

PENGEMBANGAN DAN PENGGUNAAN INSTRUMEN *TWO-TIER MULTIPLE CHOICE* PADA MATERI TERMOKIMIA UNTUK MENGUKUR KEMAMPUAN BERPIKIR KRITIS

Pandu Jati Laksono

Universitas Islam Negeri Raden Fatah Palembang

E-mail: pandujati_uin@radenfatah.ac.id

Abstrak: Penelitian ini bertujuan untuk mengembangkan instrumen tes *Two-Tier Multiple Choice* pada materi termokimia menjadi instrumen penilaian yang mampu mengukur keterampilan berpikir kritis. Instrumen yang dikembangkan perlu diujikan kelayakannya ditinjau dari validitas logis para ahli dan instrumen penilaian *Two-Tier Multiple Choice* yang dikembangkan ditinjau dari validitas empiris yaitu reliabilitas, analisis kualitas butir soal, dan analisis pengecoh yang dapat dikategorikan soal yang baik. Penelitian ini dikategorikan sebagai penelitian pengembangan. Instrumen tes yang dikembangkan adalah *Two-Tier Multiple Choice* pada materi termokimia untuk mengukur kemampuan berpikir kritis. Model pengembangan yang digunakan yaitu model pengembangan 4 D (*four D model*) oleh S. Thigarajan, Dorothy S. Semmel, dan Melvyn I. Semmel. Model pengembangan 4D terdiri atas 4 tahap utama yaitu: (1) *Define* (Pendefinisian), (2) *Design* (Perancangan), (3) *Develop* (Pengembangan), dan (4) *Disseminate* (Penyebaran). Berdasarkan hasil penelitian yang telah dilakukan kesimpulan yang didapatkan adalah Instrumen penilaian *Two-tier Multiple Choice* yang dikembangkan dinyatakan layak dan memenuhi kriteria sebagai suatu soal yang baik dengan hasil validitas aiken rata-rata 0,84. Instrumen penilaian *Two-tier Multiple Choice* yang dikembangkan dinyatakan layak ditinjau dari realibilitas tes rata-rata 0,737 yang tergolong tinggi, memiliki daya beda dengan persentase 41,67% sangat baik, 40,00% baik, 18,33% cukup baik dan 0% jelek, berdasarkan tingkat kesukaran 8,33% mudah, 80,00% sedang dan 11,67% sukar, berdasarkan indeks pengecoh disimpulkan pengecoh tidak ada yang bekerja lebih baik daripada kunci jawaban dan nilai pengecoh sebagian besar lebih dari 5% dipilih sehingga dinyatakan efektif. Kata kunci: Penelitian Pengembangan, *Two-Tier Multiple Choice*, Termokimia, Kemampuan Berpikir Kritis

PENDAHULUAN

Pendidikan memiliki peranan yang sangat penting dengan kemajuan suatu bangsa. Pendidikan di Indonesia menjadi perhatian serius bagi pemerintah, hal ini dapat dilihat dari pengalokasian dana bagi pendidikan yang sangat besar dibandingkan bidang lain. Pendidikan yang baik memberikan peningkatan pada kualitas sumber daya manusia yang ada di Indonesia. Peningkatan kualitas sumber daya manusia menjadi hal penting yang tengah digerakkan di era globalisasi ini. Pendidikan yang baik adalah pendidikan yang menjadikan orang yang menerima ilmu pendidikan menjadi lebih baik daripada sebelumnya.

Ilmu kimia sebagai bagian dari IPA memiliki karakteristik sebagai produk berupa fakta, konsep, prinsip hukum, dan teori, dan sebagai proses yang meliputi kerja ilmiah. Kebanyakan konsep dalam ilmu kimia bersifat abstrak sehingga sulit dipahami oleh siswa. Salah satu materi dalam kimia yang dianggap sulit adalah termokimia, karena selain menuntut pemahaman konsep juga mengharuskan siswa mengoperasikan perhitungan matematis. Sebagian besar siswa telah memiliki pengetahuan awal tentang konsep yang akan dipelajari, selanjutnya akan dibangun pemahaman sesuai konsep yang akan dipelajari.

Pendekatan saintifik menekankan penumbuhan dan pengembangan keterampilan peserta didik sehingga mampu memproses informasi untuk mendapatkan fakta, konsep, pengembangan konsep, dan nilai. Penerapan

kurikulum 2013 yang berbasis pada pendekatan saintifik sangatlah sesuai dengan ilmu kimia. Dalam penyelesaian pendekatan saintifik diperlukan proses berpikir, berpikir adalah kegiatan mental dalam memecahkan masalah (Gagne, 1980). Sejalan dengan pemikiran tersebut, dalam jurnalnya Qing menjelaskan bahwa dalam proses belajar diperlukan kemampuan berpikir sebagai tujuan utama dalam pendidikan (Qing, Jing, Yazhuan, Ting, & Junping, 2010).

Kurikulum 2013 yang dicanangkan oleh pemerintah Indonesia mengacu pada pemikiran tingkat tinggi atau lebih kita kenal dengan *Higher Order Thinking Skill (HOTS)*. Dalam *HOTS* diperlukan kemampuan yang berpikir diantara berpikir tingkat tinggi, seseorang dapat dikatakan dapat berpikir kritis apabila dapat menguji pengalaman, menilai pengetahuan dan gagasan dan menimbang argumen-argumen sebelum sampai pada penilaian (Fisher, 1992: 65-66). Hal ini sesuai dengan penekanan yang ada dalam kurikulum 2013.

Krulik dan Rudnick (1993) mendefinisikan berpikir kritis adalah berpikir yang menguji, menghubungkan, dan mengevaluasi semua aspek dari situasi masalah. Termasuk di dalam berpikir kritis adalah mengelompokkan, mengorganisasikan, mengingat dan menganalisis informasi. Berpikir kritis memuat kemampuan membaca dengan pemahaman dan mengidentifikasi materi yang diperlukan dengan yang tidak ada hubungan. Hal ini juga berarti dapat menggambarkan kesimpulan dengan sempurna dari data yang diberikan, dapat menentukan ketidakkonsistenan dan kontradiksi di dalam sekelompok data. Berpikir kritis adalah analitis dan refleksif.

Proses berpikir kritis seseorang dapat memperoleh informasi dengan benar, mengevaluasinya dan memproses informasi tersebut sehingga diperoleh suatu kesimpulan yang terpercaya. Swart dan Perkin (Hassoubah, 2004) menyatakan bahwa berpikir kritis berarti mencari dan menghimpun informasi yang dapat dipercaya untuk dipakai sebagai bukti yang dapat mendukung suatu penilaian. Dengan demikian berpikir kritis sebagian besar terdiri dari mengevaluasi argumen atau informasi dan membuat keputusan yang dapat membantu mengembangkan kepercayaan dan mengambil tindakan serta membuktikan

Penggunaan instrumen yang dipakai oleh guru pada umumnya adalah soal essay dan soal pilihan ganda. Soal essay memiliki kelemahan pada subjektivitas sedangkan soal pilihan ganda biasa memiliki kelemahan pada sulitnya mengukur kemampuan berpikir siswa dan pemilihan jawaban yang berpotensi asal menjawab atau *gambling*. Melihat kelemahan dari kedua instrumen tersebut maka dari itu perlu adanya pengembangan instrumen tes untuk mengukur keterampilan berpikir kritis. Pengembangan instrumen tes masih jarang ditemui di Indonesia, meskipun di tingkat universitas sudah banyak penelitian pengembangan tapi kebanyakan pengembangannya adalah mengenai perangkat pembelajaran

Karakteristik berpikir kritis memerlukan latihan yang salah satunya dengan membiasakan mengerjakan soal-soal yang mengembangkan keterampilan berpikir kritis (Kartimi & Liliyasi, 2012). Pada analisis awal diketahui bahwa kendala penerapan pembelajaran yang menekankan pada kemampuan berpikir kritis sudah dilakukan tapi dirasa kurang optimal. Kurang optimalnya penerapan kemampuan berpikir kritis belum dapat diketahui dengan pasti dikarenakan belum adanya alat ukur untuk mengukur kemampuan berpikir kritis siswa. Keterampilan

berpikir kritis dapat diukur dengan menggunakan instrumen yang khusus diperuntukkan untuk mengukur keterampilan berpikir kritis dengan berorientasi pada aspek-aspek yang ada di dalamnya.

Instrumen penilaian *Two-Tier Multiple Choice* terdiri dari dua tingkat soal pilihan ganda pada setiap nomornya. *Two-Tier Multiple Choice (TTMC)* adalah bentuk pertanyaan yang lebih canggih dari pertanyaan pilihan ganda (Adodo, 2013). Pada tingkat pertama di *Two-Tier Multiple Choice* ini nantinya siswa dituntut untuk bisa mengidentifikasi soal yang disajikan. Kemudian menganalisis dan menarik kesimpulan dengan memilih salah satu jawaban yang telah disajikan. Pada tingkat kedua, diberi beberapa pilihan pernyataan alasan atas jawaban dari tingkat pertama. Tingkat ini dapat mewakili analisis dan refleksi pengetahuan dari siswa.

Pengembangan instrumen tes perlu dilakukan sebagai masukan untuk meningkatkan mutu dan kualitas pendidikan nasional. Instrumen tes yang dikembangkan adalah *Two-Tier Multiple Choice* yang nantinya akan menjadi alat ukur kemampuan berpikir kritis siswa.

METODE PENELITIAN

1. Jenis Penelitian

Penelitian yang dilakukan termasuk pada jenis penelitian pengembangan. Model pengembangan instrumen ini menggunakan tahapan 4D dari Thiagarajan (1974). Nana Syaodih Sukmadinata (2009: 206) Penelitian dan pengembangan atau *Research and Development (R&D)* adalah rangkaian proses atau langkah-langkah dalam rangka mengembangkan suatu produk baru atau menyempurnakan produk yang telah ada agar dapat dipertanggungjawabkan.

2. Prosedur Pengembangan

Metode penelitian dan pengembangan (*Research and Development*) merupakan metode penelitian yang digunakan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut. (Sugiyono, 2009) Pengembangan Penggunaan Instrumen tes *Two-Tier Multiple Choice* Pada Materi Termokimia untuk Mengukur Kemampuan Berpikir Kritis ini didasarkan pada langkah pelaksanaan penelitian dan pengembangan 4-D Thiagarajan. Hal ini meliputi 4 tahap yaitu tahap pendefinisian (*define*), perancangan (*design*), pengembangan (*develop*) dan diseminasi (*disseminate*)

3. Uji coba Produk

Desain uji coba produk pada pelaksanaan penelitian dan pengembangan ini meliputi 2 tahap, diantaranya:

- a. Pada tahap 1, validasi dilakukan oleh validator ahli. Validasi ini dikenal sebagai validitas logis mengandung kata logis berasal dari kata logika, yang berarti penalaran. Validitas logis sebuah instrumen evaluasi menunjuk pada kondisi bagi sebuah instrumen yang memenuhi persyaratan valid berdasarkan hasil penalaran. (Arikunto, 2011:65) evaluasi penilaian, validator ahli materi, dan guru mata pelajaran kimia untuk menilai hasil pengembangan instrumen tes awal (draft 1). Hasil validasi pengembangan instrumen tes draft 1 digunakan untuk revisi pertama (draft 2). Draft produk pengembangan yang berupa instrumen penilaian *Two-Tier Multiple Choice* dibuat kemudian dilakukan validasi oleh para ahli. Hal ini dilakukan untuk mengetahui validitas isi dari

instrumen penilaian dengan kata lain untuk mengetahui apakah instrumen tersebut mempunyai kualitas yang baik. Hasil dari validasi isi oleh para ahli dihitung menggunakan formula Aiken (Aiken, 1985),

- b. Tahap 2, Instrumen tes *Two-Tier Multiple Choice* pada draft 2 digunakan atau diujikan dalam pembelajaran di kelas guna mengetahui validitas, reliabilitas, analisis kualitas butir soal, dan analisis pengecoh yang kemudian dilihat instrumen yang baik dan buruk yang selanjutnya dilakukan perbaikan produk.

4. Jenis data

Jenis data dalam penelitian dan pengembangan ini berupa: (1) Data test instrumen yang dihasilkan oleh siswa pada uji coba instrumen penilaian *Two-Tier Multiple Choice*. (2) Data hasil wawancara dalam bentuk wawancara tak terstruktur untuk mengetahui kondisi awal dan analisis kebutuhan. (3) Data penilaian angket siswa.

5. Instrumen pengumpulan data

Data yang diperoleh dengan menggunakan instrumen antara lain:

- a. Pedoman wawancara terhadap guru dan siswa untuk mengetahui keadaan awal dan untuk analisis kebutuhan.
- b. Instrumen validasi oleh ahli berupa angket validasi instrumen penilaian *Two-Tier Multiple Choice*.
- c. Angket tanggapan instrumen penilaian *Two-Tier Multiple Choice* yang sudah dikembangkan.
- d. Pedoman wawancara terbuka untuk mengetahui tanggapan guru dan siswa terhadap instrumen penilaian *Two-Tier Multiple Choice*.
- e. Instrumen penilaian *Two-Tier Multiple Choice* yang telah dikembangkan.

6. Teknik analisis data

Data hasil uji coba lapangan dideskripsikan secara kuantitatif dan kualitatif dengan melakukan analisis (menggunakan program *software ITEMAN*) untuk mengetahui: 1) Reliabilitas; 2) Daya pembeda; 3) Tingkat kesukaran soal; 4) Indeks pengecoh

HASIL DAN PEMBAHASAN

1. Hasil Penelitian

Penelitian yang dilakukan termasuk pada jenis penelitian pengembangan. Model pengembangan instrumen ini menggunakan tahapan 4D dari Thiagarajan (1974). Tahapan 4D ini adalah mendefinisikan, mendesain, mengembangkan, dan diseminasi. Pada penelitian ini akan dikembangkan instrumen tes *Two-Tier Multiple Choice* pada materi Termokimia. Pada *define* (pendefinisian) meliputi tahapan analisis awal akhir, analisis siswa, analisis tugas dan analisis materi serta merumuskan tujuan pembelajaran. Tahap *design* (perancangan) meliputi Penetapan bentuk penyajian dan perancangan awal. Pada tahap *develop* (pengembangan) penyusunan draf produk, validasi produk dan pengujian produk. Pada tahap terakhir *disseminate* (penyebaran). Berikut deskripsi dari tiap tahapan yang dilakukan:

A. *Define* (Pendefinisian)

Define (pendefinisian) adalah tahap pertama yang dilakukan peneliti dalam proses penelitian dan pengembangan model 4-D. Pada tahap ini langkah yang

dilakukan adalah sebagai berikut:

1. Analisis ujung depan (*front and analysis*)

Analisis ujung depan bertujuan untuk memunculkan dan menetapkan masalah dasar yang dihadapi dalam evaluasi/penilaian pembelajaran kimia di sekolah. Pengidentifikasian masalah-masalah ditemukan dengan cara bertanya kepada guru atau wawancara dan menyebarkan angket kepada siswa. Dari pengidentifikasian ini didapatkan beberapa masalah yang dihadapi siswa yaitu data yang diperoleh menyatakan bahwa pada pembelajaran, guru sudah menerapkan pembelajaran dua arah. Sistem pembelajaran pada kedua sekolah sudah menerapkan kurikulum 2013 dengan pendekatan saintifik.

Penilaian merupakan penafsiran angka atau skor yang didapatkan melalui pengukuran. Dalam melakukan pengukuran digunakan alat ukur yang disesuaikan dengan kriteria yang akan diukur. Alat ukur ini bisa dalam bentuk tes maupun non tes. Pada penilaian pembelajaran guru pada umumnya lebih sering menggunakan soal essay, karena soal jenis ini lebih dapat mengungkap pengetahuan yang dimiliki siswa secara lebih mendalam dibandingkan dengan soal pilihan ganda, soal jenis ini dianggap dapat mengurangi tindakan curang dari peserta didik. Dalam beberapa ulangan harian terkadang guru juga menggunakan soal jenis pilihan ganda untuk mengungkap pengetahuan siswanya, karena mempermudah proses penilaian dan mengurangi subjektifitas dalam penilaian. Selain itu, instrumen yang digunakan untuk melakukan penilaian pada dua sekolah ini lebih banyak mencakup tingkat C1, C2, dan C3 pada ranah kognitif Bloom, ini menunjukkan masih minimnya penggunaan instrumen untuk mengungkap keterampilan berpikir kritis.

Berdasarkan kelemahan dari masing-masing tipe soal tersebut perlu itu dikembangkan instrumen tes yang dapat membantu guru dalam melaksanakan penilaian dan evaluasi lebih mendalam tanpa menjadi subjektif. Salah satu alternatifnya adalah dengan pengembangan instrumen tes.

2. Analisis siswa (*learner analysis*)

Analisis siswa bertujuan untuk mengetahui apa yang dibutuhkan oleh siswa yang berhubungan dengan proses dan evaluasi pembelajarannya. Pada kurikulum 13 Berpikir tingkat tinggi atau *Higher Order Thinking Skills (HOTS)* sangat ditekankan untuk menjadi substansi yang penting dalam pembelajaran. Berpikir tingkat tinggi diantaranya terdapat berpikir kritis. Berpikir kritis adalah mode berpikir mengenai hal, substansi atau masalah apa saja dimana si pemikir meningkatkan kualitas pemikirannya dengan menangani secara terampil struktur-struktur yang melekat dalam pemikiran dan menerapkan standar-standar intelektual padanya (Fisher, 2009).

Dari hasil angket menunjukkan bahwa lebih dari 40% siswa merasa guru cukup baik dalam membangkitkan berpikir kritis siswa, akan tetapi belum sepenuhnya dari siswa memahami berpikir kritis yang dimaksudkan oleh peneliti. Analisis diketahui dengan dilakukan persebaran angket. Hasil angket dari responden pengisi angket ini adalah siswa kelas XI IPA/ MIPA yang dipilih secara acak, hasil dari angket kebutuhan siswa ini dapat kesimpulan:

- a) Sebanyak 80,8% siswa memilih tertarik dengan kemampuan berpikir kritis yang digunakan sebagai acuan untuk evaluasi hasil belajar agar dapat mengetahui sejauh mana kemampuan berpikir kritis mereka.

b) Sebanyak 72,2% responden menginginkan adanya pengembangan soal dengan alasan secara umum dari responden angket yang mengacu pada keterampilan berpikir tingkat tinggi, baik untuk siswa, karena dapat merangsang kemampuan siswa untuk berpikir tingkat kritis sehingga siswa dapat melatih kemampuan berpikir kreatif juga dapat mengembangkan kemampuan dan pengetahuan mereka dalam bidang ilmu kimia.

3. Analisis tugas (*task analysis*)

Pada tahap awal analisis tugas, perlu dilakukan analisis kompetensi-kompetensi yang ingin dicapai sesuai dengan kurikulum, yang bertujuan untuk menentukan langkah pengembangan instrumen tes. Selain itu juga menganalisis indikator yang harus dikuasai siswa agar dapat mencapai kompetensi minimal. Berdasarkan analisis tugas kurikulum yang sesuai mulai dari bahan kajian, pokok bahasan, subpokok bahasan serta garis besar perincian isi pokok bahasan. Adapun kurikulum yang dipakai adalah kurikulum 2013 dengan bahan kajian yang sesuai dengan kondisi berdasarkan analisis ujung depan dan analisis siswa Analisis tugas pada penelitian pengembangan adalah sebuah alat penilaian yang mengakomodasi kekurangan dari soal *essay* dan soal *multiple choice* yang sekaligus mengetahui kemampuan berpikir kritis siswa agar dapat diketahui sejauh mana pemahaman siswa tentang materi pembelajaran dan perbaikan apa yang harus dilakukan oleh guru setelah melakukan evaluasi. Dalam hal ini evaluasi sebagai alat pengukur ketercapaian tujuan mata pelajaran, sebagai alat pengukur tujuan proses belajar mengajar, mengetahui kelemahan siswa dan dapat menyelesaikan kesulitan belajar siswa.

4. Analisis konsep (*concept analysis*)

Analisis konsep ini berupa analisis penerapan kemampuan berpikir kritis siswa yang dimasukkan dalam instrumen tes *Two-Tier Multiple Choice*. Berdasarkan analisis konsep didapat materi-materi utama yang akan dimasukkan dalam pengembangan instrumen tes. Pada kompetensi dasar pertama membedakan reaksi eksoterm dan reaksi endoterm berdasarkan hasil percobaan dan diagram tingkat energi. Pada kompetensi dasar kedua menentukan ΔH reaksi berdasarkan hukum Hess, data perubahan entalpi pembentukan standar, dan data energi ikatan.

Kompetensi dasar diatas akan menjadi dasar dalam penentuan indikator kompetensi dan indikator soal yang dikembangkan. Dalam instrumen tes yang dikembangkan harus memiliki indikator keterampilan berpikir kritis menurut (Facione, 2015) memiliki enam aspek yaitu: 1) interpretasi (*interpretation*) 2) analisis (*analysis*); 3) evaluasi (*evaluation*); 4) kesimpulan (*inference*); 5) penjelasan (*explanation*); 6) pengaturan diri (*self regulation*)

5. Perumusan tujuan pembelajaran (*specifying instructional objectives*)

Perumusan tujuan pembelajaran berdasarkan hasil analisis ujung depan, analisis siswa, analisis tugas dan analisis konsep. Adapun pengalaman belajar yang diharapkan dalam tujuan pembelajaran adalah siswa mampu membedakan reaksi eksoterm dan reaksi endoterm berdasarkan hasil percobaan dan diagram tingkat energi dan menentukan ΔH reaksi berdasarkan hukum Hess, data perubahan entalpi pembentukan standar, dan data energi ikatan. Perumusan tujuan pembelajaran berguna untuk merangkum hasil dari analisis konsep dan analisis tugas untuk menentukan pengembangan objek penelitian. Selain itu untuk membatasi agar

penelitian dan pengembangan dari instrumen tes tidak menyimpang dari tujuan semula. Pada perumusan tujuan pembelajaran ini penyusunan tes menyesuaikan alokasi waktu siswa disekolah dan karakteristik materi yang digunakan dalam pembelajaran.

B. *Design (Perancangan)*

Tujuan dari tahap ini adalah untuk merancang instrumen tes yang akan dikembangkan sehingga diperoleh *prototipe* instrumen tes *Two-Tier Multiple Choice* (instrumen tes *two-tier multiple choice* draft I). Tahap design ini dilakukan untuk menyiapkan bahan dan mendesain produk yang akan dikembangkan. Thiagarajan (1974: 7) mengemukakan bahwa tahap *design* ini meliputi:

1. Penetapan bentuk penyajian

Penetapan bentuk penyajian disesuaikan dengan kebutuhan yang akan diukur, dimana dalam penelitian pengembangan ini instrumen yang dihasilkan adalah *Two-Tier Multiple Choice*. Menurut Adodo (2013) *Two-Tier Multiple Choice (TTMC)* adalah bentuk pertanyaan yang lebih canggih dari pertanyaan pilihan ganda. Tingkat pertama menyerupai pilihan ganda tradisional, yang biasanya berkaitan dengan pernyataan pengetahuan. Tingkat kedua menyerupai format dari soal pilihan ganda tradisional tetapi bertujuan untuk mendorong pemikiran dan penalaran keterampilan yang lebih tinggi. Penetapan bentuk penyajian dalam pengembangan instrumen ini dimaksudkan untuk merancang isi soal/instrumen agar dapat mengukur kemampuan berpikir kritis.

2. Perancangan awal

Pada langkah ini peneliti membuat rancangan awal yang berisi rancangan seluruh kegiatan yang harus dilakukan sebelum dilaksanakannya uji coba. Hasil tahap ini berupa rancangan awal instrumen tes yang merupakan draft I berserta instrumen penelitian. Berikut uraian singkat mengenai rancangan awal instrumen tes dan instrumen penelitian. Instrumen tes *Two-Tier Multiple Choice* ini berupa rancangan keseluruhan tes *Two-Tier Multiple Choice* yang telah dikembangkan sebelum dilakukan penilaian atau validasi oleh validator ahli dan guru kimia, kemudian dilakukan penyempurnaan atas masukan dan saran oleh validator.

C. *Development (tahap Pengembangan)*

Tujuan dari tahap pengembangan adalah untuk menghasilkan instrumen tes *Two-Tier Multiple Choice* yang telah direvisi berdasarkan masukan para ahli dan data yang diperoleh dari uji coba. Kegiatan pada tahap ini adalah penilaian para ahli (validator) dan uji coba lapangan instrumen yang akan dikembangkan. Langkah pengembangan pada tahap ini antara lain :

1. Penyusunan draf produk

Pengembangan instrumen *Two-Tier Multiple Choice* penilaian ini, digunakan prosedur yang diusulkan oleh Treagust dalam Chandrasegaran. Tahap pertama adalah identifikasi area konten penelitian. Dalam hal ini kontennya adalah materi termokimia. Materi ini meliputi membedakan reaksi eksoterm dan reaksi endoterm berdasarkan hasil percobaan dan diagram tingkat energi, menentukan ΔH reaksi berdasarkan hukum Hess, data perubahan entalpi pembentukan standar, dan data energi ikatan. Tahap kedua adalah identifikasi konsepsi siswa berdasarkan hasil analisis pada tahap pendefinisian. Tahap terakhir adalah merancang item/soal tes dan validasi yang berdasarkan pada tahap *design*/perancangan.

Penyusunan draf awal instrumen adalah menentukan tujuan dari penilaian. Ada empat tujuan penilaian, yaitu formatif, sumatif, diagnostik, dan penempatan. Pada penelitian ini, instrumen penilaian disusun untuk tujuan formatif, yaitu hasil penilaian digunakan sebagai umpan balik baik guru maupun siswa. Hal ini untuk mengetahui berhasil tidaknya suatu proses pembelajaran yang telah dilaksanakan. Apabila hasilnya menunjukkan ketidakberhasilan proses pembelajaran, maka perlu adanya program remedial bagi siswa tapi dalam penelitian ini adalah menguji kelayakan dari instrumen yang dikembangkan.

Setelah menentukan tujuan dari penilaian, maka dibuat kisi-kisinya. Kisi-kisi disusun berdasarkan kompetensi apa saja yang harus dinilai. Penelitian dan pengembangan ini bertujuan untuk mengembangkan instrumen penilaian untuk mengukur keterampilan berpikir kritis siswa. Maka oleh itu, indikator yang dibuat tidak hanya berdasar pada indikator kompetensi dari materi termokimia, tetapi juga dihubungkan dengan indikator dari keterampilan berpikir kritis. Indikator keterampilan berpikir kritis yang digunakan ada enam, yaitu interpretasi, analisis, kesimpulan, evaluasi, penjelasan, dan pengaturan diri.

Setelah menyusun kisi-kisi, kemudian disusun soal sesuai dengan bentuk soal yang diinginkan, yaitu berupa pilihan ganda. Soal tes pilihan ganda dipilih karena efisiensi dalam pengerjaan dan pengoreksian. Pemberian skor pun tidak dipengaruhi oleh subjektivitas guru karena bersifat objektif. Kelemahan dari soal pilihan ganda adalah sulitnya untuk menyusun konstruksi soal dan tidak bisa mengetahui bagaimana proses siswa dalam menyelesaikan permasalahan, apakah siswa benar-benar memahami materi, atau sekedar menebak, atau bahkan mencontek pekerjaan teman. Kelemahan tersebut akan muncul apabila soal pilihan ganda biasa yang digunakan. Untuk mengurangi dan menghindari kelemahan tersebut, maka dikembangkanlah instrumen penilaian *Two-Tier Multiple Choice*. Jumlah soal disesuaikan dengan indikator kompetensi materi termokimia serta ditinjau dari alokasi waktu untuk mengerjakannya.

2. Validasi draf produk

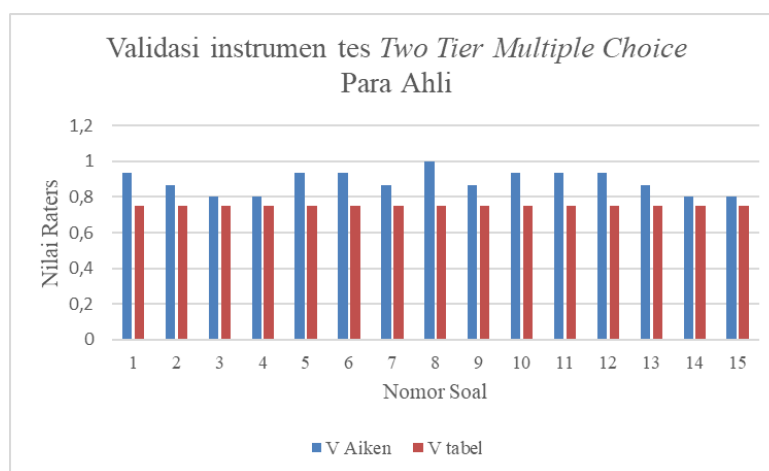
Instrumen penilaian harus memiliki kualitas yang baik. Instrumen penilaian dapat dikatakan memiliki kualitas yang baik apabila memenuhi persyaratan yang tercantum pada Permendiknas Nomor 20 Tahun 2007 tentang Standar Penilaian. Menurut Sukiman, agar suatu instrumen penilaian memenuhi persyaratan tersebut dapat dilakukan dua cara, yaitu 1) analisis soal secara teoritik atau kualitatif dan 2) analisis soal secara empiris atau kuantitatif (Sukiman, 2012).

Validasi dengan para ahli merupakan cara analisis secara teoritik yang dilakukan sebelum pelaksanaan uji coba. Cara ini untuk menganalisis instrumen penilaian dilihat dari segi substansi, konstruksi dan bahasa. Analisis dilakukan dengan validasi terhadap para ahli. Draft produk yang telah selesai dibuat kemudian divalidasi. Tujuan dari dilakukannya validasi ini adalah untuk mengetahui apakah instrumen tersebut mempunyai kualitas yang baik. Kualitas yang baik ditentukan oleh isi dari instrumen tersebut yaitu meliputi kesesuaian indikator keterampilan berpikir kritis dan indikator soal dengan soal, penggunaan bahasa yang mudah dimengerti, tidak adanya kesalahan penulisan dan tanda baca, penentuan kunci jawaban yang benar.

Validasi ini dilakukan oleh para ahli yang menguasai dan mengerti tentang

instrumen yang dibuat. Maka validatornya terdiri dari empat dosen pendidikan kimia yang menguasai tentang pendidikan, evaluasi dan materi kimia, serta guru yang sudah berpengalaman mengajar dan memahami karakteristik siswa dengan baik. Validitas dipengaruhi oleh dua variabel, yaitu jumlah validator dan jumlah kriteria validitas yang digunakan. Valid tidaknya instrumen ditentukan dengan membandingkan nilai validitas hasil perhitungan dengan nilai validitas standar (V tabel). Hasil dari validasi isi oleh para ahli dihitung menggunakan formula Aiken (Aiken, 1985).

Setelah mengalami validasi produk instrumen tes *Two Tier Multiple Choice* Tes diperbaiki sesuai saran yang direkomendasikan oleh para validator. Hasil dari validasi isi oleh para ahli dihitung menggunakan formula Aiken. Validator berjumlah 5 orang para ahli yang terdiri dari 4 orang dosen dan 1 orang guru serta terdapat 4 skala pengukuran. Tiap soal dikatakan valid apabila memiliki nilai validitas hitung lebih dari nilai validitas tabel. Berdasarkan jumlah validator yaitu 5 orang dan skala pengukuran yaitu 4, maka nilai validitas tabelnya adalah 0,75. Hasil perhitungan validitas terhadap 15 soal menunjukkan nilai validitasnya lebih dari 0,75 sehingga dapat dikatakan bahwa semua soal sudah valid. Akan tetapi perbaikan soal juga dilakukan sesuai dengan saran dari masing-masing validator agar soal yang telah dibuat menjadi lebih baik lagi. Rata-rata validitas instrumen *Two-Tier Multiple Choice* yang dipakai dalam validasi ini adalah 0,844, sehingga termasuk dalam kategori valid dan layak digunakan dalam segi analisis logis.



Gambar 1. Validitas Aiken Instrumen TTMC

3. Pengujian Produk

Instrumen yang telah diperbaiki sesuai dengan saran dari validator dan diperoleh produk awal kemudian diujicobakan. Uji coba instrumen tes *Two-Tier Multiple Choice* dilakukan selama dua kali. Pengujian produk ini adalah analisis secara empiris. Cara empiris dilakukan dengan cara uji coba untuk mengetahui reliabilitas, daya pembeda, tingkat kesukaran, dan indeks pengecoh soal. Uji coba terdiri dari 2 kali pelaksanaan. Sebuah instrumen dapat dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman. Analisis soal secara kuantitatif menekankan pada analisis karakteristik internal tes melalui data yang diperoleh secara empiris. Karakteristik internal secara kuantitatif dimaksudkan meliputi

parameter soal tingkat kesukaran, daya pembeda dan reliabilitas.

- a. Reliabilitas tes Suatu tes dapat dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda. Reliabilitas alat penilaian adalah ketepatan atau keajegan alat tersebut dalam menilai apa yang dinilainya”. (Sudjana, 2014:16). Reliabilitas rata-rata yang diperoleh pada dua kali ujicoba adalah 0,737 sehingga dinyatakan reliabilitasnya dikategorikan tinggi, sehingga instrumen tes *Two-Tier Multiple Choice* yang dikembangkan dapat diujikan dalam kondisi yang berbeda dan diperoleh hasil yang relatif sama.
- b. Daya pembeda suatu soal tes yaitu bagaimana kemampuan soal tersebut untuk membedakan siswa-siswa yang termasuk kelompok pandai (*upper group*) dengan siswa-siswa yang termasuk kelompok kurang (*lower group*) (Purwanto, 2002:120). Daya beda dalam instrumen tes ini dihitung berdasarkan rata-rata dari kedua tingkatan soal yaitu tier 1 dan tier 2, dimana terdapat dua kali uji coba diperoleh hasil daya beda sebagai berikut ini : 41,67% dikategorikan sangat baik, 40,00% dikategorikan baik, dan 18,33% dikategorikan cukup baik. Tidak ada soal yang dinyatakan jelek dalam kedua kali ujicoba ini.
- c. Tingkat kesukaran (*difficulty index*) dalam evaluasi hasil belajar umumnya dilambangkan dengan huruf P (proporsi), didefinisikan sebagai proporsi siswa peserta tes yang menjawab benar. Butir-butir item tes hasil belajar dinyatakan sebagai butir-butir item yang baik apabila butir-butir item tersebut tidak terlalu sukar dan tidak terlalu mudah, dengan kata lain derajat kesukaran item tersebut adalah sedang atau cukup (Sudijono, 2009:370). Tingkat kesukaran instrumen tes ini dihitung berdasarkan rata-rata dari kedua tingkatan soal yaitu tier 1 dan tier 2, dimana terdapat dua kali uji coba diperoleh hasil tingkat kesukaran sebagai sebagai berikut ini : 8,33% dikategorikan soal yang mudah, 80,00% dikategorikan soal yang sedang dan 11,67% dikategorikan sukar.
- d. Menurut Anas Sudijono (2009:410), Distraktor yang berfungsi dengan baik yaitu distraktor yang memiliki daya tarik atau daya rangsang, sehingga *testee* merasa bimbang, dan ragu-ragu yang pada akhirnya mereka menjadi terkecoh untuk memilih distraktor sebagai jawaban betul. Berdasarkan hasil *ITEMAN* untuk indeks pengecoh disimpulkan bahwa pengecoh tidak ada yang bekerja lebih baik daripada kunci jawaban dan nilai pengecoh sebagian besar lebih dari 5% dipilih sehingga dinyatakan efektif.

4. *Disseminate* (Penyebaran)

Penyebaran dilakukan kepada skala yang lebih luas artinya disebarikan pada bukan hanya satu sekolah tapi sekolah lain dan pada guru lain untuk lebih melihat efektifitas instrumen yang dikembangkan.

2. Pembahasan

Pada proses pengembangan instrumen *Two-Tier Multiple Choice* pada materi termokimia memiliki beberapa fungsi. Instrumen penilaian Two-Tier MC memiliki fungsi yaitu pertama fungsi diagnostik yaitu untuk mengetahui kelemahan siswa dan penyebabnya sehingga dapat diselesaikan masalahnya. Kedua, fungsi formatif yaitu untuk menentukan *feedback* bagi bagi guru maupun siswa, sebagai

dasar untuk memperbaiki proses pembelajaran dan remedial bagi siswa. Ketiga, fungsi penempatan yaitu mengetahui tingkat keterampilan berpikir kritis dari masing-masing siswa, karena pada dasarnya siswa memiliki kemampuan yang berbeda-beda dalam keterampilan berpikir kritis sehingga guru dapat menempatkan siswa dalam situasi pembelajaran yang tepat sesuai dengan tingkat kemampuan siswa. Pada penelitian ini tiga fungsi ini dapat menjadi acuan guru dalam mengetahui sejauh mana materi pelajaran dapat diterima siswa, apakah tujuan pembelajaran sudah tercapai atau belum, serta membantu guru untuk menentukan langkah yang tepat untuk memperbaikinya.

Instrumen *Two-tier Multiple Choice* yang dikembangkan mengandung dua tingkat pertanyaan yang saling terhubung. Tujuan dari lapis kedua ini adalah untuk mendorong siswa untuk berpikir tingkat tinggi dan keterampilan penalaran. Tingkat pertama dari pertanyaan biasanya berkaitan dengan pernyataan pengetahuan sedangkan unsur kedua dari pertanyaan memfasilitasi pengujian siswa belajar di tingkat berpikir yang lebih tinggi. Instrumen pertanyaan ini membuat lebih mudah untuk menguji tingkat pemikiran siswa yang lebih tinggi dibandingkan dengan soal pilihan ganda konvensional. Berpikir kritis merupakan bagian dari berpikir tingkat tinggi. Pada penelitian Pengembangan instrumen penilaian *Two-tier Multiple Choice* untuk mengukur keterampilan berpikir tingkat tinggi ini juga sejalan dengan teori-teori pembelajaran, sebagai contoh teori teori perkembangan kognitif dari Piaget (Shidiq, 2014). Tujuan dari pertanyaan tingkat kedua ini adalah untuk membantu siswa dan guru untuk mengidentifikasi masalah siswa sehingga mereka dapat berpikir ulang untuk memperbaiki kesalahan atau kesulitan dan mengembangkan pemahaman yang lebih dalam tentang topik (Cullinane, 2011).

Hasil persentase keterampilan berpikir kritis siswa menunjukkan keterkaitan antara keterampilan berpikir kritis siswa dengan kemampuan atau hasil belajar siswa. Persentase paling banyak adalah yaitu 55,71 % dengan kategori sedang, kategori tinggi 24,29% dan kategori rendah sebesar 20%. Pada penelitian ini kemampuan berpikir kritis siswa perlu dianalisis untuk melihat bagaimana instrumen ini bekerja dalam mengukur kemampuan berpikir kritis pada sampel yang lebih besar dalam uji efektifitas instrumen tes. Pada keterampilan berpikir kritis yang diperlihatkan diatas memakai sampel sebesar 70 siswa. Pada ujicoba sampel ini kemampuan berpikir kritis yang dikategorikan tinggi memiliki nilai yang lebih baik dari dua kategori yang lain. Menurut Kartimi berpikir kritis termasuk karakteristik psikologis seseorang yang dapat diketahui kualifikasinya (rendah, sedang, atau tinggi) dan hal itu bisa diketahui apabila diadakan pengukuran dengan aturan dan formula yang jelas (Kartimi & Liliyasi, 2012).

Pada penelitian ini nilai ujicoba siswa dengan nilai yang tinggi sesuai dengan ketrampilan berpikir kritisnya tinggi. Penilaian dalam penelitian mengacu pada penggunaan *Grade Response Measurement (GRM)* sesuai untuk diterapkan pada penskoran soal yang dapat klasifikasikan kategorinya atau memiliki tingkatan yang cenderung meningkat (Yamtinah, 2016). Pengembangan instrumen penilaian ini dapat digunakan sebagai alternatif penilaian untuk mengukur keterampilan berpikir kritis siswa. Pada penggunaan instrumen ini sesuai dengan pendapat yang dikemukakan oleh Qing bahwa di dalam proses belajar diperlukan kemampuan berpikir. Ketika siswa-siswinya memiliki keterampilan berpikir kritis yang baik

atau tinggi maka hasil belajarnya juga akan baik atau tinggi (Qing, Jing, Yazhuan, Ting, & Junping, 2010). Instrumen penilaian TTMC untuk mengukur keterampilan berpikir tingkat tinggi ini membantu siswa yang tergolong usia remaja menuju dewasa untuk mengembangkan tahapan perkembangannya dengan melatih untuk berpikir kritis dalam pemecahan masalah (King, 2004).

KESIMPULAN

1. Instrumen penilaian *Two-Tier Multiple Choice* yang dikembangkan dinyatakan layak dan memenuhi kriteria sebagai suatu soal yang baik dengan hasil validitas aiken rata-rata 0,84
2. Instrumen penilaian *Two-Tier Multiple Choice* yang dikembangkan dinyatakan layak ditinjau dari realibilitas tes rata-rata 0,737 yang tergolong tinggi, memiliki daya beda dengan persentase 41,67% sangat baik, 40,00% baik, 18,33% cukup baik dan 0% jelek., berdasarkan tingkat kesukaran 8,33% mudah, 80,00% sedang dan 11,67% sukar, berdasarkan indeks pengecoh disimpulkan pengecoh tidak ada yang bekerja lebih baik daripada kunci jawaban dan nilai pengecoh sebagian besar lebih dari 5% dipilih sehingga dinyatakan efektif.

DAFTAR RUJUKAN

- Adodo, S. O. (2013). *Effects of Two-Tier Multiple Choice Diagnostic Assessment items on Students' Learning Outcome in Basic Science Technology*. Ondo State: Academic Journal of Interdisciplinary Studies by MC SER-CEMAS-Sapienza University of Rome. E-ISSN 2281-4612. ISSN 2281-3993. Vol 2 no 2.
- Aiken, L. R. (1985). Three Coefficients for Analyzing The Reliability and Validity of Ratings. *Education and Psychological Measurement*, 45, 131–142.
- Chandrasegaran, A. L., Treagus, D. F. Mucerino, M. (2007). *The Development of a Two-tier Multiple-Choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Level of Representation*. The Royal Society of Chemistry. Chemistry Education Research and Practice, 8(3), 293-307
- Cullinane, A., Liston, M. (2011). *Two-tier Multiple Choice Question (MCQs)-How Effective are they: A Pre-servis Teachers' Perspective*. UK: IOSTE-NW Europe
- Facione, P. A. (2015). *Critical Thinking : What It Is and Why It Counts*. Insight Assessment.
- Fisher, A. (2009). *Berpikir Kritis*. Jakarta: Erlangga.
- Fisher, Alec., Spiker, A, and Riedel, S. L. (1992). *Critical Thinking Training for Army Officers Volume Two: A Model of Critical Thinking (Research Report)*. US Army Research Institute for the Behavioral and Social Science.

- Gagne, R.M. (1980). Learnable Aspects of Human Thinking. In: Lawson, A. E. (Ed). Science Education Information Report. (hlm. 1-28.) New York: The Eric Science, Mathematics and Environmental Education Clearing House.
- Hassoubah, Z. I. (2004). *Developing Creative & Critical Thinking : Cara Berpikir Kreatif & Kritis*. Bandung : Nuansa.
- Kartimi, & Liliyasi. (2012). Pengembangan Alat Ukur Berpikir kritis pada Konsep Termokimia untuk Siswa SMA Peringkat Atas dan Menengah. *Jurnal Pendidikan IPA Indonesia (JPPI)*, 1(1), 21–26.
- Kartimi, Liliyasi, & Permanasari, A. (2012). Pengembangan Alat Ukur Berpikir Kritis pada Konsep Senyawa Hidrokarbon untuk Siswa SMA di Kabupaten Kuningan. *Jurnal Pendidikan MIPA*, 13(1), 18–25.
- King, F. J., Goodson, L., Rohani, F.,(2004). Higher Order Thinking Skill. A publication of the Educational Services Program, now known as the Center for Advancement of Learning and Assessment. <http://www.Cala.Fsu.Edu>
- Krulik, S dan Rudnick, J.A (1993). *The New Sourcebook for Teaching Reasoning and Problem Solving in Elementary School*. Massachusetts: Allyn & Bacon A Simon & Schuster Company.
- Purwanto, Ngilim (2002) *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: PT Remaja Rosdakarya.
- Shidiq, A.S. 2014. *Pengembangan Instrumen Penilaian Two-Tier Multiple Choice Untuk Mengukur Keterampilan Berpikir tingkat Tinggi (Higher Order Thinking Skills) Pada Materi Kelarutan Dan Hasil Kali Kelarutan Untuk Siswa SMA/MA Kelas XI*. *Jurnal Pendidikan Kimia (JPK)*,3(4),83-92.
- Qing, Z., Jing, G., Yazhuan, L., Ting, W., & Junping, M. (2010). Promoting Preservice Teacher's Critical Thinking Disposition by Inquiry-Based Chemical Experiment. *Procedia - Social and Behavioral Sciences*, 9, 1429–1436.
- Sudijono, Anas. (2009). *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Pers
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Mandiri.
- Thiagarajan, Sivasailam, Dorothy S. Semmel, and Melvyn I. Semmel.(1974). *Instructional Development for Training Teachers of Exceptional Children*. Washinton DC: National Center for Improvement Educational System.
- Yaminah, S., Haryono, Saputro, S., Mulyani, B., & BU, S. (2016). Item Discrimination of Two Tier Test. *Proceeding International Conference on Educatonal Research and Evaluating (ICERE)*. 29-31 May 2016, Yogyakarta State University.