



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

Ari Syahidul Shidiq^{*1}, Mohammad Masykuri², dan Elfi Susanti V. H.²

¹ Mahasiswa Program Studi Pendidikan Kimia PMIPA, FKIP, UNS Surakarta, Indonesia

² Dosen Program Studi Pendidikan Kimia PMIPA, FKIP, UNS Surakarta, Indonesia

*Keperluan korespondensi, telp: 08-7829-5050-03, email: arisyahidul@yahoo.co.id

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian *Two-tier Multiple Choice* yang memiliki validitas, reliabilitas, tingkat kesukaran, daya pembeda dan indeks pengecoh yang memenuhi kriteria sebagai suatu soal yang baik dan mengembangkan instrumen *Two-tier Multiple Choice* pada materi kelarutan dan hasil kali kelarutan menjadi instrumen penilaian yang mampu mengukur keterampilan berpikir tingkat tinggi. Penelitian ini merupakan penelitian pengembangan yang terdiri dari analisis kebutuhan dan penelitian awal, perencanaan, pengembangan bentuk awal produk, revisi dan uji lapangan awal, revisi dan uji lapangan utama, dan revisi dan uji pelaksanaan lapangan. Subjek penelitian pada penelitian ini adalah siswa kelas XI IPA SMAN 1 Surakarta dan SMAN Sragen Bilingual Boarding School tahun ajaran 2013/2014. Data penelitian diperoleh melalui wawancara, tes dan angket. 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 isi (CV) 1,00, memiliki reliabilitas tes rata-rata 0,92 yang tergolong sangat tinggi, memiliki tingkat kesukaran dengan persentase 4,1% (mudah), 79,2% (sedang), dan 16,7% (sukar), memiliki daya pembeda dengan persentase 18,7% (sangat baik), 39,6% (baik), 14,6% (cukup), 27,1% (jelek), dan memiliki indeks pengecoh dengan persentase 26,1% (sangat baik), 21,9% (baik), 30,2% (kurang), 18,2% (buruk), dan 3,6% (sangat buruk) dan instrumen *Two-tier Multiple Choice* untuk mengukur keterampilan berpikir tingkat tinggi pada materi kelarutan dan hasil kali kelarutan yang dikembangkan cukup mampu mengukur keterampilan berpikir tingkat tinggi dengan angka konsistensi 41,6%.

Kata Kunci: Penelitian dan pengembangan (R&D), *Two-tier multiple choice*, *Higher order thinking skills*, kelarutan dan hasil kali kelarutan

PENDAHULUAN

Proses pendidikan yang berkualitas akan menghasilkan pendidikan yang sesuai dengan fungsi dan tujuannya. Proses pendidikan merupakan kegiatan sosial atau pergaulan antara pendidik dan peserta didik dengan menggunakan isi atau materi pendidikan, metode, dan alat pendidikan tertentu yang berlangsung dalam suatu lingkungan untuk mencapai tujuan pendidikan yang telah ditetapkan [1]. Kualitas pendidikan sangat ditentukan

oleh kemampuan satuan pendidikan dalam mengelola proses pembelajaran. Penilaian merupakan bagian yang penting dalam proses pembelajaran. Dengan melakukan penilaian, pendidik sebagai pengelola kegiatan pembelajaran dapat mengetahui kemampuan yang dimiliki peserta didik, ketepatan metode mengajar yang digunakan, dan keberhasilan peserta didik dalam meraih kompetensi yang telah ditetapkan.

Standar Kompetensi Lulusan (SKL) pada peraturan menteri pendidikan nasional nomor 23 tahun 2006 untuk SMA/MA yang antara lain menyebutkan bahwa lulusan SMA/MA harus dapat membangun dan menerapkan informasi dan pengetahuan secara logis, kritis, kreatif, dan inovatif juga harus dapat menunjukkan kemampuan berpikir logis, kritis, kreatif, dan inovatif dalam pengambilan keputusan. Demi terciptanya SKL yang telah ditetapkan, instrumen penilaian haruslah berorientasi pada keterampilan berpikir tingkat tinggi sehingga mampu mendorong siswa untuk mengembangkan kemampuan berpikirnya dan mampu mengikuti perkembangan pengetahuan dan teknologi [2] [3].

Berdasarkan hasil observasi dan wawancara yang dilakukan dengan guru kimia di SMA N 1 Surakarta dan SMA N Sragen Bilingual Boarding School (SBBS) materi kelarutan dan hasil kali kelarutan merupakan materi yang dianggap sulit oleh siswa. Kedua sekolah ini memiliki cara yang berbeda untuk melakukan penilaian. SMA N 1 Surakarta lebih sering menggunakan soal jenis *essay* untuk menilai pengetahuan siswanya, karena soal jenis ini lebih dapat mengungkap pengetahuan yang dimiliki siswa secara lebih mendalam dan soal jenis ini dianggap dapat mengurangi tindakan curang dari peserta didik. Sedangkan SMA N SBBS lebih sering menggunakan soal jenis pilihan ganda, karena mempermudah proses penilaian dan mengurangi subjektivitas dalam penilaian. Selain itu, dari analisis terhadap soal Ulangan Tengah Semester yang digunakan di masing-masing sekolah untuk SMA N 1 Surakarta instrumen yang digunakan memiliki komposisi jenjang 2,8% C1, 37,2% C2, dan 60% C3 sedangkan untuk SMA N SBBS memiliki komposisi 3,3% C1, 56,7% C2, dan 40% C3, pada jenjang ranah kognitif Bloom. Hal ini menunjukkan masih minimnya penggunaan instrumen untuk mengungkap keterampilan berpikir tingkat tinggi siswa karena berdasarkan ranah kognitif Bloom tingkat C1, C2, dan C3 masih tergolong kemampuan berpikir tingkat rendah.

Hasil observasi di dua sekolah ini menunjukkan bahwa untuk memaksimalkan hasil dari penilaian dibutuhkan instrumen evaluasi yang dapat mengungkap pengetahuan yang dimiliki siswa secara lebih mendalam, mampu mengukur keterampilan berpikir tingkat tinggi siswa dan dapat digunakan secara praktis tanpa takut adanya subjektivitas penilaian dan juga dapat mengurangi tindakan curang siswa dalam menjawab, untuk itu dibutuhkan penelitian guna menghasilkan suatu instrumen penilaian tersebut. Penelitian untuk menghasilkan suatu instrumen penilaian tersebut dilakukan dengan penelitian jenis *Research and Development (R&D)*.

Penggunaan instrumen pilihan ganda hanya dapat menilai konten pengetahuan tanpa mempertimbangkan alasan dibalik pilihan tersebut. Sebagai perbaikan dari hal ini dikembangkan suatu instrumen pilihan ganda yang mencakup respon dan alternatif konsepsi siswa. Siswa diharuskan untuk membenarkan pilihan jawaban mereka dengan memberikan alasan. Penggunaan alasan ketika menjawab item tes pilihan ganda menjadi cara yang sensitif dan efektif untuk menilai pembelajaran yang bermakna. Hasil positif temuan yang terkait dengan alasan siswa menyebabkan perkembangan tes diagnostik pilihan ganda dua tingkat (*Two-tier Multiple Choice Diagnostic Test*) yang bertujuan khusus untuk mengidentifikasi alternatif konsepsi siswa dalam konten yang jelas dan terbatas [4].

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 berpikir yang lebih tinggi (*Higher Order Thinking Skill*) [5].

Higher Order Thinking Skill meliputi berpikir kritis, logis, reflektif, metakognitif, dan kreatif. Semuanya diaktifkan ketika individu mendapatkan masalah yang tidak familiar, tidak tentu, penuh pertanyaan

dan dilematis. Sementara itu, banyak guru tidak dapat menyiapkan instrumen *Two-tier Multiple Choice* dengan versi mereka sendiri karena kurangnya waktu atau kurangnya pengetahuan tentang pengembangan tes jenis ini. Disisi lain untuk mendorong siswa melampaui kapasitas pengetahuan mereka, mereka perlu mengembangkan keterampilan berpikir yang lebih tinggi seperti berpikir kritis, pengambilan keputusan, dan pemecahan masalah oleh karena itu para guru sendiri perlu memperoleh dan mempraktekkan keterampilan ini [6,7,8].

Berdasarkan uraian diatas, perlu adanya 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.

Ada teknis kriteria dalam penilaian secara profesional yang digunakan untuk mengevaluasi kualitas dari suatu tes dan prosedur pengukuran lainnya yaitu "*psychometric soundness*" dari suatu tes, ada dua aspek yaitu reliabilitas dan validitas. Disisi lain, identifikasi terhadap seluruh instrumen hasil belajar (item tes) itu dilakukan dengan harapan akan menghasilkan berbagai informasi berharga yang pada dasarnya merupakan umpan balik guna melakukan perbaikan, pembenahan, dan penyempurnaan kembali terhadap butir-butir item yang dikeluarkan dalam tes hasil belajar. Rangkaian kegiatan seperti itu disebut dengan istilah analisis item (item analysis) yang meliputi analisis derajat kesukaran, daya pembeda item, dan analisis fungsi distraktor atau indeks pengecoh.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan instrumen penilaian *Two-tier Multiple Choice* yang memiliki validitas, reliabilitas, tingkat kesukaran, daya pembeda dan indeks pengecoh yang memenuhi kriteria sebagai suatu soal yang baik dan mengembangkan instrumen *Two-tier Multiple Choice* pada materi kelarutan dan hasil kali kelarutan menjadi instrumen penilaian yang mampu mengukur keterampilan berpikir tingkat tinggi [9,10].

METODE PENELITIAN

Penelitian ini merupakan jenis penelitian dan pengembangan (*Research and Development*) yang mengacu pada Borg and Gall (1983) yang diadaptasi sesuai dengan kebutuhan penelitian [11]. Penelitian ini dilakukan di SMA N 1 Surakarta dan SMA N Sragen Bilingual Boarding School (SBBS).

Penelitian ini dibagi kedalam lima tahap utama yaitu tahap pertama merupakan *Planning*. Tahap ini bertujuan untuk merencanakan penelitian yang didahului oleh wawancara dan analisis kebutuhan di dua sekolah tempat penelitian.

Tahap kedua yaitu mengembangkan bentuk awal produk (*Develop of the Preliminary Form of the Product*). Pada tahap ini mulai dikembangkan produk yaitu berupa instrumen penilaian dengan bentuk *Two-tier Multiple Choice* sesuai rencana. Sebelum memasuki tahap berikutnya, produk yang telah dikembangkan terlebih dahulu divalidasi oleh 2 orang ahli. Hasil validasi ini digunakan sebagai landasan revisi produk.

Tahap ketiga yaitu uji coba lapangan awal dan revisi (*Preliminary Field Test and Revision*). Uji coba ini dimaksudkan untuk menguji keterbacaan produk instrumen penilaian dengan bentuk *Two-tier Multiple Choice*. Tahap ini dilakukan dengan melakukan tes terhadap 12 orang siswa untuk mengerjakan soal *Two-tier Multiple Choice*. Kemudian diberikan angket balikan untuk mengetahui tanggapan mereka tentang instrumen *Two-tier Multiple Choice*. Hasil tes dan angket ini digunakan sebagai landasan untuk merevisi produk.

Tahap keempat yaitu uji coba lapangan utama dan revisi (*Main Field Test and Revision*). Pada tahap ini dilakukan uji coba produk dengan melakukan tes formatif kepada 1 kelas dari 2 sekolah yang berbeda dengan jumlah siswa 53 orang. Dari uji coba ini, reliabilitas, tingkat kesukaran, daya pembeda dan indeks pengecoh yang memenuhi kriteria sebagai suatu soal yang baik. Hasilnya digunakan sebagai

landasan revisi produk sehingga menjadi suatu produk yang lebih baik.

Tahap kelima yaitu uji coba operasional (*Operational Field Test and Revision*). Tahap ini merupakan uji coba terakhir yang dilakukan, instrumen diujikan pada 2 kelas dari dua sekolah yang berbeda dengan jumlah 95 orang siswa. Setelah mengerjakan soal bentuk *Two-tier Multiple Choice* siswa diminta untuk mengisi angket guna mengungkap tanggapan mereka tentang instrumen yang sedang dikembangkan. Hasil dari uji coba ini digunakan untuk penyempurnaan produk.

Pedoman penskoran yang digunakan dalam penelitian menggunakan instrumen *Two-tier Multiple Choice* ini mengacu pada pedoman penskoran Beyrak (2013) yang telah diadaptasi, yaitu seperti pada Tabel 1.

Tabel 1. Pedoman Penskoran Instrumen *Two-tier Multiple Choice*

Kriteria	Skor
Tidak ada jawaban	0
Menjawab lebih dari satu	0
Satu Jawaban benar pada <i>Second Tier</i>	0
Satu Jawaban benar pada <i>First Tier</i>	1
Dua Jawaban benar pada <i>First and Second tier</i>	2

HASIL DAN PEMBAHASAN

A. Hasil Studi Pendahuluan

1. Hasil Wawancara Guru Mata Pelajaran

Berdasarkan hasil wawancara yang dilakukan dengan guru kimia di SMA N 1 Surakarta dan SMA N Sragen Bilingual Boarding School (SBBS) pembelajaran kimia di masing-masing sekolah sudah menerapkan berbagai macam pendekatan seperti ekspositori, inkuiri, dan kooperatif. Hal ini dimaksudkan guna memberikan variasi dalam proses pembelajaran dan untuk memacu siswa agar turut aktif dalam proses belajar mengajar. Dari pengamatan yang dilakukan oleh guru, materi kelarutan dan hasil kali kelarutan merupakan materi yang dianggap sulit oleh siswa. Kedua sekolah ini memiliki cara yang berbeda

untuk melakukan penilaian. SMA N 1 Surakarta lebih sering menggunakan soal jenis *essay* untuk menilai pengetahuan siswanya, karena soal jenis ini lebih dapat mengungkap pengetahuan yang dimiliki siswa secara lebih mendalam dan soal jenis ini dianggap dapat mengurangi tindakan curang dari peserta didik. Sedangkan SMA N SBBS lebih sering menggunakan soal jenis pilihan ganda, karena mempermudah proses penilaian dan mengurangi subjektivitas dalam penilaian.

2. Hasil Analisis Soal Ulangan Tengah Semester

Analisis terhadap soal Ulangan Tengah Semester yang digunakan di masing-masing sekolah mendapatkan hasil untuk SMA N 1 Surakarta instrumen yang digunakan memiliki komposisi jenjang 2,8% C1, 37,2% C2, dan 60% C3 sedangkan untuk SMA N SBBS memiliki komposisi 3,3% C1, 56,7% C2, dan 40% C3, pada jenjang ranah kognitif Bloom.

3. Hasil Angket Kebutuhan Siswa

Responden pengisi angket ini adalah 86 siswa kelas XI IPA yang dipilih secara acak dari berbagai kelas di SMA N 1 Surakarta dan SMA N SBBS. Hasilnya menunjukkan sebanyak 91,8% siswa memilih *Higher Order Thinking Skills* yang digunakan sebagai acuan untuk evaluasi hasil belajar dan sebanyak 87,2% responden menginginkan adanya pengembangan soal *Higher Order Thinking Skills* dengan alasan secara umum dari responden soal yang mengacu pada keterampilan berpikir tingkat tinggi, baik untuk siswa, karena dapat merangsang kemampuan siswa untuk berpikir tingkat tinggi sehingga siswa dapat melatih kemampuan berpikir kreatif dan kritisnya juga dapat mengembangkan kemampuan dan pengetahuan mereka dalam bidang ilmu kimia.

B. Hasil desain Pengembangan Produk

1. Perencanaan (*Planning*)

Perencanaan pengembangan instrumen meliputi perencanaan tujuan tes, perencanaan kawasan tes,

perencanaan kompetensi yang akan diuji, perencanaan kisi-kisi, dan perencanaan pembuatan produk awal yang terdiri dari 25 butir soal instrumen *Two-tier Multiple Choice* (TTMC).

2. Mengembangkan bentuk awal produk (*Develop of the Preliminary Form of the Product*)

Pengembangan bentuk awal produk instrumen penilaian *Two-tier Multiple Choice* untuk mengukur keterampilan berpikir tingkat tinggi meliputi pembuatan kisi-kisi soal, pembuatan 25 item soal, pembuatan jawaban dari soal yang telah dibuat, dan rubrik penilaian soal. Sebelum memasuki tahap berikutnya, produk yang telah dikembangkan terlebih dahulu divalidasi oleh 2 orang ahli yaitu 1 orang ahli materi dan 1 orang ahli pembelajaran atau praktisi di sekolah. Hasil validasi ini digunakan sebagai landasan revisi produk awal. Hasil perhitungan *Content Validity* disajikan pada Tabel 2.

Tabel 2. Perhitungan *Content Validity*

Jumlah Soal	Panelis 1		Panelis II	
	Valid	Tidak Valid	Valid	Tidak Valid
25	25	0	25	0
<i>Content Validity</i> = 1 (> 0,700 maka analisis dapat dilanjutkan)				

C. Hasil Uji Coba Produk

Uji coba ini dimaksudkan untuk menguji keterbacaan dan performa dari produk instrumen penilaian dengan bentuk *Two-tier Multiple Choice*. Uji coba dilakukan sebanyak 3 kali dengan responden yang berbeda tiap ujinya. Tahap uji yang pertama yaitu uji coba dan revisi produk awal (*Primary Field Test and Product Revision*). Tahap uji yang kedua yaitu uji coba lapangan utama dan revisi (*Main Field Test and Revision*). Tahap uji yang ketiga yaitu uji coba operasional (*Operational Field Test and Revision*). Dari ketiga uji coba tersebut dilakukan analisis terhadap

3. Daya Pembeda

Hasil pengujian daya pembeda terhadap instrumen penilaian *Two-tier Multiple Choice* disajikan dalam Gambar 1.

kualitas instrumen yang diujikan. Kualitas instrumen yang dianalisis meliputi:

1. Reliabilitas

Pada penelitian ini pengujian reliabilitas menggunakan program AnatesV4. Hasil pengujian reliabilitas instrumen *Two-tier Multiple Choice* disajikan dalam Tabel .

Tabel 3. Hasil Pengujian Reliabilitas Instrumen penilaian *Two-tier Multiple Choice*

Tahap Uji	Nilai Reliabilitas	Keterangan
<i>Primary Field Test</i>	0,90	Sangat Tinggi
<i>Main Field Test</i>	0,93	Sangat Tinggi
<i>Operational Field Test</i>	0,93	Sangat Tinggi

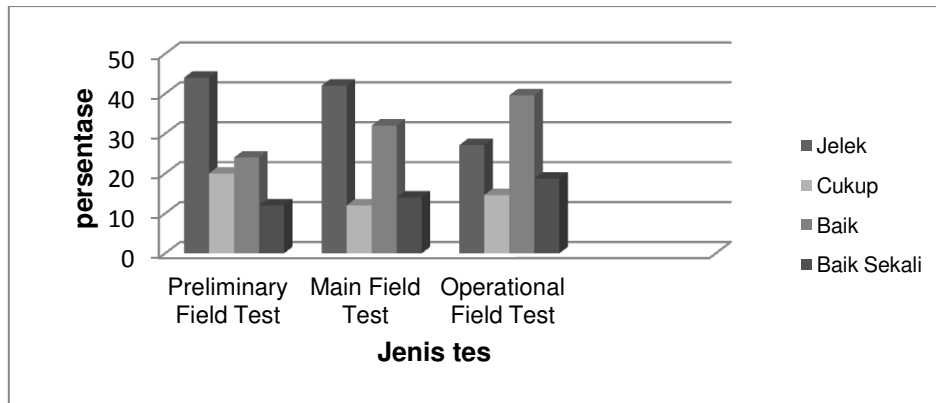
Tabel 2 menunjukkan nilai reliabilitas dari ke tiga tahap uji berada pada rentang 0,80-1,00 yang dikategorikan sangat tinggi, artinya instrumen penilaian *Two-tier Multiple Choice* memiliki keajegan atau konsistensi yang sangat tinggi.

2. Tingkat Kesukaran

Tingkat kesukaran diperlukan untuk mengetahui seberapa sukar instrumen yang diujikan berdasarkan hasil tes yang dikerjakan oleh siswa. Perhitungan tingkat kesukaran pada penelitian ini menggunakan program Anates V. Hasil pengujian tingkat kesukaran terhadap instrumen penilaian *Two-tier Multiple Choice* disajikan dalam Tabel 4.

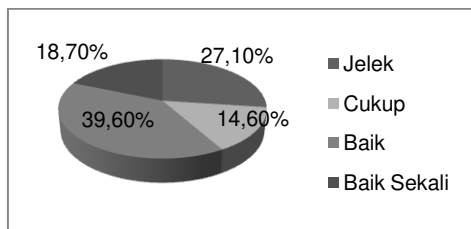
Tabel 4 Persentase Tingkat Kesukaran

Kategori	Persentase Tingkat Kesukaran		
	<i>Preliminary Field Test</i> (%)	<i>Main Field Test</i> (%)	<i>Operational Field Test</i> (%)
Mudah	16	12	4,1
Sedang	64	62	79,2
Sukar	20	26	16,7



Gambar 1. Grafik Persentase Daya Pembeda

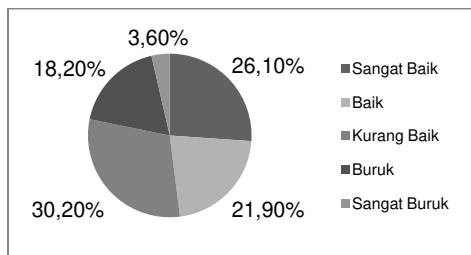
Grafik pada Gambar 1 menunjukkan bahwa persentase soal yang memiliki daya pembeda berkualitas jelek, dari mulai *Preliminary field test*, *main field test*, dan *operational field test*, memberikan penurunan yang cukup signifikan. Berbeda halnya dengan persentase soal dengan kualitas daya pembeda baik, dan baik sekali yang mengalami kenaikan dari setiap tes yang dilakukan. Persentase daya pembeda pada *operational field test*, disajikan dalam Gambar 2.



Gambar 2. Persentase daya pembeda pada *operational field test*

4. Indeks Pengecoh

Kualitas Indeks pengecoh dari hasil tes pada tahap *Operational Field Test* disajikan dalam Gambar 3.



Gambar 3. Persentase Indeks Pengecoh

D. Pembahasan

1. Kualitas Tes

Kualitas suatu tes dapat ditentukan dengan memperhatikan beberapa indikator kualitas soal yang digunakan dalam tes, yaitu validitas, reliabilitas, daya pembeda, tingkat kesukaran, dan indeks pengecoh. Dalam penelitian pengembangan instrumen penilaian *Two-tier Multiple Choice* (TTMC) yang dilakukan, telah dilakukan analisis terhadap indikator kualitas soal tersebut.

Uji validitas yang telah dilakukan menunjukkan hasil bahwa nilai validitas isi (*Content Validity*) dari instrumen penilaian TTMC adalah 1, karena validitasnya > 0,700 maka instrumen soal dapat digunakan. Hal ini juga menunjukkan bahwa instrumen ini relevan untuk mengukur variabel yang diteliti.

Uji reliabilitas yang telah dilakukan sebanyak tiga kali menunjukkan hasil 0,90, 0,93, dan 0,93. Nilai reliabilitas dari ke tiga tahap uji berada pada rentang 0,80-1,00 yang dikategorikan sangat tinggi, artinya instrumen penilaian *Two-tier Multiple Choice* memiliki keajegan atau konsistensi yang sangat tinggi.

Butir soal instrumen penilaian TTMC yang dikembangkan dan diujikan pada 148 siswa dengan tiga kali tahap uji, memberikan respon yang berbeda-beda. Dari 25 butir yang diberikan pada tahap awal yaitu *Preliminary field test*, berdasarkan hasil yang diperoleh dan dengan mempertimbangkan kualitas daya pembeda, tingkat kesukaran, indeks pengecoh, komposisi indikator soal *Higher Order Thinking Skills* (HOTS) dan konsistensi hasil analisis butir soal dari

setiap tes yang dilakukan, sebanyak 21 butir soal layak untuk digunakan, karena memberikan hasil yang baik dan konsisten. Namun 4 butir soal, dinyatakan tidak layak untuk digunakan karena memberikan hasil yang buruk. Keempat butir soal tersebut ialah, butir soal nomor 1, 14, 17, dan 23.

Sehingga tersisa butir soal nomor 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 24, dan 25. Dengan demikian tes yang diberikan akan memberikan kualitas yang baik karena menggunakan butir soal yang berkualitas baik dan layak untuk digunakan.

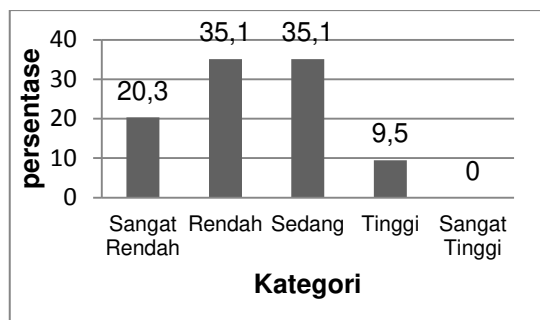
2. Analisis Tingkat Berpikir Siswa

Tingkat berpikir siswa dibagi kedalam 5 kategori, yaitu sangat rendah, rendah, sedang, tinggi, dan sangat tinggi. Tujuan pengelompokan ini adalah guna menempatkan individu kedalam kelompok-kelompok yang posisinya berjenjang menurut suatu kontinum, berdasarkan atribut yang diukur [12]. Adapun pengelompokan siswa kedalam 5 kategori berdasarkan nilai yang diperoleh dari tes pada tahap *operational field test* menggunakan instrumen penilaian *Two-tier Multiple Choice* untuk mengukur keterampilan berpikir tingkat tinggi adalah seperti ditunjukkan dalam Tabel 5.

Tabel 5. Skala Kategori Keterampilan Berpikir Tingkat Tinggi

Skala	Interpretasi
$X \leq 12$	Sangat Rendah
$12 < X \leq 20$	Rendah
$20 < X \leq 28$	Sedang
$28 < X \leq 36$	Tingg
$36 < X$	Sangat Tinggi

Berdasarkan skala kategori keterampilan berpikir tingkat tinggi pada Tabel 5., maka hasil yang didapatkan untuk pengelompokan sebanyak 148 siswa disajikan dalam Gambar 4.



Gambar 4. Grafik Keterampilan Berpikir Tingkat Tinggi

Hasil tes dengan menggunakan instrumen *Two-tier Multiple Choice* yang digunakan untuk mengelompokkan siswa kedalam kategori tingkat berpikir dapat dibandingkan dengan hasil Ujian Tengah Semester (UTS) yang telah dilakukan sebelumnya. Perbandingan ini bertujuan untuk melihat kesamaan dan konsistensi hasil yang diperoleh oleh siswa. Dengan membandingkan dan mengelompokkan siswa yang memiliki nilai terendah dan tertinggi berdasarkan tes dengan menggunakan instrumen *Two-tier Multiple Choice* dan hasil UTS, diperoleh bahwa frekuensi munculnya nama yang sama pada 5 orang siswa dengan kategori terendah dari setiap kelas adalah 43,3% dan frekuensi munculnya nama yang sama pada 5 orang siswa dengan kategori tertinggi dari setiap kelas adalah 40% atau dengan kata lain konsistensi pengukuran yang diberikan instrumen untuk mengukur keterampilan berpikir tingkat tinggi memiliki rata-rata 41,6%. Hasil ini menunjukkan bahwa instrumen penilaian TTMC ini dapat dikatakan cukup berfungsi untuk mengukur keterampilan berpikir tingkat tinggi.

Karakteristik dari instrumen penilaian TTMC ini ialah mengandung dua tingkat pertanyaan yang saling terhubung. Tingkat pertama (*First tier*) berkaitan dengan pernyataan pengetahuan, sedangkan tingkat kedua (*Second tier*) memfasilitasi pengujian siswa belajar di tingkat berpikir yang lebih tinggi. Tujuan dari pertanyaan 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 [13]. Hasil dari analisis pola jawaban siswa pada tes tahap *operational field test* menggunakan instrumen TTMC disajikan dalam Tabel 6

Tabel 6. Persentase Jawaban Siswa terhadap Instrumen TTMC

No Soal	First Tier (%)	Second Tier (%)
1	-	-
2	33,7	28,4
3	52,6	37,9
4	55,8	43,2
5	51,6	46,3
6	35,8	35,8
7	48,4	34,7
8	62,1	50,5
9	66,3	53,7
10	67,4	57,9
11	37,9	35,8
12	52,6	38,9
13	49,5	41
14	-	-
15	53,7	27,4
16	32,6	18,9
17	-	-
18	11,6	9,5
19	51,6	43,2
20	54,7	31,6
21	28,4	21
22	41	5,3
23	-	-
24	55,6	54,7
25	55,6	20

Tabel 6. menunjukkan bahwa siswa lebih dapat menjawab pertanyaan pada *first tier* dibandingkan dengan menjawab pada *first* dan *second tier*. Hal ini membuktikan bahwa siswa masih kesulitan untuk menemukan alasan dari jawaban yang telah mereka pilih. Setiap nomor pada instrumen penilaian TTMC merupakan penjabaran dari indikator keterampilan berpikir tingkat tinggi yang diadaptasi dari Brookhart (2010) [14], pola jawaban pada Tabel 6 memberikan Gambaran tentang kemampuan yang dimiliki siswa ditinjau dari setiap indikator yang ada. Persentase jawaban siswa berdasarkan indikator Keterampilan Berpikir Tingkat Tinggi (*Higher Order Thinking Skills*) disajikan dalam Tabel 7.

Tabel 7. Persentase Jawaban Siswa Berdasarkan Indikator HOTS

Indikator HOTS	No Soal	First Tier (%)	Second Tier (%)
	3	52,6	37,9
Menggunakan keterampilan berpikir analisis	7	48,4	34,7
	9	66,3	53,7
	10	67,4	57,9
	13	49,5	41
Rata-rata		56,8	46,7
	2	33,7	28,4
Menggunakan keterampilan berpikir evaluatif	6	35,8	35,8
	4	55,8	43,2
	8	62,1	50,5
	12	52,6	38,9
Rata-rata		48	39,4
	1	-	-
Menggunakan keterampilan berpikir kreatif	14	-	-
	21	28,4	21
	5	51,6	46,3
	18	11,6	9,5
Rata-rata		30,5	25,6
	15	53,7	27,4
Menggunakan keterampilan berpikir kritis	19	51,6	43,2
	25	55,6	20
	22	41	5,3
	24	55,6	54,3
Rata-rata		51,5	30
Menggunakan keterampilan berpikir logis yang beralasan (<i>Logical Reasoning</i>) untuk memecahkan masalah	16	32,6	18,9
	11	37,9	35,8
	23	-	-
	17	-	-
	20	54,7	31,6
Rata-rata		31,7	28,7

Dari 5 indikator yang ada pada Tabel 7, siswa masih mengalami kesulitan untuk menggunakan keterampilan berpikir kreatif dan keterampilan berpikir logis yang

beralasan (*logical reasoning*) untuk memecahkan masalah, hal ini dibuktikan dengan persentase rata-rata pada kedua indikator ini masih rendah.

Keterampilan berpikir tingkat tinggi siswa dapat ditingkatkan, hasil penelitian yang dilakukan Madhuri, et al. (2012) menunjukkan bahwa pembelajaran berbasis pendekatan inkuiri lebih baik dibandingkan dengan pendekatan konvensional untuk meningkatkan keterampilan berpikir tingkat tinggi pada siswa. Hal ini dapat lebih efektif jika terintegrasi dengan pembelajaran berbasis masalah [15].

Jensen, et al (2014) berpendapat bahwa banyak pendidik yang gagal karena hanya memberikan pertanyaan tentang isi untuk mengetahui keterampilan berpikir siswa, untuk itu harus dibuat pertanyaan yang benar-benar mengukur keterampilan berpikir tingkat tinggi. penilaian tingkat tinggi mungkin faktor kunci dalam mendorong siswa untuk secara efektif memperoleh pemahaman yang mendalam mengenai materi, pemahaman yang mendukung, tidak hanya aplikasi, analisis dan evaluasi, tetapi juga tentang fakta [15].

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. Menurut Piaget tahapan perkembangan adalah kunci untuk perkembangan kognitif. Pada usia sekolah, anak-anak mengembangkan pemikiran operasional yang logis dan sistematis dengan manipulasi simbol. Pada saat remaja yang beranjak dewasa, mereka mengembangkan keterampilannya seperti menggunakan pemikiran yang logis dari simbol yang berhubungan dengan konsep-konsep abstrak, penalaran ilmiah, dan pengujian hipotesis. Keterampilan ini merupakan dasar untuk pemecahan masalah, refleksi diri, dan berpikir kritis [6]. Implikasi dari teori ini pada instrumen yang dikembangkan adalah, 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 sesuai dengan indikator HOTS yang telah dibuat.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan:

1. Instrumen penilaian *Two-tier Multiple Choice* yang dikembangkan dinyatakan layak dan memenuhi kriteria sebagai suatu soal yang baik dengan hasil validitas isi (CV) 1,00, memiliki reliabilitas tes rata-rata 0,92 yang tergolong sangat tinggi, memiliki tingkat kesukaran dengan persentase 4,1% mudah, 79,2% sedang dan 16,7% sukar, memiliki daya pembeda dengan persentase 18,7% sangat baik, 39,6% baik 14,6% cukup, 27,1% jelek, dan memiliki indeks pengecoh dengan persentase 26,1% sangat baik, 21,9% baik, 30,2% kurang, 18,2% buruk, dan 3,6% sangat buruk.
2. Instrumen *Two-tier Multiple Choice* untuk mengukur keterampilan berpikir tingkat tinggi pada materi kelarutan dan hasil kali kelarutan yang dikembangkan cukup mampu mengukur keterampilan berpikir tingkat tinggi dengan angka konsistensi sebesar 41,6%.

DAFTAR RUJUKAN

- [1] Wahyudin, D., Kurniasih, Saripudin, T., Setiasih, O, 2008, *Pengantar Pendidikan*, Jakarta: Penerbit Universitas Terbuka.
- [2] Depdiknas, 2006, *Standar Kompetensi Lulusan untuk Satuan Pendidikan Dasar dan Menengah*, Jakarta: Departemen Pendidikan Nasional, Direktorat Jendral Pendidikan Dasar dan Menengah. Direktorat Pembinaan Pendidikan Menengah Atas.
- [3] Lissa., Prasetyo, Andreas P. B., Indriyanti, Dyah R, 2012, *Pengembangan Instrumen*

- Penilaian Keterampilan Berpikir Tingkat Tinggi Materi Sistem Respirasi dan Ekskresi, *Jurnal Lembar Ilmu Kependidikan*, Vol. 41, No. 1. ISSN 0216-0847
- [4] 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*, Vol. 8, No. 3, 293-307
- [5] 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 MCSER-CEMAS-Sapienza University of Rome*. E-ISSN 2281-4612. ISSN 2281-3993, Vol. 2 No. 2
- [6] 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>
- [7] Beyrak, B. K., 2013, Using Two-Tier to Identify Primary Students' Conceptual Understanding and Alternative Conception in Acid Base, *Mevlana International Journal of Education*, Vol. 3, No. 2, pp. 19-26
- [8] Barak, M., and Dori, Y. J, 2009, Enhancing Higher Order Thinking Skills Among Inservice Science Teachers Via Embedded Assessment. *J Sci Teacher Educ*, Vol. 20: 459-474 DOI 10.1007/s10972-009-9141-
- [9] Cohen, R. J., Swerdlik, M. E., Sturman, E. D, 2013, *Psychological Testing and Assessment*. NY: Mc Graw-Hill
- [10] Sudijono, A, 2005, *Pengantar Evaluasi Pendidikan*, Jakarta: PT Raja Grafindo Persada
- [11] Gregory, R.J, 2007. *Psychological Testing Histor. Principles, and Applications*, United States of America.
- [12] Azwar, S, 2013, *Penyusunan Skala Psikologi*, Yogyakarta: Pustaka pelajar
- [13] Cullinane, A., Liston, M, 2011, *Two-tier Multiple Choice Question (MCQs)-How Effective are they: A Pre-servis Teachers' Perspective*. UK: IOSTE-NW Europe
- [14] Brookhart, S. M, 2010, *How to Assess Higher-Order Thinking Skill in Your Classroom*, Virginia: ASCD
- [15] Madhuri, G. V., Kantamreddi, V. S. S. N., dan Goteti, L. N. S. P, 2012, Promoting higher order thinking skills using inquiry-based learning. *European Journal of Engineering Education*. Vol. 37, No. 2, 117-123
- [16] Jensen, J. L., McDaniel, M. A., Woodard, S. M., and Kummer, T. A, 2014, Teaching to the Test or Testing to Teach: Exams Requiring Higher Order Thinking Skills Encourage Greater Conceptual Understanding. *Educ Psychol Rev* DOI 10.1007/s10648-013-9248-9