

PENGEMBANGAN *TWO-TIER MULTIPLE-CHOICE DIAGNOSTIC INSTRUMENT* PADA MATERI LARUTAN PENYANGGA SISWA SMA NEGERI KELAS XI

Development of a Two-Tier Multiple-Choice Diagnostic Instrument on Buffer Solution Material for XI-Grade Students in State Senior High School

Mahdian* & Muhayar

Program Studi Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan,
Universitas Lambung Mangkurat

Jl. Brigjen H. Hasan Basri, Banjarmasin 70123, Kalimantan Selatan, Indonesia

*email: mahdian_kimia@ulm.ac.id

Informasi Artikel	Abstrak
Kata kunci: Instrumen diagnostic, <i>two-tier multiple choice</i> , kesulitan belajar, larutan penyangga.	Abstrak. Penelitian ini bertujuan untuk mengevaluasi (1) kualitas butir soal pada instrumen tes diagnostik <i>two-tier multiple choice</i> yang dikembangkan dan (2) karakteristik kesulitan belajar siswa yang dapat teridentifikasi melalui instrumen tersebut pada materi larutan penyangga. Uji coba instrumen dilaksanakan melalui dua tahap, yaitu uji coba lapangan awal dan uji coba lapangan pada beberapa sekolah menengah atas. Teknik pengumpulan data meliputi penggunaan instrumen tes diagnostik <i>two-tier multiple choice</i> , wawancara, serta angket tanggapan siswa. Analisis kualitas instrumen dilakukan berdasarkan kriteria validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh. Hasil penelitian menunjukkan bahwa dari 31 butir soal yang dikembangkan, sebanyak 20 soal dinyatakan layak digunakan sebagai instrumen tes diagnostik setelah melalui proses analisis dan revisi. Instrumen memiliki reliabilitas tinggi dengan koefisien berkisar antara 0,62–0,77, tingkat kesukaran bervariasi dari kategori mudah hingga sulit, serta daya beda yang tersebar dari kategori sangat baik hingga jelek. Sebagian besar pengecoh berfungsi dengan baik, meskipun beberapa masih perlu perbaikan. Instrumen ini mampu mengidentifikasi kesulitan belajar siswa baik pada aspek konseptual maupun algoritmik, sehingga dapat digunakan sebagai alat diagnostik yang efektif untuk mendukung perbaikan proses pembelajaran kimia.
Keywords: <i>diagnostic instrument, two-tier multiple choice, learning difficulties, buffers</i>	Abstract. This study aimed to evaluate (1) the item quality of a developed <i>two-tier multiple-choice diagnostic test</i> and (2) the characteristics of students' learning difficulties identified through the instrument in the buffer solution topic. The instrument was implemented in two phases, namely a preliminary field test and a main field test conducted in several senior high schools. Data were collected using the <i>two-tier diagnostic test</i> , interviews, and student response questionnaires. Item analysis was carried out based on validity, reliability, difficulty index, discrimination power, and distractor effectiveness. The findings revealed that, out of 31 developed items, 20 were considered appropriate for use as a diagnostic instrument after undergoing analysis and revision. The instrument demonstrated high reliability, with coefficients ranging from 0.62 to 0.77. The difficulty levels varied from easy to difficult,

Copyright © JCAE- Jurnal Tugas Akhir Mahasiswa

How to cite: Mahdian & Muhayar (2025). PENGEMBANGAN *TWO-TIER MULTIPLE-CHOICE DIAGNOSTIC INSTRUMENT* PADA MATERI LARUTAN PENYANGGA SISWA SMA NEGERI KELAS XI. JCAE (Journal of Chemistry And Education), 9(2), 56-68.

while discrimination indices ranged from very good to poor. Most distractors functioned effectively, although several required refinement. Furthermore, the instrument was able to identify students' learning difficulties in both conceptual understanding and algorithmic problem solving. Therefore, the developed test can be regarded as a valid and reliable diagnostic tool that supports teachers in detecting learning obstacles and improving chemistry instruction.

PENDAHULUAN

Ilmu kimia merupakan ilmu yang mempelajari materi dan sifat-sifatnya, perubahan materi, serta energi yang menyertai perubahan tersebut, dan merupakan salah satu mata pelajaran yang memiliki banyak konsep kompleks dan abstrak (Maharani dkk., 2012). Kesulitan siswa dalam memahami konsep pada pelajaran kimia menyebabkan kompetensi dasar belum tercapai, meskipun alasannya bervariasi, mulai dari sifat konsep kimia yang abstrak hingga kesulitan penggunaan bahasa kimia (Ozmen, 2004). Untuk memahami kimia dengan baik, siswa harus memahami tiga aspek, yaitu makroskopik, mikroskopik, dan simbolik (Gabel, 1998 dalam Wu et al., 2000). Aspek makroskopik menunjukkan fenomena yang terjadi dalam kehidupan sehari-hari maupun yang dipelajari di laboratorium dalam bentuk yang dapat diamati secara langsung. Aspek mikroskopik merupakan representasi yang menjelaskan fenomena yang diamati sehingga dapat dipahami, misalnya pergerakan elektron, molekul, atau atom. Tingkat mikroskopik ini lebih menekankan pada partikel dan sifat-sifatnya. Aspek simbolik digunakan untuk mewakili fenomena makroskopik melalui persamaan kimia, persamaan matematika, grafik, dan mekanisme reaksi. Oleh karena itu, untuk memahami konsep kimia secara utuh, ketiga aspek representasi kimia tersebut harus disampaikan dalam proses pembelajaran secara terintegrasi dan proporsional (Middlecamp dan Kean, 1985 dalam Jefriadi dkk., 2013).

Peserta didik perlu mengintegrasikan ketiga aspek tersebut untuk memperoleh pemahaman yang utuh sehingga banyak siswa beranggapan bahwa kimia sulit karena kesulitan memahami konsep yang kompleks dan abstrak dapat menimbulkan kesulitan belajar serta miskonsepsi, dan salah satu materi yang sering dianggap sulit adalah larutan penyangga karena materi ini mensyaratkan penguasaan konsep asam-basa dan kesetimbangan kimia (Johnstone dalam Fauziah, 2013). Marsita dkk. (2010) dan Mohtar dkk. (2011) melaporkan kesulitan siswa untuk konsep pada materi larutan penyangga adalah (1) konsep pengertian larutan penyangga (2) konsep perhitungan pH dan pOH larutan penyangga (3) konsep fungsi larutan penyangga dalam tubuh makhluk hidup dan dalam kehidupan sehari-hari. Kurniawan dkk. (2013) juga melakukan penelitian untuk menggali pemahaman siswa dan ditemukan 14 pola miskonsepsi siswa pada materi larutan penyangga. Yunitasari dkk. (2013) melaporkan bahwa terjadi miskonsepsi siswa pada konsep-konsep materi pokok larutan penyangga, terutama pada konsep sifat, komposisi, prinsip kerja, pH, dan peran larutan penyangga pada siswa SMAN 2 Sragen.

Pada umumnya, bentuk instrumen yang digunakan guru untuk mengevaluasi hasil belajar adalah bentuk tes pilihan ganda biasa (*traditional multiple choice*) dan essay. Dalam hal soal pilihan ganda, dipilih dalam hal kepraktisan penggunaan dan pemeriksaan dalam penskoran hasil belajar siswa namun pada bentuk tes seperti ini memiliki kesempatan bagi siswa untuk menebak jawaban (*chance of guessing*) sedangkan menggunakan soal bentuk essay, guru bermaksud untuk mengetahui pemahaman siswa dan penyelesaian namun pada pengoreksian hasil jawaban siswa cenderung menjadi sulit apabila jumlah siswa banyak (Arifin, 2009). Selain dari alasan di atas, masih langkanya instrumen tes untuk mengidentifikasi kesulitan belajar

pada materi larutan penyangga yang terstandar. Oleh sebab itu perlu dikembangkan suatu tes untuk dapat menganalisis kesulitan belajar yang dialami siswa.

Salah satu jenis tes instrumen diagnostik yang banyak digunakan dalam penelitian pendidikan adalah instrumen two-tier yang dikembangkan oleh Treagust (Tüysüz, 2009). Dalam instrument two-tier terdapat dua bagian, bagian pertama berisi pertanyaan yang mengandung berbagai pilihan jawaban, bagian ke dua berisi alasan-alasan yang mengacu pada jawaban-jawaban yang terdapat pada bagian pertama. Hal ini menjadikan instrumen diagnostik lebih efektif dalam memberikan pengetahuan sebagai alasan yang mendasari jawaban siswa (Tan dkk., 2005).

Tes diagnostik two-tier telah terbukti dapat mendeteksi kesulitan belajar, miskonsepsi, kesalahan pemahanan, serta pemahaman konseptual yang terjadi pada berbagai materi kimia seperti hasil kali kelarutan (Maharani dkk., 2012), kalor (Tiyas dkk., 2013), reaksi kimia (Chandrasegaran, dkk., 2007), energi ionisasi (Tan dkk., 2005) dan larutan asam basa (Bayrak, 2013; Fauziah, 2013). Berdasarkan hal-hal yang telah diuraikan, maka peneliti melakukan penelitian mengenai “Pengembangan Two-Tier Multiple Choice Diagnostic Instrument untuk Menganalisis Kesulitan Belajar Kimia pada Materi Larutan Penyangga terhadap Siswa SMA Negeri Kelas XI di Kota Banjarmasin”.

METODE PENELITIAN

Penelitian ini menggunakan metode penelitian dan pengembangan (research and development), yaitu metode yang bertujuan menghasilkan produk tertentu sekaligus menguji keefektifannya (Sugiyono, 2013). Model pengembangan yang digunakan mengacu pada langkah-langkah Borg and Gall (1989 dalam Sukmadinata, 2013) yang meliputi sepuluh tahap: (1) penelitian dan pengumpulan data, (2) perencanaan, (3) pengembangan draf produk, (4) uji coba lapangan awal, (5) revisi hasil uji coba, (6) uji coba lapangan utama, (7) penyempurnaan produk, (8) uji pelaksanaan lapangan, (9) penyempurnaan produk akhir, dan (10) diseminasi serta implementasi. Penelitian ini dibatasi sampai tahap ketujuh, yaitu penyempurnaan produk, agar ruang lingkup penelitian fokus diarahkan pada pengembangan serta validasi produk awal.

Desain uji coba dilakukan dengan 2 tahap yaitu tahap validasi ahli dan tahap uji coba produk. Validator dalam penelitian ini berjumlah 6 orang. Uji coba produk dilakukan 3 kali yakni uji keterbacaan soal, uji lapangan awal dan uji lapangan. Subjek penelitian ini yaitu siswa kelas XI pada tahun ajaran 2014/2015 yang berasal dari tiga sekolah berbeda yaitu SMAN 7, SMAN 4 dan SMAN 13 Banjarmasin. Instrumen dalam penelitian ini adalah wawancara, angket respon siswa, dan instrumen tes tertulis yang digunakan adalah tes diagnostik pilihan ganda dua tingkat (*two tier multiple choice*). Analisis data dilakukan dengan mencari kualitas butir tes berupa validitas dan reliabilitas. Validitas yang digunakan adalah Content Validity Ratio (CVR) dan uji reliabilitas soal yang digunakan adalah teknik *Kuder-Richarson* atau KR_{20} . Kualitas butir soal berupa tingkat kesukaran, daya beda dan efektivitas pengecoh. Kriteria penafsiran tingkat kesukaran soal dan daya pembedannya adalah sebagai berikut:

Tabel 1. Tingkat kesukaran soal

Tingkat Kesukaran Soal	Kategori
$p > 0,70$	Mudah
$0,30 \leq p \leq 0,70$	Sedang
$p < 0,30$	Sulit

(Arifin, 2009)

Tabel 2. Penafsiran indeks daya pembeda

Indeks Daya Pembeda	Kategori
$0,40 < D$	Sangat baik
$0,30 < D \leq 0,40$	Baik
$0,20 < D \leq 0,30$	Cukup
$D \leq 0,20$	Jelek

(Arikunto, 2009)

Suatu pengecoh dapat dikatakan berfungsi dengan baik apabila jika paling sedikit dipilih oleh 5% pengikut tes (Arikunto, 2012). Analisis kesulitan siswa dilakukan pertama kali adalah memberikan skor pada jawaban siswa. Pedoman penskoran yang digunakan dalam penelitian menggunakan instrumen *Two-tier Multiple Choice* ini mengacu pada pedoman penskoran Bayrak (2013) yang telah diadaptasi, yaitu seperti pada Tabel 5.

Tabel 3. Pedoman Penskoran Instrumen *Two-tier Multiple Choice*

Kriteria	Skor
Jawaban benar dan alasan benar	1
Jawaban benar dan alasan salah	0
Jawaban salah dan alasan benar	0
Jawaban salah dan alasan salah	0

(Jauhariansyah, 2014)

Setelah dilakukan penskoran kemudian dilakukan pengkategorian terhadap pemahaman siswa dengan kategori berikut:

Tabel 4. Kemungkinan pola jawaban siswa dan kategorinya

No	Pola Jawaban Siswa	Kategori Tingkat Pemahaman
1	Jawaban inti tes benar – alasan benar	Memahami (M)
2	Jawaban inti tes benar – alasan salah	Miskonsepsi (Mi-1)
3	Jawaban inti tes salah – alasan benar	Miskonsepsi (Mi-2)
4	Jawaban inti tes salah – alasan salah	Tidak memahami (TM-1)
5	Jawaban inti tes salah – alasan tidak diisi dan sebaliknya	Tidak memahami (TM-2)
6	Tidak menjawab inti tes dan alasan	Tidak memahami (TM-3)

(Salirawati, 2010 dalam Maunah, 2014)

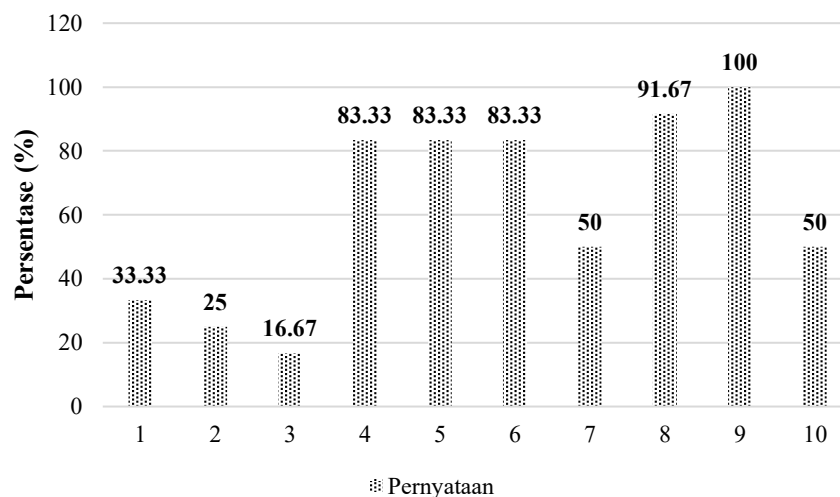
HASIL PENELITIAN DAN PEMBAHASAN

Kualitas butir tes dan butir soal

Tahap awal penelitian ini meliputi penelitian pendahuluan, pengumpulan data, dan perencanaan. Tahap tersebut terdiri atas dua kegiatan, yaitu studi kepustakaan dan studi lapangan. Studi kepustakaan dilakukan untuk memperoleh literatur pendukung serta menentukan SK/KD dan indikator, sedangkan studi lapangan dilakukan melalui wawancara dengan guru mata pelajaran di tiga sekolah lokasi uji coba. Hasil wawancara menunjukkan bahwa siswa mengalami kesulitan

pada materi yang bersifat teoretis, khususnya dalam mengidentifikasi asam–basa dan pasangan asam–basa konjugasi. Selain itu, siswa juga mengalami kesulitan pada materi algoritmik, yang mengindikasikan bahwa kemampuan matematika dasar masih rendah. Tahap selanjutnya adalah pengembangan draf awal instrumen. Penyusunan draf didasarkan pada taksonomi Bloom dengan memilih empat tingkat kognitif awal, yaitu C1 (mengingat), C2 (memahami), C3 (mengaplikasikan), dan C4 (menganalisis), untuk dimasukkan dalam tes diagnostik two-tier. Pemilihan tingkat kognitif tersebut mengacu pada penelitian terdahulu oleh Marsita (2010) serta didukung hasil observasi peneliti yang menunjukkan bahwa siswa masih mengalami kesulitan dalam menjawab soal pada tingkat C4.

Soal yang telah dikembangkan selanjutnya divalidasi, dan hasil validasi menunjukkan bahwa dua butir soal dinyatakan tidak valid karena tidak sesuai dengan indikator serta penggunaan tata bahasa yang kurang tepat. Setelah proses validasi, dilakukan uji coba terhadap soal yang telah direvisi. Uji coba tahap awal berupa uji keterbacaan yang bertujuan mengevaluasi aspek teknis soal, meliputi kejelasan redaksi, ketepatan alokasi waktu, dan kemungkinan adanya kata atau kalimat yang sulit dipahami siswa. Uji keterbacaan dilaksanakan pada 12 siswa kelas XII IPA 1 SMAN 4 Banjarmasin yang telah mempelajari materi larutan penyangga; kelas tersebut berjumlah 36 siswa. Hasil uji coba disajikan pada Gambar 1.



Gambar 1. Hasil tanggapan peserta didik

Keterangan:

1. Terdapat kata-kata atau kalimat yang tidak jelas dalam soal-soal
2. Terdapat istilah-istilah yang tidak Anda mengerti artinya di dalam soal-soal
3. Terdapat Tabel yang sulit dimengerti di dalam soal-soal
4. Terdapat soal yang sulit peserta didik mengerti
5. Terdapat soal yang tingkat kesulitannya terlalu mudah untuk dijawab
6. Terdapat soal yang tingkat kesulitannya terlalu sulit untuk dijawab
7. Terdapat soal yang membingungkan
8. Tes diagnostik *two-tier multiple choice* cocok digunakan untuk mengevaluasi pemahaman siswa
9. Alokasi waktu yang diberikan untuk menjawab semua soal belum sesuai
10. Saran terhadap soal tes diagnostik *two-tier multiple choice*

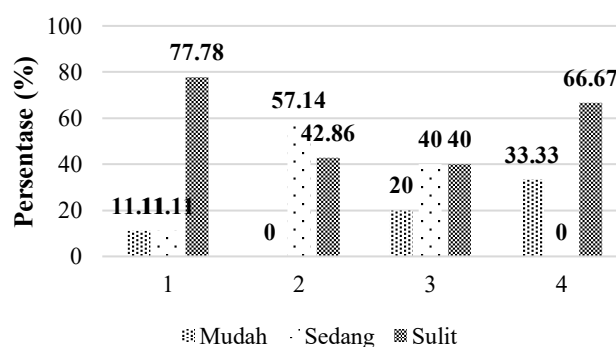
Berdasarkan Gambar 1, hasil tanggapan peserta didik menunjukkan bahwa aspek keterbacaan dan kejelasan redaksi soal secara umum tergolong baik, ditunjukkan oleh persentase rendah pada pernyataan adanya kata atau kalimat yang tidak jelas (33,33%), istilah yang tidak dipahami (25%), serta tabel yang sulit dimengerti (16,67%). Namun demikian, sebagian besar siswa menyatakan bahwa terdapat soal yang sulit dipahami (83,33%), bahkan pada saat yang sama 83,33% siswa menilai ada soal yang terlalu mudah dan 83,33% lainnya menilai ada soal yang terlalu sulit. Temuan ini mengindikasikan adanya variasi tingkat kemampuan peserta didik sekaligus menunjukkan bahwa tingkat kesukaran butir soal belum sepenuhnya merata. Selain itu, 50% siswa menilai beberapa soal masih membingungkan dan 50% memberikan saran perbaikan. Di sisi lain, mayoritas siswa (91,67%) menyatakan bahwa tes diagnostik two-tier multiple choice sesuai digunakan untuk mengevaluasi pemahaman, tetapi seluruh siswa (100%) menyatakan alokasi waktu 90 menit belum mencukupi untuk menyelesaikan 31 soal.

Temuan tersebut sejalan dengan hasil observasi bahwa sebagian siswa belum memahami istilah “akuades”, sehingga peneliti menambahkan keterangan istilah tersebut sebagai tindak lanjut perbaikan. Uji coba lapangan awal ini bertujuan memperoleh data kualitas butir soal meliputi reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh (distractor), yang selanjutnya digunakan sebagai dasar revisi sebelum uji coba pada tiga sekolah. Uji coba lapangan awal dilaksanakan pada kelas XII IPA 2 SMAN 4 Banjarmasin..

Reliabilitas

Instrumen tes diagnostik yang dikembangkan memiliki koefisien reliabilitas sebesar 0,75, yang menunjukkan bahwa instrumen memiliki tingkat konsistensi tinggi sehingga layak digunakan sebagai alat ukur.

Tingkat kesukaran



Gambar 2. Tingkat kesukaran berdasarkan indikator

Indikator 1 : Menganalisis larutan penyangga dan bukan penyangga

Indikator 2 : Menghitung pH atau pOH larutan penyangga

Indikator 3 : Menghitung pH larutan penyangga dengan penambahan sedikit asam atau sedikit basa atau dengan pengenceran

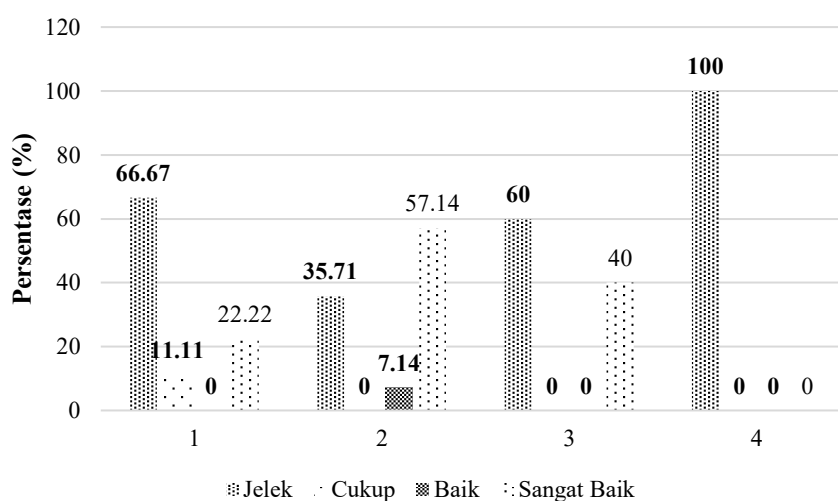
Indikator 4 : Menjelaskan fungsi larutan penyangga dalam tubuh makhluk hidup

Tingkat kesukaran butir soal berdasarkan indikator disajikan pada Gambar 2. Indikator yang diukur meliputi: (1) menganalisis larutan penyangga dan bukan penyangga, (2) menghitung pH atau pOH larutan penyangga, (3) menghitung pH larutan penyangga setelah penambahan

sedikit asam, sedikit basa, atau pengenceran, serta (4) menjelaskan fungsi larutan penyangga dalam tubuh makhluk hidup. Hasil analisis menunjukkan bahwa distribusi tingkat kesukaran bervariasi pada setiap indikator, yang mengindikasikan bahwa instrumen telah mencakup rentang tingkat kesulitan soal dari mudah hingga sulit.

Daya beda

Pengujicobaan 31 soal tes diagnostik *two-tier multiple choice* yang dikembangkan adalah sebagai berikut. Adapun hasil ujicoba berdasarkan indikator adalah sebagai berikut:



Gambar 3. Persentase daya beda berdasarkan indikator

Berdasarkan grafik persentase daya beda tiap indikator, kualitas butir soal masih bervariasi. Indikator 1 didominasi kategori jelek (66,67%) sehingga kemampuan membedakan siswa belum optimal. Indikator 2 menunjukkan kualitas terbaik karena sebagian besar butir berkategori sangat baik (57,14%). Indikator 3 hanya terdiri atas kategori jelek (60%) dan sangat baik (40%), yang menandakan kualitas soal belum merata. Sementara itu, seluruh butir pada indikator 4 berkategori jelek (100%) sehingga memerlukan revisi menyeluruh. Secara umum, hasil ini menunjukkan bahwa perbaikan butir soal masih diperlukan agar instrumen mampu membedakan kemampuan siswa secara lebih efektif.

Pengecoh

Setelah dilakukan pengujicobaan pada kelas XII IPA SMAN 4 Banjarmasin yang terdiri dari 32 orang siswa didapati bahwa masih banyak pengecoh yang tidak berfungsi. Selanjutnya pengecoh yang tidak berfungsi diganti dengan pengecoh baru.

Berdasarkan hasil yang didapatkan pada pengujicobaan keterbacaan dan lapangan awal terhadap siswa kelas XII IPA di SMAN 4 Banjarmasin didapati hasil reliabilitas, tingkat kesukaran, daya beda dan pengecoh seperti yang telah dibahas di atas maka hasil itu menjadi landasan bagi peneliti untuk merevisi produk tes diagnostik yang dikembangkan yang selanjutnya akan diujicobakan terhadap 3 sekolah yaitu SMAN 7, SMAN 4 dan SMAN 13. Berdasarkan hasil yang telah dipaparkan maka dilakukan pengrevisian soal sebagai berikut:

Tabel 5. Kualitas butir soal tes diagnostik yang dikembangkan

Nomor Soal	Validitas	Tingkat Kesukaran	Daya Pembeda	Keterangan
1	Valid	Sulit	Jelek	Dibuang
2	Valid	Sulit	Jelek	Dipakai
3	Valid	Sulit	Jelek	Dibuang
4	Valid	Sulit	Jelek	Dibuang
5	Valid	Sulit	Jelek	Dibuang
6	Valid	Sulit	Sangat Baik	Dipakai
7	Valid	Sedang	Sangat Baik	Dipakai
8	Valid	Sulit	Jelek	Dibuang
9	Valid	Mudah	Cukup	Dipakai
10	Valid	Sulit	Sangat Baik	Dipakai
11	Valid	Sulit	Jelek	Dibuang
12	Valid	Sulit	Baik	Dipakai
13	Valid	Sedang	Sangat Baik	Dipakai
14	Valid	Sedang	Jelek	Dibuang
15	Valid	Sedang	Sangat Baik	Dipakai
16	Valid	Sedang	Jelek	Dibuang
17	Valid	Sedang	Sangat Baik	Dipakai
18	Valid	Sedang	Sangat Baik	Dipakai
19	Valid	Sedang	Jelek	Dibuang
20	Valid	Sedang	Sangat Baik	Dipakai
21	Valid	Sedang	Sangat Baik	Dipakai
22	Valid	Sulit	Jelek	Dibuang
23	Valid	Sedang	Sangat Baik	Dipakai
24	Valid	Mudah	Jelek	Dipakai
25	Valid	Sedang	Sangat Baik	Dipakai
26	Valid	Sedang	Sangat Baik	Dipakai
27	Valid	Sulit	Jelek	Dipakai
28	Valid	Sulit	Jelek	Dipakai
29	Valid	Mudah	Jelek	Dipakai
30	Valid	Sulit	Jelek	Dipakai
31	Valid	Sulit	Jelek	Dipakai

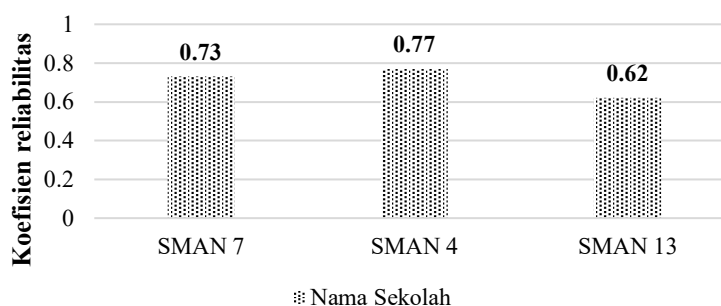
Hasil analisis kualitas butir soal (Tabel 5) menunjukkan bahwa tidak semua soal memenuhi kriteria kelayakan. Beberapa soal dinyatakan gugur karena memiliki daya pembeda jelek meskipun valid secara isi. Setelah proses eliminasi, tersisa 20 soal yang memenuhi kriteria kualitas instrumen. Distribusi soal berdasarkan indikator dan tingkat kognitif (Tabel 6) menunjukkan dominasi pada ranah C3, yang menandakan bahwa instrumen lebih banyak mengukur kemampuan aplikasi konsep dibandingkan hafalan atau pemahaman dasar.

Tabel 6. Kisi-kisi soal setelah eliminasi

Kompetensi Dasar	Indikator	Tingkat Kognitif				Jumlah Soal
		C1	C2	C3	C4	
4.3 Mendeskripsikan sifat larutan penyangga dan peranan larutan penyangga dalam tubuh makhluk hidup	(1) Menganalisis larutan penyangga dan bukan penyangga	1	1	2	0	4
	(2) Menghitung pH atau pOH larutan penyangga	0	0	7	2	9
	(3) Menghitung pH larutan penyangga dengan penambahan sedikit asam atau sedikit basa atau dengan pengenceran	0	1	3	1	5

(4) Menjelaskan fungsi larutan penyangga dalam tubuh makhluk hidup	1	1	0	0	2
Jumlah	2	3	12	3	20

Uji coba lapangan dilakukan pada tiga sekolah untuk memperoleh kualitas akhir butir soal setelah revisi serta menentukan kelayakan instrumen sebagai alat analisis kesulitan belajar siswa. Hasil reliabilitas uji lapangan (Gambar 4) menunjukkan bahwa seluruh koefisien reliabilitas berada pada kategori tinggi (Arifin, 2009), sehingga instrumen dinilai konsisten digunakan pada populasi berbeda.



Gambar 4. Reliabilitas uji coba lapangan

Tafsiran dari ketiga koefisien reliabilitas di atas adalah tinggi (Arifin, 2009).

Data pada Tabel 7 menunjukkan bahwa tingkat kesukaran soal bervariasi pada setiap sekolah. Secara umum terlihat kecenderungan peningkatan tingkat kesukaran dari SMAN 7 ke SMAN 4 hingga SMAN 13. Variasi ini mengindikasikan adanya perbedaan kemampuan akademik antar kelompok sampel, sekaligus menegaskan bahwa instrumen cukup sensitif dalam mendeteksi variasi kemampuan peserta didik.

Tabel 7. Tingkat kesukaran uji coba lapangan

Nomor Soal	Tafsiran tingkat Kesukaran		
	SMAN 7	SMAN 4	SMAN 13
1	Sedang	Sulit	Sulit
2	Sedang	Sulit	Sulit
3	Sulit	Sulit	Sulit
4	Sedang	Sedang	Mudah
5	Mudah	Sedang	Sulit
6	Mudah	Sulit	Sedang
7	Mudah	Mudah	Sedang
8	Mudah	Mudah	Sulit
9	Sedang	Sedang	Sulit
10	Sedang	Sedang	Sedang
11	Sedang	Sedang	Sedang
12	Sedang	Sedang	Sedang
13	Sedang	Sedang	Sedang
14	Sedang	Sedang	Sulit
15	Sedang	Sedang	Sulit
16	Sulit	Sulit	Sulit

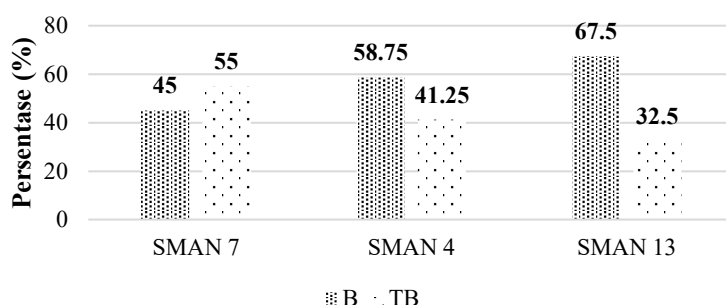
Nomor Soal	Tafsiran tingkat Kesukaran		
	SMAN 7	SMAN 4	SMAN 13
17	Sulit	Sulit	Sulit
18	Sedang	Sulit	Sulit
19	Sulit	Sulit	Sulit
20	Sedang	Sulit	Sulit

Berdasarkan Tabel 8, sebagian besar butir soal memiliki daya beda sangat baik, meskipun masih terdapat beberapa soal dengan kategori cukup atau jelek pada sekolah tertentu. Hal ini menunjukkan bahwa kualitas diskriminatif butir soal dipengaruhi oleh karakteristik responden, namun secara keseluruhan instrumen telah memenuhi syarat sebagai alat evaluasi diagnostik.

Tabel 8. Tafsiran Daya Beda

Nomor Soal	Tafsiran Daya Beda		
	SMAN 7	SMAN 4	SMAN 13
1	Sangat Baik	Sangat Baik	Cukup
2	Sangat Baik	Sangat Baik	Jelek
3	Sangat Baik	Jelek	Jelek
4	Sangat Baik	Sangat Baik	Sangat Baik
5	Sangat Baik	Cukup	Jelek
6	Sangat Baik	Jelek	Jelek
7	Cukup	Sangat Baik	Sangat Baik
8	Sangat Baik	Jelek	Baik
9	Jelek	Baik	Jelek
10	Sangat Baik	Sangat Baik	Sangat Baik
11	Sangat Baik	Cukup	Sangat Baik
12	Sangat Baik	Sangat Baik	Sangat Baik
13	Sangat Baik	Sangat Baik	Sangat Baik
14	Sangat Baik	Sangat Baik	Sangat Baik
15	Cukup	Sangat Baik	Jelek
16	Jelek	Sangat Baik	Cukup
17	Jelek	Sangat Baik	Jelek
18	Sangat Baik	Sangat Baik	Sangat Baik
19	Jelek	Jelek	Jelek
20	Jelek	Sangat Baik	Jelek

Persentase efektivitas pengecoh pada masing-masing sekolah disajikan pada Gambar 5. Hasil analisis menunjukkan bahwa sebagian besar pengecoh telah berfungsi dengan baik, ditandai dengan adanya distribusi pilihan jawaban pada alternatif selain kunci. Pengecoh yang belum efektif menjadi dasar perbaikan lanjutan untuk meningkatkan kualitas instrumen.



Gambar 5. Pengecoh uji coba lapangan

SIMPULAN

Berdasarkan hasil penelitian pada siswa kelas XI IPA di SMAN 7, SMAN 4, dan SMAN 13 Banjarmasin Tahun Pelajaran 2014/2015, dapat disimpulkan bahwa instrumen tes diagnostik two-tier multiple choice yang dikembangkan layak digunakan untuk menganalisis kesulitan belajar pada materi larutan penyangga karena memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas pengecoh. Instrumen memiliki reliabilitas tinggi (0,62–0,77), tingkat kesukaran dan daya beda yang bervariasi, serta pengecoh yang umumnya berfungsi, meskipun masih terdapat beberapa butir yang perlu revisi untuk penyempurnaan.

DAFTAR RUJUKAN

- Arifin, Z. 2009. *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, S. 2009. *Manajemen Penelitian*. Jakarta, Rineka Cipta.
- Arikunto, S. 2012. *Dasar-dasar Evaluasi Pendidikan Edisi 2*. Jakarta, Bumi Aksara.
- Bayrak, B.K. 2013. *Using Two-Tier Test to Identify Primary Students' Conceptual Understanding and Alternative Conceptions in Acid Base*. Mevlana International Journal of Education (MIJE) Vol. 3(2), pp. 19-26
- Chandrasegaran, A. L., David F. T., and Mauro 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 Levels of Representation. *Chemistry Education Research and Practice*, 2007, 8 (3), 293-307
- Cohen, R. J. 2010. *Psychological Testing and Assessment*. New York, McGraw-Hill.
- Fauziah, N.E. 2013. *Pengembangan Instrumen Tes Diagnostik TwoTier Untuk Mengidentifikasi Miskonsepsi Siswa Kelas XI dalam Memahami Materi Larutan Penyangga*. Bandung, Universitas Pendidikan Indonesia.
- Jauhariansyah, S. 2014. *Pengembangan dan Penggunaan Tes Diagnostik Pilihan Ganda Dua Tingkat (Two Tier Multiple Choice) untuk Mengungkap Pemahaman Siswa Kelas X pada Materi Konsep Redoks dan Larutan Elektrolit*. Bengkulu, Universitas Bengkulu.
- Jefriadi, R. S. dan Erlina. 2013. *Deskripsi Kemampuan Representasi Mikroskopik dan Simbolik Siswa SMA Negeri di Kabupaten Sambas Materi Hidrolisis Garam*. Pontianak, FKIP Untan.
- Kurniawan, M.A., Prayitno dan Yahmin. 2013. *Menggali Pemahaman Siswa Sma Pada Konsep Larutan Penyangga Menggunakan Instrumen Diagnostik Two-Tier*. Diakses melalui <http://jurnalonline.um.ac.id/data/artikel/artikeIF5E18132B5AA6AD2873C07F36708438C.pdf>. Pada tanggal 3 November 2014.

- Maharani, T.R P dan Yahmin. 2012. *Menggali Pemahaman Siswa SMA Pada Konsep Kelarutan Dan Hasil Kali Kelarutan dengan Menggunakan Tes Diagnostik Two-Tier*. Diakses melalui http://jurnalonline.um.ac.id/data/artikel/artikel/16DD5292F2EA988FB46FC08A2_F98579A9.pdf. Pada tanggal 11 Oktober 2014.
- Marsita, R.A., Sigit P. dan Ersanghono K. 2010. *Analisis Kesulitan Belajar Kimia Siswa SMA dalam Memahami Materi Larutan Penyangga Dengan Menggunakan Two-Tier Multiple Choice Diagnostic Instrument*. Diakses melalui <http://journal.unnes.ac.id/nju/index.php/JIPK/article/viewFile/1308/1378>. Pada tanggal 11 Oktober 2014.
- Maunah., W.N. 2014. Pengembangan Two-Tier Multiple Choice Diagnostic Test untuk Menganalisis Kesulitan Belajar Siswa Kelas X pada Materi Suhu dan Kalor. *Jurnal Inovasi Pendidikan Fisika (JIPF)*. 195200.
- Mohtar, H., Nurhayati, B., & Mangara, S. 2011. *Identifikasi Kesulitan Siswa Kelas XI IPA SMA Negeri 3 Gorontalo dalam Menyelesaikan Soal-soal Larutan Penyangga*. Diakses melalui http://repository.ung.ac.id/get/simlit_res/1/525/IdentifikasiKesulitan-Siswa-Kelas-XI-IPASMA-Negeri-3-Gorontalo-dalamMenyelesaikan-Soal-SoalLarutan-Penyangga-Penulis-3.pdf. Pada tanggal 19 Oktober 2014.
- Ozmen, H. 2004. *Some Student Misconceptions in Chemistry: A Literature Review of Chemical Bonding*. Journal of Science Education and Technology, Vol. 13, No. 2, June 2004. Diakses melalui <http://content.ebscohost.com/ContentServer.asp?T=P&P=AN&K=16859058&S=R&D=ehh&EbscoContent=dGJyMNHX8kSeqLU4wtvhOLCmr0yepRBSsKm4SbWWxWXS&ContentCustomer=dGJyMOzprkixr69MuePfgeyx44Dt6fIA>. Pada tanggal 27 Desember 2014.
- Sukmadinata, N.S. 2013. *Metode Penelitian Pendidikan*. Bandung, Remaja Rosdakarya.
- Sugiyono, 2013. *Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif dan R&D)*. Bandung, Alfabeta.
- Tan K.C.D., Keith S. Taberb, Ngoh- Khang Goha, dan Lian-Sai Chiaa. 2005. *The Ionisation Energy Diagnostic Instrument: A TwoTier Multiple-Choice Instrument to Determine High School Students' Understanding of Ionisation Energy*. Chemistry education Research and Practice, 180-197. Diakses melalui http://www.rsc.org/images/Tanpaper_tcm1841069.pdf. Pada tanggal 11 oktober 2014
- Tiyas, K.N.RN., Muhardjito, & Kadim M. 2013. *Pengembangan Instrumen Penilaian Diagnostik Bentuk Pilihan Ganda 2 Tingkat untuk Mengetahui Kelemahan Pemahaman Konsep Materi Kalor Siswa Kelas X-7 SMA Laboratorium UM*. Malang, Universitas Malang.
- Tüysüz, C. 2009. *Development of TwoTier Instrument and Assess Students' Understanding in Chemistry*. Scientific Research and Essay, 4 (6): 626-631.
- Wu, Hsin-kai, Krajcik, J. S. & Soloway, E., (2000). *Promoting Conceptual Understanding of Chemical Representations: Students' Use of a Visualization Tool in the Classroom*. Paper presented at the annual meeting of the National Association of Research in Science Teaching. New Orleans, LA.

Yunitasari, W., Endang, S., dan Nanik D.N. 2013. *Pembelajaran Direct Instruction disertai Hierarki Konsep untuk Mereduksi Miskonsepsi Siswa pada Materi Larutan Penyangga Kelas XI IPA Semester Genap SMA Negeri 2 Sragen Tahun Ajaran 2012/2013*. Jurnal Pendidikan Kimia (JPK), Vol. 2 No. 3. Surakarta, Universitas Sebelas Maret.