

**PENGEMBANGAN INSTRUMEN MULTIPLE CHOICE FOUR-TIER DIAGNOSTIC TEST PADA MATERI HIDROLISIS GARAM UNTUK PESERTA DIDIK KELAS XI SMA**

**Dina Azizatul Islamiah**

Universitas Wahidiyah, Kediri, Indonesia, Email: [dinaazizatulislamiah@gmail.com](mailto:dinaazizatulislamiah@gmail.com)

**Firmansyah**

Universitas Wahidiyah, Kediri, Indonesia, Email: [firmansyah@uniwa.ac.id](mailto:firmansyah@uniwa.ac.id)

**Abstrak**

Pengembangan ini didasarkan pada masih seringnya terjadi miskonsepsi peserta didik terhadap konsep tersebut. Pengembangan ini bertujuan untuk mengembangkan dan menguji kevalidan serta keterbacaan instrumen *multiple choice four-tier diagnostic test* pada materi hidrolisis garam sebagai alat identifikasi miskonsepsi peserta didik kelas XI SMA. Tes diagnostik *four-tier* dikembangkan sebagai alternatif evaluasi pembelajaran yang lebih efektif dibandingkan tes tradisional. Pengembangan dilakukan dengan pendekatan *Research and Development (R&D)* dengan model 4D (*Define, Design, Develop, Disseminate*) berdasarkan model Sugiyono (2016), mencakup proses validasi ahli, uji keterbacaan, dan analisis kualitas soal. Instrumen pengumpulan data terdiri dari angket validasi ahli, angket keterbacaan, dan tes diagnostik *four-tier* yang mencakup empat tingkatan: pilihan jawaban, keyakinan terhadap jawaban, alasan, dan keyakinan terhadap alasan. Data dianalisis menggunakan uji validitas isi dan validitas butir soal, reliabilitas (dengan koefisien Alpha), tingkat kesukaran, dan daya pembeda. Setelah direvisi, instrumen diuji pada 33 pesertadidik kelas XI SMA Wahidiyah Kepanjen. Hasil validasi oleh tiga ahli materi menunjukkan bahwa instrumen tergolong sangat valid dengan revisi minor pada redaksi soal dan kesesuaian indikator. Uji coba kepada 33 peserta didik menghasilkan 25 soal valid dan 5 soal tidak valid berdasarkan nilai  $r_{\text{hitung}} > r_{\text{tabel}}$  (0,3338). Keterbacaan produk berada pada kategori sangat layak dengan skor rata-rata 87%. Uji reliabilitas menghasilkan nilai  $r_{11} = 0,969$ , menunjukkan bahwa instrumen tergolong sangat reliabel. Analisis tingkat kesukaran menunjukkan 16 soal mudah, 10 sedang, dan 4 sulit. Daya pembeda menunjukkan 19 soal sangat baik, 5 baik, 5 cukup, dan 1 jelek. Dengan demikian, instrumen ini dinyatakan valid, reliabel, layak digunakan, dan efektif untuk mendeteksi miskonsepsi pesertadidik dalam materi hidrolisis garam.

**Kata Kunci:** *Diagnostic Test Multiple Choice, Four-Tier, Hidrolisis Garam.*

**PENDAHULUAN**

Pendidikan merupakan proses penting dalam membentuk pribadi yang berkualitas, berpengetahuan luas, serta mampu menghadapi tantangan kehidupan yang semakin kompleks. Salah satu indikator keberhasilan pendidikan adalah terciptanya pembelajaran yang efektif, yakni pembelajaran yang mampu mengembangkan potensi peserta didik secara maksimal melalui kombinasi materi, fasilitas, metode, dan peran aktif guru. Hal ini sejalan dengan amanat Undang-Undang No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional, yang menegaskan bahwa pendidikan nasional bertujuan mengembangkan potensi peserta didik agar menjadi pribadi yang beriman, cerdas, mandiri, dan bertanggung jawab sebagai warga negara.

Dalam kerangka pencapaian tujuan tersebut, kurikulum memiliki peran vital sebagai pedoman dalam proses pembelajaran. Kurikulum Merdeka yang saat ini diterapkan di Indonesia menekankan pada pembelajaran yang berpusat pada peserta didik, dengan guru berperan sebagai fasilitator. Pembelajaran tidak hanya bertumpu pada penguasaan materi, tetapi juga pengembangan

keterampilan berpikir kritis, komunikasi, dan pemecahan masalah.

Ilmu kimia sebagai bagian dari rumpun sains memiliki karakteristik tersendiri yang menuntut kemampuan peserta didik dalam memahami konsep secara abstrak dan multidimensional. Salah satu materi yang menantang dalam pembelajaran kimia adalah hidrolisis garam. Materi ini menuntut keterampilan dalam mengintegrasikan tiga level representasi kimia makroskopik, mikroskopik, dan simbolik yang kerap menjadi hambatan bagi peserta didik. Ketidakmampuan dalam menghubungkan ketiga representasi ini seringkali menjadi sumber munculnya *miskonsepsi*, yakni pemahaman yang keliru dan tidak sesuai dengan konsep ilmiah sebenarnya.

Hasil wawancara dengan guru kimia di SMA Wahidiyah Kepanjen menunjukkan bahwa peserta didik mengalami kesulitan dalam memahami konsep hidrolisis garam,

terutama dalam aspek perhitungan dan perbedaan konsep antara larutan penyangga dan larutan garam. Banyak peserta didik juga mengalami kesulitan dalam menyelesaikan soal-soal hitungan karena lemahnya pemahaman konsep dasar asam-basa serta kompleksitas rumus dan langkah-langkah penyelesaian soal. Miskonsepsi yang tidak segera diidentifikasi dan

diluruskan dapat menghambat pemahaman lanjutan dan menurunkan minat belajar peserta didik terhadap kimia.

Dalam rangka mengatasi masalah tersebut, diperlukan instrumen yang mampu mengidentifikasi secara akurat letak miskonsepsi yang dimiliki peserta didik. Salah satu pendekatan yang efektif adalah dengan menggunakan *multiple choice four-tier diagnostic test*, yaitu alat evaluasi yang tidak hanya menguji jawaban peserta didik, tetapi juga mengungkap tingkat keyakinan dan alasan di balik jawaban tersebut. Instrumen ini telah terbukti lebih akurat dalam mengidentifikasi miskonsepsi dibandingkan tes dua tingkat atau tiga tingkat, karena mampu memetakan struktur kognitif peserta didik secara lebih menyeluruh.

Dengan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan instrumen *multiple choice four-tier diagnostic test* pada materi hidrolisis garam yang dapat digunakan sebagai alat diagnostik untuk mengidentifikasi dan menganalisis miskonsepsi peserta didik kelas XI SMA. Instrumen ini diharapkan dapat membantu guru dalam merancang strategi pembelajaran yang lebih tepat sasaran dan efektif dalam memperbaiki pemahaman konsep kimia peserta didik.

## METODE

Pada dasarnya bagian ini menjelaskan bagaimana penelitian itu dilakukan. Materi pokok bagian ini adalah: (1) rancangan penelitian; (2) populasi dan sampel (sasaran penelitian); (3) teknik pengumpulan data dan pengembangan instrumen; (4) dan teknik analisis data. Untuk penelitian yang menggunakan alat dan bahan, perlu dituliskan spesifikasi alat dan bahannya. Spesifikasi alat menggambarkan kecanggihan alat yang digunakan sedangkan spesifikasi bahan menggambarkan macam bahan yang digunakan.

Penelitian ini menggunakan metode Research and Development (R&D) mengacu pada model pengembangan 4D (*Define, Design, Develop, Disseminate*) yang dikemukakan oleh Thiagarajan dkk. (1974). Penelitian bertujuan untuk mengembangkan instrumen *multiple choice four-tier diagnostic test* pada materi hidrolisis garam bagi peserta didik kelas XI SMA.

Tahap Penelitian

### 1) Define (Pendefinisian)

Tahap ini mencakup analisis kurikulum, wawancara dengan guru kimia, serta identifikasi miskonsepsi melalui studi pendahuluan di SMA Wahidiyah Kepanjen. Hasil analisis digunakan untuk merumuskan indikator dan kisi-kisi soal yang sesuai dengan Capaian Pembelajaran (CP) dan Tujuan Pembelajaran (TP) Kurikulum Merdeka.

### 2) Design (Perancangan)

Perancangan instrumen dilakukan dengan menyusun kisi-kisi, kartu soal, dan naskah soal *four-tier diagnostic test* sebanyak 30 butir. Setiap soal terdiri dari empat tingkatan: pilihan jawaban, tingkat keyakinan, alasan jawaban, dan keyakinan terhadap

alasan.

### 3) Develop (Pengembangan)

Instrumen divalidasi oleh tiga ahli materi (dua guru kimia dan satu dosen). Validasi dilakukan melalui angket yang menilai aspek materi, konstruksi, dan bahasa. Revisi dilakukan berdasarkan saran validator. Selanjutnya, uji keterbacaan dilakukan terhadap 33 peserta didik kelas XI.

### 4) Disseminate (Penyebaran)

Tahap penyebaran tidak dilakukan dalam penelitian ini karena keterbatasan waktu. Penelitian berakhir pada tahap pengembangan.

Instrumen dan teknik pengumpulan data yang digunakan meliputi:

- 1) Angket validasi ahli dan angket keterbacaan peserta didik
- 2) Tes diagnostik *four-tier* untuk mendeteksi miskonsepsi secara mendalam.
- 3) Wawancara terstruktur untuk menggali informasi pendukung.
- 4) Dokumentasi untuk melengkapi data pengembangan.

Teknik Analisis Data

- 1) Validasi isi diuji melalui penilaian ahli, sedangkan validitas butir soal dianalisis menggunakan korelasi biserial.
- 2) Reliabilitas diuji menggunakan rumus KR-20 (Kuder Richardson).
- 3) Tingkat kesukaran dihitung dengan rumus proporsi jawaban benar.
- 4) Daya pembeda dianalisis berdasarkan perbedaan proporsi jawaban benar antara kelompok atas dan bawah.
- 5) Kriteria analisis mengacu pada pedoman Arikunto, Riduwan, dan sumber-sumber terkait lainnya.

## HASIL DAN PEMBAHASAN

### Hasil Pengembangan

Hasil pengembangan dan pengembangan yang telah dilakukan oleh peneliti yaitu instrumen *multiple choice four-tier diagnostic test* untuk mengidentifikasi miskonsepsi peserta didik pada materi hidrolisis garam. Instrumen *multiple choice four-tier diagnostic test* merupakan pengembangan lebih lanjut dari tes diagnostik *three-tier*. Pengembangan instrumen tes diagnostik *four-tier* dengan penambahan tingkat keyakinan pilihan jawaban dan alasan yang dipilih oleh peserta didik (Caleon & Subramaniam, 2010). Tes diagnostik *four-tier* terdiri dari empat tingkat yaitu tingkat pertama pilihan jawaban, tingkat kedua tingkat keyakinan pilihan jawaban, tingkat

ketiga pilihan alasan, dan tingkat keempat tingkat keyakinan pilihan alasan.

Komponen utama instrumen tes diagnostik *four-tier* yang dikembangkan adalah kisi-kisi soal, petunjuk pengerjaan soal, kartu soal, kunci jawaban, dan pedoman interpretasi hasil tes diagnostik *four-tier*. Instrumen tes diagnostik *four-tier* divalidasi oleh tiga ahli materi (validator) dan di ujikan kepada pesertadidik kelas XI SMA Wahidiyah Kepajen. Hasil pengembangan yang telah dilakukan melalui model 4-D menghasilkan suatu produk berupa soal *multiple choice four-tier diagnostic tes*.

### Uji Kelayakan

Hasil pengembangan telah melalui beberapa tahap pengumpulan data dengan beberapa teknik analisis data menghasilkan data sebagai berikut:

#### 1) Hasil uji validitas

Hasil angket kartu soal yang telah dibuat dan divalidasi oleh para ahli dengan mengisi angket yang telah diberikan. Hasil angket validasi yang telah divalidasi oleh para ahli menghasilkan data sebagai berikut:

**Tabel 4. 1 Rekapitulasi Angket Hasil Validator**

| No        | Aspek yang di nilai | Total skor |            |            | Rata-rata | Persentase validitas | Kelayakan    |
|-----------|---------------------|------------|------------|------------|-----------|----------------------|--------------|
|           |                     | Validasi 1 | Validasi 2 | Validasi 3 |           |                      |              |
| 1         | Materi              | 109        | 148        | 133        | 130       | 86%                  | Sangat layak |
| 4         | Konstruksi          | 104        | 140        | 131        | 125       | 83%                  | Sangat layak |
| 3         | Bahasa              | 104        | 149        | 142        | 132       | 88%                  | Sangat layak |
| Total     |                     |            |            |            | 387       | 257%                 | Sangat layak |
| Rata-rata |                     |            |            |            | 129       | 86%                  | Sangat layak |

Berdasarkan hasil dari penilaian validator pada tabel 4.2 menyatakan bahwa produk yang di buat layak untuk digunakan untum mengidentifikasi miskonsepsi. Akan tetapi dalam hal kelayakan ini produk juga ada beberapa soal yang harus revisi bedasarkan kritik dan saran dari validator, adapun soal-soal yang harus direvisi dan tampilan yang sudah di revisi.

#### 2) Uji validitas butir soal

Hasil uji coba instrumen tes digunakan untuk memvalidasi instrumen tes diagnostik *four-tier*. Produk yang telah selesai divalidasi dan direvisi selanjutnya akan diujikan kepada pesertadidik. Tujuan uji validitas butir soal adalah untuk mengetahui valid atau tidaknya setiap butir soal yang dikembangkan. Uji coba ini dilakukan kepada pesertadidik-siswi kelas XI SMA Wahidiyah Kepanjen. Jumlah responden dalam uji coba ini adalah pesertadidik yang berjumlah 33 orang. Dengan mengerjakan soal *multiple choice four-tier diagnostic test*.

Penilaian setiap item instrumen tes diagnostik *four-*

*tier* menggunakan skala nol hingga satu sesuai petunjuk pengisian. Penskoran lembar validasi dijelaskan pada petunjuk pengisian lembar validasi instrumen tes diagnostik *four-tier* yang terdapat pada Lampiran 7. Penilaian jika tier 1 benar dan tier 3 benar berarti nilainya adalah 1, jika tier-1 salah dan tier-3 benar maka nilainya adalah 0, jika tier-1 benar dan tier-3 salah maka nilainya 0, jika tier-1 dan tier-3 salah maka nilainya adalah 0.

Uji validitas soal dihitung dengan menggunakan koefisien korelasi *point biserial* antara skor butir soal data yang berskala diskrit dengan skor total yang diperoleh tes, kriteria soal dikatakan valid jika  $r_{hitung} \geq r_{tabel}$ , sedangkan  $r_{hitung} \leq r_{tabel}$  dikatakan tidak valid dengan tingkat signifikansi 5% (Arikunto, S. 2012). Nilai  $r_{tabel}$  untuk 33 sampel dengan taraf signifikansi 5% yaitu 0,3338. Hasil analisis uji validitas soal dapat dilihat pada Lampiran 7.

Berdasarkan hasil analisis validitas terhadap 30 butir soal menggunakan bantuan perangkat lunak SPSS versi 0.30, diperoleh nilai koefisien korelasi ( $r_{hitung}$ ) dan nilai signifikansi (*Sig. 2-tailed*). Kriteria pengambilan keputusan adalah sebagai berikut.

**Tabel 4. 2 Rekapitulasi Uji Butir Soal**

| No Soal | Pearson Correlation | $r_{tabel}$ | Nilai Sig. (2-tailed) | Taraf signifikan | Valid/Tidak |
|---------|---------------------|-------------|-----------------------|------------------|-------------|
| 1       | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 2       | 0.847               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 3       | 0.757               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 4       | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 5       | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 6       | 0.386               | 0,3338      | 0.026                 | 0,05             | Valid       |
| 7       | 0.292               | 0,3338      | 0.100                 | 0,05             | Tidak       |
| 8       | 0.847               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 9       | 0.757               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 10      | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 11      | 0.292               | 0,3338      | 0.100                 | 0,05             | Tidak       |
| 12      | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 13      | 0.805               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 14      | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 15      | 0.942               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 16      | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 17      | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 18      | 0.292               | 0,3338      | 0.100                 | 0,05             | Tidak       |
| 19      | 0.526               | 0,3338      | 0.002                 | 0,05             | Valid       |
| 20      | 0.292               | 0,3338      | 0.100                 | 0,05             | Tidak       |
| 21      | 0.084               | 0,3338      | 0.643                 | 0,05             | Tidak       |
| 22      | 0.819               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 23      | 0.779               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 24      | 0.805               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 25      | 0.526               | 0,3338      | 0.002                 | 0,05             | Valid       |
| 26      | 0.932               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 27      | 0.779               | 0,3338      | 0.000                 | 0,05             | Valid       |
| 28      | 0.526               | 0,3338      | 0.002                 | 0,05             | Valid       |
| 29      | 0.526               | 0,3338      | 0.002                 | 0,05             | Valid       |
| 30      | 0.779               | 0,3338      | 0.000                 | 0,05             | Valid       |

Dari hasil tabel 4.3 rekapitulasi uji butir soal berdasarkan hasil pengerjaan soal oleh pesertadidik

kelas XI SMA Wahidiyah Kepanjen dengan jumlah 33 pesertadidik, menyatakan bahwa nilai  $t_{hitung} > t_{tabel}$  menyatakan bahwa soal yang di katakan valid, dan hasil dari nilai ( $Sig.2-tailed$ )  $> 0,05$  menyatakan setiap butir soal valid, maka butir soal yang di ujikan layak untuk di gunakan untuk mendeteksi miskonsepsi pada materi hidrolisis garam. Terdapat 5 butir soal yang tidak valid kelima soal tersebut tidak dapat digunakan untuk langkah selanjutnya, jadi kelima soal tersebut harus dibuang.

### 3) Uji keterbacaan

Uji coba ini dilakukan kepada pesertadidik-siswi kelas XI SMA Wahidiyah Kepanjen. Jumlah responden dalam uji coba ini adalah pesertadidik satu kelas yang berjumlah 33 orang. Uji coba keterbacaan produk dilakukan untuk mengetahui respon pesertadidik terhadap produk yang dikembangkan. Respon pesertadidik akan menunjukkan tingkat kelayakan produk tersebut untuk digunakan.

Uji coba keterbacaan produk dilakukan oleh pesertadidik dengan mengisi angket yang telah disediakan. Pengisian lembar angket keterbacaan dilakukan oleh pesertadidik setelah mengerjakan soal *four-tier* dan membaca produk yang diujikan. Kemudian mengisi angket keterbacaan untuk menilai produk yang telah mereka kerjakan dan baca. Angket keterbacaan tersebut berisi tentang petunjuk pengisian angket dan pedoman pemberian skor yang dimuat dalam judul, dan dibawahnya berisi kriteria penilaian yang berisi 10 aspek penilaian seperti yang terdapat pada lampiran 8. Angket keterbacaan pesertadidik dibuat dalam *g-form* dengan 10 aspek yang dinilai.

**Tabel 4.3 Rekapitulasi Hasil Keterbacaan Peserta didik**

| Aspek            | Rata-rata skor | Nilai tertinggi | Persentase  | Kriteria            |
|------------------|----------------|-----------------|-------------|---------------------|
| 1                | 4,15           | 5               | 83%         | Sangat layak        |
| 2                | 4,36           | 5               | 87,27%      | Sangat layak        |
| 3                | 4,21           | 5               | 84,24%      | Sangat layak        |
| 4                | 4,27           | 5               | 85,45%      | Sangat layak        |
| 5                | 4,36           | 5               | 87,27%      | Sangat layak        |
| 6                | 4,21           | 5               | 84,24%      | Sangat layak        |
| 7                | 4,42           | 5               | 88,48%      | Sangat layak        |
| 8                | 4,48           | 5               | 90%         | Sangat layak        |
| 9                | 4,39           | 5               | 88%         | Sangat layak        |
| 10               | 4,36           | 5               | 87,27%      | Sangat layak        |
| <b>Jumlah</b>    | <b>43.21</b>   | <b>50</b>       | <b>865%</b> |                     |
| <b>Rata-rata</b> | <b>4.321</b>   | <b>5</b>        | <b>87%</b>  | <b>Sangat Layak</b> |

Berdasarkan rekapitulasi hasil uji keterbacaan produk oleh pesertadidik pada tabel 4.4 diperoleh skor rata-rata dan persentase pada setiap butir aspek penilaian. Skor rata-rata dan persentase dihitung untuk setiap aspek berdasarkan jumlah responden yang mengisi angket yakni 33 pesertadidik. Dapat dilihat bahwa masing-masing butir aspek penilaian didapatkan hasil skor rata-rata 4 termasuk dalam kriteria baik sebagaimana yang terdapat pada lampiran 9.

Kemudian dari keseluruhan aspek dihitung dan mendapatkan skor rata-rata 4,32 yang termasuk dalam kriteria baik. Persentase setiap aspek penilaian yang diisi oleh pesertadidik menunjukkan bahwa nilai terendah berada pada butir aspek 1,3 dan 6 dengan persentase 83% dan 84% yang tergolong kriteria layak sedangkan butir aspek lainnya memiliki persentase 85% keatas yang tergolong kriteria sangat layak. Secara keseluruhan tampak bahwa rata-rata respon pesertadidik terhadap produk sebesar 87%. Jika menggunakan kriteria pada tabel 3.2 maka produk yang dikembangkan termasuk dalam kriteria sangat layak.

### 4) Uji reliabilitas

Uji reliabilitas berfungsi untuk mengetahui reliabel atau tidaknya suatu instrumen soal. Jika nilai koefisien reliabilitas  $> 0,6$  maka instrumen memiliki reliabilitas yang baik dan dapat dipercaya (reliabel). Jika nilai koefisien reliabilitas  $< 0,6$  maka instrumen tidak memiliki reliabilitas yang baik dan tidak dapat dipercaya (tidak reliabel) (Arikunto, S. 2012). Interpretasi nilai koefisien reliabilitas  $0.60 < r_{11} \leq 0.80$ : Reliabilitas tinggi (instrumen cukup konsisten). Nilai  $t_{tabel}$  untuk 33 sampel dengan taraf signifikansi 5% yaitu 0,3338.

**Tabel 4. 4 Hasil Uji Reliabilitas**

| Reliability Statistics |            |
|------------------------|------------|
| Cronbach's Alpha       | N of Items |
| 0.969                  | 30         |

Berdasarkan output *Reliability Statistics*, diperoleh nilai *Cronbach's Alpha* menggunakan bantuan perangkat lunak SPSS versi 0.30, menghasilkan nilai sebesar **0,969** yang hasilnya  $> 0,60$  dengan jumlah butir instrumen sebanyak **30 item**. Hasil hitung  $r_{11} > t_{tabel}$  menunjukkan bahwa instrumen tes yang dikembangkan dikatakan **sangat reliabel** dengan kriteria sangat tinggi. Nilai ini menunjukkan bahwa instrumen memiliki tingkat reliabilitas yang sangat tinggi. Nilai reliabilitas di atas 0,90 dikategorikan sebagai **sangat reliabel**, yang berarti bahwa item-item dalam instrumen tersebut memiliki konsistensi internal yang sangat kuat dalam mengukur konstruk yang dimaksud (Ghozali. 2016). Dengan demikian, instrumen ini dapat dinyatakan layak dan andal untuk digunakan dalam pengumpulan data pengembangan.

### 5) Tingkat kesukaran

Uji tingkat kesukaran soal berguna untuk menentukan tingkatan soal yang mudah, sedang, dan sukar (Arikunto, S. 2012). Uji tingkat kesukaran soal dihitung menggunakan aplikasi SPSS versi 0.30 tingkat kesukaran dengan hasil analisis tingkat kesukaran soal tes diagnostik *four-tier* dapat dilihat pada Lampiran.

**Tabel 4. 5 Rekapitulasi Tinglat Kesukaran**

| Kategori | No soal                                                   | Jumlah soal |
|----------|-----------------------------------------------------------|-------------|
| Mudah    | 1, 4, 5, 6, 7, 10, 12, 14, 15, 16, 17, 19, 25, 26, 28, 29 | 16          |
| Sedang   | 2, 3, 8, 9, 13, 21, 22, 23, 24, 27                        | 10          |
| Sulit    | 11, 18, 20, 30                                            | 4           |

Berdasarkan tabel 4.27 dengan bantuan aplikasi SPSS versi 0.30 menentukan beberapa butir soal dengan 3 kriteria. Soal terdiri dari 4 butir berkategori sulit, 10 butir soal sedang, 16 soal mudah. Hal ini menunjukkan kemampuan soal dalam membedakan peserta didik berkemampuan tinggi dan rendah

#### 6) Daya Pembeda

Daya pembeda soal digunakan untuk mengetahui kemampuan suatu soal dalam memilah tingkat kemampuan peserta didik (Arikunto, S. 2012). Hasil tersebut dapat mengungkapkan kemampuan peserta didik yang sudah atau belum menguasai materi. Uji daya pembeda soal dihitung menggunakan rumus daya pembeda dengan hasil analisis daya pembeda soal dapat dilihat pada Lampiran.

**Tabel 4. 6 Rekapitulasi Daya Pembeda**

| No Soal | Corrected Item-Total Correlation | Kriteria    |
|---------|----------------------------------|-------------|
| 1       | 0.926                            | Baik sekali |
| 2       | 0.832                            | Baik sekali |
| 3       | 0.735                            | Baik sekali |
| 4       | 0.926                            | Baik sekali |
| 5       | 0.926                            | Baik sekali |
| 6       | 0.359                            | Cukup       |
| 7       | 0.268                            | Cukup       |
| 8       | 0.832                            | Baik sekali |
| 9       | 0.735                            | Baik sekali |
| 10      | 0.926                            | Baik sekali |
| 11      | 0.268                            | Cukup       |
| 12      | 0.926                            | Baik sekali |
| 13      | 0.786                            | Baik sekali |
| 14      | 0.926                            | Baik sekali |
| 15      | 0.936                            | Baik sekali |
| 16      | 0.926                            | Baik sekali |
| 17      | 0.926                            | Baik sekali |
| 18      | 0.268                            | Cukup       |
| 19      | 0.490                            | Baik        |
| 20      | 0.268                            | Cukup       |
| 21      | 0.048                            | Jelek       |
| 22      | 0.801                            | Baik sekali |
| 23      | 0.758                            | Baik sekali |
| 24      | 0.786                            | Baik sekali |
| 25      | 0.490                            | Baik        |
| 26      | 0.926                            | Baik sekali |
| 27      | 0.758                            | Baik sekali |
| 28      | 0.490                            | Baik        |
| 29      | 0.490                            | Baik        |
| 30      | 0.758                            | Baik        |

Dengan bantuan aplikasi SPSS versi 0.30 dapat direkapitulasi hasil daya pembeda menghasilkan soal dengan kriteria jelek, cukup, baik dan baik sekali. Terdapat 1 butir soal dengan kriteria jelek, terdapat 5 butir soal dengan kriteria cukup, terdapat 5 dengan kriteria baik, dan terdapat 19 butir soal dengan kriteria baik sekali.

## PENUTUP

### Simpulan

Simpulan menyajikan ringkasan dari uraian mengenai hasil dan pembahasan, mengacu pada tujuan penelitian. Berdasarkan kedua hal tersebut dikembangkan pokok-pokok pikiran baru yang merupakan esensi dari temuan penelitian.

Berdasarkan hasil pengembangan dan pembahasan yang telah dilakukan mengenai pengembangan instrumen *multipel choice four-tier diagnostic test* pada materi hidrolisis garam, maka dapat disimpulkan bahwa instrumen *multipel choice four-tier diagnostic test* yang dikembangkan dalam pengembangan ini terbukti valid, reliabel, dan layak digunakan sebagai alat diagnosis miskonsepsi peserta didik pada materi hidrolisis garam. Pengembangan instrumen ini merupakan inovasi dari model three-tier dengan menambahkan aspek tingkat keyakinan terhadap jawaban dan alasan yang dipilih peserta didik, sehingga mampu memberikan informasi diagnostik yang lebih mendalam. Validasi oleh ahli materi dan instrumen menunjukkan bahwa sebagian besar butir soal memenuhi kriteria kelayakan setelah dilakukan revisi berdasarkan saran para validator. Hasil uji coba pada 33 peserta didik, kelas XI menunjukkan bahwa dari 30 butir soal, sebanyak 25 soal dinyatakan valid dengan  $r_{tt} > r_{tabel}$ , dan instrumen memiliki reliabilitas sangat tinggi dengan nilai  $r_{tt}$  sebesar 0,969. Tingkat kesukaran soal terbagi dalam kategori mudah (16 soal), sedang (10 soal), dan sukar (4 soal), sedangkan daya pembeda soal menunjukkan 19 butir soal dalam kategori sangat baik, 5 baik, 5 cukup, dan 1 soal berkategori jelek. Selain itu, hasil uji keterbacaan menunjukkan bahwa instrumen dipahami dengan baik oleh peserta didik, dengan skor rata-rata 4,32 dan persentase keterbacaan sebesar 87% yang termasuk dalam kategori sangat layak. Berdasarkan hasil tersebut, dapat disimpulkan bahwa instrumen diagnostik four-tier ini memenuhi kriteria kevalidan, keandalan, keterbacaan, dan efektivitas dalam mengidentifikasi miskonsepsi peserta didik secara akurat dan komprehensif.

#### DAFTAR PUSTAKA

- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974) .  
*Instructional Development for Training Teachers of  
Exceptional Children: A Sourcebook*. Indiana  
University: Center for Innovation in Teaching the  
Handicapped.
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974).  
*The 4-D Model: Aframework for Developing  
Educational Materials*. In S. Thiagarajan, D. S.  
Semmel, & M. I. Semmel, *Istructional Development  
for Training Teachers of Exceptional Children*.  
Indiana University.
- Arikunto, S. (2012). Dasar-dasar evaluasi pendidikan  
(edisi revisi). Jakarta: Bumi Aksara.
- Arikunto, S, (2012) Dasar-dasar Evaluasi Pendidikan:  
Bumi Aksara.

