

## Pengembangan Instrumen Tes Diagnostik *Four-tier* untuk Mengidentifikasi Model Mental Peserta Didik pada Materi Larutan Penyangga

Wardah Halizah<sup>1</sup> and Fajriah Azra<sup>1\*</sup>

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

\*Email: [bunda\\_syasfa@yahoo.com](mailto:bunda_syasfa@yahoo.com)

### ABSTRACT

Mental models are referred to as ideas in a person's mind in describing and explaining a phenomenon. Students' mental models can be explored by administering a test, which is referred to as a mental model diagnostic test. The mental model diagnostic test aims to measure students' ability to explain data, predict and explain chemical phenomena macroscopically by involving reasoning at the submicroscopic level using chemical symbolic language. Therefore, the purpose of this study was to develop a Four-tier Diagnostic Test instrument to identify students' mental models in buffer solution material that met good criteria. This research is Research and Development (R&D) using the diagnostic test development model that has been carried out by Treagust (1988). Based on the results of logical validity analysis, it was found that the average value of validity (V) was 0.84 with a valid category. Based on the results of trials on students, the first-tier and third-tier reliability values were 0.88 and 0.89 in the very high category. The results showed that the developed four-tier diagnostic test instrument was valid and met good criteria.

*Keywords:* Diagnostic Test, Four-tier Multiple Choice, Mental Models, Buffer Solution

### ABSTRAK

Model mental disebut sebagai ide-ide dalam pikiran seseorang dalam menggambarkan dan menjelaskan suatu fenomena. Model mental peserta didik dapat digali dengan pemberian suatu tes, yang disebut sebagai tes diagnostik model mental. Tes diagnostik model mental bertujuan untuk mengukur kemampuan peserta didik dalam menjelaskan data, memprediksi dan menjelaskan fenomena kimia secara makroskopik dengan melibatkan penalaran pada level submikroskopik menggunakan bahasa simbolik kimia. Oleh karena itu, tujuan dari penelitian ini adalah untuk mengembangkan instrumen tes diagnostik *four-tier* untuk mengidentifikasi model mental peserta didik pada materi larutan penyangga yang memenuhi kriteria yang baik. Penelitian ini merupakan penelitian pengembangan atau *Research and Development (R&D)* dengan menggunakan model pengembangan tes diagnostik yang telah dilakukan oleh Treagust (1988). Berdasarkan hasil analisis validitas logis didapatkan bahwa nilai rata-rata validitas (V) sebesar 0,84 dengan kategori valid. Berdasarkan hasil uji coba terhadap peserta didik, didapatkan nilai reliabilitas pada *first-tier* dan *third-tier* sebesar 0,88 dan 0,89 dengan kategori sangat tinggi. Hasil penelitian menunjukkan instrumen tes diagnostik *four-tier* yang dikembangkan sudah valid dan memenuhi kriteria yang baik.

**Kata Kunci:** Tes Diagnostik, Pilihan Ganda *Four-tier*, Model Mental, Larutan Penyangga

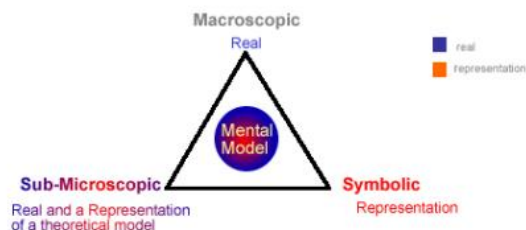
## PENDAHULUAN

Kimia adalah ilmu sains yang mempelajari tentang materi dan sifat-sifatnya, perubahan materi, dan perubahan energi yang menyertai perubahan materi tersebut (Brady dkk., 2012). Konsep-konsep kimia yang bersifat abstrak dan kompleks menimbulkan kesulitan peserta didik dalam mempelajari kimia (Ni Made, 2022). Materi kimia dengan konsep abstrak dapat digambarkan dengan multirepresentasi kimia. Multirepresentasi kimia perlu diperhatikan dan dipahami dalam mempelajari kimia. Multirepresentasi tersebut yaitu representasi makroskopik, submikroskopik, dan simbolik (Chandrasegaran dkk., 2007).

Materi larutan penyangga merupakan salah satu materi yang memiliki konsep abstrak sehingga peserta didik sulit untuk memahami materi tersebut. Berdasarkan penelitian terdahulu oleh Orgill dan Sutherland (2008) ditemukannya kesulitan peserta didik dalam menjawab soal-soal dan adanya beberapa miskonsepsi pada konsep larutan penyangga. Hasil penelitian tersebut menyatakan bahwa pada konsep larutan penyangga peserta didik menemukan beberapa kesulitan diantaranya sulit membedakan asam kuat dan asam lemah, tidak mampu merepresentasikannya dalam bentuk simbolik serta tidak mampu mengkoneksikan representasi makroskopik dan submikroskopik.

Multirepresentasi kimia berkontribusi pada pengembangan model mental peserta didik (Chittleborough, 2004). Model mental sebagai pusat segitiga yang menghubungkan ketiga representasi kimia. Chittleborough, (2004) mengungkapkan hubungan antara model mental dan multirepresentasi kimia terlihat pada

Gambar 1. Hubungan multirepresentasi kimia dengan model mental (Chittleborough, 2004).



Gambar 1. Hubungan multirepresentasi kimia dengan model mental (Chittleborough, 2004).

Model mental disebut sebagai ide-ide dalam pikiran seseorang dalam menggambarkan dan menjelaskan suatu fenomena (Jansoon dkk., 2009). Model mental pada dasarnya digunakan untuk memprediksi dan memecahkan masalah dalam kimia (Chittleborough, 2004). Model mental peserta didik dapat dibangun melalui pengalaman, interpretasi, dan penjelasan mereka ketika dilibatkan dalam proses pembelajaran (Sagita dkk., 2017). Model mental penting untuk diketahui dan diteliti karena ada dua alasan. Pertama, model mental mempengaruhi fungsi kognitif. Kedua, model mental dapat memberikan informasi yang berharga untuk para peneliti pendidikan sains tentang konsep-konsep yang dimiliki peserta didik (Laliyo, 2018).

Model mental peserta didik dapat digali dengan pemberian suatu tes, yang disebut sebagai tes diagnostik model mental (Wang, 2007; Kania dkk., 2020). Instrumen tes diagnostik merupakan instrumen yang dapat mengidentifikasi miskonsepsi dan model mental yang dimiliki peserta didik terhadap suatu konsep (Kania dkk., 2020). Tes diagnostik model mental bertujuan untuk mengukur kemampuan peserta didik dalam menjelaskan data, memprediksi dan menjelaskan fenomena kimia secara makroskopik dengan melibatkan penalaran

pada level submikroskopik menggunakan bahasa simbolik kimia. Dengan pemberian tes diagnostik model mental ini, diharapkan dapat mengukur pemahaman peserta didik mengenai konsep kimia (Coll & Taylor, 2002).

Tes diagnostik pilihan ganda merupakan salah satu tes diagnostik. Tes diagnostik pilihan ganda memiliki beberapa jenis tingkatan, terdiri dari tes diagnostik *two-tier*, *three-tier*, dan *four-tier* (Tumanggor dkk., 2020). Tes diagnostik *two-tier* dan *three-tier* dapat mendeteksi miskonsepsi, tetapi kedua tes diagnostik ini memiliki keterbatasan. Kelemahan tes *two-tier* adalah tidak dapat menentukan jawaban yang diberikan oleh peserta didik disebabkan oleh pemahaman konsep peserta didik itu sendiri atau hanya tebakan. Kelemahan dari tes *three-tier* adalah tidak diketahui apakah responden memiliki tingkat kepercayaan yang berbeda antara tier pertama dan kedua (Pujayanto dkk., 2018). Penelitian telah dikembangkan dan menghasilkan tes *four-tier* dengan tingkat kepercayaan jawaban dan alasan tingkatan secara terpisah. Tes diagnostik *four-tier* adalah salah satu jenis tes diagnostik pilihan ganda multi-tingkat (Tumanggor dkk., 2020).

Namun, berdasarkan hasil wawancara dengan 2 orang guru kimia di masing-masing sekolah yaitu: SMAN 2 Kota Padang, SMAN 3 Kota Padang, dan SMAN 7 Kota Padang, bahwa soal evaluasi yang digunakan pada materi larutan penyangga untuk melihat hasil belajar peserta didik masih menggunakan soal pilihan ganda biasa dan belum menerapkan multirepresentasi kimia, hanya memuat representasi simbolik dan sedikit representasi makroskopik.

Penelitian mengenai pengembangan tes diagnostik *four-tier* sebelumnya telah

dilakukan oleh Habiddin dan Page (2019) untuk mengidentifikasi pemahaman peserta didik pada materi kinetika kimia. Berdasarkan hasil penelitian tersebut didapatkan bahwa hasil analisis butir soal secara komprehensif menunjukkan instrumen tes diagnostik *four-tier* valid dan reliabel serta cocok untuk mengidentifikasi pemahaman peserta didik tentang kinetika kimia.

Selanjutnya, penelitian telah dilakukan oleh Kafiyani dkk., (2019) pada materi fluida statis. Berdasarkan hasil penelitian tersebut didapatkan bahwa tes diagnostik *four-tier* (FTDT) layak digunakan sebagai instrumen yang dapat mengidentifikasi model mental peserta didik pada konsep fluida statis. Kemudian, penelitian instrumen tes diagnostik *four-tier* telah dilakukan oleh Kania dkk., (2020) pada konsep efek rumah kaca. Berdasarkan hasil penelitian tersebut didapatkan bahwa instrumen tes diagnostik *four-tier* dapat diterima dan reliabel serta layak digunakan untuk mengidentifikasi model mental peserta didik pada konsep efek rumah kaca. Tes diagnostik *four-tier* yang dikembangkan pada ketiga penelitian terdahulu di atas belum tersedia pada materi larutan penyangga.

Berdasarkan uraian di atas, maka perlu dikembangkan instrumen tes diagnostik *four-tier* berbasis multirepresentasi kimia untuk mengidentifikasi model mental peserta didik pada materi larutan penyangga yang memenuhi kriteria yang baik dari segi validitas, reliabilitas, indeks kesukaran, daya beda soal, dan fungsi distraktor.

## METODE

Penelitian ini merupakan penelitian pengembangan atau *Research and Development (R&D)* dengan menggunakan

model pengembangan tes diagnostik yang telah dilakukan oleh Treagust (1988).

Penelitian ini dilakukan di kampus FMIPA UNP dan SMAN 7 Padang pada tahun ajaran 2023/2024 semester ganjil. Subjek dari penelitian ini adalah 3 orang dosen kimia FMIPA UNP, 2 orang guru kimia SMA dan 30 orang peserta didik SMAN 7 Padang. Objek penelitian adalah *Test Diagnostic Four-tier* pada materi larutan penyangga.

Langkah-langkah pengembangan instrumen tes diagnostik *four-tier* adalah sebagai berikut:

#### 1. Menentukan isi

Terdapat 4 langkah yang dilakukan pada tahap ini yaitu: a) Mengidentifikasi pernyataan proposisional; b) Mengembangkan peta konsep; c) Menghubungkan pernyataan proposisional dan peta konsep; d) Memvalidasi konten.

#### 2. Memperoleh informasi tentang model mental peserta didik pada materi larutan penyangga

Pada tahap ini terdapat 2 langkah, yaitu : a) Melakukan tinjauan studi literatur terkait model mental peserta didik pada materi larutan penyangga; b) Wawancara guru terkait model mental peserta didik.

#### 3. Mengembangkan tes diagnostik

Pada tahap ini terdapat 3 langkah, yaitu a) Mengembangkan tes diagnostik *four-tier*; b) Merancang spesifikasi kisi; c) Tahap penyempurnaan.

## HASIL DAN DISKUSI

### 1. Menentukan Isi

Pernyataan proposisional materi larutan penyangga diidentifikasi berdasarkan Capaian Pembelajaran (CP) pada Kurikulum Merdeka. Capaian Pembelajaran (CP) pada materi larutan penyangga yaitu “Menggunakan konsep

asam-basa dalam keseharian”. Tabel pernyataan proposisi memuat beberapa konsep larutan penyangga serta definisinya dan konsep prasyarat. Konsep prasyarat larutan penyangga terdiri dari konsep reaksi kimia, kesetimbangan, stoikiometri, dan asam basa.

Hasil analisis konsep larutan penyangga digambarkan dalam bentuk peta konsep. Peta konsep disusun mengikuti hierarki konsep yang terdapat dalam tabel pernyataan proposisional. Agar produk yang dihasilkan konsisten dan sesuai dengan kurikulum merdeka, maka dipastikan semua konsep yang ada dalam pernyataan proposisional saling berhubungan dan tergambar dalam peta konsep.

Validasi pernyataan proposisional dan peta konsep dilakukan oleh 3 orang dosen kimia FMIPA UNP serta 2 orang guru kimia SMA. Validasi ini dilakukan dengan menggunakan angket validasi dengan skala Likert. Pada angket validasi terdapat 11 pernyataan yang menilai sejauh mana tabel pernyataan proposisional, peta konsep dan konsep prasyarat yang telah dikembangkan sesuai dengan Capaian Pembelajaran (CP) dan Tujuan Pembelajaran (TP), serta sejauh mana kesesuaian antara fakta/konsep/prinsip berdasarkan keilmuan kimia. Selain itu, angket validasi ini juga menilai sejauh mana konsep prasyarat dapat membantu peserta didik dalam menjawab tes diagnostik *four-tier*. Hasil validasi tersebut dianalisa menggunakan teknik data *Aiken's V*. Hasil pengolahan data validasi pertama didapatkan nilai rata-rata validitas (V) sebesar 0,77, menunjukkan bahwa pernyataan proposisional dan peta konsep yang telah dikembangkan tidak valid karena terdapat beberapa item yang tidak valid.

Setelah dilakukan perbaikan sesuai saran validator, dilakukan validasi kedua.

Hasil pengolahan data validasi kedua didapatkan nilai rata-rata validitas ( $V$ ) sebesar 0,81, menunjukkan bahwa pernyataan proposisional dan peta konsep yang dikembangkan telah valid. Hal ini sesuai dengan Aiken (1985) bahwa suatu instrumen dikatakan valid dengan 5 orang validator dengan skala penilaian 5, apabila mendapatkan nilai validitas ( $V$ )  $\geq 0,8$ .

## 2. Memperoleh Informasi tentang Model Mental Peserta Didik pada Materi Larutan Penyangga

Berdasarkan hasil kajian literatur didapatkan bahwa peserta didik kesulitan dalam memahami spesi apa yang ada di dalam larutan yang menentukan sifat asam atau basa dari larutan penyangga. Peserta didik juga mengalami kesulitan dalam menjelaskan pergeseran kesetimbangan yang terjadi, hal ini menunjukkan bahwa penggunaan representasi submikroskopik untuk menjelaskan prinsip kerja larutan penyangga masih kurang bisa dimaknai oleh peserta didik. Peserta didik masih belum familiar dengan representasi yang digunakan dan hubungan antara interaksi antar molekul terhadap pergeseran kesetimbangan dan persamaan reaksi yang terjadi (Widarti dkk., 2020).

Hasil penelitian Sariati dkk., (2020) juga mendapatkan bahwa peserta didik mengalami kesulitan dalam menjelaskan konsep larutan penyangga, kesulitan dalam memahami sifat dan komponen-komponen larutan penyangga, serta pemahaman peserta didik terhadap konsep asam basa dan stoikiometri pun masih rendah. Penelitian oleh Genes et al (2021) juga menyatakan bahwa kesulitan peserta didik pada konsep asam basa sangat tinggi. Selain itu, peserta didik juga mengalami kesulitan pada prinsip kerja larutan penyangga,

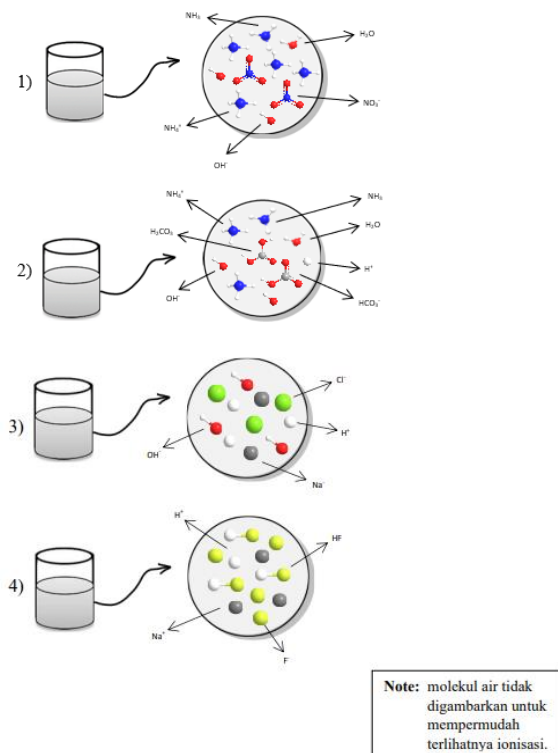
menghitung pH larutan penyangga, dan fungsi larutan penyangga.

Berdasarkan hasil wawancara terhadap beberapa guru kimia SMA, peserta didik masih kesulitan dalam membedakan asam kuat dan asam lemah serta basa kuat dan basa lemah. Peserta didik juga mengalami kesulitan pada soal yang berkaitan dengan perhitungan serta menerapkan stoikiometri dalam menentukan pH larutan penyangga.

## 3. Mengembangkan Tes Diagnostik

Instrumen tes diagnostik *four-tier* dikembangkan berdasarkan kisi-kisi soal yang telah dirancang sebelumnya. Soal yang dirancang terdiri dari 20 soal pilihan ganda 4 tingkat berbasis multirepresentasi kimia, dimana pada tingkat pertama terdiri dari pertanyaan dengan 5 pilihan jawaban, pada tingkat kedua merupakan tingkat keyakinan peserta didik dalam menjawab pertanyaan pada tingkat pertama, pada tingkat ketiga terdiri dari 5 pilihan alasan yang mengacu pada jawaban pada tingkat pertama, dan pada tingkat keempat merupakan tingkat keyakinan peserta didik dalam memilih alasan pada tingkat ketiga. Salah satu soal yang dirancang dapat dilihat pada Gambar 2.

Pak Adi menyediakan beberapa jenis larutan yang dimasukkan ke dalam gelas beaker dengan komposisi berbeda yang ada di dalam laboratorium. Beliau memberi tugas kepada para siswa untuk mengidentifikasi larutan yang termasuk penyangga dan larutan yang bukan termasuk penyangga. Berikut beberapa larutan yang telah disediakan oleh Pak Adi:



Campuran yang **bukan** penyangga terdapat pada gelas beaker yang bernomor.....

- 1 dan 2
- 4 saja
- 2 dan 3
- 1 dan 4
- 1 saja

Keyakinan anda terhadap pilihan jawaban .....

- Yakin
- Tidak yakin

Alasan yang tepat untuk jawaban anda adalah.....

- Campuran asam lemah dan basa lemah tidak termasuk larutan penyangga, sehingga gelas beaker 2 bukan larutan penyangga
- Campuran asam kuat dengan basa kuat bukan termasuk larutan penyangga, sehingga gelas beaker 3 bukan termasuk larutan penyangga
- Larutan yang ada pada gelas beaker 1 terdiri dari campuran asam kuat dan garamnya, sedangkan pada gelas beaker 4 terdiri dari campuran asam lemah dan garamnya, sehingga keduanya bukan termasuk larutan penyangga

- Larutan pada gelas beaker 1 terdiri dari campuran asam kuat dan basa lemah sehingga tidak termasuk ke dalam larutan penyangga
- Larutan pada gelas beaker 2 adalah campuran asam lemah dengan basa lemah sedangkan gelas beaker 3 adalah campuran asam kuat dengan basa kuat sehingga kedua larutan tersebut bukan larutan penyangga

Keyakinan anda terhadap pilihan alasan .....

- Yakin
- Tidak yakin

Gambar 2. Contoh Soal Tes Diagnostik *Four-tier*

Soal yang telah dirancang tersebut dilakukan validasi logis oleh para ahli dan untuk melihat validitas empiris, nilai reliabilitas, indeks kesukaran, daya beda soal, dan fungsi distraktor dilakukan uji coba pada peserta didik.

### 3.3.1. Validitas dan Reliabilitas

Validasi yang dilakukan pada instrument tes diagnostik *four-tier* yang dikembangkan adalah validasi logis dan validasi empiris. Validitas logis terdapat dua jenis yaitu validitas isi dan validitas konstruk (Kurniawan et al., 2022). Nilai validitas logis diperoleh dari hasil validasi oleh 5 orang validator. Validasi logis dilakukan dengan menggunakan angket validasi dengan skala Likert. Hasil validasi tersebut dianalisa menggunakan rumus *Aiken's V* (Aiken, 1985).

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

$$s = r - l_o$$

Keterangan:

$V$  = indeks validitas isi Aiken's  $V$

$n$  = banyak penilai

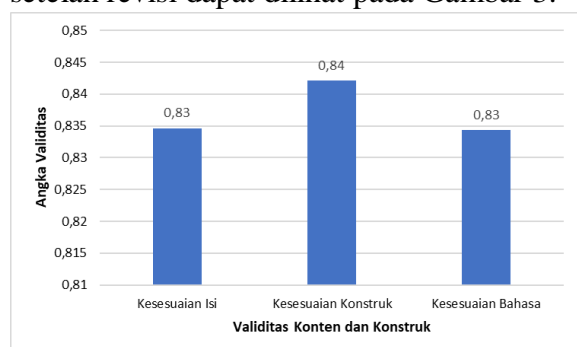
$c$  = angka penilaian tertinggi

$l_o$  = angka penilaian terendah

$r$  = angka yang diberikan oleh penilai

Validasi logis dilakukan dua kali, karena terdapat beberapa item yang tidak

valid dengan nilai validitas  $< 0,8$ . Butir soal yang tidak valid dikarenakan beberapa hal, antara lain kejelasan gambar, tidak adanya gambar makroskopik dan submikroskopik serta butir soal tergantung pada jawaban soal lain. Setelah dilakukan perbaikan sesuai dengan saran dari validator, dilakukan validasi kedua. Semua soal yang dirancang sudah valid karena nilai validitas ( $V$ )  $\geq 0,8$ . Nilai rata-rata validitas logis setelah revisi dapat dilihat pada Gambar 3.



Gambar 3. Grafik Rata-rata Validitas Logis Setelah Revisi

Validitas empiris didapatkan dengan melakukan uji coba instrumen tes diagnostik *four-tier* pada 30 orang peserta didik SMAN 7 Padang yang telah mempelajari materi larutan penyangga. Validitas empiris dianalisa menggunakan uji validitas korelasi biserial. Berdasarkan hasil pengolahan data, didapatkan bahwa pada *first-tier* 19 soal dengan kategori valid dan 1 soal dengan kategori tidak valid, sedangkan pada *third-tier* 18 soal dengan kategori valid dan 2 soal dengan kategori tidak valid. Soal nomor 4 yang dapat dilihat pada

Gambar 2. Contoh Soal Tes Diagnostik Four-tier merupakan salah satu soal dengan kategori valid dan memiliki nilai validitas pada *first-tier* dan *third-tier* yaitu sebesar 0,58 dan 0,79. Hal ini menunjukkan bahwa informasi yang diperoleh dari suatu tes mewakili pemahaman peserta didik yang

sebenarnya (Habiddin & Page, 2019). Rekap validitas empiris pada *first-tier* dan *third-tier* dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. Validitas *First-tier*

No. Soal	Kategori
1,2,3,4,5,7,8,9,10,11,12,13,14,15,16,17,18,19,20	Valid
6	Tidak Valid

Tabel 2. Validitas *Third-tier*

No. Soal	Kategori
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18	Valid
19,20	Tidak Valid

Uji reliabilitas dilakukan dengan teknik KR-20, didapatkan nilai reliabilitas instrumen tes diagnostik *four-tier* sebesar 0,88 pada *first-tier* dan 0,89 pada *third-tier*. Hal ini berarti pada *first-tier* dan *third-tier* telah reliabel dengan kriteria sangat tinggi (Arikunto, 2016). Suatu tes dinyatakan reliabel apabila dilakukan pengambilan data secara berulang terhadap subjek yang sama, akan menghasilkan data yang konsisten (Arikunto, 2016).

### 3.3.2. Indeks Kesukaran, Daya Beda Soal, dan Distraktor

Untuk mengetahui tingkat kesukaran soal dilakukan uji indeks kesukaran soal. Indeks kesukaran adalah bilangan yang menunjukkan sukar atau mudahnya suatu soal. Indeks kesukaran ini menunjukkan taraf kesukaran soal (Arikunto, 2016). Berdasarkan hasil analisis indeks kesukaran pada *first-tier* diperoleh bahwa 15 soal dengan kategori sedang dan 5 soal dengan kategori sulit. Sedangkan pada *third-tier* diperoleh bahwa 1 soal dengan kriteria mudah, 10 soal dengan kriteria sedang, dan 9 soal dengan kriteria sulit.

Jika sebagian besar peserta didik dapat menjawab soal dengan benar, maka soal

tersebut dikatakan mudah, sedangkan jika sebagian besar peserta didik tidak dapat menjawab soal dengan benar, maka soal tersebut dikatakan sulit (Arikunto, 2016). Soal nomor 4 yang dapat dilihat pada

Gambar 2. Contoh Soal Tes Diagnostik Four-tier memiliki nilai indeks kesukaran sebesar 0,37 pada *first-tier* dan 0,3 pada *third-tier*. Nilai indeks kesukaran pada *third-tier* soal tersebut lebih kecil dibandingkan pada *first-tier*, hal ini dikarenakan dari 11 orang peserta didik yang mampu menjawab dengan benar pada *first-tier* hanya 9 orang peserta didik yang mampu menjawab dengan benar pada *third-tier*. Sehingga, indeks kesukaran pada *first-tier* termasuk kategori sedang sedangkan pada *third-tier* termasuk kategori sulit. Hal ini menunjukkan bahwa peserta didik hanya mampu menerapkan pengetahuan konten saja, namun belum mampu menerapkan penalaran konseptual dengan baik (Habiddin & Page, 2019). Rekap indeks kesukaran soal pada *first-tier* dan *third-tier* dapat dilihat pada Tabel 3 dan Tabel 4.

Tabel 3. Indeks Kesukaran *First-tier*

No. Soal	Kategori
1,2,3,4,5,6,7,8,9,10,13,14,15,17,18	Sedang
11,12,16,19,20	Sulit

Tabel 4. Indeks Kesukaran *Third-tier*

No. Soal	Kategori
3	Mudah
1,2,6,7,8,9,10,13,16,17	Sedang
4,5,11,12,14,15,18,19,20	Sulit

Daya beda soal diartikan sebagai kemampuan butir item soal untuk menunjukkan perbedaan kemampuan dari peserta tes. Analisis daya beda soal

dikategorikan menjadi empat, yaitu baik sekali, baik, cukup dan jelek (Arikunto, 2016). Berdasarkan hasil pengolahan data daya pembeda diperoleh bahwa pada *first-tier* terdapat 1 soal dengan kategori baik sekali, 16 soal dengan kategori baik, 2 soal dengan kategori cukup, dan 1 soal dengan kategori jelek. Sedangkan pada *third-tier* didapatkan hasil bahwa 1 soal dengan kategori baik sekali, 16 soal dengan kategori baik, 1 soal dengan kategori cukup, dan 2 soal dengan kategori jelek. Rekap daya beda soal pada *first-tier* dan *third-tier* dapat dilihat pada Tabel 5 dan Tabel 6.

Tabel 5. Daya Beda *first-tier*

No. Soal	Kategori
2	Baik Sekali
1,3,4,5,7,8,9,10,11,13,14,15,16,17,18,19	Baik
6,12	Cukup
20	Jelek

Tabel 6. Daya Beda *Third-tier*

No. Soal	Kategori
2	Baik Sekali
1,3,4,5,6,7,8,9,10,11,13,14,15,16,17,18	Baik
12	Cukup
19, 20	Jelek

Soal dengan daya beda yang baik akan memiliki tingkat kemampuan yang baik pula dalam membedakan kelompok yang berprestasi tinggi (kelompok atas) dengan kelompok yang berprestasi rendah (kelompok bawah) (Arikunto, 2016). Soal nomor 4 yang dapat dilihat pada

Gambar 2. Contoh Soal Tes Diagnostik Four-tier merupakan salah satu soal dengan daya beda yang baik pada *first-tier* dan *third-tier* dengan nilai daya beda yang sama sebesar 0,47. Hal ini menunjukkan bahwa



soal nomor 4 memiliki tingkat kemampuan yang baik dalam membedakan kelompok yang berprestasi tinggi dengan kelompok yang berprestasi rendah. Sedangkan soal nomor 19 dan 20 memiliki daya beda yang rendah dengan kategori daya beda jelek, ini menunjukkan bahwa soal tersebut belum mampu membedakan kelompok yang berprestasi tinggi (kelompok atas) dengan kelompok yang berprestasi rendah (kelompok bawah). Oleh karena itu, semakin tinggi nilai daya beda soal, maka semakin baik soal dalam membedakan peserta didik yang berprestasi tinggi dengan peserta didik yang berprestasi rendah (Habiddin & Page, 2019).

Distraktor dibuat untuk mengecoh peserta didik yang kurang mampu (tidak tahu) dan untuk membedakan dengan peserta didik yang mampu (lebih tahu) (Devi & Azra, 2023). Distraktor dapat dikatakan berfungsi dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih oleh 5% dari seluruh peserta tes (Yusrizal, 2016). Berdasarkan hasil analisis distraktor didapatkan hasil bahwa pada *first-tier* dan *third-tier* distraktor pada semua soal berfungsi dengan baik, karena lebih dari 5% peserta didik memilih jawaban selain dari kunci jawaban.

### KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dihasilkan instrumen tes diagnostik *four-tier* yang telah valid dan reliabel dengan nilai rata-rata validitas  $V$  sebesar 0,84 dan nilai realibilitas *first-tier* dan *third-tier* dengan kategori sangat tinggi sebesar 0,88 dan 0,89 serta memiliki indeks kesukaran dan daya beda soal yang baik.

### REFERENSI

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings, *Educational and Psychological Measurement. Journal Articles; Reports - Research; Numerical/Quantitative Data*, 45(1), 131–142.
- Arikunto, S. (2016). *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara.
- Brady, J. E., Hyslop, A., & Jespersen, N. D. (2012). *Chemistry*. Wiley. <https://books.google.co.id/books?id=wCRHbwAACAAJ>
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293–307. <https://doi.org/10.1039/B7RP90006F>
- Chittleborough, G. D. (2004). The Role of Teaching Models and Chemical Representations in Developing Students' Mental Models of Chemical Phenomena. *Curtin University of Technology, May*, 1–494.
- COLL, R. K., & TAYLOR, N. (2002). Mental Models in Chemistry: Senior Chemistry Students' Mental Models of Chemical Bonding. *Chem. Educ. Res. Pract.*, 3(2), 175–184. <https://doi.org/10.1039/b2rp90014a>
- Devi, N. A., & Azra, F. (2023). Pengembangan Instrumen Tes Diagnostik Untuk Melihat Gambaran Model Mental Peserta Didik Pada Materi Asam Basa. *Entalpi Pendidikan Kimia*, 16–26.
- Genes, A. J., Lukum, A., & Laliyo, L. A. R. (2021). Identifikasi Kesulitan Pemahaman Konsep Larutan

- Penyangga Siswa Di Gorontalo. *Jambura Journal of Educational Chemistry*, 3(2), 61–65. <https://doi.org/10.34312/jjec.v3i2.1191>
- Habiddin, & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, 19(3), 720–736. <https://doi.org/10.22146/ijc.39218>
- Jansoon, N., Cooll, R. K., & Somsook, E. (2009). Understanding Mental Models of Dilution in Thai Students. *International Journal of Environmental & Science Education. International Journal of Environmental and Science Education*, 4(2), 147–168.
- Kafiyani, F., Samsudin, A., & Saepuzaman, D. (2019). Development of four-tier diagnostic test (FTDT) to identify student's mental models on static fluid. *Journal of Physics: Conference Series*, 1280(5), 0–8. <https://doi.org/10.1088/1742-6596/1280/5/052030>
- Kania, V. I., Samsudin, A., Purwanto, Aminudin, A. H., Rasmitadila, Rachmadtullah, R., Jermisittiparsert, K., & Nurtanto, M. (2020). Multitier of greenhouse effect (Moge) instrument development to identify middle school students' mental model in Thailand with rasch analysis. *International Journal of Advanced Science and Technology*, 29(7), 3223–3237.
- Kurniawan, A., Febrianti, A. N., Hardianti, T., Risan, R., Sari, D. M. M., Sitopu, J. W., Dewi, R. S., & Sianipar, D. (2022). *Evaluasi Pembelajaran*. Get Press. <https://books.google.co.id/books?id=nKeAEAAAQBAJ>
- Laliyo, L. A. R. (2018). Model Mental Siswa Dalam Memahami Perubahan Wujud Zat. *Jurnal Penelitian Dan Pendidikan*, 8(1), 1–12.
- Ni Made, A. S. (2022). Analisis Reduksi Miskonsepsi Kimia dengan Pendekatan Multi Level Representasi: Systematic Literature Review. *Jurnal Pendidikan Mipa*, 12(2), 341–348. <https://doi.org/10.37630/jpm.v12i2.59>
- Pujayanto, Budiharti, R., Radiyono, Y., Amalia Nuraini, N. R., Vernanda Putri, H., Eko Saputro, D., & Adhitama, D. E. (2018). Developing four tier misconception diagnostic test about kinematics. *Cakrawala Pendidikan*, 37(2), 237–249. <https://doi.org/10.21831/cp.v37i2.1649>
- Sagita, R., Azra, F., & Azhar, M. (2017). Pengembangan Modul Konsep Mol Berbasis Inkuiri Terstruktur Dengan Penekanan Pada Interkoneksi Tiga Level Representasi Kimia Untuk Kelas X Sma. *Jurnal Eksakta Pendidikan (Jep)*, 1(2), 25. <https://doi.org/10.24036/jep.v1i2.48>
- Sariati, Ni Kadek, Suardana, Nyoman, Wiratini, N. M. (2020). *Materi Larutan Penyangga*. 4(April 2020), 76–87.
- 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/10.1080/095006988010204>
- Tumanggor, A. M. R., Supahar, Kuswanto, H., & Ringo, E. S. (2020). Using four-tier diagnostic test instruments to detect physics teacher candidates' misconceptions: Case of mechanical wave concepts. *Journal of Physics: Conference Series*, 1440(1). <https://doi.org/10.1088/1742-6596/1440/1/012059>
- Wang, C.-Y. (2007). The role of mental-modeling ability, content knowledge, and mental models in general chemistry students' understanding

- about molecular polarity. *Disertasi*, 70(3-A), 835.
- Widarti, H., Sigit, D., & Irianti, D. (2020). Pengaruh kemampuan awal terhadap kemampuan interkoneksi multi representasi siswa pada materi larutan penyangga. *J-PEK (Jurnal Pembelajaran Kimia)*, 5(1), 40–46.
- <https://doi.org/10.17977/um026v5i12020p040>
- Yusrizal, M. P. (2016). *Pengukuran & Evaluasi Hasil dan Proses Belajar*. Penerbit Pale Media Prima. <https://books.google.co.id/books?id=Sj6VEAAAQBAJ>