



PENGEMBANGAN E-INSTRUMEN FOUR-TIER MULTIPLE CHOICE TEST BERBANTUAN GOOGLE FORM UNTUK MENGUKUR KETERAMPILAN BERPIKIR KRITIS PADA MATERI STOIKIOMETRI

Development of E-Instrument Four-Tier Multiple Choice Test Assisted by Google Form to Measuring Critical Thinking Skills on Stoichiometry Material

Risa Nur Pahlevi*, Mahdian, Muhammad Kusasi, Rilia Iriani

Program Studi Pendidikan Kimia FKIP Universitas Lambung Mangkurat,
Jl. Brigjend. H. Hasan Basry Banjarmasin 70123 Kalimantan Selatan Indonesia

*email: risapahlevin@gmail.com

Informasi Artikel

Kata kunci:
e-instrumen tes,
keterampilan berpikir
kritis, four-tier
multiple choice test,
stoikiometri, validitas,
daya pembeda soal

Keywords:
*e-instrument test,
critical thinking skills,
four-tier multiple choice
test, stoichiometry,
validity, distinguishing
power of questions*

Abstrak

Telah dilakukan penelitian mengenai pengembangan instrumen *Four-Tier Multiple Choice Test* untuk mengukur keterampilan berpikir kritis peserta didik. Penelitian ini bertujuan untuk mengetahui tingkat kevalidan, reliabilitas, daya beda, dan tingkat kesukaran dari butir soal yang dikembangkan. Penelitian ini menggunakan metode Research and Development (R&D). Model pengembangan penelitian ini mengikuti model ADDIE. Subjek penelitian terdiri dari 145 peserta didik XI MIPA SMAN 4 Banjarmasin. Data dikumpulkan dengan menggunakan instrumen tes dan angket. Instrumen tes diuji coba dan dianalisis melalui tahap uji coba kelompok kecil dan uji coba lapangan. Diperoleh sebanyak 12 butir soal dinyatakan valid, memiliki reliabilitas instrumen sebesar 0,726 (reliabel), daya pembeda memiliki nilai diskriminasi lebih dari 0,30 (sedang-baik sekali), dan tingkat kesukaran yang baik (rendah-sedang). Instrumen tes memiliki tingkat keterbacaan yang baik sekali.

Abstract. Research has been conducted on the development of a Four-Tier Multiple Choice Test instrument to measure students' critical thinking skills. This study aims to determine the level of validity, reliability, differentiation, and difficulty of the items developed. This research uses the Research and Development (R&D) method. The model development model follows the ADDIE model. The research subjects consisted of 145 students of XI MIPA SMAN 4 Banjarmasin. Data were collected using test instruments and questionnaires. The test instrument was tested and analyzed through small group trials and field trials. It was obtained that 12 items were declared valid, had instrument reliability of 0.726 (reliable), differentiating power had a discrimination value of more than 0.30 (medium-excellent), and a good level of difficulty (low-medium). The test instrument has a very good readability level.

PENDAHULUAN

Peran seorang guru dalam mengarahkan proses pembelajaran haruslah kreatif, inovatif, serta memiliki kemampuan dalam mengonsep kegiatan evaluasi pembelajaran. Evaluasi dalam proses pembelajaran diperlukan seorang guru untuk

Copyright © JCAE-Jurnal Tugas Akhir Mahasiswa, e-ISSN 2613-9782

How to cite: Pahlevi, R. N., Mahdian, Kusasi, M., Iriani, R. (2024). PENGEMBANGAN E-INSTRUMEN FOUR-TIER MULTIPLE CHOICE TEST BERBANTUAN GOOGLE FORM UNTUK MENGUKUR KETERAMPILAN BERPIKIR KRITIS PADA MATERI STOIKIOMETRI. *JCAE (Journal of Chemistry And Education)*, 8(1), 1-13.

mengetahui sejauh mana ketercapaian kompetensi yang diharapkan kepada peserta didik. Seorang guru juga memerlukan proses evaluasi pembelajaran untuk mengetahui keefektifan dan efisiensi semua komponen yang digunakan dalam proses pembelajaran berlangsung. Selain itu, peserta didik juga dapat mengetahui sejauh mana hasil belajarnya sendiri.

Seorang guru memerlukan sebuah alat ukur untuk mengetahui kuantitas suatu ketercapaian dalam proses pembelajaran. Tes adalah alat ukur yang dapat menjadi salah satu cara dalam melaksanakan kegiatan evaluasi pembelajaran. Alat ukur tersebut harus standar, yaitu memiliki derajat validitas dan reliabilitas yang tinggi (Arifin, 2012). Oleh karena itu, seorang guru harus memiliki kemampuan dalam membuat alat ukur yang sesuai dengan kaidah penilaian tersebut. Menurut (Pranata & Dwijayanti, *et al.*, 2021), guru masih banyak kesulitan dalam membuat alat ukur atau instrumen penilaian yang tepat. Dalam penelitian ini, dikembangkan instrumen yang di dalamnya melalui tahap pengembangan hingga instrumen tersebut sesuai dengan kaidah penilaian.

Instrumen yang dikembangkan digunakan untuk mengukur keterampilan berpikir kritis, karena keterampilan berpikir kritis merupakan salah satu keterampilan yang dituntut dalam pembelajaran Abad 21. Keterampilan Abad 21 tersebut diantaranya, yaitu keterampilan berpikir kritis (*critical thinking*), keterampilan berpikir kreatif (*creative thinking*), keterampilan komunikasi (*communication*) serta keterampilan kerja sama (*collaboration*). Selain itu, pembelajaran Abad 21 juga menuntut seorang guru semakin inovatif dalam menggunakan teknologi dalam proses pembelajaran yang berlangsung.

Keterampilan berpikir kritis yang dimiliki peserta didik membuat dirinya akan mampu memilih dan menganalisis kebenaran suatu informasi dan membuat keputusan dengan baik. Pada kenyataannya tingkat keterampilan berpikir kritis peserta didik masih rendah. Hal ini ditunjukkan oleh penelitian TIMSS (*Trends in International Mathematics and Science Study*) pada tahun 2015, Indonesia berada pada peringkat ke-45 dari 50 negara peserta dengan skor 397 jauh di bawah standar skor rata-rata 600 (Anggraeny *et al.*, 2019). Hal ini juga sejalan dari hasil observasi dan wawancara yang dilakukan dengan guru mata pelajaran kimia di SMAN 4 Banjarmasin, yaitu keterampilan berpikir kritis ada dalam ketercapaian kurikulum merdeka yang mereka terapkan pada profil pelajar pancasila. Di sekolah tersebut tidak mengetahui sejauh mana ketercapaian keterampilan berpikir kritis peserta didik karena belum pernah mengukur tingkat keterampilan berpikir kritis peserta didik menggunakan instrumen yang relevan.

Dalam pembelajaran kimia, memiliki keterampilan berpikir kritis sangat membantu dalam proses pembelajaran. Oleh karena itu, diperlukan seorang guru yang mampu mengarahkan peserta didik untuk dapat menggunakan keterampilan berpikir kritis. Proses evaluasi menggunakan instrumen yang sesuai dengan permasalahan tersebut merupakan salah satu solusi. Instrumen dalam pengukuran tingkat berpikir kritis dikembangkan dengan tingkat kesukaran dan parameter yang tepat, yaitu diperuntukkan untuk dapat mengukur keterampilan berpikir kritis yang berorientasi pada aspek-aspek indikator di dalamnya (Laksono, 2018).

Instrumen tes yang digunakan pada penelitian ini adalah instrumen tes empat tingkat atau *Four-Tier Multiple Choice Test* yang di dalamnya memuat indikator berpikir kritis. Instrumen *Four-Tier* disusun dalam 4 tingkatan, yaitu tingkat pertama soal pilihan ganda, tingkat kedua keyakinan jawaban, tingkat ketiga alasan jawaban pada tingkat pertama dan satu pilihan kosong yang dapat diisi sendiri, tingkat keempat berisi tingkat keyakinan siswa memilih alasan tersebut (Aprita *et al.*, 2018).

Instrumen tes tersebut merupakan jenis rancangan tes yang bertujuan untuk

mengidentifikasi kelemahan peserta didik sebagai bahan acuan untuk menentukan tindak lanjut perlakuan yang tepat pada peserta didik ke depannya (Prianti et al., 2020). Jenis instrumen tes ini juga memiliki kelebihan dibanding jenis tes lainnya, yaitu guru dapat membedakan tingkat berpikir peserta didik melalui keyakinan jawaban dan tingkat keyakinan alasan, sehingga dapat mendiagnosis lebih dalam pemahaman dan tingkat berpikir yang dialami oleh peserta didik (Saputri et al., 2021).

Berdasarkan uraian di atas, perlu dilakukan penelitian agar dikembangkannya instrumen tes untuk mengukur ketereampilan berpikir kritis dengan berkonten stoikiometri sehingga harapannya peserta didik juga dapat terlatih dalam menghadapi soal – soal yang menggunakan keterampilan berpikir kritisnya.

METODE PENELITIAN

Penelitian *Research & Development* (R&D) ini menggunakan model pengembangan ADDIE (*Analysis, Design, Development, Implementation, Evaluation*). Prosedur pengembangan model ADDIE yang sistematis dimana setiap tahapan yang akan dilalui mengacu pada tahapan sebelumnya yang telah dievaluasi menghasilkan produk yang efektif (Juannita & Adhi, 2017).

Tahap dalam pengembangan ini terdiri dari lima tahap, di antaranya *Analysis, Design, Development, Implementation, Evaluation*. Pada tahap Analisis (*Analysis*), dilakukan analisis kajian pustaka dan analisis kurikulum. Analisis kajian pustaka dilakukan dengan mencari, membaca dan menelaah sumber – sumber ilmiah yang relevan berkaitan dengan penelitian yang akan dilakukan. Analisis kurikulum dilakukan dengan menelaah terhadap kurikulum sekolah yang dijadikan tempat penelitian untuk mengetahui pokok bahasan materi yang nantinya sebagai acuan dalam membuat instrumen tes. Tahapan ini dilakukan untuk sebagai landasan dasar dalam melakukan penelitian (Ridwan et al., 2021).

Pada tahap *design* (desain), dilakukan kegiatan penyusunan kisi – kisi sebagai batasan dalam menyusun soal dan merancang butir soal instrumen tes. Pada tahap *development* (pengembangan) merupakan kegiatan pembuatan dan pengujian produk yang sebelumnya telah dirancang. Kegiatan yang dilakukan pada tahap ini meliputi penulisan soal tes ke dalam *google form*, uji validasi produk oleh ahli dan uji keterbacaan kepada peserta didik. Pada tahap *implementation* (implementasi) merupakan tahapan mengimplementasikan produk yang telah divalidasi untuk uji coba butir soal kepada peserta didik secara uji coba kelompok kecil dan uji coba lapangan. Instrumen yang dikembangkan akan dilakukan uji validitas butir soal, reliabilitas instrumen, daya pembeda dan tingkat kesukaran butir soal. Pada tahap *evaluation* (evaluasi) dilakukan untuk melihat kualitas instrumen yang dikembangkan. Evaluasi dilakukan berdasarkan dari analisis hasil uji validitas instrumen oleh ahli dan uji keterbacaan.

Populasi pada penelitian ini ialah peserta didik dari SMAN 4 Banjarmasin kelas XI MIPA. Sampel uji coba produk dilakukan uji keterbacaan terhadap 10 peserta didik kelas XI MIPA 1. Uji coba produk pada kelompok kecil melibatkan 36 peserta didik kelas XI MIPA 3. Untuk uji coba lapangan melibatkan 35 peserta didik kelas XI MIPA 2 dan 35 peserta didik kelas XI MIPA 4, sehingga total pada uji coba lapangan ini sebanyak 70 peserta didik, serta 20 orang peserta didik kelas XI MIPA 1 sebagai sampel untuk analisis keterampilan berpikir kritis. Angket respon peserta didik diberikan kepada peserta didik pada setiap kali uji coba produk dalam uji coba kelompok kecil dan uji coba lapangan

Lembar validasi digunakan dalam menentukan kelayakan suatu instrumen sebelum digunakan. Lembar validasi yang digunakan berupa lembar validasi instrumen *Four-Tier Multiple Choice Test* untuk mengukur keterampilan berpikir

kritis, lembar validasi angket uji keterbacaan produk dan angket respon peserta didik. Lembar angket uji keterbacaan produk untuk mengetahui tingkat keterbacaan instrumen tes oleh peserta didik sebelum dilakukan uji coba produk. Lembar angket respon peserta didik digunakan untuk mengetahui respon dari peserta didik terhadap instrumen tes yang diberikan.

Instrumen *Four-Tier Multiple Choice Test* yang dikembangkan untuk mengukur keterampilan berpikir kritis yang memuat materi stoikiometri, berbentuk pilihan ganda empat tingkat dengan jumlah awal soal sebanyak 20 butir soal. Instrumen tes digunakan dalam memperoleh data berupa jawaban peserta didik yang digunakan dalam menganalisis butir soal.

Instrumen dikatakan valid jika instrumen tersebut dapat mengukur dengan tepat apa yang hendak diukur. Instrumen yang telah dikembangkan akan diuji validitas isi dan validitas konstruk menggunakan pendapat ahli (*expert judgement*) (Sugiono, 2019). Instrumen tes penelitian ini akan dinilai menggunakan lembar validasi yang diisi oleh 5 ahli selaku validator dibidang pendidikan kimia. Hasil penilaian dihitung menggunakan indeks kesepakatan ahli dengan indeks *Aiken's V* dan hasil perhitungan akan dibandingkan dengan persamaan tabel kriteria pada tabel 1 (Retnawa, 2016).

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

Keterangan:

- V = indeks kesepakatan rater mengenai validitas butir
- s = skor yang ditetapkan setiap rater dikurang skor terendah dalam kategori yang dipakai ($s = r - l_0$)
- r = skor kategori pilihan rater
- l_0 = skor terendah dalam kategori penskoran (dalam hal ini = 1)
- c = banyak kategori yang dapat dipilih rater (dalam hal ini = 5)
- n = banyaknya rater

Tabel 1. Klasifikasi validitas Aiken's V

Skala <i>Aiken's V</i>	Validitas
$V < 0,4$	Kurang
$0,4 < V < 0,8$	Sedang
$V > 0,8$	Sangat Valid

Hasil dari penilaian oleh ahli tersebut kemudian dilanjutkan dengan uji coba lapangan. Uji coba di lapangan kemudian perlu dianalisis kembali mengenai validitas butir soalnya, karena pada saat diujicoba kepada peserta didik, terdapat butir soal yang tidak valid yang mengakibatkan soal tersebut tidak dapat digunakan (Widoyoko, 2020). Validitas butir soal dihitung menggunakan rumus korelasi point biserial karena data penelitian yang diperoleh bersifat dikotomi.

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Keterangan:

- r_{pbi} = koefisien korelasi *point biserial*
- M_p = Rata-rata skor total yang menjawab benar pada butir soal
- M_t = rata-rata skor total
- SD_t = standar deviasi skor total
- p = proporsi siswa yang menjawab benar
- q = proporsi siswa yang menjawab salah (1-p)

(Guilford, 1973)

Hasil dari harga koefisien korelasi tersebut kemudian dibandingkan dengan membandingkan harga r_{pbi} dengan harga r dalam tabel *product moment*. Apabila harga r_{pbi} hitung lebih besar atau sama dengan r tabel ($r_h \geq r_t$) berarti ada korelasi bersifat signifikan sehingga butir soal dianggap valid, begitupun sebaliknya (Widoyoko, 2020). Rusmayani (2020) menjelaskan bahwa butir tes yang baik haruslah mampu membedakan antara peserta didik yang menguasai materi dengan yang tidak. Besarnya daya pembeda disebut indeks diskriminasi (D) yang berkisar antara 0,00 – 1,00. Daya pembeda soal dihitung menggunakan rumus dan diklasifikasikan menurut Fatimah (2019) berikut ini:

$$D = P_a - P_b$$

Keterangan:

D = Daya Beda

P_a = Proporsi Peserta Kelompok Atas yang Menjawab Benar

P_b = Proporsi Peserta Kelompok Bawah yang Menjawab Benar

Tabel 2. Klasifikasi daya pembeda butir soal

Skala Indeks Diskriminasi Item	Klasifikasi Indeks Diskriminasi Item
$D = -$ (bertanda negatif)	Jelek Sekali
$D < 0,2$	Rendah
0,20–0,40	Sedang
0,40–0,70	Baik
0,80–1,00	Baik Sekali

Indeks kesukaran ialah angka proporsi peserta didik menjawab betul dalam soal yang menunjukkan tingkat kesukaran suatu butir. Indeks kesukaran dapat dihitung dan diklasifikasikan menurut Rusmayani (2020) berikut ini:

$$I = \frac{B}{J}$$

Keterangan:

I = Tingkat Kesukaran Butir Soal

B = Jumlah Peserta Didik yang Menjawab Benar

J = Jumlah Peserta Didik yang Mengikuti Tes

Tabel 3. Klasifikasi tingkat kesukaran butir soal

Skala Indeks Kesukaran Butir Soal	Klasifikasi Indeks Kesukaran Soal
0,00 – 0,30	Sukar
0,31 – 0,71	Sedang
0,71 – 1,00	Rendah

Magdalena et al. (2021) menjelaskan bahwa instrumen pengukuran dikatakan reliabel jika pengukurannya konsisten, cermat, dan akurat. Reliabilitas instrumen penelitian ini dihitung menggunakan metode belah dua (*split half method*) dengan formulasi Spearman – Brown karena data penelitian yang diperoleh bersifat dikotomi. Rumus yang digunakan sebagai berikut:

$$r_i = \frac{2r_b}{1 + r_b}$$

$$V = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma XY^2 - (N\Sigma X)^2\}\{N\Sigma XY^2 - (N\Sigma Y)^2\}}}$$

Keterangan:

r_i = koefisien reabilitas skor instrumen

r_b = koefisien korelasi antara dua belahan 1

N = banyaknya responden
 X = belahan pertama
 Y = belahan kedua

Koefisien reliabilitas yang diperoleh kemudian dikonsultasikan dengan tabel r product moment pada taraf signifikan 5%. Apabila r hitung lebih besar atau sama dengan dengan r tabel ($r_h \geq r_t$) berarti ada korelasi bersifat signifikan sehingga instrumen tes dianggap reliabel, begitupun sebaliknya (Widoyoko, 2020). Hasil koefisien reliabilitas kemudian diklasifikasikan menurut tabel 4 yang telah disusun (Ndiung & Jediut, 2020) berikut ini:

Tabel 4. Klasifikasi reliabilitas instrumen

Skala Perhitungan Validitas	Validitas Butir Soal
$r_{ii} \leq 0,20$	Sangat rendah
$0,20 < r_{ii} \leq 0,40$	Rendah
$0,40 < r_{ii} \leq 0,60$	Sedang
$0,60 < r_{ii} \leq 0,80$	Tinggi
$0,80 < r_{ii} \leq 1,00$	Sangat Tinggi

Respon jawaban atau alasan yang benar diberi skor '1' dan respon jawaban atau alasan yang salah diberi skor '0'. Adapun persamaan yang digunakan untuk penskoran peserta didik dapat dilihat sebagai berikut.

$$S = R$$

Keterangan:

S = Skor yang diperoleh
 R = Jawaban atau alasan yang benar

HASIL PENELITIAN DAN PEMBAHASAN

Tahap *Analysis* (Analisis)

Berdasarkan analisis kurikulum, kurikulum yang diterapkan di SMA Negeri 4 Banjarmasin khususnya kelas XI adalah kurikulum Merdeka, sehingga instrumen tes yang dikembangkan mengacu pada Keputusan Kepala Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 033/H/KR/2023 dan Nomor 008/H/KR/2022 tentang Capaian Pembelajaran Pada Pendidikan Anak Usia Jenjang Pendidikan Dasar, Dan Jenjang Pendidikan Menengah Pada Kurikulum Merdeka.

Berdasarkan hasil kajian pustaka yang didapat, karakteristik dari keterampilan berpikir kritis dan kaitannya dengan materi stoikiometri dapat dilihat pada tabel 5 berikut.

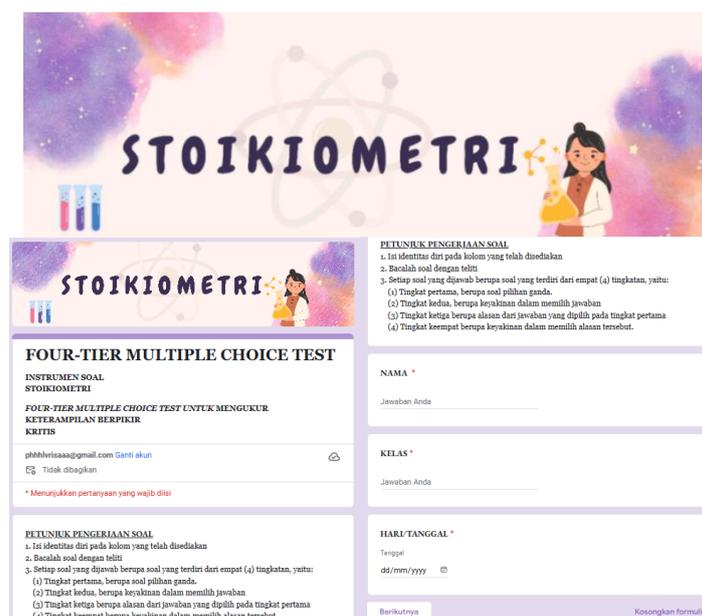
Tabel 5. Karakteristik keterampilan berpikir kritis dengan materi stoikiometri

No	Karakteristik	Definisi Operasional	Keterkaitan materi stoikiometri
1	Jelas	Permasalahan yang disajikan dapat diidentifikasi dengan secara jelas dan fokus	Materi yang bersifat riil dan perlu menghubungkan antara pemahaman konsep yang jelas dan fokus
2	Memiliki solusi tertentu	Permasalahan ditemukan dengan cara yang tepat yang dapat dipakai untuk menyelesaikan masalah tersebut	Memiliki urutan materi dari teori hingga rumus yang dapat digunakan sebagai informasi untuk menentukan solusi dari permasalahan

No	Karakteristik	Definisi Operasional	Keterkaitan materi stoikiometri
3	Keterkaitan antara masalah dan solusi	Mengenal hubungan yang logis antara masalah, asumsi, dan nilai (<i>inference</i>)	Mempelajari hubungan kuantitatif pereaksi dan hasil reaksi
4	Informasi terdapat dalam permasalahan	Permasalahan memuat informasi yang dapat dikumpulkan, disusun, dan digunakan untuk memecahkan permasalahan	Membutuhkan kemampuan berpikir yang sangat baik dalam mencari sebuah informasi untuk menyelesaikan soal
5	Mengevaluasi	Permasalahan dapat dievaluasi dengan cara menilai fakta-fakta yang ada	pemahaman konsep dan nalar logika yang tinggi dalam penyelesaian soal-soalnya

Tahap *Design* (Desain)

Analisis kurikulum yang didapat pada tahap sebelumnya menjadi acuan untuk menyusun kisi-kisi awal pada tahap desain. Kisi-kisi tersebut digunakan untuk menjadi acuan dalam membuat soal. Soal yang disusun memuat aspek indikator berpikir kritis dan memuat materi stoikiometri. Pada tahap ini juga menghasilkan desain *header google form* dan desain tampilan yang dapat dilihat pada Gambar 1 berikut.



Gambar 1 Desain *Header* dan Tampilan

Tahap *Development* (Pengembangan)

Instrumen tes yang dibuat kemudian dimasukkan ke dalam *google form*. Pada tahap ini dilakukan kegiatan validasi oleh lima orang ahli. Adapun hasil dari penilaian lima ahli dapat dilihat pada tabel 6 berikut.

Tabel 6. Rekapitulasi penilaian instrumen tes oleh ahli

Nomor Butir Soal	Validator					Indeks Aiken's V	Kriteria
	I	II	III	IV	V		
1	5	5	5	4	5	0,95	Sangat Valid
2	5	5	5	4	5	0,95	Sangat Valid
3	5	5	5	4	5	1	Sangat Valid
4	5	4	5	4	5	0,95	Sangat Valid
5	5	5	5	4	5	0,95	Sangat Valid
6	5	4	5	4	5	0,9	Sangat Valid
7	5	5	5	4	5	1	Sangat Valid
8	5	5	5	4	5	0,9	Sangat Valid
9	5	5	5	4	5	0,9	Sangat Valid
10	5	5	5	4	5	1	Sangat Valid
11	5	3	5	4	5	1	Sangat Valid
12	5	5	5	4	5	1	Sangat Valid
13	5	3	5	4	5	0,85	Sangat Valid
14	5	3	5	4	5	0,95	Sangat Valid
15	5	5	5	4	5	0,85	Sangat Valid
16	5	5	5	4	5	0,85	Sangat Valid

Keterangan pada tabel 6: Validator I (Dr. H. Rusmasyah, M.Pd.), Validator II (Dr. Arif Sholahuddin, S.Pd, M.Sc), Validator III (Drs. Rilia Iriani, M.Si), Validator IV (Yogo Dwi Prasetyo, M.Pd., M.Sc.), Validator V (Dra. Hj. Noor Jennah)

Hasil uji validasi oleh ahli menunjukkan bahwa instrumen tes dinyatakan sangat valid. Instrumen yang valid tersebut selanjutnya dilakukan uji keterbacaan instrumen tes. Uji keterbacaan dilakukan kepada peserta didik kelas XI MIPA 1 SMAN 4 Banjarmasin sebanyak 10 orang. Hasil menunjukkan bahwa instrumen tes memiliki rata-rata skor sebesar 97, sehingga instrumen dapat digunakan untuk tahap uji coba selanjutnya.

Tahap *Implementation* (Implementasi)

Tahap implementasi dilakukan kegiatan pengimplementasian produk yang sudah divalidasi oleh ahli dan uji keterbacaan. Tahap ini terbagi menjadi dua kegiatan uji coba, yaitu uji coba kelompok kecil dan uji coba lapangan.

Pertama dilakukan tahap uji coba kelompok kecil. Uji coba kelompok kecil dilakukan untuk menganalisis validitas butir, reliabilitas instrumen tes, tingkat kesukaran butir, dan daya pembeda butir. Dengan jumlah responden (N) sebanyak 36 peserta didik maka diketahui r_{tabel} sebesar 0,339 pada taraf signifikansi 5%. Berikut rekapitulasi hasil analisis validitas butir soal pada uji coba kelompok kecil.

Tabel 7. Hasil analisis validitas uji coba kelompok kecil

No soal	Total	Klasifikasi
1,2,4,5,7,9,10,11,13,14,15,16	12	Valid
3,6,8,12	4	Tidak Valid

Hasil analisis validitas tersebut didapatkan bahwa sebanyak 4 butir soal tidak valid, sehingga 4 butir soal tersebut akan dibuang dan tidak diikutkan dalam uji reliabilitas. Empat butir soal tersebut tidak dapat mengukur apa yang seharusnya diukur sesuai dengan variabel yang akan digunakan (Hamimi et al., 2020). Empat butir soal yang tidak valid tersebut sejalan dengan hasil analisis daya pembeda dari butir soal. Berikut rekapitulasi hasil analisis daya pembeda butir soal pada uji coba kelompok kecil.

Tabel 8. Hasil analisis daya pembeda butir soal uji coba kelompok kecil

No soal	Total	Klasifikasi
4,5,10	3	Baik Sekali-Baik
1,2, 7,9, 11,13,14,15,16	9	Sedang
3,6,8,12	4	Rendah

Butir soal yang memiliki daya pembeda dan memiliki klasifikasi tidak valid memiliki korelasi yang searah, sesuai dengan penelitian Nurhalimah et al. (2022), bahwa daya pembeda suatu butir soal memiliki korelasi searah dengan nilai koefisien validitas. Dimana nilai indeks diskriminan 4 butir soal tersebut di bawah 0,30 ($D < 0,30$) yang artinya soal tersebut tidak mampu dengan baik membedakan kemampuan peserta didik yang menguasai materi dengan peserta didik yang tidak menguasai materi yang diujikan (Retnawati, 2016). Rendahnya nilai indeks diskriminan dapat terjadi akibat tidak berfungsinya pengecoh, soal terlalu sulit sehingga peserta didik banyak menebak dan sebagian siswa yang menguasai materi berpikir ada yang salah informasi dalam butir soal tersebut (Depdiknas, 2008).

Kualitas butir soal juga diukur dengan mengetahui tingkat kesukaran butir soal. Adapun hasil dari analisis tingkat kesukaran butir soal dari uji coba kelompok kecil dapat dilihat pada tabel 9 berikut.

Tabel 9. Hasil analisis tingkat kesukaran butir soal uji coba kelompok kecil

No soal	Total	Klasifikasi
4,5,10	3	Baik Sekali-Baik
1,2, 7,9, 11,13,14,15,16	9	Sedang
3,6,8,12	4	Rendah

Tingkat kesukaran dari soal terjadi karena banyaknya peserta didik yang tidak mampu menjawab soal dan sebaliknya. Tindak lanjut dari hasil klasifikasi tingkat butir soal yang didapat adalah menelusuri penyebab, memperbaiki dan menguji ulang butir soal tersebut. Menurut Essen (2018), tidak ada hubungan yang linear dari butir soal dengan indeks kesukaran yang bagus dalam membedakan peserta didik yang pandai dan tidak. Sehingga butir soal yang terbukti valid dan memiliki daya pembeda dengan indeks $D > 0,30$ akan diikutsertakan dalam analisis reliabilitas. Instrumen tes yang memiliki jumlah soal sebanyak 12 butir dianalisis reliabilitasnya. Analisis reliabilitas pada tahap uji coba kelompok kecil menunjukkan bahwa instrumen tes dianggap reliabel karena memiliki nilai koefisien reliabilitas (r_{hitung}) lebih besar dari r_{tabel} ($r_{hi} > r_t$), yaitu sebesar 0,726. Butir soal sebanyak 12 tersebut kemudian dilanjutkan pada tahap uji coba selanjutnya, yaitu uji coba lapangan.

Uji coba lapangan dilakukan untuk mengidentifikasi kualitas instrumen tes lebih dalam dengan menggunakan lebih banyak sampel. Pada uji coba ini, kemungkinan membantu dalam menemukan permasalahan atau kekurangan yang belum diketahui pada tahap sebelumnya. Uji coba lapangan bertujuan untuk menganalisis validitas butir, reliabilitas instrumen tes, tingkat kesukaran butir dan daya pembeda butir dari instrumen tes yang telah mengalami revisi dan perbaikan sebelumnya. Dengan jumlah responden (N) sebanyak 70 peserta didik maka diketahui r_{tabel} sebesar 0,239 pada taraf signifikansi 5%. Berikut rekapitulasi hasil analisis validitas butir soal pada uji coba lapangan.

Tabel 10. Rekapitulasi validitas butir soal uji coba lapangan

No soal	Total	Klasifikasi
1,2,3,4,5,6,7,8,9,10,11,12	12	Valid

Dari hasil rekapitulasi analisis validitas butir soal di atas didapat bahwa semua soal dinyatakan valid. Hasil analisis validitas pada butir soal tersebut menunjukkan bahwa nilai koefisien korelasi (r_{pbis}) lebih dari r_{tabel} hingga bernilai positif. Semua soal yang valid tersebut sejalan dengan hasil analisis daya pembeda butir soal bernilai baik. Adapun rekapitulasi dari hasil analisis daya pembeda uji coba lapangan dilihat pada tabel 11 berikut.

Tabel 11. Rekapitulasi analisis daya pembeda butir soal uji coba lapangan

No soal	Total	Klasifikasi
2,3,5,7,8,10,12	7	Baik
1,4,6,9,11	5	Sedang

Berdasarkan hasil pada tabel 11 di atas, didapatkan bahwa semua butir soal memiliki nilai diskriminan lebih dari 0,30. Oleh karena itu, semua soal dinyatakan baik untuk digunakan. Kualitas butir soal juga diukur dengan mengetahui tingkat kesukaran butir soal. Berikut rekapitulasi hasil analisis tingkat kesukaran butir soal pada uji coba lapangan.

Tabel 12. Hasil analisis tingkat kesukaran butir soal uji coba kelompok kecil

No soal	Total	Klasifikasi
2,3,5,7,8,9,10,11,12	9	Sedang
1,4,6	3	Rendah

Pendapat butir soal yang dianggap baik adalah butir soal yang diklasifikasikan sedang, yaitu indeks kesukarannya 0,30 – 0,70. Pada hasil analisis Tingkat kesukaran tahap ini, semua soal memiliki tingkat kesukaran yang baik. Tetapi perlu diketahui bahwa soal-soal yang terlalu mudah atau terlalu sukar tidak berarti tidak boleh digunakan karena tergantung kegunaannya (Widiyanto, 2018). Tujuan dari penggunaan tes dikaitkan dengan hasil tingkat kesukaran butir soal (Depdiknas, 2008).

Tingkat kesukaran yang tinggi terjadi karena banyaknya peserta didik yang tidak mampu menjawab butir soal. Menurut Sudijono (1996) tindak lanjut yang dapat dilakukan terhadap butir soal yang sukar dengan menelusuri penyebab, memperbaiki dan menguji ulang butir soal tersebut. Berdasarkan hasil analisis yang telah dilakukan, semua soal dapat dilanjutkan untuk dianalisis tingkat reliabilitasnya. Hasil menunjukkan bahwa instrumen tes dianggap reliabel karena nilai koefisien reliabilitasnya lebih besar dari r_{tabel} . Nilai reliabilitasnya sebesar 0,848 sedangkan nilai r_{tabel} sebesar 0,239. Klasifikasi nilai reliabilitasnya adalah sangat tinggi.

Tahap *Evaluation* (Evaluasi)

Berdasarkan hasil analisis yang telah dilakukan dan direvisi pada uji coba lapangan, diperoleh produk akhir berjumlah 12 butir soal. Adapun rekapitulasi dari hasil analisis kualitas butir soal yang telah dilakukan dapat dilihat pada tabel 13 berikut.

Tabel 13. Rekapitulasi hasil analisis kualitas butir soal

Nomor	Validitas	Daya Pembeda Butir Soal	Kesukaran Butir	Reliabilitas Instrumen
1	Valid	Sedang	Rendah	
2	Valid	Baik	Sedang	
3	Valid	Baik	Sedang	
4	Valid	Sedang	Rendah	
5	Valid	Baik	Sedang	

Nomor	Validitas	Daya Pembeda Butir Soal	Kesukaran Butir	Reliabilitas Instrumen
6	Valid	Sedang	Rendah	Sangat Tinggi
7	Valid	Baik	Sedang	
8	Valid	Baik	Sedang	
9	Valid	Sedang	Sedang	
10	Valid	Baik	Sedang	
11	Valid	Sedang	Sedang	
12	Valid	Baik	Sedang	

Sebaran soal berdasarkan indikator keterampilan berpikir kritis dan materi stoikiometri dari produk akhir butir soal pada instrumen tes yang dikembangkan dapat dilihat pada tabel berikut.

Tabel 14. Pemetaan materi pada instrumen tes

No	Aspek Indikator Berpikir Kritis	Nomor Soal	Jumlah Soal
1	Indikator <i>Focus</i>	1,2	2
2	Indikator <i>Reason</i>	3,4	2
3	Indikator <i>Inference</i>	5,6,10	2
4	Indikator <i>Situation</i>	7,8	1
5	Indikator <i>Clarity</i>	9	2
6	Indikator <i>Overview</i>	11,12	3

Hasil evaluasi pada tahap implementasi ini, analisis butir soal menggunakan pendekatan *classical test theory* (CTT), dimana hasil sangat dipengaruhi oleh pada sampel yang dipergunakan. Dimana kelemahan ini disebut dengan group dependent (Retnawati, 2016). Hasil analisis secara klasik memiliki keterbatasan, yaitu tingkat kesukaran sangat sulit untuk diestimasi secara tepat karena estimasi tingkat kesukaran dibiaskan oleh sampel. Oleh karena itu, kelemahan penggunaan pendekatan CTT hendaknya dapat diatasi dengan analisis item *response theory* (IRT), karena IRT dapat mengestimasi tingkat kesukaran soal tanpa menentukan siapa peserta tesnya (invariance) (Depdiknas, 2008).

SIMPULAN

Hasil pengembangan instrumen *Four-Tier Multiple Choiche Test* untuk mengukur keterampilan berpikir kritis berbantuan google form dengan materi stoikiometri terdiri dari 12 butir soal instrumen yang dapat digunakan. Soal – soal yang telah dikembangkan tersebut mampu mengukur keterampilan berpikir kritis didasarkan pada hasil uji validitas secara empiris yang menunjukkan 12 butir soal tersebut dianggap valid dan hasilnya dapat diandalkan yang dapat dilihat dari hasil uji reliabilitas instrumen dengan klasifikasi sangat tinggi. Analisis kualitas butir lainnya menunjukkan soal yang dikembangkan memiliki daya pembeda butir soal dan tingkat kesukaran butir soal yang dianggap baik.

DAFTAR RUJUKAN

- Anggraeny, T. K., Rohana, & Jayanti. (2019). Pengaruh Pendekatan Metaphorical Thinking Terhadap Kemampuan Berpikir Kritis Dan Kecerdasan Emosional Siswa SMAN 4 Kayuagung. *Jurnal Pendidikan Matematika Rafa*, 5(1), 57–69. <https://doi.org/10.19109/jpmrafa.v5i1.3001>

- Aprita, D. F., Supriadi, B., & Prihandono, T. (2018). Identifikasi Pemahaman Konsep Fluida Dinamis Menggunakan Four Tier Test Pada Siswa Sma. *Jurnal Pembelajaran Fisika*, 7(3), 315–321. <https://doi.org/10.19184/jpf.v7i3.8607>
- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Direktorat Jenderal Pendidikan Islam Kementerian Ri. www.Diktis.Kemenag.Go.Id
- Depdiknas. (2008). *Panduan analisis butir soal*. Jakarta: Departemen Pendidikan Nasional.
- Essen, C., & Akpan, G. (2018). Analysis of difficulty and point-biserial correlation indices of 2014 Akwa Ibom State Mock Multiple Choices Mathematics Test. *International Journal of Education and Evaluation*, 4(5), 1-11.
- Fatimah, L. U., & Kharuddin, A. (2019). Analisis Kesukaran Soal, Daya Pembeda Dan Fungsi Distraktor. *Jurnal Komunikasi Dan Pendidikan Islam*, 8(2), 37–64. <https://doi.org/10.36668/jal.v8i2.115>
- Guilford, J. P. (1973). *Fundamental statistics in psychology and education (5nd ed.)*. New York: McGraw-Hill.
- Hamimi, L., Zamharirah, R., & Rusydy, R. (2020). Analisis butir soal ujian Matematika kelas VII semester ganjil tahun pelajaran 2017/2018. *Mathema: Jurnal Pendidikan Matematika*, 2(1), 57-66. <https://doi.org/10.33365/jm.v2i1.459>
- Hidayah, F. N., Bakti, I., Iriani, R., & Hamid, A. ANALISIS MISKONSEPSI PESERTA DIDIK MENGGUNAKAN INSTRUMEN TES DIAGNOSTIK FOUR TIER MULTIPLE CHOICE PADA MATERI HIDROLISIS GARAM. *Quantum: Jurnal Inovasi Pendidikan Sains*, 14(2), 279-287. <http://dx.doi.org/10.20527/quantum.v14i2.15097>
- Juannita, J., & Adhi, B. P. (2017). Pengembangan media pembelajaran sistem pencernaan manusia untuk Kelas 8 SMP dengan fitur augmented reality 76 berbasis android (Studi Kasus: SMPN 7 Depok). *PINTER: Jurnal Pendidikan Teknik Informatika dan Komputer*, 1(1), 76-81.
- Laksono, P. J. (2019). PENGEMBANGAN DAN PENGGUNAAN INSTRUMEN TWO-TIER MULTIPLE CHOICE PADA MATERI TERMOKIMIA UNTUK MENGUKUR KEMAMPUAN BERPIKIR KRITIS. *Orbital: Jurnal Pendidikan Kimia*, 2(2), 80-92. <https://doi.org/10.19109/ojpk.v2i2.2646>
- Magdalena, I., Fauziah, S. N., Faziyah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas Iii Sdn Karet 1 Sepatan. *Bintang : Jurnal Pendidikan Dan Sains*, 3(2), 198–214.
- Ndiung, S., & Jediut, M. (2020). Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi. *Premiere Educandum: Jurnal Pendidikan Dasar Dan Pembelajaran*, 10(1), 94 – 111. <https://doi.org/10.25273/pe.v10i1.6274>
- Pitriani, N. R. V., Wahyuni, I. G. A. D., & Gunawan, I. K. P. (2021). Penerapan Model Addie Dalam Pengembangan Media Pembelajaran Interaktif Menggunakan Lectora Inspire Pada Program Studi Pendidikan Agama Hindu. *Cetta: Jurnal Ilmu Pendidikan*, 4(3), 515-532. <https://doi.org/10.37329/cetta.v4i3.1417>
- Pranata, A. H., & Dwijayanti, I. (2021). Analisis Kesulitan Penyusunan Instrumen Penilaian Bagi Guru Sekolah Dasar Di Gugus Dewi Sartika Kecamatan Reban. *Jurnal Pendidikan Sekolah Dasar*, 3(2), 61-74.
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian (panduan peneliti, mahasiswa dan psikometrian)*. Yogyakarta: Parama Publishing.

- Ridwan, M., Suhar, A. M., Ulum, B., & Muhammad, F. (2021). Pentingnya penerapan literature review pada penelitian ilmiah. *Jurnal Masohi*, 2(1), 42-51.
- Rushapiana, R., Mahdian, M., & Rusmansyah, R. (2018). Penerapan model auditoy intellectually repetition (AIR) dalam pembelajaran kelarutan dan hasil kali kelarutan untuk meningkatkan kemampuan berpikir kritis dan hasil belajar. *JCAE (Journal of Chemistry and Education)*, 1(3), 218-224.
- Rusmansyah, R., Huda, N., Safitri, L., Heldaniah, H., Isnawati, I., & Mahdian, M. (2022). Melatihkan Keterampilan Berpikir Kritis Dan Berpikir Kreatif Peserta Didik dengan Model Scientific Critical Creative Thinking. *Orbital: Jurnal Pendidikan Kimia*, 6(2), 124-135. <https://doi.org/https://doi.org/10.19109/ojpk.v6i2.14930>
- Rusmayani, R. (2020). Analisis butir soal penilaian akhir semester genap mata pelajaran pendidikan agama islam di SMP Bintang Persada Tabanan – Bali. *Widya Balina*, 5(1), 41-49.
- Saputri, L., Maison, & Kurniawan, W. (2021). Pengembangan Four-Tier Diagnostic Test Berbasis Website Untuk Mengidentifikasi Miskonsepsi Pada Materi Suhu Dan Kalor. *Jurnal Ilmiah Teknologi Informasi Asia*, 15(1), 61– 68.
- Sholeha, D., Kusasi, M., & Bakti, I. (2024). PENGEMBANGAN INSTRUMEN SOAL HIGHER-ORDER THINKING SKILLS (HOTS) BERBASIS GOOGLE FORM UNTUK MENGUKUR HASIL BELAJAR PESERTA DIDIK PADA MATERI HIDROLISIS GARAM. *JCAE (Journal of Chemistry And Education)*, 7(1), 32-48. <https://doi.org/10.20527/jcae.v7i1.2548>
- Sugiono. (2019). *Metode Penelitian Pendidikan (Kuantitatif, Kualitatif, Kombinasi, R & D Dan Penelitian Pendidikan)*. Bandung: Alfabeta.
- Widoyoko, E. P. (2020). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Belajar.