

Pengembangan Instrumen Tes Diagnostik *Four-Tier* Hukum Dasar Kimia untuk Mengidentifikasi Model Mental Peserta Didik

Development of a Four-Tier Test Instrument on Basic Law of Chemistry to Identify Students' Mental Models

T Fadhillah¹, F Azra^{1*}

¹ Pendidikan Kimia, Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar Barat, Padang Utara, Sumatera Barat, Indonesia. 25171.

* bunda_syasfa@yahoo.com

Received on:

30th October 2024

Revised till:

30th November 2024

Accepted on:

1st December 2024

Publisher version

published on:

31th December
2024

ABSTRACT

Students' understanding of chemical concepts, especially the basic laws of chemistry, is very important for successful learning. The representation of this understanding can be seen in the form of mental models. This research aims to develop a Four-Tier diagnostic test instrument to identify students' mental models on basic laws of chemistry and test the validity and reliability of the instrument. This research used the Research and Development (R&D) method with the Treagust development model. The subjects of this research were 3 chemistry lecturers of FMIPA UNP and 2 teachers and 30 students of class X Phase E at MAN 2 Kota Padang. The results showed that the diagnostic test instrument developed had an average logical validity value of 0.96 and reliability at the first tier of 0.89 and the third tier of 0.86 which was categorized as very high. The results of this diagnostic test instrument that has been developed are valid and have good instrument criteria.

KEYWORDS

Diagnostic test, Mental models, Basic laws of chemistry

ABSTRAK

Pemahaman siswa terhadap konsep kimia, terutama hukum dasar kimia, sangat penting untuk keberhasilan pembelajaran. Penggambaran dari pemahaman tersebut dapat dilihat dalam bentuk model mental. Penelitian ini bertujuan mengembangkan instrumen tes diagnostik *Four-Tier* untuk mengidentifikasi model mental siswa pada materi hukum dasar kimia dan menguji validitas serta reliabilitas instrumen tersebut. Penelitian ini menggunakan metode *Research and Development* (R&D) dengan model pengembangan Treagust. Subjek dari penelitian ini adalah 3 orang dosen kimia FMIPA UNP dan 2 orang guru serta 30 orang siswa kelas X Fase E di MAN 2 Kota Padang. Hasil penelitian menunjukkan bahwa instrumen tes diagnostik yang dikembangkan memiliki rata-rata nilai validitas logis sebesar 0,96 dan reliabilitas pada tingkat pertama sebesar 0,89 dan pada tingkat ketiga sebesar 0,86 yang dikategorikan sangat tinggi. Hasil Instrumen tes diagnostik yang telah dikembangkan ini telah valid dan memiliki kriteria instrumen yang baik.

KATA KUNCI

Tes diagnostik, Model mental, Hukum dasar kimia



<https://doi.org/10.24036/ekj.v6.i3.a568>

2024 • Vol.6, No. 3

T Fadhillah¹, F Azra^{1*}

1. PENDAHULUAN

Kimia merupakan pelajaran dengan konsep yang abstrak dan kompleks^[1], sehingga tidak cukup dipahami melalui teori saja. Hal ini menyebabkan pemahaman dalam kimia memerlukan penguasaan tiga level representasi, yaitu makroskopik, submikroskopik, dan simbolik^{[1][2]}. Guru menghadapi tantangan untuk membantu siswa memahami dan menghubungkan ketiga level ini secara menyeluruh, terutama pada materi seperti hukum dasar kimia yang bersifat abstrak dan berisi prinsip-prinsip penting. Ketidapahaman siswa pada materi ini dapat menghambat pemahaman mereka terhadap konsep kimia yang lebih kompleks.

Hukum dasar kimia adalah salah satu materi yang mengandung prinsip-prinsip fundamental dalam pembelajaran kimia^[3]. Hal ini karena materi tersebut mencakup hukum dasar kimia yang menjadi acuan dalam pembelajaran kimia. Penelitian Krisnawati & Fajaroh (2013) mengatakan bahwa persentase peserta didik yang mampu memahami materi hukum dasar kimia tergolong rendah^[4]. Ulva, dkk. (2016) juga menyatakan bahwa pemahaman konsep pada materi hukum dasar kimia masih kurang^[5]. Dengan demikian, hukum dasar kimia dianggap sebagai materi yang cukup sulit untuk dipelajari. Sejalan dengan penelitian sebelumnya, hasil wawancara dengan tiga guru kelas X Fase E dari SMAN 3 Padang, SMAN 13 Padang, dan MAN 2 Kota Padang, mengungkapkan bahwa hukum dasar kimia termasuk materi yang sulit dipahami oleh peserta didik. Materi hukum dasar kimia tersebut dinilai sulit dikarenakan berisi konsep-konsep yang abstrak dan kompleks sehingga peserta didik bingung dalam membedakan dan mengimplementasikan hukum-hukum yang telah dipelajari. Ketidakmampuan memahami materi ini dapat berdampak pada kesulitan peserta didik dalam mempelajari konsep kimia lebih lanjut. Untuk mengatasi hal tersebut, maka diperlukan sebuah instrumen untuk mengidentifikasi pemahaman peserta didik, yang dapat diwujudkan melalui model mental kimia^[6].

Model mental yang dapat diwujudkan melalui model mental kimia^[2,6-8]. Pemahaman yang mendalam muncul saat ketiga level representasi kimia saling terhubung, didukung oleh elemen visualisasi, didukung oleh visualisasi dalam memori kerja^[9]. Ketika peserta didik mampu memahami konsep kimia melalui ketiga level representasi kimia, yaitu makroskopik, submikroskopik, dan simbolik, pengembangan model mental peserta didik terhadap fenomena ilmiah akan meningkat, Peningkatan ini berkontribusi pada terciptanya model mental ilmiah peserta didik^[10]. Hal ini terjadi karena model mental ilmiah hanya dapat dibangun jika peserta didik

memiliki keterampilan berpikir kritis yang dikembangkan melalui proses pembelajaran kimia yang baik^[2].

Analisis model mental peserta didik dapat dilakukan dengan memahami bagaimana peserta didik memahami konsep kimia melalui tiga tingkat representasi kimia dan bagaimana tiga tingkat tersebut berhubungan satu sama lain^[7]. Dengan mengidentifikasi model mental peserta didik, guru akan mendapatkan informasi yang akurat mengenai konsep yang dimiliki peserta didik. Informasi ini memungkinkan guru merancang pembelajaran yang lebih efektif dan memberikan umpan balik yang spesifik. Umpan balik tersebut diharapkan dapat membantu peserta didik membangun pemahaman yang benar dan berkelanjutan.

Coll & Treagust (2003) mengungkapkan bahwa model mental dapat dianalisis melalui model mental ekspresif (*expressed mental model*), yaitu model yang dapat diidentifikasi dan dipahami melalui pancaindra. Oleh karena itu, pemetaan model mental peserta didik memerlukan instrumen yang tepat, seperti wawancara^{[7][11][12]}, tes diagnostik dengan pertanyaan terbuka^{[12][13]}, tes diagnostik *two-tier*^{[2][6][11]}, tes diagnostik *three-tier*^[14] ataupun tes diagnostik *four-tier*^{[15][16][17]}.

Penggunaan instrumen dengan soal model biasa atau satu tingkat kurang efektif untuk mengidentifikasi model mental peserta didik. Hal ini disebabkan soal satu tingkat memungkinkan peserta didik menjawab dengan menebak, sehingga alasan di balik jawaban mereka sulit diidentifikasi secara mendalam^[17]. Oleh karena itu, diperlukan instrumen yang lebih efektif untuk mengungkap pemahaman peserta didik secara komprehensif

Tes diagnostik *four-tier* saat ini dianggap sebagai alat yang paling efektif untuk mengidentifikasi model mental. Penelitian Mi, dkk. (2023) berhasil mengembangkan dan memvalidasi tes diagnostik *four-tier* pada materi elektrostatis. Tes ini memungkinkan penilaian yang lebih mendalam, tidak hanya menilai jawaban benar atau salah tetapi juga meminta peserta didik memberikan alasan atas jawaban mereka^[18]. Namun, penelitian Mi, dkk. (2023) belum menggunakan tes tersebut untuk memetakan model mental.

Kafiyani, dkk. (2019) menggunakan tes diagnostik *four-tier* untuk mengidentifikasi model mental pada fluida statis^[15]. Meskipun berhasil, penelitian ini tidak menekankan aspek tiga level representasi dalam pengembangannya, padahal aspek tersebut penting dalam pemetaan model mental. Tes diagnostik *four-tier* juga digunakan dalam penelitian

Fakhriyah & Masfuah (2021); Mi, dkk. (2023); Habiddin, dkk. (2019) untuk mengidentifikasi miskonsepsi pada konsep kimia^{[18][19][20]}. Mengidentifikasi miskonsepsi dan model mental peserta didik merupakan tugas penting bagi guru. Identifikasi miskonsepsi melalui model mental memberikan wawasan berharga tentang cara peserta didik dalam memahami konsep. Informasi ini membantu guru merancang strategi pembelajaran yang lebih efektif^[21]. Identifikasi model mental memungkinkan peneliti mengenali miskonsepsi yang dimiliki peserta didik secara efektif. Hal ini dapat menjadi dasar yang kuat untuk meningkatkan kualitas pembelajaran.

Keterbatasan instrumen yang akurat untuk memetakan model mental peserta didik pada materi hukum dasar kimia menjadi tantangan bagi guru. Penelitian ini bertujuan untuk mengembangkan tes diagnostik *four-tier* yang dapat mengidentifikasi model mental siswa, serta menguji validitas, reliabilitas, daya beda, tingkat kesukaran, dan fungsi distraktornya. Hasil penelitian ini diharapkan dapat memberikan informasi terkait model mental peserta didik, merancang pembelajaran yang lebih efektif, dan memberikan umpan balik yang lebih spesifik

2. METODE

Penelitian ini menggunakan jenis penelitian pengembangan R&D (*Research and Development*) dengan model pengembangan Treagust. Lokasi penelitian berada di MAN 2 Kota Padang pada semester genap tahun ajaran 2023/2024. Subjek penelitian melibatkan 3 orang dosen kimia FMIPA UNP, 2 orang guru MAN 2 Kota Padang, dan 30 orang peserta didik pada peserta didik kelas X Fase E di MAN 2 Kota Padang. Penelitian ini mengikuti prosedur dari model pengembangan Treagust (1988), yang mencakup beberapa langkah utama:

2.1 Menentukan Konten

Tahapan ini bertujuan untuk menentukan batasan konsep berdasarkan capaian pembelajaran^[22]. Analisis capaian pembelajaran menghasilkan tujuan pembelajaran yang digunakan sebagai menyusun pernyataan proposisional dan peta konsep. Validasi pernyataan proposisional dan peta konsep dilakukan oleh lima validator menggunakan indeks validitas Aiken.

2.2 Mengumpulkan Informasi terkait Peserta Didik

Informasi dikumpulkan dapat berupa melalui penelitian terkait model mental peserta didik pada materi hukum dasar kimia, wawancara dengan guru untuk mengidentifikasi materi yang dipelajari, serta kesulitan yang dihadapi peserta didik.

Tahap ini bertujuan untuk memahami struktur kognitif dan permasalahan konseptual yang umum dialami peserta didik.^[22] Informasi yang dikumpulkan dapat berupa penelitian terkait model mental peserta didik pada materi hukum dasar kimia, wawancara dengan guru untuk mengidentifikasi materi yang dipelajari peserta didik, dan kesulitan peserta didik pada materi hukum dasar kimia.

2.3 Mengembangkan Tes Diagnostik

Tes diagnostik dikembangkan berdasarkan hasil dari dua tahap sebelumnya. Tes ini dirancang untuk mengukur capaian pembelajaran dengan memperhatikan validitas logis, validitas empiris, reliabilitas, daya pembeda, tingkat kesukaran, dan efektivitas distraktor. Berikut adalah rincian dari masing-masing analisis:

2.3.1 Validitas Logis

Instrumen yang telah dirancang diuji validitas logisnya dengan memberikan angket skala likert 5, yang meliputi kesesuaian isi, kesesuaian penyajian dan aspek Bahasa. Hasil angket dianalisis menggunakan indeks validitas aiken untuk memastikan keabsahan.

2.3.2 Validitas Empiris

Pengujian validitas empiris dilakukan pada 30 peserta didik, menggunakan rumus koefisien korelasi biserial (r_{pbi}) pada tiap butir soal. Hasil validitas dikategorikan berdasarkan nilai r tabel. Dengan kategori valid jika $r_{pbi} > r$ table dan tidak valid jika $r_{pbi} < r$ table.

2.3.3 Reliabilitas

Pengujian reliabilitas yang dilakukan adalah dengan menggunakan metode Split Half. Pembagian soal adalah dengan pembagian soal ganjil dan soal genap, yang kemudian dicari dengan menggunakan korelasi *product moment*, berikut:

$$r_{1/2^{1/2}} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Keterangan:

- $r_{1/2^{1/2}}$: Koefisien reliabilitas yang sudah disesuaikan
 X : Banyak jawaban siswa yang benar pada soal ganjil
 Y : Banyak jawaban siswa yang benar pada soal genap

Lalu, akan didapatkan reliabilitas separo dari tes, yang kemudian digunakan rumus Spearman Brown untuk mengetahui reliabilitas seluruh tes, berikut:

$$r_{11} = \frac{2r_{1/2^{1/2}}}{1 + r_{1/2^{1/2}}}$$

Keterangan:

- r_{11} : Korelasi antar skor-skor setiap belahan tes
 $r_{1/2^{1/2}}$: Koefisien reliabilitas yang sudah disesuaikan
 Hasil dari reliabilitas seluruh tes akan dikategorikan sesuai dengan kategori pada Tabel 1.

Tabel 1. Indeks koefisien korelasi

Daya Pembeda	Kategori
$0,80 \leq r < 1,00$	Sangat Tinggi
$0,60 \leq r < 0,80$	Tinggi
$0,40 \leq r < 0,60$	Cukup
$0,20 \leq r < 0,40$	Rendah
$0,00 \leq r < 0,20$	Sangat Rendah

2.3.4 Daya Pembeda

Daya pembeda dianalisis dengan rumus berikut :

$$DP = \frac{J_p}{N_p} - \frac{J_k}{N_k}$$

Keterangan:

Jp : Banyak peserta didik kelompok pintar yang menjawab benar

Jk : Banyak peserta didik kelompok kurang pintar yang menjawab benar

Np : Banyak peserta didik kelompok pintar

Nk : Banyak peserta didik kelompok kurang pintar

Setelah dilakukan analisis daya pembeda, maka akan dikategorikan daya pembeda pada tiap butir soal, sesuai dengan Tabel 2.

Tabel 2. Interpretasi indeks daya pembeda

Daya Pembeda	Kategori
Diatas 0,40	Baik sekali
0,30 – 0,39	Baik
0,20 – 0,29	Sedang
0,01 – 0,19	Kurang

2.3.5 Tingkat Kesukaran

Pada tingkat kesukaran dilakukan analisis data dengan menggunakan rumus berikut:

$$P = \frac{B}{Js}$$

Keterangan:

P : Tingkat Kesukaran

B : Banyak siswa yang menjawab soal dengan benar

Js : Jumlah seluruh peserta didik yang mengikuti tes

Selanjutnya dikategorikan sesuai dengan acuan pada Tabel 3.

Tabel 3. Interpretasi tingkat kesukaran

Tingkat Kesukaran	Kategori
0,00 – 0,30	Soal Sulit
0,31 – 0,70	Soal Sedang
0,71 – 1,00	Soal Mudah

2.3.6 Distraktor

Distraktor dianalisis dengan menghitung persentase siswa yang memilih setiap opsi jawaban pada instrumen. Distraktor dianggap efektif jika dipilih oleh minimal 5% peserta didik.

3. HASIL DAN PEMBAHASAN

3.1 Menentukan Konten

Analisis capaian pembelajaran menghasilkan menjadi lima tujuan pembelajaran, yang menjadi dasar penyusunan tabel pernyataan proposisional. Tabel ini mencakup enam konsep hukum dasar kimia serta tiga materi prasyarat, yaitu materi konsep ilmu kimia (lima konsep), struktur atom (satu konsep), dan reaksi kimia (satu konsep). Berdasarkan tabel tersebut, peta konsep dirancang untuk menghubungkan dan menyusun konsep-konsep yang ada secara konsisten agar instrumen sesuai dengan area konsepnya.

Pengujian validitas logis dilakukan oleh lima validator, menghasilkan rata-rata validitas sebesar 0,92. Nilai validitas ini lebih tinggi dari indeks Aiken (0,8 untuk lima validator dengan skala lima), sehingga tabel dan peta konsep dinyatakan valid.

3.2 Mengumpulkan Informasi terkait Peserta Didik

Hasil wawancara dengan guru kelas X fase E di MAN 2 Kota Padang, SMAN 3 Padang dan SMAN 13 Padang menunjukkan bahwa hukum dasar kimia merupakan salah satu materi yang cukup sulit dipahami siswa, karena konsepnya yang abstrak dan kompleks. Peserta didik bingung dalam membedakan dan mengimplementasikan hukum-hukum yang telah dipelajari.

Hal ini didukung oleh penelitian Ulva dkk (2016) yang mencatat rata-rata nilai siswa pada hukum-hukum dasar kimia hanya 48,15 atau dapat dikatakan bahwa rata-rata tersebut tergolong kurang. Krisnawati & Fajaroh (2013) juga menemukan bahwa pemahaman siswa terhadap materi hukum dasar kimia tergolong rendah, terutama pada materi hukum perbandingan tetap dan hipotesis avogadro.

Sebagai contoh, beberapa siswa salah dalam memahami massa abu pembakaran pita magnesium memiliki massa yang lebih kecil daripada massa magnesium sebelum dibakar. Hal ini menunjukkan bahwa pemahaman siswa sangat rendah terutama pada level submikroskopik.

3.3 Mengembangkan Tes Diagnostik

Tes diagnostik empat tingkat (*four-tier*) dikembangkan untuk mengukur capaian pembelajaran hukum dasar kimia. Tes ini terdiri dari 20 soal: 16 soal pilihan ganda dan 4 soal benar-salah. Tingkat pertama berupa pilihan ganda, tingkat kedua keyakinan jawaban, tingkat ketiga alasan pilihan, dan tingkat keempat keyakinan terhadap alasan. Instrumen ini mencakup representasi kimia di tiga level.

Dalam sebuah tabung berisi 1 g gas oksigen, dimasukkan pita magnesium dengan hasil setelah reaksi adalah berupa magnesium oksida. Setelah bereaksi magnesium oksida yang dihasilkan adalah 2,5 g. Seperti Gambar 9!

Gambar 9
Sumber: Silberberg, 2010

Massa pita magnesium yang bereaksi adalah ...

A. 0,5 g
B. 1,0 g
C. 1,5 g
D. 2,0 g
E. 2,5 g

Keyakinan terhadap pilihan jawaban:
A. Yakin
B. Tidak Yakin

Alasan:
A. Massa Magnesium yang bereaksi + Massa Oksigen yang bereaksi = Massa Magnesium Oksida yang dihasilkan
B. Massa Magnesium yang bereaksi - Massa Oksigen yang bereaksi = Massa Magnesium Oksida yang dihasilkan
C. Massa Oksigen yang bereaksi - Massa Magnesium yang bereaksi = Massa Magnesium Oksida yang dihasilkan
D. Massa Oksigen yang bereaksi + Massa Magnesium Oksida yang bereaksi = Massa Magnesium yang dihasilkan

Keyakinan terhadap pilihan alasan:
A. Yakin
B. Tidak Yakin

Gambar 1. Contoh Soal Pilihan Ganda Instrumen Tes Diagnostik *Four-Tier*

Seorang anak menganalisis sebuah data percobaan pada tabel dibawah ini.

Nama Senyawa	Submikroskopik	Perbandingan Massa	
		Unsur N	Unsur O
N ₂ O		20 g	12 g
N ₂ O ₃		20 g	36 g

Sumber: <https://pubchem.ncbi.nlm.nih.gov/>

Dari data tersebut, anak itu menarik kesimpulan bahwa kedua senyawa tersebut termasuk contoh dari hukum perbandingan berganda.
Pernyataan diatas :

A. Benar
B. Salah

Keyakinan terhadap pilihan jawaban:
A. Yakin
B. Tidak Yakin

Alasan:
A. Dengan perbandingan unsur N yang sama, didapatkan perbandingan unsur O pada senyawa N₂O dan N₂O₃ berturut-turut adalah 1:3
B. Dengan perbandingan unsur N yang sama, didapatkan perbandingan unsur O pada senyawa N₂O dan N₂O₃ berturut-turut adalah 3:1
C. Perbandingan unsur N dan unsur O pada kedua senyawa bukan perbandingan bilangan bulat sederhana
D. Perbandingan unsur N dan unsur O pada senyawa N₂O bukan perbandingan bilangan bulat sederhana

Keyakinan terhadap pilihan alasan:
A. Yakin
B. Tidak Yakin

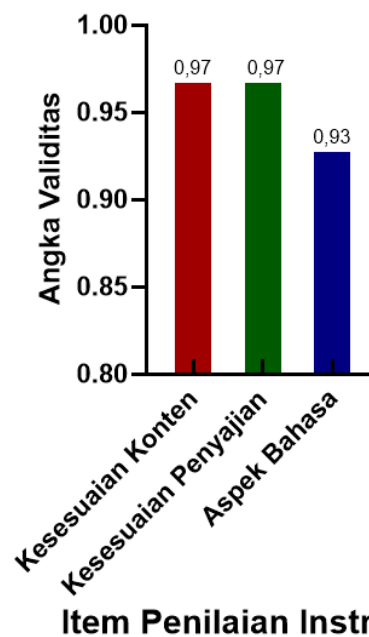
Gambar 2. Contoh Soal Benar Salah Instrumen Tes Diagnostik *Four-tier*

Instrumen ini dikembangkan sesuai dengan capaian dan tujuan pembelajaran yang dapat dilihat melalui kisi-kisi. Kisi-kisi instrumen memuat tujuan pembelajaran, indikator soal, level kognitif, bentuk soal dan nomor soal. Menurut Kurniawan (2022), kelayakan instrumen yang akan digunakan dapat ditentukan melalui analisis kualitatif maupun analisis kuantitatif^[23], sehingga untuk menyempurnakan instrumen tersebut, tahap berikutnya adalah melakukan analisis kualitatif dan kuantitatif.

3.3.1 Validitas Logis

Hasil validitas logis menunjukkan bahwa beberapa soal tidak valid ($V < 0,8$), terutama soal yang tidak memiliki multi representasi kimia. Soal tersebut diperbaiki sesuai saran validator dan divalidasi ulang. Hasil akhirnya menunjukkan instrumen telah valid dengan nilai rata-rata yang terlihat pada Gambar 3.

Validitas Logis Instrumen Tes Diagnostik Four-Tier



Gambar 3. Validitas Logis Instrumen Tes Diagnostik *Four-Tier*

Berdasarkan hasil analisis, diperoleh data rata-rata untuk setiap aspek sebagai berikut: kesesuaian konten sebesar 0,97, kesesuaian penyajian sebesar 0,97, dan kesesuaian bahasa sebesar 0,93. Nilai-nilai tersebut menunjukkan bahwa instrumen memiliki tingkat validitas yang sangat tinggi, khususnya dalam hal kesesuaian konten dan penyajian, yang mendekati kesempurnaan. Meskipun kesesuaian bahasa sedikit lebih rendah dibandingkan aspek lainnya, nilai 0,93 masih berada dalam kategori sangat baik. Hal ini menunjukkan bahwa instrumen telah dirancang dengan baik dan sesuai dengan tujuan pembelajaran, serta layak digunakan dengan sedikit perbaikan pada aspek bahasa untuk lebih meningkatkan kualitasnya.

3.3.2 Uji coba

Tujuan dari melakukan uji coba instrumen tes soal diagnostik ini adalah untuk melihat analisa kuantitatif dari instrumen yang telah dikembangkan. Analisa ini hanya dilakukan pada tingkat pertama dan tingkat ketiga. Analisa kuantitatif ini mencakup validitas empiris instrumen, reliabilitas, indeks kesukaran, daya pembeda dan distraktor.

Hasil analisis menunjukkan bahwa bahwa beberapa soal memiliki validitas empiris di bawah nilai r tabel (0,36), sehingga dianggap tidak valid. Soal-soal tersebut memerlukan revisi dan uji coba ulang untuk meningkatkan kualitasnya. Rekap hasil validitas tersebut dapat dilihat pada Tabel 4.

Tabel 4. Rekap Hasil Validitas Empiris

No Soal	Tier-1		Tier-3	
	Nilai	Kategori	Nilai	Kategori
1	0,48	Valid	0,50	Valid
2	0,65	Valid	0,49	Valid
3	0,64	Valid	0,46	Valid
4	0,84	Valid	0,46	Valid
5	0,19	Invalid	0,53	Valid
6	0,37	Valid	0,38	Valid
7	0,64	Valid	0,38	Valid
8	0,41	Valid	0,37	Valid
9	0,47	Valid	0,44	Valid
10	-0,29	Invalid	0,16	Invalid
11	0,75	Valid	0,62	Valid
12	0,42	Valid	0,31	Invalid
13	0,61	Valid	0,76	Valid
14	-0,25	Invalid	0,39	Valid
15	0,39	Valid	0,52	Valid
16	0,51	Valid	0,43	Valid
17	0,70	Valid	0,44	Valid
18	0,59	Valid	0,12	Invalid
19	0,32	Invalid	0,43	Valid
20	0,38	Valid	0,39	Valid

Berdasarkan tabel 4 tersebut, terdapat empat soal pada *tier-1* yang tidak valid dan tiga soal pada *tier-3*. Rendahnya nilai validitas ini menunjukkan bahwa soal tersebut tidak layak digunakan sebagai tes diagnostik. Soal tersebut akan layak digunakan jika dilakukan perbaikan maupun perubahan pada soal dan dilakukan uji coba ulang^[25].

Selain validitas, reliabilitas instrumen juga menjadi aspek penting dalam penilaian soal. Hasil penelitian menunjukkan bahwa reliabilitas pada *tier-1* didapatkan angka 0,89, sedangkan *tier-3* sebesar 0,86. Kedua nilai ini masuk ke dalam kategori sangat tinggi.

Penilaian instrumen tidak hanya bergantung padavaliditas dan reliabilitas, tetapi juga memerlukan analisis tambahan seperti daya pembeda dan tingkat kesukaran. Analisis ini tidak membantu menentukan kualitas soal, tetapi juga memberikan informasi terkait kekurangan dari soal yang akan menjadi petunjuk untuk perbaikan untuk kedepannya^[25]. Rekapitulasi hasil analisis daya pembeda dan tingkat kesukaran dapat dilihat pada Tabel 5.

Tabel 5. Rekap Hasil Analisis Tingkat Kesukaran dan Daya Pembeda

No Soal	Tingkat Kesukaran				Daya Pembeda			
	Tier-1		Tier-3		Tier-1		Tier-3	
	Nilai	Kategori	Nilai	Kategori	Nilai	Kategori	Nilai	Kategori
1	0,93	Mudah	0,70	Sedang	0,13	Kurang	0,60	Baik Sekali
2	0,77	Mudah	0,53	Sedang	0,47	Baik Sekali	0,53	Baik Sekali
3	0,73	Mudah	0,63	Sedang	0,40	Baik Sekali	0,33	Baik
4	0,70	Sedang	0,67	Sedang	0,60	Baik Sekali	0,40	Baik Sekali
5	0,93	Mudah	0,47	Sedang	0,13	Kurang	0,53	Baik Sekali
6	0,93	Mudah	0,47	Sedang	0,13	Kurang	0,27	Sedang
7	0,53	Sedang	0,50	Sedang	0,40	Baik Sekali	0,33	Baik
8	0,60	Sedang	0,57	Sedang	0,53	Baik Sekali	0,20	Sedang
9	0,67	Sedang	0,70	Sedang	0,40	Baik Sekali	0,47	Baik Sekali
10	0,37	Sedang	0,30	Sulit	-0,07	Kurang	-0,07	Kurang
11	0,60	Sedang	0,67	Sedang	0,67	Baik Sekali	0,53	Baik Sekali
12	0,07	Sulit	0,13	Sulit	0,13	Kurang	0,00	Kurang
13	0,70	Sedang	0,33	Sedang	0,33	Baik	0,67	Baik Sekali
14	0,53	Sedang	0,60	Sedang	-0,13	Kurang	0,40	Baik Sekali
15	0,53	Sedang	0,73	Mudah	0,27	Sedang	0,53	Baik Sekali
16	0,33	Sedang	0,30	Sulit	0,40	Baik Sekali	0,20	Sedang
17	0,60	Sedang	0,33	Sedang	0,53	Baik Sekali	0,40	Baik Sekali
18	0,77	Mudah	0,80	Mudah	0,33	Baik	0,00	Kurang
19	0,10	Sulit	0,73	Mudah	0,20	Sedang	0,27	Sedang
20	0,30	Sulit	0,43	Sedang	0,20	Sedang	0,33	Baik

Tingkat kesukaran soal merupakan salah satu indikator untuk menentukan kualitas suatu soal. Arikunto (2018), menyatakan bahwa soal yang baik adalah soal yang tidak terlalu mudah ataupun terlalu sulit^[25]. Hasil penelitian ini menunjukkan bahwa distribusi tingkat kesukaran soal pada *tier* pertama adalah 3 butir soal sulit, 11 butir soal sedang, dan 6 butir soal mudah. Sementara itu, pada *tier* ketiga didapatkan 3 butir soal sulit, 14 butir soal sedang, 3 butir soal mudah. Berdasarkan pendistribusian tersebut diketahui bahwa lebih dari setengah dari total soal merupakan soal sedang. Namun, hal ini tidak secara langsung menunjukkan bahwa soal-soal tersebut memiliki daya pembeda yang baik. Daya pembeda adalah kemampuan suatu soal untuk membedakan tingkat kemampuan peserta didik^[25]. Artinya, soal yang memiliki daya pembeda yang baik adalah soal yang mampu membedakan siswa dengan kemampuan tinggi dan siswa dengan kemampuan rendah. Dalam

Yusuf (2017), menyatakan butir soal yang paling baik adalah soal yang memiliki tingkat kesukaran 0,50 dan daya pembeda maksimum^[26]. Berdasarkan penelitian ini, beberapa soal memenuhi kriteria tersebut, yaitu memiliki tingkat kesukaran sedang dan daya pembeda yang sangat baik.

Indikator lain yang digunakan untuk menganalisis kualitas soal, khususnya pilihan berganda, adalah dengan analisis distraktor. Distraktor atau pengecoh adalah pilihan jawaban yang dirancang untuk menarik peserta didik yang kurang yakin atau tidak memahami materi yang diujikan. distraktor^[27]. Jika distraktor memiliki daya tarik yang besar, maka distraktor tersebut dapat dikatakan berfungsi dengan baik karena peserta didik yang kurang yakin dan tidak memahami materi yang diujikan akan terkecoh dan memilihnya. Distraktor dianggap efektif jika dipilih oleh lebih dari 5% peserta tes^[28]. Rekap hasil analisis distraktor dapat dilihat pada Tabel 6

Tabel 6. Rekap hasil analisis distraktor

No Soal	Tier-1										Tier-3							
	A		B		C		D		E		A		B		C		D	
	%	Ket	%	Ket	%	Ket	%	Ket	%	Ket	%	Ket	%	Ket	%	Ket	%	Ket
1	93	Good	7	Good	0	Poor	0	Poor	0	Poor	70	Good	10	Good	17	Good	3	Poor
2	3	Poor	77	Good	13	Good	0	Poor	7	Good	53	Good	7	Good	7	Good	33	Good
3	0	Poor	10	Good	73	Good	17	Good	0	Poor	3	Poor	10	Good	63	Good	23	Good
4	0	Poor	3	Poor	10	Good	17	Good	70	Good	13	Good	7	Good	67	Good	13	Good

5	93	Good	7	Good	-	-	-	-	-	-	47	Good	3	Poor	13	Good	37	Good
6	0	Poor	93	Good	0	Poor	0	Poor	7	Good	27	Good	47	Good	10	Good	17	Good
7	0	Poor	0	Poor	13	Good	53	Good	33	Good	0	Poor	17	Good	50	Good	33	Good
8	60	Good	40	Good	-	-	-	-	-	-	23	Good	7	Good	57	Good	13	Good
9	7	Good	13	Good	67	Good	10	Good	3	Poor	70	Good	20	Good	0	Poor	10	Good
10	37	Good	63	Good	-	-	-	-	-	-	0	Poor	30	Good	37	Good	30	Good
11	10	Good	20	Good	60	Good	3	Poor	7	Good	67	Good	20	Good	13	Good	0	Poor
12	50	Good	3	Poor	37	Good	7	Good	3	Poor	17	Good	13	Good	37	Good	33	Good
13	0	Poor	10	Good	70	Good	0	Poor	20	Good	33	Good	27	Good	37	Good	3	Poor
14	53	Good	27	Good	0	Poor	20	Good	0	Poor	60	Good	20	Good	3	Poor	13	Good
15	53	Good	47	Good	-	-	-	-	-	-	73	Good	17	Good	0	Poor	10	Good
16	40	Good	0	Poor	27	Good	33	Good	0	Poor	23	Good	30	Good	47	Good	0	Poor
17	3	Poor	13	Good	60	Good	10	Good	13	Good	3	Poor	33	Good	0	Poor	63	Good
18	17	Good	77	Good	0	Poor	7	Good	0	Poor	10	Good	80	Good	0	Poor	10	Good
19	10	Good	57	Good	0	Poor	20	Good	13	Good	0	Poor	73	Good	17	Good	10	Good
20	23	Good	37	Good	7	Good	30	Good	3	Poor	43	Good	23	Good	0	Poor	33	Good

Berdasarkan hasil analisis dapat disimpulkan bahwa distraktor sudah berfungsi dengan baik, karena pada tingkat pertama dan tingkat ketiga, distraktor

dipilih oleh 5% siswa. kesimpulan ini dapat dilihat pada Tabel 7.

Tabel 7. Rekap hasil analisis soal berdasarkan tujuan pembelajaran

No Soal	Validitas				Tingkat Kesukaran				Daya Pembeda			
	Tier-1		Tier-3		Tier-1		Tier-3		Tier-1		Tier-3	
	Nilai	Ket	Nilai	Ket	Nilai	Ket	Nilai	Ket	Nilai	Ket	Nilai	Ket
TP 1. Peserta didik mampu menerapkan hukum kekekalan massa dalam perhitungan kimia												
1	0,48	Valid	0,50	Valid	0,93	Mudah	0,70	Sedang	0,13	Kurang	0,60	Baik Sekali
8	0,41	Valid	0,37	Valid	0,60	Sedang	0,57	Sedang	0,53	Baik Sekali	0,20	Sedang
9	0,47	Valid	0,44	Valid	0,67	Sedang	0,70	Sedang	0,40	Baik Sekali	0,47	Baik Sekali
10	-0,29	Invalid	0,16	Invalid	0,37	Sedang	0,30	Sulit	-0,07	Kurang	-0,07	Kurang
TP 2. Peserta didik mampu menerapkan hukum perbandingan tetap dalam perhitungan kimia												
2	0,65	Valid	0,49	Valid	0,77	Mudah	0,53	Sedang	0,47	Baik Sekali	0,53	Baik Sekali
11	0,75	Valid	0,62	Valid	0,60	Sedang	0,67	Sedang	0,67	Baik Sekali	0,53	Baik Sekali
12	0,42	Valid	0,31	Invalid	0,07	Sulit	0,13	Sulit	0,13	Kurang	0,00	Kurang
13	0,61	Valid	0,76	Valid	0,70	Sedang	0,33	Sedang	0,33	Baik	0,67	Baik Sekali
TP 3. Peserta didik mampu menerapkan hukum perbandingan berganda dalam perhitungan kimia												
3	0,64	Valid	0,46	Valid	0,73	Mudah	0,63	Sedang	0,40	Baik Sekali	0,33	Baik
5	0,19	Invalid	0,53	Valid	0,93	Mudah	0,47	Sedang	0,13	Kurang	0,53	Baik Sekali
6	0,37	Valid	0,38	Valid	0,93	Mudah	0,47	Sedang	0,13	Kurang	0,27	Sedang
14	-0,25	Invalid	0,39	Valid	0,53	Sedang	0,60	Sedang	-0,13	Kurang	0,40	Baik Sekali
15	0,39	Valid	0,52	Valid	0,53	Sedang	0,73	Mudah	0,27	Sedang	0,53	Baik Sekali
TP 4. Peserta didik mampu menerapkan hukum perbandingan volume dalam perhitungan kimia												
7	0,64	Valid	0,38	Valid	0,53	Sedang	0,50	Sedang	0,40	Baik Sekali	0,33	Baik
16	0,51	Valid	0,43	Valid	0,33	Sedang	0,30	Sulit	0,40	Baik Sekali	0,20	Sedang
18	0,59	Valid	0,12	Invalid	0,77	Mudah	0,80	Mudah	0,33	Baik	0,00	Kurang
19	0,32	Invalid	0,43	Valid	0,10	Sulit	0,73	Mudah	0,20	Sedang	0,27	Sedang
TP 5. Peserta didik mampu menerapkan hipotesis Avogadro dalam perhitungan kimia												
4	0,84	Valid	0,46	Valid	0,70	Sedang	0,67	Sedang	0,60	Baik Sekali	0,40	Baik Sekali
17	0,70	Valid	0,44	Valid	0,60	Sedang	0,33	Sedang	0,53	Baik Sekali	0,40	Baik Sekali
20	0,38	Valid	0,39	Valid	0,30	Sulit	0,43	Sedang	0,20	Sedang	0,33	Baik

Berdasarkan kesimpulan pada Tabel 7 tersebut, didapatkan hasil bahwa beberapa soal yang tidak valid

mesti diubah, direvisi maupun dibuang. Perubahan maupun perevisian dapat dipertimbangkan melalui isi

dari konten, bahasa soal yang digunakan, maupun berdasarkan tingkat kesukaran dan daya pembeda dari soal. Namun, dari Tabel 7, terdapat 14 soal yang dapat digunakan dan memiliki kriteria yang baik, serta 6 soal merupakan soal dengan kriteria tidak cukup baik karena soal tersebut memiliki kategori tidak valid, sehingga jika soal tersebut tetap ingin dipergunakan maka diperlukan dilakukannya perubahan maupun perbaikan agar soal tersebut valid dan dapat digunakan oleh guru maupun siswa dengan baik.

Instrumen ini dibuat dengan tujuan agar dapat memetakan model mental peserta didik pada materi hukum dasar kimia. Dalam Kania dkk. (2020), telah dikembangkan rubrik untuk memetakan model mental tersebut melalui tingkat pemahaman konsep^[29].

Tabel 8. Kriteria model mental peserta didik

Model Mental	Tingkat Pemahaman Konsep
<i>Scientific Model</i>	<i>Sound Understanding (SU)</i>
<i>Synthesis-Partial Understanding A</i>	<i>Partial Understanding (PU)</i>
<i>Synthesis-Partial Understanding B</i>	<i>Partial Understanding with Alternative Conception (PU-AC)</i>
<i>Synthesis-Misconception</i>	<i>Misconception (MC)</i>
<i>Initial Model</i>	<i>No Response (NC)</i> <i>No Understanding (NU)</i>

Dengan pengkriteriaan model mental berdasarkan tingkat pemahaman diatas, didapatkan penjelasan pada tingkat pemahaman sebagai berikut.

Tabel 10. Pengkodean instrumen *four-tier*

Tier	Kategori																
	SU		PU		PU-AC				MC	NU		NC					
1	B	B	B	B	B	B	B	B	S	S	S	S	S	S	S	S	S
2	Y	T	Y	T	Y	T	Y	T	Y	T	Y	T	Y	Y	T	T	NR
3	B	B	B	B	S	S	S	S	B	B	B	B	S	S	S	S	S
4	Y	Y	T	T	Y	Y	T	T	Y	Y	Y	T	Y	T	Y	T	NR

Keterangan:

- B : Benar
- Y : Yakin
- S : Salah
- T : Tidak Yakin

Penggunaan model mental tersebut akan membantu guru dalam memetakan model mental dengan soal instrumen yang telah dikembangkan tersebut. Jika peserta didik pada salah satu soal menjawab dengan tepat dan alasan yang tepat serta peserta didik yakin dengan jawabannya maka dapat dikategorikan bahwa peserta didik tersebut memiliki tingkat pemahaman *Sound Understanding* dengan model mental saintifik. Namun, jika peserta didik tidak menjawab dengan benar dan alasan yang diberikanpun salah namun keyakinan pada jawaban yang diberikan adalah yakin, maka kategori dari peserta didik ini adalah pada tingkat pemahaman *misconception* dengan

Tabel 9. Kategori tingkat pemahaman peserta didik

Tingkat Pemahaman Konsep	Keterangan
<i>Sound Understanding (SU)</i>	Jawaban terisi dengan semua komponen jawaban yang dapat diterima secara ilmiah
<i>Partial Understanding (PU)</i>	Jawaban yang diisi mengandung beberapa komponen respon yang dapat diterima dengan ilmiah
<i>Partial Understanding with Alternative Conception (PU-AC)</i>	Jawaban yang diisi memperlihatkan bahwa konsep telah dipahami tetapi juga mengandung konsep lain
<i>Misconception (MC)</i>	Jawaban yang diisi salah secara ilmiah serta informasi yang ada berisi informasi yang tidak masuk akal atau tidak benar
<i>No Response (NC), No Understanding (NU)</i>	Jawaban kosong, tidak relevan, atau tidak jelas.

Sehingga, didapatkan pengkodean untuk instrumen *four-tier* pada Tabel 10.

model mental sintesis-miskonsepsi. Dengan pengkategorian melalui rubik oleh kania, dkk. (2020) maka guru akan mampu melakukan pemetaan model mental dengan menggunakan soal instrumen yang dikembangkan ini.

4. KESIMPULAN

Berdasarkan penelitian, telah dikembangkan instrumen tes diagnostik *four-tier* yang terbukti valid dan reliabel, serta memenuhi standar terkait tingkat kesukaran, daya pembeda, dan kualitas distraktor. Instrumen ini memiliki validitas logis sebesar 0,96 dan reliabilitas pada tingkat pertama dan tingkat ketiga

yang sangat tinggi, yaitu berturut-turut 0,89 dan 0,86. Dari hasil ujicoba, terdapat 6 soal tidak valid, sedangkan 14 soal lainnya valid dan dapat digunakan.. Hasil ini juga didukung dengan tingkat kesukaran pada soal yang memiliki sebaran soal yang cukup baik dengan daya pembeda soal yang valid berkategori baik dan sebaran distraktor yang cukup baik.

REFERENSI

- [1] Treagust DF, Chittleborough G, Mamiala TL. The Role of Submicroscopic and Symbolic Representations In Chemical Explanations. *Int J Sci Educ.* 2003; 25(11): 1353–68.
- [2] Redhana IW, Sudria IB, Suardana IN, Suja IW, Putriani VD. Students' Mental Models in Acid-Base Topic. *J Phys Conf Ser.* 2020.
- [3] Herron JD, Cantu LL, Ward R, Srinivasan V. Problems Associated with Concept Analysis. *Sci Educ.* 1977; 61(2): 185–99.
- [4] Krisnawati I, Fajaroh F. Menggali Pemahaman Konsep Siswa Madrasah Aliyah tentang Stoikiometri dengan Menggunakan Instrumen Diagnostik Two-Tier. *J Kim FMIPA Univ Negeri Malang.* 2013; (2006): 1–7.
- [5] Ulva Y., Santosa A, Parlan. Identifikasi Tingkat Pemahaman Konsep Larutan Penyangga. *J Pembelajaran Kim.* 2016; 1(2): 69–75.
- [6] Yoni AAS, Suja IW, Karyasa IW. Profil Model Mental Siswa SMA Kelas X Tentang Konsep-Konsep Dasar Kimia pada Kurikulum Sains SMP. *J Pendidik Kim Indones.* 2018; 2(2): 64–9.
- [7] Laliyo LAR. Model Mental Siswa Dalam Memahami. *J Penelit dan Pendidik.* 2011; 8(1).
- [8] Atikah, Habiddin, Nazriati, Rahayu S, Dasna IW. A Systematic Literature Review: Model Mental pada Konsep-Konsep Kimia. *J Inov Pendidik Kim.* 2023; 17(2).
- [9] Devetak I, Lorber Drogenik ED, Jurišević M, Glažar SA. Comparing Slovenian Year 8 and Year 9 Elementary School Pupils' Knowledge of Electrolyte Chemistry and Their Intrinsic Motivation. *Chem Educ Res Pract.* 2009; 10(4): 281–90.
- [10] Devi NA, Azra F. Pengembangan Instrumen Tes Diagnostik untuk Melihat Gambaran Model Mental Peserta Didik pada Materi Asam Basa. *Entalpi Pendidik Kim.* 2023; 16–26.
- [11] Dewi KM, Suja IW, Sastrawidana IDK. Model Mental Siswa Tentang Termokimia. *J Pendidik Kim Undiksha.* 2018; 2(2): 45–52.
- [12] Handayanti Y, Setiabudi A, Nahadi. Analisis Profil Model Mental Siswa SMA. *J Penelit Pendidik IPA.* 2015; 1(1): 107–22.
- [13] Kurnaz MA, Eksi C. An Analysis of High School Students' Mental Models of Solid Friction in Physics. *Educ Sci Theory Pract* 2015;15(3):787–95.
- [14] Suja IW, Suidiana IK, Redhana IW, Sudria IBN. Mental Model of Prospective Chemistry Teachers on Electrolyte and Nonelectrolyte Solutions. *IOP Conf Ser Mater Sci Eng* 2021;1115(1):012064.
- [15] Kafiyan F, Samsudin A, Saepuzaman D. Development of four-tier diagnostic test (FTDT) to identify student 's mental models on static fluid. *J Phys Conf Ser* 2019;
- [16] Dewi FH, Samsudin A, Chandra DT, Aminudin AH, Kunaedi J, Astuti IRW, et al. How is students ' mental model in the post-pandemic era? Work and Energy Concept Analysis using the Wright Map. *J Penelit Pendidik IPA* 2023;9(3):1260–8.
- [17] Aagustin U, Susilaningsih E, Nurhayati S, Wijayati N. Pengembangan Instrumen Tes Diagnostik *Four-Tier Multiplechoice* Untuk Identifikasi Miskonsepsi Siswa Pada Materi Kesetimbangan Kimia. *Chem Educ* 2022;11:1–7.
- [18] Mi S, Ye J, Yan L, Bi H. Development and validation of a conceptual survey instrument to evaluate senior high school students' understanding of electrostatics. *Phys Rev Phys Educ Res.*
- [19] Habiddin, Page EM. Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indones J Chem* 2019;19(3):720–36.
- [20] Fakhriyah F, Masfuah S. The Development of a Four Tier-Based Diagnostic Test Diagnostic Assessment on Science Concept Course. *J Phys Conf Ser* 2021;1842(1).
- [21] Ni Made Ary Suparwati. Analisis Reduksi Miskonsepsi Kimia dengan Pendekatan Multi Level Representasi: Systematic Literature Review. *J Pendidik Mipa* 2022;12(2):341–8.
- [22] Treagust DF. Development and use of diagnostic tests to evaluate students' misconceptions in science. *Int J Sci Educ* 1988;10(2):159–69.
- [23] Kurniawan A, Febrianti AN, Hardianti T, Ichsan I, Desy D, Risan R, et al. Evaluasi pembelajaran. Padang: PT Global Eksekutif Teknologi; 2022.
- [24] Supardi. Statistik Penelitian Pendidikan. Depok: PT Raja Grafindo Persada; 2017.
- [25] Arikunto S. Dasar-Dasar Evaluasi Pendidikan. Edisi 3. Jakarta: Bumi Aksara; 2018.
- [26] Yusuf AM. Asesmen dan Evaluasi Pendidikan. Jakarta: Kencana; 2017.
- [27] Maulidiana CV, Azra F. Development of Four-tier Diagnostic Test Instrument to Identify Students' Mental Models on Salt Hydrolysis Materials. *J Pendidik dan pembelajaran Kim* 2023;12(3):116–26.
- [28] Rahmi R, Kustati M, Hardeli H. Evaluasi Pendidikan Perspektif Islam. Yogyakarta: deepublish; 2022.
- [29] Kania VI, Samsudin A, Aminudin AH. Multitier of Greenhouse Effect (MoGE) Instrument Development to Identify Middle School Students Mental Model in Thailand with Rasch Analysis Multitier of Greenhouse Effect (MoGE) Instrument Development to Identify Middle School Students' Mental Mod. 2020;(May).