tinggi   |
| 15         | 0,82 | Tinggi   |
| 16         | 0,74 | Sedang   |
| 17         | 0,82 | Tinggi   |
| 18         | 0,89 | Tinggi   |
| 19         | 0,87 | Tinggi   |
| 20         | 0,84 | Tinggi   |
| 21         | 0,76 | Sedang   |
| 22         | 0,76 | Sedang   |
| 23         | 0,78 | Sedang   |
| 24         | 0,83 | Tinggi   |
| 25         | 0,84 | Tinggi   |
| 26         | 0,85 | Tinggi   |
| 27         | 0,78 | Sedang   |
| 28         | 0,84 | Tinggi   |
| 29         | 0,78 | Sedang   |
| 30         | 0,86 | Tinggi   |

Keterangan V = Indeks *Aiken's V*

Berdasarkan hasil analisis dengan menggunakan Indeks *Aiken's V* terlihat bahwa butir soal pada instrumen tes berbasis *HOTS*

pada materi hidrolisis garam memiliki tingkat kevalidan yang sangat tinggi secara konstruk.

#### *Prototipe IV*

Pada tahap ini dilakukan uji coba kelompok kecil (*small group*) untuk menentukan reliabilitas instrumen tes berbasis HOTS yang telah valid. Peserta tes berjumlah 6 orang dengan masing-masing memiliki kemampuan yang berbeda-beda yaitu tinggi, sedang, dan rendah. Berdasarkan hasil analisis diperoleh bahwa nilai reliabilitas suatu instrumen tes sebesar 0,95 yang menunjukkan suatu instrumen tes memiliki nilai reliabilitas yang tinggi. Suatu tes dikatakan memiliki nilai reliabilitas yang tinggi apabila nilainya sama dengan atau lebih besar dari 0,70 (Sudjiono, 2011).

#### **Tahap Penilaian**

Pada tahap ini dilakukan uji lapangan terhadap 26 orang siswa SMAN 1 Padang untuk menentukan kualitas instrumen tes yang telah dihasilkan. Penjabarannya sebagai berikut:

##### 1. Validitas Empiris

Instrumen tes dikatakan memiliki validitas empiris apabila tes tersebut memiliki korelasi yang tinggi dengan suatu alat ukur lain pada bidang yang sama dan pada waktu yang sama pula (Purwanto, 2010). Nilai validitas empiris dianalisis dengan bantuan *software* ANATES. Berdasarkan hasil yang diperoleh, didapatkan 22 soal yang sangat signifikan dengan korelasi  $XY = 0,91$  yang menunjukkan bahwa kriteria validitas instrumen tes tersebut sangat tinggi.

##### 2. Reliabilitas

Suatu instrumen tes dikatakan reliabel jika instrumen tes tersebut akan memberikan hasil yang sama jika digunakan berulang-ulang kali untuk mengevaluasi objek yang sama (Latisma, 2011). Nilai reliabilitas dianalisis dengan bantuan *software* ANATES versi 4.0.9. Berdasarkan perhitungan, diperoleh bahwa nilai reliabilitas suatu instrumen tes yang diperoleh sebesar 0,95 yang menunjukkan suatu instrumen tes tersebut memiliki reliabilitas sangat tinggi.

##### 3. Indeks Kesukaran

Soal yang baik merupakan soal dengan tingkat kesukaran sedang (Arifin, 2012). Dalam penyusunan suatu instrumen tes perlu diperhatikan tingkat kesukarannya. Tingkatan indeks kesukaran terbagi tiga yaitu mudah, sedang dan sukar. Indeks kesukaran instrumen tes berbasis *HOTS* dianalisis menggunakan *software* ANATES versi 4.0.9. Berdasarkan hasil analisisnya diperoleh hasil bahwa semua

butir soal yang ada pada instrumen tes berbasis *HOTS* pada materi hidrolisis garam memiliki indeks kesukaran sedang.

##### 4. Daya Pembeda

Analisis daya pembeda soal bertujuan untuk membedakan siswa yang memiliki kemampuan tinggi dengan siswa yang berkemampuan rendah (Arikunto, 2015). Butir soal yang memiliki daya pembeda sangat baik berada pada rentang 0,70-1,00. Berdasarkan hasil analisis menggunakan *software* ANATES versi 4.0.9 diperoleh hasil bahwa 15 butir soal memiliki nilai indeks daya pembeda 1,00 yang menandakan butir soal tersebut memiliki daya pembeda yang baik sekali (Latisma, 2011). 7 butir soal memiliki nilai indeks daya pembeda sebesar 0,5 yang berarti bahwa butir soal tersebut memiliki daya pembeda yang baik, satu butir memiliki nilai indeks daya pembeda sebesar 0,00 yang menandakan bahwa butir soal tersebut memiliki daya pembeda yang lemah (jelek), sedangkan 3 butir soal memiliki nilai daya pembeda sebesar -0,5 yang menandakan bahwa butir soal tersebut memiliki kualitas daya pembeda yang sangat jelek.

#### *Diskusi*

Analisis kualitas tes berdasarkan reliabilitas, validitas empiris, indeks kesukaran, daya pembeda, dan fungsi diksator didapatkan setelah uji coba lapangan dengan 26 orang siswa. Nilai reliabilitas instrumen tes yang dihasilkan adalah sebesar 0,89 yang menunjukkan reliabilitas instrumen tes berbasis *HOTS* tersebut sangat tinggi. Reliabilitas yang didapatkan setelah uji lapangan dengan uji coba kelompok kecil memiliki perbedaan yang tidak signifikan sehingga dapat diabaikan. Instrumen tes berbasis *HOTS* pada materi hidrolisis garam telah baik dan dapat digunakan pada tempat dan waktu yang berbeda dan akan memberikan hasil yang relatif sama (Latisma, 2011).

Validitas empiris suatu instrumen tes berbasis *HOTS* pada materi hidrolisis garam menunjukkan bahwa 21 butir soal yang diujikan berkorelasi positif yang signifikan antara setiap butir soal dengan tes secara keseluruhan. 80% butir soal berkorelasi positif dengan instrumen tes secara keseluruhan sehingga instrumen tes berbasis *HOTS* ini telah dapat mengukur kemampuan *HOTS* siswa. 4 soal yang tidak signifikan menunjukkan bahwa butir soal tersebut tidak ada korelasi dengan tes secara keseluruhan.

Berdasarkan analisis indeks kesukaran instrumen tes, sebanyak 23 soal memiliki indeks kesukaran sedang. 1 butir soal berada pada kategori mudah dan 2 butir berada pada kategori sukar. 88% butir soal yang terdapat pada instrumen tes berbasis *HOTS* yang telah valid memiliki indeks kesukaran sedang sehingga soal tersebut dapat dikerjakan pada seluruh kalangan. Butir soal yang memiliki indeks kesukaran yang efektif jika soal tersebut dapat dijawab dengan benar oleh peserta tes dengan kemampuan tinggi, sedang dan rendah (Husnawati A, 2019). Butir soal yang memiliki kategori soal sukar yaitu butir soal nomor 11 dan 24 akan disimpan dan dikeluarkan pada tes-tes yang sifatnya ketat seperti seleksi. Soal dengan kategori mudah dibuang dan tidak lagi digunakan pada penilaian hasil belajar yang akan datang (Latisma, 2011).

Analisis dari segi daya pembeda didapatkan 16 soal dengan daya pembeda sangat baik dengan persentase 61%, 4 soal dengan daya pembeda baik dengan persentase 15%, 4 soal dengan kategori daya pembeda sedang dengan persentase 15% serta 2 soal dengan daya pembeda jelek yaitu bada butir soal dengan persentase 7%. Soal yang mempunyai daya pembeda yang baik adalah soal yang dapat membedakan kemampuan siswa yang berkemampuan tinggi dengan berkemampuan rendah, sehingga butir soal tersebut memiliki kategori sangat baik (Arikunto, 2015). Soal dengan diskriminasi yang buruk dapat disebabkan oleh kegagalan fungsi diktator (Latisma, 2015). Analisis fungsi diktator menunjukkan bahwa faktor interferensi yang ditetapkan ke opsi jawaban di setiap soal dapat berfungsi secara normal. Jika setidaknya 5% siswa memilih diktator pada setiap soal, hal ini menunjukkan para pengecoh dapat menjalankan fungsinya dengan benar (Sujiono, 2009).

### KESIMPULAN

Berdasarkan penelitian yang telah dilakukan diperoleh kesimpulan bahwa Instrumen tes berbasis *higher order thinking skill (HOTS)* pada materi hidrolisis garam untuk siswa SMA/MA yang dikembangkan telah valid secara logis dan empiris, memiliki reliabilitas yang sangat tinggi, serta menghasilkan kualitas item yang baik.

### DAFTAR PUSTAKA

- Arifin, Z. 2012. *Evaluasi Pembelajaran (Prinsip, Teknik, Prosedur)*. Bandung: Remaja Rosdakarya
- Arikunto, Suharsimi. 2015. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Hanifah, Nurdinah. 2019. Pengembangan Intrumen Penilaian Higher Order Thinking Skill (HOTS) disekolah dasar. *Current Research in Education: Conference Series Journal*. Vol. 1. No. 1.
- Husnawati, A. 2019. Pengembangan Soal Higher Order Thinking Skill (HOTS) Fisika Kelas VIII SMP Materi Gerak Pada Benda. *UPEJ Unnes Phys Educ J*. Vol. 8. No. 2.
- Kemendikbud. 2019. *Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skill)*. Jakarta: Direktorat pembinaan SMA Direktorat Jendral Pendidikan Dasar dan Menengah Departemen Pendidikan dan Kebudayaan.
- Kemendikbud. 2019. *Panduan Penulisan Soal HOTS (Higher Order Thinking Skill)*. Jakarta: Pusat Penilaian Pendidikan Badan Penelitian dan Pengembangan Kementerian Pendidikan dan Kebudayaan.
- Latisma. 2011. *Evaluasi Pendidikan*. Padang: UNP Press.
- Lawshe, C. H. 1985. *A Quantitative Approach to Content Validity*. Personel Psyclogy.
- Lynn, M. 1986. Determination and Quantification of Content Validity. *Nursing Research*, 35, 382-386
- OECD. 2019. *PISA 2018: Assesment and Analytical Frame Work: Science, Reading, Mathematic and Financial Literacy*. Paris: OECD Publishing.
- Plomp, T. dan N. N. 2013. *Education Design Research*. Enschede, Netherland: National Institute for Curriculum Development (SLO).
- Purwanto, N. 2012. *Prinsip-prinsip Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya.
- Puspendik. 2011. *Instrumen Penilaian Hasil Belajar Matematika SMP: Belajar dari PISA dan TIMMS*. Jakarta: Balitbang Depdiknas.

- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian*. Yogyakarta: Parama Publishing.
- Sudjiono, A. 2009. *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Press.
- Sudjiono, A. 2011. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
- Sugiyono. 2012. *Metode Penelitian Pendidikan*. Bandung: Alfabeta.
- Wilson, F. R, et al. 2012. "Recalculation of the Critical Values For Lawshe's Content Validity Ratio". *Measurement and Evaluation In Counseling and Development*. 45(3): 197-210.