

Development of Four-Tier Multiple Choice Diagnostic Test to Determine Students' Concept Understanding Level On Thermochemical Material

Sanaa Jauza Roghdah¹, Muhammad Zammi¹, Julia Mardhiya¹

¹Pendidikan Kimia, UIN Walisongo Semarang

Abstract

The difficulty of students in understanding thermochemical material through a learning process has the potential to cause difficulties for students to understand other chemical materials. To avoid this, it is necessary to find the location of the difficulties of students with tests that are diagnostic to know the level of understanding and weaknesses of students in learning. The purpose of this study was to determine the characteristics and quality of the developed four-level multiple-choice diagnostic test instrument. This study uses the Design-Based Research (DBR) method with a development model adapted by Reeves. The subjects of this study were students of class XI MAN 2 Semarang. The instruments developed to consist of a grid of test questions, test instructions, test questions, answer keys, scoring guidelines, and guidelines for interpreting test results. Through the trials conducted, this diagnostic test instrument was declared valid and reliable with a reliability value of 0.985

Keywords: Diagnostic test, level of understanding, thermochemistry.

Pengembangan Four-Tier Multiple Choice Diagnostic Test Untuk Mengetahui Tingkat Pemahaman Konsep Peserta Didik Pada Materi Termokimia

Abstrak

Kesulitan peserta didik pada pemahaman materi termokimia melalui suatu proses pembelajaran berpotensi menyebabkan kesulitan peserta didik dalam

memahami materi kimia yang lain. Untuk menghindari hal ini, maka perlu dicari letak kesulitan peserta didik dengan tes yang bersifat diagnostik agar dapat diketahui tingkat pemahaman dan kelemahan peserta didik dalam pembelajaran. Penelitian ini bertujuan untuk mengetahui karakteristik dan kualitas instrumen tes diagnostik pilihan ganda empat tingkat yang dikembangkan. Penelitian ini menggunakan metode Design Based Research (DBR) dengan model pengembangan yang diadaptasi oleh Reeves. Subjek dari penelitian ini yaitu peserta didik kelas XI MAN 2 Kota Semarang. Instrumen yang dikembangkan terdiri atas kisi-kisi soal tes, petunjuk pengerjaan tes, soal tes, kunci jawaban, pedoman penskoran, dan pedoman interpretasi hasil tes. Melalui uji coba yang dilakukan, instrumen tes diagnostik ini dinyatakan valid dan reliabel dengan nilai reliabilitas sebesar 0,985.

Kata kunci: Tes diagnostik, tingkat pemahaman, termokimia.

PENDAHULUAN

Ketidakmampuan siswa dalam mencapai tujuan belajar sebagaimana yang diharapkan dapat disebabkan karena kecenderungan untuk membentuk konsep berdasarkan pemahaman mereka sendiri pada suatu proses pembelajaran, yang terkadang hasilnya tidak sesuai menurut teori aslinya Hal tersebut seringkali terjadi pada siswa yang belum mempunyai kerangka pengetahuan sebagai acuan (Septiana et al., 2014). Keberhasilan suatu proses pembelajaran dapat ditentukan berdasarkan pemahaman siswa terhadap konsep yang telah disampaikan oleh guru.

Termokimia pada umumnya merupakan cabang ilmu kimia mengenai energi yang menyertai perubahan fisika dan reaksi kimia. Pada proses pembelajarannya di sekolah jenjang menengah atas, materi termokimia memiliki tujuan yaitu untuk memberikan siswa kemampuan dalam pemahaman konsep perubahan entalpi reaksi pada persamaan termokimia dan penentuan berbagai kalor reaksi. Pemahaman materi termokimia dapat dikaitkan dengan fakta yang dilihat dalam kehidupan sehari-hari (makroskopis) seperti peristiwa pembakaran atau penguapan. Kemudian melalui peristiwa itu dapat dikaitkan lagi dengan fakta yang tidak bisa dilihat secara langsung (mikroskopis) seperti bentuk energi yang diserap atau dibebaskan dalam persamaan termokimia, yang selanjutnya agar dapat menentukan berbagai kalor reaksi fase-fase pembentukan perubahan entalpi, atau

menghitung penentuan perubahan entalpi (simbolik) (Irby et al., 2016; Talanquer, 2011)

Apabila siswa tidak mampu memahami materi termokimia, hal tersebut dapat berdampak pada kemampuan siswa dalam memahami materi kimia yang lain. Karena pada sistem reaksi kimia, perubahan fase, dan pembentukan larutan dapat melibatkan data perubahan energi yang diperkirakan menggunakan ilmu termokimia. Sehingga banyak dinyatakan bahwa setiap materi dalam pelajaran kimia memiliki keterkaitan satu sama lain. Maka, apabila siswa tidak memahami konsep materi sebelumnya maka dimungkinkan akan sulit untuk memahami materi selanjutnya.

Kesulitan siswa pada pemahaman materi termokimia telah dibuktikan oleh beberapa peneliti dalam penelitiannya, diantaranya; kesulitan siswa dalam membedakan kalor dan suhu, melakukan identifikasi reaksi eksoterm dan endoterm, serta melakukan identifikasi sistem dan lingkungan pada reaksi yang berlangsung menggunakan kalorimeter (Yalçinkaya et al., 2009). Menurut hasil observasi pada beberapa siswa di MAN 2 Kota Semarang, diketahui bahwa anggapan kimia merupakan mata pelajaran yang sulit oleh sebagian siswa menyebabkan kurangnya keterlibatan mereka dalam proses pembelajarannya. Meskipun siswa melakukan penafsiran sendiri, dipastikan bahwa konsep yang tertanam pada diri siswa merupakan konsep kimia yang sebenarnya. Namun ada kalanya konsep kimia yang tertanam pada diri siswa itu menyalahi konsep kimia yang sebenarnya, hal ini bisa mempengaruhi proses belajar siswa secara berkelanjutan. Berdasarkan keragaman keadaan tersebut maka diketahui terdapat empat kategori pemahaman konsep yang mungkin dialami oleh siswa, yaitu; memahami konsep, miskonsepsi, tidak memahami konsep, dan error (Yakubi & Hanum, 2017). Miskonsepsi yaitu suatu konsepsi yang menyimpang dari teori ilmiah menurut para ahli pada bidang tertentu (Shefityawan et al., 2018). Dalam suatu pembelajaran hal tersebut bisa terjadi jika pengetahuan siswa yang diterima dalam proses pembelajaran tidak menyeluruh, sehingga ada kemungkinan siswa membangun pengetahuan konsep berdasarkan pemahamannya sendiri (Rohmah, 2019).

Perbedaan kemampuan siswa dalam memahami materi menunjukkan bahwa untuk mengidentifikasi tingkat pemahaman konsep tidak dapat dilakukan dengan menggunakan tes objektif biasa yang memungkinkan siswa dapat menebak jawaban. Melainkan tes yang harus bersifat diagnostik agar dapat diketahui kelemahan-kelemahan siswa sehingga bisa dilakukan pemberian penanganan yang tepat, serta dapat diketahui secara pasti

bagaimana tingkat pemahaman siswa pada materi tersebut (Gurel et al., 2015). Nurhidayatullah & Prodjosantoso (2018), menyampaikan dalam penelitiannya bahwa untuk melakukan perbaikan pengajaran (*learning corrective*) secara efektif yang terpenting adalah melakukan kegiatan diagnosis kesulitan belajar untuk menemukan letak kesulitan siswa. Karena suatu proses pembelajaran yang tidak berhasil dalam mencapai ketuntasan materinya dapat disebabkan oleh salah satu faktor seperti, siswa yang mengalami kesulitan dalam proses belajar.

Tes diagnostik disajikan dalam berbagai metode berupa pilihan ganda dengan alasan tertutup, pilihan ganda dengan alasan terbuka, dan esai tertulis (Nugraeni et al., 2013). Dalam tes pilihan ganda, telah disampaikan oleh Reynolds dalam (Caleon & Subramaniam, 2010) bahwa tes pilihan ganda mempunyai kelebihan serbaguna, mudah digunakan, objektif, dan efisien. Untuk pengujian suatu populasi, penggunaan tes pilihan ganda lebih mudah jika dibandingkan dengan metode lainnya. Namun dalam penggunaannya sebagai tes diagnostik, tes pilihan ganda mempunyai kelemahan yaitu tidak bisa membedakan jawaban benar atau salah (Caleon & Subramaniam, 2010). Oleh karena itu dikembangkan tes pilihan ganda dua tingkat untuk mengurangi kekurangan tes pilihan ganda dalam mengetahui tingkat pemahaman siswa.

Pengembangan tes pilihan ganda dua tingkat (*two-tier multiple choice*) dilakukan dengan menambahkan pilihan ganda di tahap kedua yang berisi alasan-alasan yang relevan dan dapat membenarkan jawaban pada tahap pertama. Namun adapun beberapa kelemahannya yang telah disampaikan oleh Tamir bahwa; (1) di tahap kedua pada pilihan alasan dapat ditemukan petunjuk jawaban yang benar yang sebelumnya tidak terpikirkan oleh peserta dalam tes jawaban terbuka atau wawancara, (2) belum bisa dibedakan jawaban peserta yang salah dikarenakan miskonsepsi atau belum paham konsep, dan belum bisa dibedakan pula jawaban yang benar dikarenakan paham konsep atau menebak (Gurel et al., 2015). Untuk mengurangi kekurangan yang telah disebutkan, maka dilakukan penambahan tahap dalam pengembangan tes diagnostik pilihan ganda selanjutnya.

Tes pilihan ganda tiga tingkat (*three-tier multiple choice*) dikembangkan dengan penambahan tahap ketiga yang berisi tingkat keyakinan peserta untuk jawaban dalam tahap pertama dan kedua. Kemudian dinyatakan oleh Caleon & Subramaniam (2010) bahwa perbedaan keyakinan pada jawaban dan alasan dimungkinkan dimiliki oleh

peserta. Sedangkan pada tes pilihan ganda tingkat tiga ini pertanyaan tingkat keyakinan digunakan secara bersamaan. Hal tersebut dapat menyebabkan hasil proporsi terlalu tinggi untuk peserta paham konsep dan terlalu rendah untuk peserta yang belum paham konsep (Gurel et al., 2015). Maka tes pilihan ganda empat tingkat dikembangkan dengan harapan dapat mengurangi kekurangan-kekurangan yang dihasilkan oleh metode tes diagnostik sebelumnya. Pada tes pilihan ganda empat tingkat dilakukan penambahan tingkat keyakinan, sehingga tingkat keyakinan ditanyakan secara terpisah pada jawaban dan alasan.

Tes diagnostik *four-tier multiple choice* merupakan tes diagnostik pilihan ganda dengan empat tahap. Tahap pertama berupa pertanyaan dan pilihan jawaban (tes pilihan ganda). Tahap kedua berupa tingkat keyakinan menurut respon dari tahap pertama. Tahap ketiga berisi pilihan alasan yang membenarkan jawaban dalam tahap pertama, lalu untuk tahap keempat berupa tingkat keyakinan menurut respon dari tahap ketiga.

Beberapa peneliti telah melakukan tes diagnostik pilihan ganda empat tingkat untuk beberapa pokok bahasan dalam pembelajaran kimia diantaranya pada materi ikatan kimia (Ritonga & Yasthophi, 2019; Yakubi & Hanum, 2017), kinetika kimia (Habiddin & Page, 2019), kesetimbangan kimia (Ardiansah et al., 2018), laju reaksi (Harahap & Novita, 2020), elektrokimia (Murniningsih et al., 2020). Maka dengan ini peneliti bermaksud untuk melakukan pengembangan Instrumen tes diagnostik pilihan ganda empat tingkat (*four-tier*) guna mengetahui tingkat pemahaman konsep siswa pada materi termokimia.

METODE PENELITIAN

Metode penelitian *Design Based Research* (DBR) digunakan dalam penelitian ini untuk mengembangkan suatu produk berupa instrumen tes diagnostik pilihan ganda empat tingkat. Metode DBR yakni metode yang diterapkan dalam perancangan, pengembanan dan perbaikan suatu produk pendidikan menurut teori dasar dan penelitian sebelumnya (Plomp, 2013). Metode penelitian ini digunakan bersama model pengembangan yang diadaptasi oleh Reeves dikutip dalam Herrington et al., (2007) dengan tahapan model pengembangan yang terdiri dari empat langkah sebagai berikut; (1) Analisis dan identifikasi masalah, (2) Perancangan solusi, (3) Siklus berulang dalam pengujian dan penyempurnaan rancangan, dan (4) Refleksi pada produk dan menguatkan implementasi solusi. Subjek dari penelitian ini yaitu siswa kelas XI MAN 2 Kota

Semarang.

Adapun indikator pemahaman konsep yang digunakan adalah pendapat Depdiknas (2006) adalah sebagai berikut : (1) menyatakan ulang sebuah konsep, (2) mengklasifikasikan objek-objek menurut sifat-sifat tertentu (sesuai dengan konsepnya) dan (3) memberikan contoh dan non contoh dari konsep. Interpretasi hasil tes dilakukan berdasarkan analisis penskoran tes yang hasilnya diinterpretasikan ke dalam empat kategori, yaitu; paham konsep, tidak paham konsep, miskonsepsi, dan *error*. Interpretasi hasil tes diagnostik dapat dilihat pada Tabel 1. Penggunaan interpretasi pada skala 1-6 tersebut yakni dengan ketentuan; skala 6 (amat sangat yakin), 5 (sangat yakin), dan 4 (yakin) digolongkan dalam jenis tingkat keyakinan yang tinggi. Sedangkan pada skala 3 (tidak yakin), 2 (sangat tidak yakin), dan 1 (menebak) digolongkan dalam jenis tingkat keyakinan yang rendah.

Tabel 1. Interpretasi Hasil Tes Diagnostik Pilihan Ganda Empat Tingkat

Kategori	Jawaban	Confidence Rating pada Jawaban	Alasan	Confidence Rating pada Alasan
Paham konsep	<i>True</i>	≥ 4	<i>True</i>	≥ 4
	<i>True</i>	1-6	<i>True</i>	≤ 3
Tidak paham konsep	<i>True</i>	≤ 3	<i>False</i>	≥ 4
	<i>False</i>	1-6	<i>True</i>	≤ 3
Miskonsepsi	<i>True</i>	1-6	<i>False</i>	≥ 4
	<i>False</i>	1-6	<i>True</i>	≥ 4
Error	<i>False</i>	1-6	<i>True</i>	≥ 4

Berdasarkan hasil interpretasi yang dilakukan, maka dapat diketahui persentase terhadap hasil penilaian di setiap kategori melalui rumus berikut (Sudijono, 2015).

$$P = \frac{f}{N} \times 100\%$$

Keterangan:

P= persentase siswa di setiap kategori

f = jumlah siswa di setiap kategori

N= jumlah siswa (responden)

Setelah dilakukan perhitungan maka akan didapatkan kriteria persentase berdasarkan Tabel 2.

Tabel 2. Kriteria Persentase pada Setiap Kategori

Persentase	Kriteria
$60\% \geq P$	Tinggi
$30\% \leq P < 60\%$	Sedang
$P < 30\%$	Rendah

HASIL DAN PEMBAHASAN

Bagian ini merupakan uraian hasil dari penelitian dan pengembangan yang telah dilakukan. Sehingga akan dijelaskan mengenai hasil pengembangan produk awal, data hasil uji coba produk yang terdiri atas uji validitas, reliabilitas, tingkat kesukaran, dan daya beda, revisi produk, serta kajian produk akhir.

Hasil Pengembangan Produk Awal

Pengembangan instrumen tes diagnostik pilihan ganda empat tingkat memerlukan suatu pengumpulan informasi serta kajian literatur. Hasil dari pengumpulan informasi dan kajian literatur tersebut dapat digunakan sebagai petunjuk bagi peneliti dalam menyusun produk awal yang dikembangkan.

Pengumpulan informasi dan studi literatur

Studi literatur dilakukan dengan pengumpulan beberapa sumber seperti buku atau jurnal tentang evaluasi pembelajaran dan tes diagnostik, sehingga diperoleh informasi digunakan untuk pengembangan ini. Sedangkan observasi lapangan dilakukan agar diperoleh informasi serta fakta evaluasi mengenai mata pelajaran kimia. Menurut hasil observasi di MAN 2 Kota Semarang diketahui bahwa; (1) Siswa mengalami kesulitan pada materi termokimia, (2) Guru hanya menggunakan soal esai atau pilihan ganda biasa sebagai alat ukur tingkat pemahaman konsep siswa, (3) Guru belum pernah menerapkan tes diagnostik pilihan ganda empat tingkat untuk identifikasi tingkat pemahaman konsep siswa.

Penetapan solusi dan design awal

Keberhasilan suatu proses pembelajaran dapat diketahui menggunakan alat ukur penilaian hasil belajar. Alat ukur yang digunakan harus bersifat diagnostik agar dapat diketahui tingkat kemampuan siswa beserta kelemahannya sebagai acuan dalam pemberian penanganan yang tepat. Melalui tes diagnostik ini akan diketahui bagaimana tingkat pemahaman konsep siswa pada materi termokimia.

Pembuatan kisi-kisi meliputi dua komponen yaitu identitas dan matrik. Komponen identitas meliputi satuan pendidikan, kelas, mata pelajaran, dan pokok bahasan. Komponen matrik meliputi uraian materi, indikator, aspek kognitif, dan jumlah

soal. Kisi-kisi sebagai acuan penyusunan butir soal tes diagnostik disesuaikan dengan indikator dan aspek kognitif yang digunakan. Indikator dalam kisi-kisi menurut acuan indikator pada silabus Kimia SMA/MA Kurikulum 2013 Revisi, kemudian aspek kognitif yang digunakan menurut acuan aspek kognitif taksonomi Bloom meliputi; C₁ (pengetahuan), C₂ (pemahaman), C₃ (pengaplikasian), dan C₄ (analisis).

Petunjuk pengerjaan soal terdiri dari; tata cara dalam pengisian lembar jawab, pengerjaan, himbauan dan larangan, serta alokasi waktu. Setiap butir soal yang digunakan memiliki empat tahap. Tahap pertama berupa soal pilihan ganda dengan 4 pengecoh dan 1 jawaban. Tahap kedua berupa pilihan tingkat keyakinan untuk respon pada tahap pertama. Tahap ketiga berupa pilihan alasan dari respon pada tahap pertama berupa 3 pengecoh dan 1 jawaban. Tahap keempat berupa pilihan tingkat keyakinan untuk respon pada tahap ketiga. Jumlah soal dalam produk awal yang dikembangkan yakni 45 butir soal. Komponen lain dalam instrumen tes adalah kunci jawaban soal tes diagnostik pilihan ganda empat tingkat, lembar jawab tes diagnostik pilihan ganda empat tingkat, pedoman penilaian tes diagnostik pilihan ganda empat tingkat dan pedoman interpretasi tes diagnostik pilihan ganda empat tingkat

Tabel 3. Kisi-Kisi Soal Tes (produk awal)

Sub Pokok Bahasan	Jumlah soal
sistem dan lingkungan; reaksi eksoterm dan endoterm	16
macam-macam perubahan entalpi standar	10
penentuan δh reaksi menurut percobaan kalorimeter	5
penentuan δh reaksi menurut hukum hess	5
penentuan δh reaksi menurut data energi ikatan	4
perbandingan kalor pembakaran suatu bahan bakar	5
total	45

Hasil Uji Coba Produk

Uji coba produk dilakukan sebagai tahap pengujian dan penyempurnaan instrumen. Hasil dari tahap ini digunakan sebagai acuan dalam memperbaiki instrumen yang dikembangkan sebagai produk awal. Pada produk awal yang dikembangkan dilakukan penilaian dosen selaku validator ahli. Pada tahap penilaian oleh validator ini peneliti mendapatkan saran dari para ahli mengenai kelayakan instrumen yang digunakan. Lembar validasi yang diserahkan kepada validator mengandung 3 aspek penilaian yakni

berupa materi, konstruksi, dan bahasa. Kemudian untuk penilaian butir soalnya dilakukan dengan pemberian skor 1 (soal yang sesuai menurut indikator dalam lembar lembar validasi) dan skor 0 (tidak sesuai indikator). Hasil dari penilaian pada seluruh indikator tersebut kemudian dijumlahkan, dan digunakan sebagai acuan valid atau tidaknya butir soal. Berikut hasil penilaian yang diperoleh (Tabel 4).

Selain melalui angket validasi, validator juga memberikan saran secara langsung sebagai masukan untuk merevisi produk awal. Adapun saran yang diberikan yakni;

- a. Disarankan untuk menambah jumlah pengecoh pada pilihan alasan (tingkat ketiga)
- b. Disarankan untuk menggunakan tabel pada penulisan pilihan tingkat keyakinan jawaban dan alasan
- c. Disarankan untuk mengacak nomor soal pada beberapa soal yang mempunyai kesamaan indikator
- d. Disarankan untuk memperbaiki penggunaan kalimat pada beberapa soal
- e. Disarankan untuk memperbaiki penggunaan kalimat pada beberapa pengecoh

Tabel 4. Rekapitulasi Angket Validasi

Kategori butir soal	Total soal	
	Validator	Validator
	1	2
Butir soal disetujui	38	39
Butir soal disetujui dengan sedikit revisi	7	6
Butir soal perlu direvisi	0	0
Butir soal tidak disetujui	0	0

Selanjutnya dilakukan ujicoba skala kecil. Tahap ini dilaksanakan dengan melibatkan 30 siswa kelas XI di MAN 2 Kota Semarang. Pelaksanaan tahap ini bertujuan untuk mengetahui validitas butir soal, reliabilitas, tingkat kesukaran, dan daya beda soal. Hasil dari tahap ini kemudian direkap dan dikalkulasi menggunakan *Microsoft Excel*.

Data yang diperoleh dalam uji coba ini dianalisis untuk mengetahui validitas butir soal. Penentuan validitas item soal objektif dalam instrumen ini menggunakan rumus korelasi point biserial. Nilai r_{tabel} menggunakan taraf signifikansi 5% sehingga r_{tabel} yang digunakan sebagai acuan dalam interpretasi yaitu 0,444.

Tabel 5. Rekapitulasi Hasil Analisis Validitas Butir Soal

Kategori	Total soal	Nomor soal
----------	------------	------------

Valid	25	2, 5, 6, 7, 8, 9, 11, 12, 16, 18, 20, 22, 23, 25, 26, 27, 28, 29, 30, 31, 34, 36, 43, 44, 45
Invalid	20	1, 3, 4, 10, 13, 14, 15, 17, 19, 21, 24, 32, 33, 35, 37, 38, 39, 40, 41, 42

Dalam perhitungan reliabilitas soal peneliti menggunakan rumus *Cronbach Alpha*. Berdasarkan hasil jawaban siswa maka diperoleh perhitungan r_{11} sebesar 0,985. Menurut hasil tersebut dapat diketahui bahwa soal tes diagnostik yang dikembangkan bersifat reliabel. Analisis tingkat kesukaran pada butir soal dilakukan agar dapat diketahui kriteria mudah atau sukarnya pengerjaan item soal yang digunakan. Hasil tingkat kesukaran butir soal dapat dilihat pada Tabel 6.

Tabel 6. Rekapitulasi Tingkat Kesukaran

Tingkat kesukaran	Nomor soal	Total
Mudah	1	1
Sedang	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19, 20, 22, 24, 27, 28, 30, 32, 33, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45	34
Sukar	12, 17, 21, 23, 25, 26, 29, 31, 34, 36	10

Perbedaan kemampuan yang dimiliki oleh siswa menyebabkan perlunya dilakukan analisis daya beda pada butir soal untuk membedakan tingkat kemampuan siswa dalam pengerjaan soal. Dalam uji ini akan diketahui kategori butir soal yang digunakan menurut tingkat kemampuan siswa. Butir soal dikatakan memiliki nilai daya beda ‘baik’ apabila hanya terjawab secara tepat oleh siswa dengan kemampuan tinggi. Sedangkan untuk butir soal yang ‘jelek’ merupakan butir soal yang terjawab maupun tidak terjawab oleh semua siswa, serta hanya terjawab dengan baik oleh siswa dengan kemampuan rendah.

Tabel 7. Rekapitulasi Daya Beda

Daya beda	Nomor soal	Total
Baik	2, 3, 5, 7, 8, 9, 18, 22, 28, 43, 45	11
Cukup	1, 6, 10, 11, 12, 14, 15, 16, 17, 19, 20, 23, 25, 26, 27, 29, 30, 31, 32, 34, 36, 37, 40, 41, 42, 44	26
Jelek	4, 13, 21, 24, 33, 35, 38, 39	8

Pemberian angket respon kepada guru dilaksanakan agar diketahui pendapat guru mengenai tes diagnostik yang dikembangkan. Hasil respon guru diperoleh kesimpulan sebagai berikut :

- 1) Soal yang disajikan pada instrumen tes diagnostik pilihan ganda empat tingkat sudah sesuai menurut kompetensi dasar, indikator pembelajaran, dan pokok bahasan termokimia dalam pembelajaran.
- 2) Penggunaan bahasa dan teknik penulisan pada soal sudah cukup bagus sehingga dapat terbaca dengan jelas
- 3) Penggunaan kalimat soal sudah baik, panjangnya sesuai, dan dapat dimengerti
- 4) Gambar, tabel dan persamaan reaksi yang disajikan sudah jelas, terbaca, dan dapat dipahami
- 5) Jumlah soal dan waktu yang disediakan cukup sesuai

Pemberian angket respon pada siswa dilakukan agar dapat diketahui penilaian siswa untuk instrumen tes yang dikembangkan. Bagian-bagian yang disajikan dalam angket yakni judul, identitas, petunjuk pengisian angket, dan 10 pertanyaan. Angket yang digunakan disajikan dalam bentuk google form.

Tabel 8. Hasil Angket Respon Siswa

No.	Aspek Penilaian	(%)	Kategori
1	Kalimat soal uji terbaca	85.33	Sangat baik
2	Kalimat soal uji dapat dipahami	78.00	Baik
3	panjang kalimat dalam soal uji disajikan secara tepat	79.33	Baik
4	Kalimat dalam soal uji terbaca	86.67	Sangat baik
5	Pertanyaan soal dapat dipahami	76.67	Baik
6	Kebebasan pertanyaan dalam soal tes untuk menimbulkan penafsiran ganda	86.00	Sangat baik
7	Gambar/tabel dan persamaan reaksi dalam soal terbaca	89.33	Sangat baik
8	Gambar/tabel dan persamaan reaksi dalam soal dapat dipahami	89.33	Sangat baik
9	Jumlah soal yang disediakan sesuai	60.00	Cukup
10	Waktu yang disediakan untuk mengerjakan soal sesuai	84.67	Sangat baik
Rata-rata		81.19	Sangat baik

Pada produk akhir tes diagnostik pilihan ganda empat tingkat yang memiliki 25 soal diujikan terhadap 60 siswa sebagai uji coba tahap skala luas dengan hasil sebagai berikut:

a. Tingkat pemahaman siswa

Persentase pemahaman siswa dari jumlah total sampel dapat dilihat pada Tabel 4.7.

Menurut tabel tersebut dapat diketahui rata-rata persentase pemahaman konsep materi

termokimia 60 siswa kelas XI di MAN 2 Kota Semarang bahwa sejumlah 22% memahami konsep (P), 26% tidak memahami konsep (TP), 39% miskonsepsi (M), dan 13% eror (E).

b. Tingkat pemahaman siswa pada setiap butir soal

Soal tes diagnostik yang digunakan pada tahap ini berjumlah 25 dengan rincian; 9 soal pada sub materi reaksi eksoterm dan endoterm, 6 soal pada sub materi macam-macam perubahan entalpi standar, 5 soal pada sub materi penentuan ΔH reaksi menurut percobaan kalorimeter, 1 soal pada sub materi penentuan ΔH reaksi menurut hukum hess, 1 soal pada sub materi penentuan ΔH reaksi menurut data energi ikatan, dan 3 soal pada sub materi perbandingan kalor pembakaran suatu bahan bakar. Persentase pemahaman siswa pada tiap butir soal bisa dilihat dalam Tabel 4.8.

c. Nilai akhir

Nilai akhir ditentukan berdasarkan ketentuan penilaian, kemudian dihitung menggunakan rumus nilai akhir. Dari hasil perhitungan skor tersebut diperoleh nilai tertinggi sebesar 100,00 dan nilai terendah sebesar 18,00.

Revisi Produk

Revisi produk instrumen tes diagnostik pilihan ganda empat tingkat dilakukan dengan melakukan perbaikan produk menggunakan hasil dari penilaian oleh validator dan data hasil uji coba skala kecil sebagai acuan dalam memperbaiki instrumen yang dikembangkan sehingga dapat dilakukan uji coba skala besar.

Penilaian dan saran dari validator terhadap instrumen tes diagnostik digunakan acuan dalam memperbaiki produk awal yang dikembangkan. Berdasarkan hasil angket validasi dan beberapa saran yang diperoleh, maka peneliti melakukan beberapa revisi sebagai berikut;

- a. Dilakukan penambahan jumlah pengecoh pada pilihan alasan (tingkat ketiga) pada semua soal. Sehingga pada tingkat ketiga terdapat 1 jawaban dan 4 pengecoh
- b. Dilakukan perbaikan pada penulisan pilihan tingkat keyakinan atau pada tahap kedua dan keempat.
- c. Dilakukan pengacakan nomor soal pada beberapa soal yang mempunyai kesamaan indikator
- d. Dilakukan perubahan kalimat soal pada butir soal nomor 7, 8, dan 9

- e. Dilakukan perubahan kalimat pada pengecoh butir soal nomor 3, 7, 8, dan 9

Revisi yang dilakukan pada produk awal instrumen tes menyebabkan beberapa perubahan pada kisi-kisi soal tes dan soal tes. Berdasarkan hasil analisis validitas butir soal pada Tabel 4.3, maka butir soal yang digunakan sebagai produk akhir yaitu berjumlah 25 soal. Hal tersebut dikarenakan pada 45 soal yang diujikan menghasilkan 20 soal tidak valid dan 25 soal valid. Berdasarkan analisis reliabilitas diketahui bahwa instrumen yang dikembangkan memperoleh nilai r sebesar 0,985. Nilai tersebut digolongkan dalam $0,70 < r \leq 1,00$ yang menunjukkan bahwa instrumen tes yang dikembangkan memiliki reliabilitas yang baik. Analisis derajat kesukaran atau taraf kesulitan dilaksanakan agar diketahui baik atau tidaknya butir item soal yang digunakan dalam tes. Pada hasil analisis tingkat kesukaran pada Tabel 6 ditemukan 76% soal berkategori sedang. Butir-butir item soal tes yang baik merupakan butir item yang tidak terlalu sulit dan tidak terlalu mudah untuk dikerjakan (Sudijono, 2015).

Pada hasil analisis daya pembeda item pada Tabel 7, ditemukan 8 butir soal berkategori jelek. Soal dengan kategori jelek tidak digunakan kembali untuk uji selanjutnya. Sehingga menurut pertimbangan tersebut, maka soal yang digunakan dalam produk akhir yaitu 10 soal kategori baik dan 15 soal kategori cukup.

Hasil respon yang sangat baik diperoleh dalam pemberian angket respon kepada pendidik sebagai penilaian instrumen tes diagnostik pilihan ganda empat tingkat yang dikembangkan. Angket penilaian siswa dilakukan agar diketahui penilaian siswa mengenai soal tes diagnostik yang dikembangkan. Hasil angket penilaian siswa memperoleh rata-rata sebesar 81,19% berkategori 'sangat baik'. Sehingga dapat diketahui bahwa tes diagnostik yang dikembangkan terbaca dengan baik oleh siswa.

Kajian Produk Akhir

Setelah dilakukan tahap revisi produk, maka akan dihasilkan produk akhir sebagai hasil dari penelitian ini. Hasil produk akhir juga digunakan kembali dalam uji skala luas untuk mengetahui tingkat pemahaman konsep siswa pada materi termokimia.

Kisi-kisi sebagai acuan penyusunan butir soal tes diagnostik disesuaikan dengan indikator dan aspek kognitif yang digunakan. Dalam produk akhir terdapat pengurangan jumlah butir soal menjadi 25 soal, sehingga terdapat perubahan dalam kisi-kisi soal. Kisi-

kisi soal pada produk akhir terdiri atas 19 indikator soal dari 6 sub pokok bahasan. Petunjuk pelaksanaan tes terdiri dari; tata cara pengisian lembar jawab dan penyelesaian soal tes diagnostik, himbauan dan larangan dalam pelaksanaan tes, serta alokasi waktu.

Setiap butir item soal yang digunakan memiliki empat tingkatan. Tahap pertama berupa soal pilihan ganda dengan 4 pengecoh dan 1 jawaban. Tahap kedua berupa tingkat keyakinan untuk respon pada tahap pertama. Tahap ketiga berupa alasan untuk respon pada tahap pertama dengan 4 pengecoh dan 1 jawaban. Tahap keempat berupa tingkat keyakinan untuk respon pada tahap ketiga. Jumlah soal dalam produk akhir yang dikembangkan yakni 25 butir soal.

Tabel 9. Kisi-Kisi Soal Tes (produk akhir)

Sub Pokok Bahasan	Jumlah soal
Sistem dan lingkungan; Reaksi eksoterm dan endoterm	9
Beraneka ragam perubahan entalpi standar	6
Penentuan ΔH reaksi menurut percobaan kalorimeter	5
Penentuan ΔH reaksi menurut hukum hess	1
Penentuan ΔH reaksi menurut data energi ikatan	1
Perbandingan kalor pembakaran pada suatu bahan bakar	3
TOTAL	25

Hasil dari tahap pengujian dan penyempurnaan instrumen yaitu berupa bentuk produk akhir dari instrumen tes diagnostik pilihan ganda empat tingkat. Produk akhir tersebut digunakan dalam uji skala besar yang hasilnya berupa data tingkat pemahaman siswa pada materi termokimia. Pada sub pokok materi reaksi eksoterm dan endoterm terdapat 9 soal yang digunakan pada tes diagnostik, yaitu butir soal nomor 1 sampai 9. Soal yang disajikan pada sub pokok materi ini merupakan soal tentang sistem terbuka dan tertutup; bentuk proses reaksi eksoterm dan endoterm pada persamaan reaksi; perubahan wujud zat pada proses pelepasan kalor. Menurut hasil penelitian diketahui bahwa rata-rata pemahaman siswa terhadap konsep reaksi eksoterm dan endoterm yaitu sebesar 30% termasuk dalam kategori sedang.

Pada sub pokok macam-macam perubahan entalpi terdapat 6 soal yang digunakan pada tes diagnostik, yaitu butir soal nomor 10 sampai 15. Soal yang disajikan pada sub pokok materi ini merupakan soal mengenai bentuk reaksi pembentukan, penguraian, dan pembakaran dengan data perubahan entalpi standar. Menurut hasil penelitian diketahui bahwa rata-rata pemahaman siswa terhadap konsep macam-macam perubahan entalpi

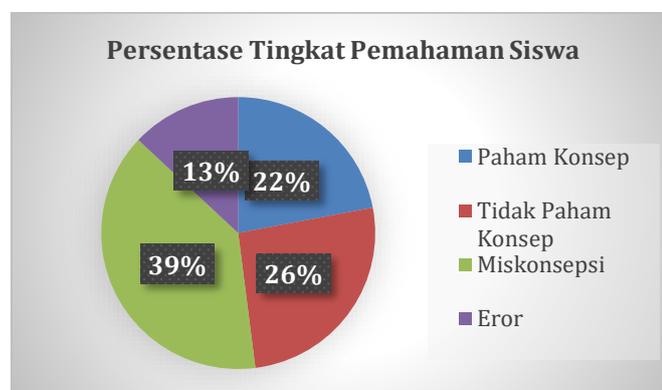
standar yaitu sebesar 22% termasuk dalam kategori rendah.

Pada sub pokok macam-macam perubahan entalpi terdapat 5 soal yang digunakan pada tes diagnostik, yaitu butir soal nomor 16 sampai 20. Soal yang disajikan pada sub pokok materi ini merupakan soal mengenai perhitungan perubahan entalpi menurut data hasil percobaan kalorimeter. Menurut hasil penelitian diketahui bahwa rata-rata pemahaman siswa terhadap konsep penentuan ΔH reaksi menurut percobaan kalorimeter yaitu sebesar 16% termasuk dalam kategori rendah.

Soal yang disajikan pada sub pokok materi ini merupakan soal dengan indikator mengkalkulasi perubahan entalpi reaksi menurut data bentuk awal dan akhir reaksi dalam sebuah diagram tingkat. Menurut hasil penelitian diketahui bahwa rata-rata pemahaman siswa terhadap konsep penentuan ΔH reaksi menurut hukum hess yaitu sebesar 12% termasuk dalam kategori rendah.

Soal yang disajikan pada sub pokok materi ini merupakan soal dengan indikator mengoreksi pernyataan yang dihasilkan dari suatu reaksi pembentukan menurut data energi ikatan. Menurut hasil penelitian diketahui bahwa rata-rata pemahaman siswa terhadap konsep penentuan ΔH reaksi berdasarkan data energi ikatan yaitu sebesar 12% termasuk dalam kategori rendah.

Pada sub pokok macam-macam perubahan entalpi terdapat 3 soal yang digunakan pada tes diagnostik, yaitu butir soal nomor 23 sampai 25. Soal yang disajikan pada sub pokok materi ini merupakan soal mengenai perhitungan kalor pembakaran pada suatu bahan bakar. Menurut hasil penelitian diketahui bahwa rata-rata pemahaman siswa terhadap konsep membandingkan kalor pembakaran suatu bahan bakar yaitu sebesar 14% termasuk dalam kriteria rendah.



Gambar 1. Diagram Persentase Tingkat Pemahaman Siswa

Dapat dilihat pada Gambar 1 bahwa persentase siswa secara keseluruhan dalam

kategori paham konsep sebanyak 22%, tidak paham konsep sebanyak 26%, miskonsepsi sebanyak 39%, dan eror sebanyak 13%. Menurut hasil tersebut dapat disimpulkan bahwa tingkat pemahaman konsep siswa memiliki kriteria 'rendah' karena hanya 22% dari jumlah total respon jawaban benar pada tahap pertama dan ketiga, serta yakin dalam pilihan jawaban maupun alasannya. Tingginya nilai persentase dengan kategori miskonsepsi menunjukkan bahwa banyaknya siswa yang mengalami miskonsepsi pada materi termokimia. Menurut Rohmah (2019), dalam suatu pembelajaran hal tersebut bisa terjadi jika pengetahuan siswa yang diterima dalam proses pembelajaran tidak menyeluruh, sehingga ada kemungkinan siswa membangun pengetahuan konsep berdasarkan pemahamannya sendiri. Berdasarkan uraian di atas maka dapat diketahui bahwa instrumen tes diagnostik dikembangkan bisa dijadikan sebagai alat untuk mengidentifikasi tingkat pemahaman siswa pada materi termokimia. Sejalan dengan hasil penelitian Gurel, Eryilmaz & McDermott (2015) yang mengatakan bahwa analisis tingkat pemahaman konsep siswa menggunakan instrumen tes diagnostik pilihan ganda empat tingkat dapat menghasilkan data yang sesuai berdasarkan faktanya, sehingga instrumen tes ini dapat dikembangkan untuk sampel yang lebih besar.

SIMPULAN

Karakteristik instrumen tes diagnostik pilihan ganda empat tingkat yang dikembangkan yaitu; Setiap butir soal tes dignostik merupakan soal pilihan ganda yang terdiri dari 25 butir soal. Butir soal tes diagnostik disusun berdasarkan 19 indikator soal yang melingkupi 6 sub materi yang terdiri dari reaksi eksoterm dan endoterm, beraneka ragam perubahan entalpi standar, penentuan ΔH reaksi menurut percobaan kalorimeter, penetapan ΔH reaksi menurut hukum hess, penentuan ΔH reaksi menurut data energi ikatan, serta perbandingan kalor pembakaran pada suatu bahan bakar. Setiap butir soal memiliki empat tingkatan pertanyaan yang saling berkaitan. Tahap pertama berupa soal yang berkaitan dengan materi termokimia dengan pilihan jawaban (4 pengecoh dan 1 jawaban). Tahap kedua berupa pilihan tingkat keyakinan untuk respon pada tahap pertama. Tahap ketiga berupa pilihan alasan dari respon pada tahap pertama (4 pengecoh dan 1 jawaban). Tahap keempat berupa pilihan tingkat keyakinan untuk respon pada tahap ketiga.

Instrumen tes diagnostik pilihan ganda empat tingkat yang dikembangkan

memiliki 25 item soal yang valid dan nilai reliabilitas sebesar 0,985 yang menyatakan bahwa instrumen tes tersebut reliabel. Dari hasil penelitian ini diketahui bahwa dari 60 peserta didik sejumlah 22% memahami konsep, 26% tidak memahami, dan 39% mengalami miskonsepsi. Sehingga tingkat pemahaman peserta didik pada pokok bahasan termokimia diperoleh kriteria yang ‘rendah.

DAFTAR PUSTAKA

- Ardiansah, A., Masykuri, M., & Rahardjo, S. B. (2018). Student certainty answering misconception question: Study of Three-Tier Multiple-Choice Diagnostic Test in Acid-Base and Solubility Equilibrium. *Journal of Physics: Conference Series*, 1006(1). <https://doi.org/10.1088/1742-6596/1006/1/012018>
- Caleon, I. S., & Subramaniam, R. (2010). Do students know What they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, 40(3), 313–337. <https://doi.org/10.1007/s11165-009-9122-4>
- Depdiknas. (2006). Standar Kompetensi dan Kompetensi Dasar. Jakarta: Departemen Pendidikan Nasional.
- Gurel, D. K., Eryilmaz, A., & McDermott, L. . (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(5), 989–1008.
- 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>
- Harahap, I. P. P., & Novita, D. (2020). Validitas dan Reliabilitas Instrumen Tes Diagnostik Four-Tier Multiple Choice (4TMC) pada Konsep Laju Reaksi. *Unesa Journal of Chemical Education*, 9(2), 222–227.
- Herrington, J., McKenney, S., Reeves, T., & Oliver, R. (2007). Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 4089–4097.
- Irby, S. M., Phu, A. L., Borda, E. J., Haskell, T. R., Steed, N., & Meyer, Z. (2016). Use of a card sort task to assess students' ability to coordinate three levels of representation in chemistry. *Chemistry Education Research and Practice*, 17(2), 337–352. <https://doi.org/10.1039/c5rp00150a>
- Murniningsih, Muna, K., & Irawati, R. K. (2020). Analysis of misconceptions by four tier tests in electrochemistry, case study on students of the chemistry education study program UIN Antasari Banjarmasin. *Journal of Physics: Conference Series*,

- 1440(1). <https://doi.org/10.1088/1742-6596/1440/1/012008>
- Nugraeni, D., Jamzuri, & Sarwanto. (2013). Penyusunan Tes Diagnostik Fisika Materi Listrik Dinamis. *Jurnal Pendidikan Fisika*, 1(2), 12.
- Nurhidayatullah, N., & Prodjosantoso, A. K. (2018). Miskonsepsi materi larutan penyangga. *Jurnal Inovasi Pendidikan IPA*, 4(1), 41–51. <https://doi.org/10.21831/jipi.v4i1.10029>
- Plomp. (2013). Educational Design Research Educational Design Research. *Educational Design Research*, 1–206.
- Ritonga, P. S., & Yasthophi, A.-. (2019). Pengembangan Instrumen Test Diagnostik Multiple Choice Four Tier Pada Materi Ikatan Kimia. *Konfigurasi : Jurnal Pendidikan Kimia Dan Terapan*, 3(1), 23. <https://doi.org/10.24014/konfigurasi.v3i1.6797>
- Rohmah, S. (2019). *Analisis konsepsi siswa kelas XI pada materi asam basa menggunakan instrumen tes four tier multiple choice: Penelitian deskriptif di kelas XI SMA Negeri Wilayah Bandung Timur* (pp. 1–8). UIN Sunan Gunung Djati Bandung.
- Septiana, D., Zulfiani, & Noor, M. F. (2014). Identifikasi Miskonsepsi Siswa Pada Konsep Archaeobacteria Dan Eubacteria Menggunakan Two-Tier Multiple Choice. *Edusains*, VI(2), 192–200.
- Sheftyawan, W. B., Prihandono, T., & Lesmono, A. D. (2018). Identifikasi Miskonsepsi Siswa Menggunakan Four-tier Diagnostic Test pada Materi Optik Geometri. *Jurnal Pembelajaran Fisika*, 7(2), 147–153.
- Sudijono, A. (2015). *Pengantar evaluasi pendidikan* (Ed. Ke-1,). Raja Grafindo Persada.
- Talanquer, V. (2011). Macro, submicro, and symbolic: The many faces of the chemistry “triplet.” *International Journal of Science Education*, 33(2), 179–195. <https://doi.org/10.1080/09500690903386435>
- Winata, B. C. (2019). *Analisis konsepsi siswa pada materi kesetimbangan kimia menggunakan instrumen tes Four-Tier Multiple Choice*. UIN Sunan Gunung Djati Bandung.
- Yakubi, M., & Hanum, L. (2017). Menganalisis Tingkat Pemahaman Siswa pada Materi Ikatan Kimia Menggunakan Instrumen Penilaian Four-Tier Multiple Choice (Studi Kasus pada Siswa Kelas X SMA Negeri 4 Banda Aceh) Abstrak Pendahuluan Metode Penelitian. *Jurnal Ilmiah Mahasiswa Pendidikan Kimia (JIMPK)*, 2(1), 19–26.
- Yalçınkaya, E., Taştan, Ö., & Boz, Y. (2009). High School Students’ Conceptions about Energy in Chemical Reactions. *Pamukkale University Journal of Education*, 26, 1–11.