



Alotrop

Jurnal Pendidikan dan Ilmu Kimia

p-ISSN 2252-8075 e-ISSN 2615-2819

PENGEMBANGAN INSTRUMEN PENILAIAN *THREE-TIER MULTIPLE CHOICE* UNTUK MENGUKUR KETERAMPILAN BERPIKIR KRITIS SISWA PADA MATERI TERMOKIMIA

Putri Nopita Sari¹, Rina Elvia^{1*}, Elvinawati¹

¹Program Studi Pendidikan Kimia Jurusan PMIPA FKIP
* For correspondence purposes, email:relvia@unib.ac.id

ABSTRACT

The purpose of this study is to develop a three-level multiple choice assessment instrument to measure critical thinking skills of students at SMA Negeri 5 Bengkulu City on thermochemical material. The development model used is the ADDIE model. The research subject is class XI MIPA 4 at SMA Negeri 5 Bengkulu City in the academic year 2022/2023. The research data were obtained from needs analysis, expert validation sheets, empirical test analysis, and analysis of students' critical thinking skills. The number of instruments developed is 20 three-tier multiple choice questions. The results of the study: (1) The three-level multiple choice test instrument on the thermochemical material developed was declared valid by the validator with a minimum CVR value of 0.78 and a CVI value of 0.91 with a very good category (2) The results of the empirical validity test obtained 14 valid questions (3) A three-level assessment instrument with several choices that can be used to measure students' thinking skills (4) The results of the implementation of developing critical thinking skills instruments obtained 2 students with very critical categories, 2 students with critical categories, 16 students in the moderately critical category, 13 students in the less critical category, and 1 student in the uncritical category. (4) The results of the development of the three-tier multiple choice assessment instrument can be used to measure students' critical thinking skills.

Keywords: *R and D; Three-Level Multiple Choice; Critical Thinking; Thermochemistry*

ABSTRAK

Tujuan penelitian ini yaitu mengembangkan instrumen penilaian three tier multiple choice untuk mengukur keterampilan berpikir kritis siswa di SMA Negeri 5 Kota Bengkulu pada materi termokimia. Model pengembangan yang digunakan adalah model ADDIE. Subjek penelitian adalah kelas XI MIPA 4 di SMA Negeri 5 Kota Bengkulu tahun ajaran 2022/2023. Data penelitian diperoleh dari analisis kebutuhan, lembar validasi ahli, dan analisis uji empiris, serta analisis kemampuan keterampilan berpikir kritis siswa. Jumlah instrument yang dikembangkan adalah 20 butir soal three tier multiple choice. Hasil penelitian diperoleh: (1) Instrument tes three tier multiple choice pada materi termokimia yang sudah dikembangkan dinyatakan valid oleh validator



dengan nilai CVR yang diperoleh minimal 0,78 dan nilai CVI yang diperoleh adalah 0,91 dengan kategori sangat baik (2) Hasil uji validitas empiris menunjukkan 14 soal yang valid (3) Instrumen penilaian *three tier multiple choice* yang dikembangkan dapat digunakan untuk mengukur keterampilan berpikir kritis siswa (4) Hasil implementasi pengembangan instrumen keterampilan berpikir kritis siswa diperoleh diperoleh 2 siswa dengan kategori sangat kritis, 2 siswa dengan kategori kritis, 16 siswa dengan kategori cukup kritis, 13 siswa dengan kategori kurang kritis, dan 1 siswa dengan kategori tidak kritis (4) Hasil pengembangan instrumen penilaian *three tier multiple choice* terbukti dapat digunakan untuk mengukur keterampilan berpikir kritis siswa

Kata kunci: Pengembangan; *Three-Tier Multiple Choice*; Berpikir Kritis; Termokimia

PENDAHULUAN

Kurikulum 2013 merupakan kurikulum yang diterapkan oleh pemerintah untuk menggantikan [Kurikulum Tingkat Satuan Pendidikan](#) (KTSP). Berdasarkan Permendikbud nomor 36 tahun 2018, tentang tantangan internal pengembangan kurikulum 2013 menghendaki suatu proses pendidikan yang memberikan kesempatan bagi siswa agar dapat mengembangkan segala potensi yang dimilikinya, salah satunya memiliki kemampuan berpikir kritis.

Berpikir kritis adalah berpikir secara mendalam untuk mengambil suatu keputusan, pemecahan masalah dan menarik kesimpulan yang tepat. Berpikir kritis merupakan berpikir reflektif dan masuk akal yang berfokus untuk menentukan apa yang harus dipercayai dan apa yang harus dilakukan [1]. Berdasarkan pengertian dari berpikir kritis, maka diperlukan suatu instrument yang dapat mengukur ketrampilan ataupun kemampuan berpikir kritis siswa. Pengukuran terkait kemampuan berpikir kritis ini masih jarang dilakukan, karena kurangnya ketersediaan soal tes khusus untuk mengukur keterampilan berpikir tingkat tinggi [2].

Berdasarkan wawancara yang dilakukan kepada guru mata pelajaran kimia di SMA Negeri 5 Kota Bengkulu, guru belum mempunyai instrumen khusus yang digunakan untuk mengukur keterampilan berpikir kritis. Guru melakukan evaluasi pembelajaran masih menggunakan soal-soal yang dominan pada ranah kognitif C1-C3 dibandingkan C4. Adapun kendala yang dihadapi guru dalam menyusun instrumen tes adalah membutuhkan waktu yang lama dalam penyusunannya, kurangnya pemahaman tentang berpikir kritis, dan kurangnya semangat siswa dalam menjawab soal yang diberikan sehingga hanya sedikit yang tuntas.

Tes pilihan ganda lebih disukai pada kelas sains karena lebih mudah diterapkan dan sifatnya yang objektif [3]. Namun, guru masih kesulitan dalam menentukan apakah jawaban siswa benar menggambarkan tingkat kemampuannya atau hanya menebak saja [4]. Solusi yang dapat digunakan yaitu menggunakan tes pilihan ganda tiga tingkat.

Instrumen yang dapat digunakan untuk mengukur keterampilan berpikir kritis adalah instrument soal pada tingkat kognitif C4 (menganalisis) dan C5



(mengevaluasi), sedangkan instrument soal yang diperlukan untuk mengukur kemampuan berpikir kreatif pada tingkat C6 (mencipta) [5]. Adanya tes pilihan ganda dua tingkat juga dapat digunakan sebagai instrumen pengukuran berpikir kritis dengan peluang siswa menebak jawaban benar yaitu sebesar 4% [6].

Tes pilihan ganda tiga tingkat (*three tier multiple choice*) merupakan instrumen penilaian yang telah dikembangkan oleh peneliti sebelumnya dengan menambahkan *tier* ketiga pada *two-tier test* sebagai tingkat keyakinan [7]. Tingkat pertama merupakan soal pilihan ganda dengan empat pengecoh dan satu kunci jawaban. Tingkat kedua merupakan alasan siswa menjawab pertanyaan berupa uraian singkat. Tingkat ketiga merupakan tingkat keyakinan siswa dalam menjawab butir soal pada tingkatan pertama dan kedua.

Penelitian ini bertujuan untuk mengembangkan instrumen *penilaian three-tier multiple choice* yang dapat digunakan untuk mengukur keterampilan berpikir kritis siswa pada materi termokimia.

METODE PENELITIAN

1. Model Pengembangan

Penelitian ini menggunakan metode penelitian dan pengembangan (*Research and Development*) dengan mengadaptasi Langkah penelitian model ADDIE. Terdapat lima tahapan dalam model tersebut, yaitu: (a) *Analysis* (penganalisisan), (b) *Design* (perencanaan), (c) *Development* (pengembangan), (d) *Implementation* (pengimplementasian), dan (e) *Evaluation* (pengevaluasian).

2. Subjek Uji Coba

Pada tahap uji coba produk, pengambilan subjek penelitian berdasarkan teknik *purposive sampling*. Teknik *purposive sampling* merupakan teknik penentuan sampel dengan menggunakan pertimbangan tertentu (Sugiyono, 2019). Berdasarkan pertimbangan tersebut maka dipilih siswa kelas XI MIPA 4 SMA Negeri 5 Kota Bengkulu yang berjumlah 34 siswa.

3. Teknik Pengumpulan Data

Wawancara merupakan cara untuk mengumpulkan data melalui tanya jawab secara langsung antara peneliti dengan subjek penelitian. Wawancara dilakukan pada guru kimia dan siswa. Wawancara pada guru kimia dilakukan di awal (sebelum penelitian) sebagai studi pendahuluan untuk memperoleh informasi tentang bagaimana cara guru dalam menilai keterampilan berpikir kritis siswa pada materi termokimia. Dalam melakukan wawancara maka perlu disiapkan lembar wawancara agar lebih fokus dalam melaksanakannya.

4. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah teknik analisis data deskriptif kuantitatif. Data yang diperoleh kemudian diolah dengan menggunakan bantuan program komputer Anates Versi 4 untuk mengetahui Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda dan Efektivitas Pengecoh.

a. Validitas Ahli

Validitas mencerminkan sejauh mana soal dapat dengan tepat mengukur apa yang

seharusnya diukur. Penentuan Validitas atau daya ketepatan mengukur dapat ditentukan berdasarkan nilai CVR yang diperoleh. Adapun ketentuan nilai CVR minimum dapat dilihat pada Tabel 1.

Tabel 1 Nilai CVR Minimum

Jumlah SME	Nilai Minimum
5	0,99
6	0,99
7	0,99
8	0,75
9	0,78

b. Validitas Empiris

Validitas empiris yaitu menguji validitas instrumen yang dilakukan setelah siswa mengerjakan tes yang sedang diuji[8]. Pengolahan nilai validitas empiris dapat menggunakan aplikasi ANATES. Berikut tafsiran korelasi untuk validitas empiris berdasarkan peneliti sebelumnya [9]:

Tabel 2. Tafsiran Koefisien Korelasi

Koefisien Korelasi	Kriteria
0,80– 1,00	Sangat tinggi
0,60 – 0,79	Tinggi
0,40 – 0,59	Cukup
0,20 – 0,39	Rendah
0,00 – 0,19	Sangat rendah

c. Analisis Uji Reliabilitas

Suatu alat ukur dikatakan dapat dipercaya (*reliable*) apabila alat ukur tersebut diujikan kepada objek atau subjek yang sama secara berulang-ulang akan memberikan hasil yang konsisten, stabil atau relatif sama (tidak berbeda secara statistik). Uji reliabilitas dapat dihitung menggunakan rumus Alpha Cronbach

untuk menemukan nilai r_{hitung} (r_{11}), dengan rumus sebagai berikut:

$$r_{11} = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_t^2} \right)$$

Keterangan:

r_{11} = koefisien reliabilitas

k = jumlah butir soal

s_i^2 = varian skor butir soal ke-i

s_t^2 = varian skor total

d. Uji Tingkat kesukaran

Tingkat kesukaran digunakan untuk mengetahui apakah soal tes termasuk soal yang mudah, sedang atau sulit untuk dikerjakan. Butir soal tes dapat dinyatakan baik apabila butir-butir soal tersebut tidak terlalu sukar dan tidak terlalu mudah, dengan kata lain tingkat kesukaran butir soal adalah sedang atau cukup [10].

Adapun rumus untuk menentukan Tingkat kesukaran soal sebagai berikut:

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

Keterangan

P : Tingkat kesukaran

B : Banyaknya siswa yang menjawab soal itu dengan benar

J_s : Banyaknya peserta didik yang mengikuti tes

Selanjutnya nilai P yang diperoleh, dikelompokkan dengan kategori sebagai berikut

Tabel 3.6 Tingkat kesukaran

Nilai P	Kategori
0,00 – 0,30	Sukar
0,31 – 0,70	Cukup
0,71 – 1,00	Mudah

e. Uji Daya Pembeda

Daya pembeda soal adalah kemampuan suatu soal yang dapat membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah [9]. Daya pembeda dapat diketahui dengan menggunakan rumus sebagai berikut:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Keterangan

- D = Daya pembeda
 J_A = Banyak siswa kelompok atas
 J_B = Banyak siswa kelompok bawah
 B_A = Banyak siswa kelompok atas yang benar
 B_B = Banyak peserta kelompok bawah yang menjawab benar
 P = Indek kesukaran

Selanjutnya untuk kriteria daya pembeda soal dapat dilihat pada Tabel 3.

Tabel 3. kriteria Daya Pembeda

Nilai P	Kategori
0,70 – 1,00	Baik sekali
0,40 – 0,69	Baik
0,20 – 0,39	Cukup
0,00 – 0,19	Kurang baik
Negative	Jelek sekali

f. Analisis Pengecoh

Analisis pengecoh mempunyai tujuan untuk mengetahui bekerja atau tidaknya pilihan-pilihan jawaban yang ada pada butir soal, bisa dilihat dengan menganalisa pola sebaran jawaban siswa. Angka yang menunjukkan besaran pengecoh adalah indeks pengecoh. Indeks pengecoh bisa

dihitung menggunakan rumus sebagai berikut:

$$IP = \frac{p}{(N - B)/(n - 1)} \times 100\%$$

Keterangan:

- IP : Indeks pengecoh
 P : Jumlah siswa yang memilih pengecoh
 N : Jumlah siswa yang mengikuti ujian
 B : Jumlah siswa yang menjawab benar pada soal
 n : Jumlah alternatif jawaban
 1 : Bilangan tetap

Hasil perhitungan setiap pengecoh diinterpretasikan menggunakan kriteria menurut peneliti sebelumnya sebagai berikut [11].

Tabel 4. kriteria perhitungan pengecoh

Kategori	Kriteria
Sangat Baik	76% - 125%
Baik	51% - 75% atau 126% - 150%
Kurang Baik	26% - 50% atau 151% - 175%
Jelek	0% - 50% atau 176% - 200%
Sangat Jelek	Dari 200%

Tabel 5. kriteria penilaian pengecoh

Pengecoh yang berfungsi	Kategori pengecoh	Keterangan
4	Sangat baik	Digunakan
3	Baik	
2	Cukup baik	Revisi
1	Kurang baik	Tidak
0	Tidak baik	Digunakan

Analisis pengecoh pada setiap butir soal kemudian disimpulkan menggunakan kriteria yang diadaptasi dari skala likert seperti pada tabel 5 [12].

g. Pengukuran Tes Keterampilan Berpikir Kritis

Pengolahan data dari jawaban siswa pada tes keterampilan berpikir kritis siswa yang telah diperoleh dari penelitian ini melalui tahapan-tahapan sebagai berikut [13]: a) Melakukan koreksi sebagai skor awal dari setiap jawaban siswa, b) Menghitung skor maksimal yang bisa didapatkan untuk setiap butir jawaban, c) Menghitung nilai tes keterampilan berpikir kritis siswa dengan memakai rumus persentase serta menentukan kriteria keterampilan berpikir kritis sesuai dengan tabel 7.

Tabel 7. Kriteria Skor Rata-rata

Skor	Kategori
$80 < P \leq 100$	Sangat Kritis
$75 < P \leq 79$	Kritis
$41 < P \leq 74$	Cukup Kritis
$21 < P \leq 40$	Kurang Kritis
$0 < P \leq 20$	Tidak Kritis

HASIL DAN PEMBAHASAN

Penelitian ini telah dilaksanakan di SMA Negeri 5 Kota Bengkulu. Produk yang dikembangkan merupakan instrument penilaian kemampuan berpikir kritis yang diterapkan pada siswa kelas XI IPA 4 SMA Negeri 5 Kota Bengkulu.

1. Tahap Analisis

Tahap analisis dilakukan untuk menganalisis kebutuhan guru terhadap instrumen penilaian keterampilan berpikir

kritis dan menganalisis indikator keterampilan berpikir kritis berdasarkan kompetensi dasar materi termokimia.

Tabel 8. Skor keterampilan berpikir kritis
Three Tier Multiple Choice Test

Tier I	Tier II	Tier III	Skor
Benar	2	Yakin	7
Benar	1	Yakin	6,5
Benar	-	Yakin	6
Salah	2	Yakin	5
Salah	1	Yakin	4,5
Salah	-	Yakin	4
Benar	2	Tidak Yakin	3
Benar	1	Tidak Yakin	2,5
Benar	-	Tidak Yakin	2
Salah	2	Tidak Yakin	1
Salah	1	Tidak Yakin	0,5
Salah	-	Tidak Yakin	0

Analisis kebutuhan bertujuan untuk mengetahui permasalahan yang ada di SMA untuk diteliti. Masalah yang ada di SMA Negeri 5 Kota Bengkulu soal evaluasi yang diberikan belum mengacu kepada instrument penilaian keterampilan berpikir kritis pada materi termokimia.

Analisis indikator keterampilan berpikir kritis bertujuan untuk merumuskan indikator berpikir kritis yang sesuai digunakan pada soal yang akan dikembangkan. Pada penelitian ini peneliti menggunakan sub indikator berpikir kritis [1]. Pada penelitian ini peneliti menggunakan sub indikator berpikir kritis sesuai dengan peneliti [1] yang terdiri dari 12 indikator dan 48 sub indikator, namun peneliti menerapkan 4 indikator dengan 5 sub indikator

Analisis kompetensi dasar termokimia bertujuan untuk menganalisis dan

mensintesis kompetensi yang cocok untuk dimasukkan kedalam instrument yang akan dikembangkan. Pada kurikulum 2013 kompetensi dasar materi termokimia yang diterapkan di sekolah.

2. Tahap Design

Pada tahap perancangan, peneliti membuat rencana desain pengembangan produk berdasarkan hasil dari tahap analisis. Produk yang dikembangkan berupa instrumen penilaian *three-tier multiple choice* untuk mengukur keterampilan berpikir kritis pada materi termokimia. Tingkat pertama merupakan soal pilihan ganda dengan empat pengecoh dan satu kunci jawaban. Tingkat kedua merupakan alasan siswa menjawab pertanyaan berupa uraian singkat. Tingkat kedua ini bertujuan untuk mendorong pemikiran dan penalaran keterampilan berpikir kritis [14]. Tingkat ketiga merupakan tingkat keyakinan siswa dalam menjawab butir soal pada tingkatan pertama dan kedua.

1.1 Butir soal terkait termokimia (Instrumen Tingkat I)

- Pilihan jawaban

1.2 Uraian Jawaban Singkat (Instrumen Tingkat II)

...

1.3 Seberapa yakin anda atas jawaban dan alasan yang anda berikan? (Instrumen Tingkat III)

(1)	(2)	(3)	(4)	(5)	(6)
-----	-----	-----	-----	-----	-----

Gambar 4.1 Rancangan Instrumen Yang Dikembangkan

3. Tahap Development

a. Hasil Uji Validasi Ahli

Uji validasi ini melibatkan 9 validator, yang terdiri dari 1 dosen pendidikan kimia FKIP UNIB dan 8 guru pada mata pelajaran kimia yang tersebar di beberapa SMA Negeri kota Bengkulu. Butir soal dikatakan valid jika nilai $CVR > 0,78$. Berdasarkan hasil validasi ahli terdapat 2 soal yang tidak valid yaitu soal nomor PG-3 dan nomor PG-8. Nilai CVI yang diperoleh pada soal tersebut ialah 0,91 dengan kategori sangat valid sehingga seluruh butir yang valid dari nilai CVR layak digunakan untuk penelitian lebih lanjut.

b. Hasil Uji Validitas Empiris

Hasil validasi empiris kelas XI SMA Negeri 5 Kota Bengkulu sebanyak 68 siswa. Tes dapat dikatakan memiliki validitas yang tinggi apabila tes tersebut menjalankan fungsi ukurnya atau memberikan hasil ukur yang tepat dan akurat sesuai dengan tujuan pengukuran.

Salah satu penyebab tidak validnya soal karena faktor dari dalam tes itu sendiri, sehingga perlu dilakukan perbaikan teknik penyusunan butir soal seperti penggunaan bahasa soal yang terlalu bertele-tele dan bentuk soal yang rumit atau tidak mudah dipahami. Hasil uji validasi empiris dapat dilihat pada tabel 9 dan 10 berikut.

Hasil uji validitas empiris terdapat 4 soal yang tidak valid. Keempat soal tersebut tidak valid dilihat dari hasil signifikansi karena pola data yang tidak terdistribusi secara merata, artinya terdapat pola jawaban dari responden yang tidak konsisten. Hal inilah yang menyebabkan

nilai koefisien korelasi rendah. Soal valid jika koefisien korelasi skor butir dengan skor total lebih besar dari koefisien korelasi dari r-tabel, koefisien korelasi butir signifikan dan butir tersebut valid secara empiris [15].

Tabel 9 Hasil Uji Validitas Empiris pada Tier I

Kriteria	Nomor Butir Soal	Kategori
r-hasil < r-tabel (0,5760)	PG-1; PG-2; PG-3; PG-4; PG-5; PG-6; PG-7; PG-10; PG-13; PG-14; PG-15; PG-16; PG-17; PG-18; PG-19; PG-20	Valid
r-hasil > r-tabel (0,5760)	PG-8; PG-9	Tidak Valid

Tabel 10 Hasil Uji Validitas Empiris pada Tier II

Kriteria	Nomor Butir Soal	Kategori
r-hasil < r-tabel (0,5760)	PG-1; PG-2; PG-3; PG-4; PG-5; PG-6; PG-7; PG-10; PG-13; PG-14; PG-15; PG-16; PG-17; PG-19	Valid
r-hasil > r-tabel (0,5760)	PG-8; PG-9; PG-18; PG-20	Tidak Valid

Salah satu penyebab tidak validnya soal bisa terjadi karena faktor dari dalam tes itu sendiri, sehingga perlu dilakukan perbaikan dalam teknik penyusunan butir-butir soal seperti penggunaan bahasa soal yang terlalu bertele-tele dan bentuk soal yang rumit atau tidak mudah dipahami. Faktor lain yang perlu diperhatikan yaitu kesesuaian antara materi dengan butir soal yang dibuat.

c. Hasil Uji Reliabilitas

Reliabilitas menunjukkan sejauh mana hasil pengukuran dengan alat tersebut dapat dipercaya [16].

Tabel 11 Hasil Uji Reliabilitas

Kelas	Nilai	
	Tier I	Tier II
XI IPA 3	0,90	0,88
XI IPA 5	0,82	0,87

Uji reliabilitas butir soal dilakukan untuk mengetahui konsistensi instrumen tes *three tier multiple choice* yang dikembangkan dalam mengukur berpikir kritis siswa pada materi termokimia [17]. Koefisien reliabilitas rata-rata yang diperoleh untuk kedua tier jawaban dan alasan sebesar 0,86 dengan kategori sangat tinggi, maka instrumen tes *three tier multiple choice* yang dikembangkan memiliki konsistensi yang sangat baik dalam mengidentifikasi keterampilan berpikir kritis.

d. Hasil Uji Tingkat Kesukaran Soal

Butir soal yang baik perlu dilakukan analisa tingkat kesukaran soal. Hasil kesukaran soal disajikan dalam tabel 12 dan 13 berikut.

Tabel 12 Hasil Tingkat Kesukaran Soal Tier I

Kriteria	Nomor Butir Soal	Kategori
>0,7	PG-15; PG-16	Mudah
0,3 < P ≤ 0,7	PG-1; PG-2; PG-3; PG-5; PG-6; PG-7; PG-8; PG-9; PG-10; PG-13; PG-14; PG-17; PG-18; PG-19	Sedang
≤ 0,3	PG-4; PG-20	Sukar

Tabel 13. Hasil Tingkat Kesukaran Soal Tier II

Kriteria	Nomor Butir Soal	Kategori
$>0,7$	PG-10	Mudah
$0,3 < P \leq 0,7$	PG-1; PG-2; PG-3; PG-4; PG-5; PG-6; PG-7; PG-8; PG-9; PG-13; PG-14; PG-15; PG-16; PG-18; PG-19;	Sedang
$\leq 0,3$	PG-17; PG-20	Sukar

Tingkat kesukaran soal digunakan untuk mengetahui derajat kesukaran butir soal yang diberikan apakah termasuk dalam kategori mudah, sedang, atau sukar. Soal yang mencapai tingkat kesukaran tepat adalah soal yang tidak terlalu mudah dan tidak terlalu sukar dalam kata lain disebut tingkat kesukaran sedang [18].

Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sulit, karena soal yang terlalu mudah tidak membuat siswa untuk menambah usaha dalam berpikir kritis memecahkan soal tersebut, dan soal yang terlalu sulit akan membuat siswa putus asa dan tidak mempunyai motivasi dalam menyelesaikan soal karena merasa di luar jangkauan kemampuannya [19].

Tabel 14. Hasil Uji Daya Pembeda Tier I

Kriteria	Nomor Butir Soal	Kategori
$0,70 - 1,00$	PG-2; PG-3; PG-4; PG-5; PG-6; PG-7; PG-8; PG-10; PG-13; PG-14; PG-17; PG-18; PG-19	Sangat baik
$0,40 - 0,69$	PG-1; PG-9; PG-15	Baik
$0,20 - 0,39$	PG-16; PG-20	Cukup
$0,01 - 0,19$		Jelek

a. Uji Daya Pembeda

Berdasarkan penelitian diperoleh hasil uji daya beda butir soal yang disajikan dalam tabel 14 dan 15.

Tabel 15. Daya Pembeda Tier II

Kriteria	Nomor Butir Soal	Kategori
$0,70 - 1,00$	PG-1; PG-2; PG-4; PG-15	Sangat baik
$0,40 - 0,69$	PG-3; PG-5; PG-6; PG-7; PG-9; PG-10; PG-13; PG-14; PG-16; PG-17; PG-18; PG-19	Baik
$0,20 - 0,39$	PG-8; PG-20	Cukup
$0,01 - 0,19$		Jelek
$<0,00$		Sangat jelek

Uji daya beda digunakan untuk mengetahui sejauh mana butir soal mampu membedakan tingkat kemampuan siswa. Daya pembeda merujuk pada kemampuan soal untuk membedakan siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah [18]. Berdasarkan hasil uji daya pembeda butir soal dengan menggunakan aplikasi Anates Pilihan Ganda dan Uraian, ditunjukkan bahwa sebanyak 18 butir soal dalam instrumen tes *three tier multiple choice* yang telah dikembangkan dapat digunakan untuk membedakan tingkat kemampuan siswa dalam keterampilan berpikir kritis pada materi termokimia.

b. Analisis Pengecoh

Berdasarkan penelitian diperoleh hasil uji pengecoh butir soal yang disajikan dalam tabel 4.10 berikut.

Tabel 16. Hasil Uji Pengecoh

Butir soal	Kategori Pengecoh	Keterangan
PG- 4; PG-9; PG-10; PG-13;	Sangat baik	Digunakan
PG-5; PG-14; PG-16; PG-18	Baik	
PG-1; PG-2; PG-3; 6; PG-7; PG-8; 17; 20	Cukup	Revisi
PG-15; PG-19	Kurang	Tidak digunakan
-	Tidak baik	

Soal dengan kategori pengecoh cukup baik harus direvisi terlebih dahulu sebelum digunakan, sedangkan soal dengan kategori pengecoh tidak baik harus diganti karena terlalu mencolok sehingga tidak dipilih oleh siswa [11]. Maknanya, Butir soal yang baik, pengecohnya akan dipilih secara merata oleh siswa yang menjawab salah. Sebaliknya, butir soal yang kurang, pengecohnya akan dipilih secara tidak merata.

Berdasarkan validasi empiris yang telah dilakukan pada siswa kelas XI MIPA 5 dan XI MIPA 3, soal yang dinyatakan valid adalah soal yang memenuhi setiap uji pada tahap empiris seperti uji validitas soal, reliabilitas, uji daya pembeda, tingkat kesukaran dan pengecoh. Jika semua soal sudah disusun dengan baik, maka perlu diujicobakan terlebih dahulu dilapangan [11]. Tujuannya untuk melihat soal-soal mana yang perlu diubah, diperbaiki, bahkan dibuang sama sekali, serta soal-soal mana yang baik untuk dipergunakan selanjutnya. Soal yang baik adalah soal yang sudah mengalami beberapa kali uji-coba dan revisi, yang didasarkan atas analisis empiris.

4. Tahap Implementasi

Pada tahap implementasi, soal yang digunakan adalah soal yang valid dan reliabel dari hasil uji empiris. Implementasi dilakukan di kelas XI MIPA 4 SMA Negeri 5 Kota Bengkulu yang berjumlah 34 siswa. Butir soal yang dipilih untuk diimplementasikan telah memenuhi setiap indikator dari materi termokimia dan sub indikator keterampilan berpikir kritis, dengan masing-masing jumlah butir soal. Proporsi tingkat kesukaran soal tersebar secara normal yaitu soal sukar 20%, soal sedang 60%, soal mudah 20% [11].

Hasil yang didapatkan dari implementasi soal tersebut kemudian dihitung untuk menentukan tingkat keterampilan berpikir kritis siswa yang dapat dilihat pada tabel 17 berikut.

Tabel 17. Hasil Keterampilan Berpikir Kritis Siswa

Skor	Kategori	Frekuensi	Persentase
$80 < P \leq 100$	Sangat Kritis	2	5,9%
$75 < P \leq 79$	Kritis	2	5,9%
$41 < P \leq 74$	Cukup Kritis	16	47%
$21 < P \leq 40$	Kurang Kritis	13	38,2%
$0 < P \leq 20$	Tidak Kritis	1	2,9%

Berdasarkan hasil implementasi 34 siswa di kelas XI MIPA 4, diperoleh bahwasannya siswa paling banyak berada pada kategori cukup kritis yakni 16 siswa (47%) dan kategori sangat kritis ada 2 siswa (5,9%), kategori kritis 2 siswa (5,9%), kategori kurang kritis 13 siswa



(38,2%), dan kategori tidak kritis 1 siswa (2,9%). Maknanya, soal yang diimplementasikan dapat digunakan dalam mengukur keterampilan berpikir kritis siswa pada kelas XI MIPA 4 di SMA Negeri 5 Kota Bengkulu.

Berdasarkan sebaran dari data yang didapatkan, kemudian dihitung normalitas dari implementasi soal tersebut. Hasil uji normalitas yang telah dilakukan berdasarkan data hasil keterampilan berpikir kritis siswa didapatkan bahwa data terdistribusi normal dengan nilai signifikansi $> 0,05$ yaitu 0,073. Makna dari data terdistribusi normal adalah soal yang dikembangkan memiliki sebaran pola yang normal/terarah sehingga dapat disimpulkan bahwa soal yang diimplementasikan dapat digunakan untuk mengukur keterampilan berpikir kritis siswa.

5. Tahap *Evaluation*

Tahap evaluasi bertujuan untuk melihat sejauh mana produk yang dihasilkan dapat mencapai tujuan yang telah ditetapkan. Evaluasi dibagi menjadi dua macam yaitu evaluasi formatif dan sumatif. Evaluasi formatif dilakukan pada setiap tahap pengembangan untuk perbaikan produk, dan evaluasi sumatif dilakukan pada akhir program untuk menentukan hasil dari tujuan yang ingin dicapai [20].

Pada tahap *analysis* telah dilakukan analisis kebutuhan. Pada tahap analisis instrumen, dalam mengumpulkan data selain menggunakan wawancara sebaiknya juga dilakukan penyebaran angket

kebutuhan guna mendapatkan hasil yang lebih lengkap.

Pada tahap perancangan, beberapa soal tes yang sudah dirancang perlu diganti dan diperbaiki susunan kalimatnya agar mudah dipahami.

Pada tahap pengembangan, evaluasi dilakukan pada tahap uji empiris, dimana petunjuk pengerjaan soal sebaiknya dijelaskan sedetail mungkin sampai semua siswa paham sehingga pada saat pengerjaan soal lebih kondusif.

Pada tahap implementasi, waktu pelaksanaan tes sebaiknya dilakukan sesudah pembelajaran pada materi tersebut dilakukan atau diberitahukan terlebih dahulu sebelumnya agar siswa dapat melakukan persiapan sehingga hasil yang didapatkan maksimal.

KESIMPULAN

Hasil pengembangan instrumen penilaian *three-tier multiple choice* yang dikembangkan terbukti dapat digunakan untuk mengukur keterampilan berpikir kritis siswa. Berdasarkan hasil penelitian telah dilakukan dapat diambil kesimpulan sebagai berikut:

1. Instrument tes *three tier multiple choice* pada materi termokimia yang sudah dikembangkan dinyatakan valid oleh validator dengan nilai CVR yang diperoleh minimal 0,78 dan nilai CVI yang diperoleh adalah 0,91 dengan kategori sangat baik. Butir soal yang dinyatakan valid oleh validator sebanyak 18 soal dan langsung diujicobakan pada validitas empiris.
2. Berdasarkan validitas empiris yang telah dilakukan kepada 34 siswa kelas

XI MIPA 5 dan 34 siswa kelas XI MIPA 3 SMA Negeri 5 Kota Bengkulu didapatkan hasil validitas empiris valid, hasil uji reliabilitas kategori sangat tinggi, hasil uji tingkat kesukaran dominan sedang, hasil uji daya pembeda 88,89%, hasil pengecoh 16 pengecoh dapat digunakan.

3. Hasil pengukuran keterampilan berpikir kritis dari 34 siswa di kelas XI MIPA 4 SMA Negeri 5 Kota Bengkulu diperoleh 2 siswa dengan kategori sangat kritis, 2 siswa dengan kategori kritis, 16 siswa dengan kategori cukup kritis, 13 siswa dengan kategori kurang kritis, dan 1 siswa dengan kategori tidak kritis.

SARAN

1. Untuk peneliti selanjutnya, hal yang perlu diperhatikan yaitu pemilihan sekolah tempat implementasi soal yang telah dirancang supaya dalam prosesnya, hasil yang didapatkan sesuai dengan harapan.
2. Untuk mengantisipasi jika soal penilaian yang sudah dirancang banyak tidak valid, alangkah lebih baik dalam pemilihan materi berupa dominan konseptual supaya pada tingkat ke dua soal yang diranjang menjadi pilihan ganda pula dan lebih sudah dalam menghitung penskorannya.

3.

DAFTAR PUSTAKA

- [1] Ennis, R. H & Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test. Test manual, criteria, scoring sheet an instrument for teaching and testing*. USA: Midwest Publication.
- [2] Hikmah, & Amin, N. (2019). Pengembangan Instrumen Untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Dalam Mata Pelajaran Matematika di SMA Kabupaten Majene. *SAINTIFIK*, 5(1), 1-7.
- [3] Scully, D. (2017). Constructing Multiple-Choice Items to Measure Higher-Order Thinking. *Practical Assessment, Research & Evaluation*, 22(4), 1-13.
- [4] Siswaningsih, W., Anisa, N., Komalasari, N. E., & Indah, R. (2014). Pengembangan Tes Diagnostik Two Tier Untuk Mengidentifikasi Miskonsepsi Pada Materi Kimia Siswa SMA. *Jurnal Pengajaran MIPA*, 19(1), 117-127.
- [5] Saraswati, P. M. S., & Agustika, G. N. S. (2020). Kemampuan berpikir tingkat tinggi dalam menyelesaikan soal HOTS mata pelajaran matematika. *Jurnal Ilmiah Sekolah Dasar*, 4(2), 257-269.
- [6] Ramadhan, G., Dwijananti, P., & Wahyuni, S. (2018). Analisis Kemampuan Berpikir Tingkat Tinggi (*High Order Thinking Skills*) Menggunakan Instrumen Tes Two Tier Multiple Choice Materi Konsep dan Fenomena Kuantum Siswa SMA di Kabupaten Cilacap. *UPEJ: Unnes Physics Educational Journal*, 7(3), 85-90.
- [7] Hasan, S., Bagayoko, D., & Kellay, E., L. (1999). Misconception And The Certainty A Response To Huffman And



- Heller. *The Physics Teacher*. 3(3), 502[16] 506.
- [8] Arikunto, S. 2010. *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta. ISBN: 9789795189985 [17]
- [9] Arikunto, S. 2018. *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta. ISBN: 978-979-5189-98-5.
- [10] Sudijono, A. 2013. *Pengantar Evaluasi pendidikan*. PT. Raja Grafindo Persada. Jakarta. ISBN: 979-421-495-7.
- [11] Arifin, Z. (2012). *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*. Bandung: PT Remaja Rosdakarya. ISBN: 979-6929-56-2.
- [12] Sugiyono 2019. *Statistika untuk Penelitian*. Bandung : CV Alfabeta. ISBN: 9789798433108
- [13] Basri, H. Purwanto, As'ari, R, dan Sisworo. (2019). Investigating critical thinking skill of junior high school in solving mathematical problem. *International Journal of Instruction*, 12(3), 745–758. [19]
- [14] Mubarak, S., Susilaningsih, E., & Cahyono, E. (2016). Pengembangan Tes Diagnostik Three Tier Multiple Choice Untuk Mengidentifikasi Miskonsepsi Peserta Didik Kelas XI. *Journal of Innovative Science Education (JISE)*, 5(2), 101-110. [20]
- [15] Sappaile, B. I. (2007). Konsep instrumen penelitian pendidikan. *Jurnal Pendidikan dan kebudayaan*, 13(66), 1-7.
- Suryabrata, S. 2004. *Metodologi Penelitian*. Yogyakarta: Pustaka Pelajar. ISBN: 979-421-104-4
- Laeli, C. M., Gunarhadi, & Muzzazinah. (2021). Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning. *Ilkogretim Online – Elementary Education Online*, 20(2), 277-287
- Putri, R. H., & Ofianto. (2019). Efektivitas Analisis Butir Menggunakan Anajohn, Anates dan Iteman Studi Soal USBN Pelajaran Sejarah Kota Padang. *Jurnal Mahasiswa Ilmu Sejarah dan Pendidikan*, 1(2), 1-11.
- Pradana, Y., Suma, K., & Sujanem, R. (2021). Pengembangan Tes Keterampilan Berpikir Tingkat Tinggi Tiga Level (Three-Tier) Berbasis Marzano Dimensions of Learning. *Wahana Matematika dan Sains : Jurnal Matematika, Sains, dan Pembelajarannya*, 15(3), 119-128.
- Tegeh, I. M., & Kirna, I. M. (2013). Pengembangan Bahan Ajar Metode Penelitian Pendidikan Dengan ADDIE Model. *Jurnal Ika*, 11(1).