

PENGEMBANGAN INSTRUMEN PENILAIAN *TESTLET* UNTUK MENGUKUR KEMAMPUAN BERPIKIR TINGKAT TINGGI PADA MATERI ELEKTROKIMIA

Indah Damayanti¹, Mohammad Masykuri², Sri Yamtinah³

¹Program Studi Magister Pendidikan Sains FKIP Universitas Sebelas Maret
Surakarta, 57126, Indonesia
indah-damayanti@kemenperin.go.id

²Program Studi Magister Pendidikan Sains FKIP Universitas Sebelas Maret
Surakarta, 57126, Indonesia
mmasykuri@staff.uns.ac.id

³Program Studi Magister Pendidikan Sains FKIP Universitas Sebelas Maret
Surakarta, 57126, Indonesia
jengtina@staff.uns.ac.id

Abstrak

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian *testlet* yang memiliki validitas, reliabilitas, tingkat kesukaran, daya pembeda dan indeks pengecoh yang memenuhi kriteria sebagai suatu soal yang baik juga untuk mengembangkan instrumen *testlet* pada materi elektrokimia, menjadi instrumen penilaian yang mampu mengukur kemampuan berpikir tingkat tinggi. Pengembangan instrumen penilaian dalam bentuk *testlet* ini menggunakan metode penelitian dan pengembangan yang terdiri dari 10 tahap. Penelitian ini dilakukan di SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, dan SMK SMAK Makassar dengan jumlah responden siswa sebanyak 96 orang. Validasi instrumen penilaian *testlet* ini menggunakan metode Aiken dengan 7 orang validator ahli. Metode penilaian pada instrumen *testlet* ini menggunakan metode *Graded Response Model* (GRM). Hasil penelitian ini adalah instrumen penilaian *testlet* yang dikembangkan dinyatakan layak dan memenuhi kriteria sebagai suatu soal yang baik dengan validitas soal lebih dari 0,76 (valid) dan memiliki reliabilitas tes pada uji coba lapangan utama 0,889 yang tergolong tinggi, memiliki daya pembeda dengan presentase 30,56% cukup dan 38,88% baik dan 30,56% baik sekali. Instrumen penilaian *testlet* ini memiliki tingkat kesukaran dengan presentase 133,88% sukar; 63,89% sedang; 16,67% mudah; dan 5,56% sangat mudah. Instrumen penilaian *testlet* yang dikembangkan mampu mengukur kemampuan berpikir tingkat tinggi siswa dengan persentase kemampuan siswa dalam menganalisis sebesar 61,15%; mengevaluasi 52,50%; mengkreasi 34,25%; berpikir kritis 42,50%; dan bernalar logis 37,25%.

Kata Kunci : Penelitian dan pengembangan, *Testlet*, Instrumen Penilaian, Kemampuan Berpikir Tingkat Tinggi, Elektrokimia.

Abstract

This study aims to develop assessment instrument *testlet* that has validity, reliability, level of difficulty, and distractor that has good criteria, and to develop *testlet* instrument in electrochemistry subject matter that can be measure student's Higher Order Thinking Abilities. Development of *testlet* assessment instrument for measuring student's Higher Order Thinking Skills procedure has 10 steps. The subjects of this research are 352 students of SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, and SMK SMAK Makassar. Validity of this research uses Aiken formula with 7 experts. Assessment method of the *testlet* instrument uses *Graded Response Model* (GRM). The result of *testlet* assessment instrument is fulfill as a good instrument which has validity more than 0.78 (valid), percentage of reliability in main field test is

0.889 (high) and percentage of reliability in operational field test is 0.802 (high), discriminator power percentage is 50% enough and 50% good. Level difficulty percentage is 8.33% difficult, 63.89% enough, 11.11% easy, and 16.67% very easy. Testlet assessment instrument could measured student's Higher Order Thinking Abilities. The student's higher order thinking abilities percentage is analysis ability 70.05%; evaluating ability 54.23%; creating ability 42.66%; critical thinking ability 54.59%; and logical reasoning ability 36.41% at testing operational field.

Keywords: Research and Development, Testlet, Assessment Instrument, Higher Order Thinking Abilities, Electrochemistry.

Pendahuluan

Sistem Pendidikan Nasional merupakan keseluruhan komponen pendidikan yang saling terkait secara terpadu untuk mencapai tujuan pendidikan nasional. Tujuan pendidikan nasional tersebut menjadi parameter utama untuk merumuskan Standar Nasional Pendidikan. Standar Nasional Pendidikan “berfungsi sebagai dasar dalam perencanaan, pelaksanaan, dan pengawasan pendidikan dalam rangka mewujudkan pendidikan nasional yang bermutu”. Standar Nasional Pendidikan terdiri atas 8 (delapan) standar, salah satunya adalah Standar Penilaian yang bertujuan untuk menjamin:

1. Perencanaan penilaian siswa sesuai dengan kompetensi yang akan dicapai dan berdasarkan prinsip-prinsip penilaian;
2. Pelaksanaan penilaian siswa secara profesional, terbuka, edukatif, efektif, efisien, dan sesuai dengan konteks sosial budaya; dan
3. Pelaporan hasil penilaian siswa secara objektif, akuntabel, dan informatif.

Untuk melakukan suatu penilaian, dibutuhkan instrumen penilaian yang sesuai dengan kompetensi yang akan diukur, efektif, efisien, edukatif, objektif dan akuntabel. Bagi siswa SMK, kategori mengingat sampai mengaplikasikan (kemampuan berpikir tingkat rendah) sebagian besar sudah dapat dikuasai, karena pada kategori mengaplikasikan, siswa sudah dibiasakan untuk menerapkan atau menggunakan suatu prosedur. Akan tetapi, untuk kemampuan berpikir tingkat tinggi, siswa harus dilatih

dan dibiasakan. Pembiasaan siswa untuk berpikir tingkat tinggi, salah satunya dapat melalui kegiatan penilaian hasil belajar siswa. Dengan instrumen penilaian yang dapat mengukur kemampuan berpikir tingkat tinggi, siswa akan diajak untuk menganalisis, mengevaluasi, mengkreasi, berpikir kritis, dan bernalar logis.

Menurut King, Goodson, dan Rohani (2004 : 1) kemampuan berpikir tingkat tinggi adalah kemampuan di mana siswa mampu berpikir kritis, logis, reflektif, metakognitif dan kreatif. Kemampuan berpikir tingkat tinggi dapat aktif ketika siswa menemui masalah yang tidak umum, tidak pasti, menjumpai pertanyaan dan mengalami kebingungan atau dilema. Kemampuan berpikir tingkat tinggi meliputi mengingat, menganalisis, membandingkan, menyimpulkan, dan mengevaluasi. Berpikir tidak sama dengan belajar meskipun kedua proses ini berkaitan erat dan saling menunjang. Good & Brophy (Gani, Auliah, & Faika, 2011: 4) berpikir lebih mengarah ke latihan atau penerapan keterampilan kognitif, seperti *posing*, pemusatan perhatian untuk menjawab pertanyaan, *seaching memory*, *processing information and evaluation potensial solution to problems*.

Hasil penelitian yang dilakukan oleh Gani, Auliah, dan Faika (2011: 8), menunjukkan bahwa kemampuan berpikir tingkat tinggi merupakan konsep kunci dalam mencapai keberhasilan dalam belajar. Pada penelitian tersebut diketahui bahwa kemampuan mahasiswa berpikir tingkat tinggi, masih rendah terutama dalam hal masalah yang sedikit

kompleks (rumit). Hasil penelitian Fischer, Bol, dan Pribesh (2011: 5) menunjukkan bahwa berpikir tingkat tinggi masih jarang digunakan di kelas kecil, yaitu kelas untuk mempersiapkan siswa yang akan melanjutkan ke akademi.

Berdasarkan hasil angket analisis kebutuhan guru dan diskusi dengan 4 guru Kimia Fisika dari SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, dan SMK SMAK Makassar dapat diketahui bahwa kemampuan berpikir tingkat tinggi pada siswa SMK belum terasah secara maksimal, sehingga diperlukan instrumen atau alat yang dapat mengasah kemampuan berpikir tingkat tinggi siswa. Alternatif dari instrumen ini adalah berupa instrumen penilaian. Instrumen penilaian yang dikembangkan diharapkan oleh guru berupa soal uraian yang dapat menunjukkan kemampuan siswa yang sebenar-benarnya tanpa melupakan konsep dasar materi.

Dari hasil angket kebutuhan siswa dapat diketahui bahwa 83,33% dari 120 siswa di SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, dan SMK SMAK Makassar menyatakan perlu dikembangkan soal untuk mengukur kemampuan berpikir tingkat tinggi. Selain itu, 77,5% siswa mengungkapkan bahwa soal yang biasa digunakan oleh guru kimia fisika berupa soal uraian. Penggunaan soal bentuk uraian dari guru, didukung dari angket siswa yang menunjukkan bahwa sebanyak 63,33% siswa menyatakan menebak jawaban ketika menemukan pertanyaan yang sulit pada soal pilihan ganda. Dengan demikian, dapat diketahui bahwa siswa akan lebih menggunakan kemampuan berpikirnya apabila menghadapi soal uraian.

Menurut Susongko (2010: 270) bentuk tes uraian lebih kaya informasi berkaitan dengan kemampuan penempuh tes dibanding dengan tes objektif. Bentuk tes uraian memberikan kebebasan kepada

setiap penempuh tes untuk mengekspresikan daya nalarnya sehingga jawaban yang diberikan oleh setiap peserta akan menunjukkan kemampuan berpikir secara kompleks. Bentuk tes uraian ini dapat menunjukkan tingkat penguasaan siswa terhadap materi yang diujikan. Siswa yang tidak menguasai atau tidak kompeten terhadap materi yang diujikan tidak akan dapat menebak jawaban. Akan tetapi, disamping memiliki banyak kelebihan, tes uraian memiliki beberapa kekurangan yaitu, pemberian skor soal uraian membutuhkan waktu yang cenderung lama, dan subjektivitas guru dapat muncul ketika pemberian skor siswa, meskipun dalam pemberian skor sudah dibantu dengan pedoman penskoran dan penskoran dilakukan lebih dari 1 guru ($rater > 1$).

Alternatif untuk bentuk tes yang dapat menjadi pilihan adalah soal tes dalam bentuk *testlet*. *Testlet* adalah suatu kelompok item (pertanyaan) yang berhubungan dengan suatu topik tertentu yang dikembangkan menjadi satu kesatuan dan berisi sejumlah langkah yang telah ditentukan sebelumnya dan yang dapat diikuti oleh peserta. Menurut Wainer, Sireci, dan Thissen (1991: 187) *testlet* termasuk ke dalam jenis super tes yang menghasilkan lebih dari satu respon, lebih lanjut *testlet* ini memiliki respons yang relatif bertingkat (hirarkis) dalam kaitannya dengan pengetahuan (*construct*) yang akan diukur.

Testlet memiliki banyak keuntungan diantaranya, sistem penskoran yang politomus (bukan skor dikotomi seperti soal pilihan ganda), pemberian skor yang tidak membutuhkan waktu terlalu lama, dan sistem penskoran yang hirarki, di mana soal pertama merupakan tahap pengerjaan dasar dari soal kedua dan begitu juga dengan soal kedua merupakan bagian tahapan untuk mengerjakan soal ketiga. Sehingga kemampuan berpikir tingkat tinggi siswa dapat terlihat dari seberapa jauh siswa

tersebut mampu menyelesaikan setiap tahapan soal, dari soal pertama sampai pada soal ketiga.

Menurut Yamtinah, Haryono, dan Martini (2014: 4) Sebuah butir soal inti pada testlet terdiri dari beberapa soal pendukung yang bersifat dependen. Soal pada tahap pertama menjadi dasar bagi soal-soal pada tahap berikutnya sehingga apabila soal tahap pertama dijawab salah oleh siswa, maka siswa tidak akan bisa menjawab benar soal pada tahap berikutnya hingga keseluruhan tahapan. Hal ini akan berpengaruh pada skor yang diperoleh siswa.

Butir-butir soal yang berkaitan akan mempengaruhi skor yang diperoleh siswa, maka proses pemberian skor akan menggunakan *Graded Respon Model* (GRM). De Ayala, Dodd, dan Koch (. Wahyuni, Yamtinah, & Utami, 2015: 228) GRM merupakan salah satu model yang dikembangkan oleh Samejima untuk menangani skoring pada butir-butir soal politomus. Dengan penskoran GRM guru dapat mencermati proses berpikir siswa seperti pada soal uraian, tetapi penskoran instrumen penilaian dalam bentuk testlet lebih efisien karena bersifat objektif dan politomus. Proses penskoran menggunakan *Graded Respon Model* (GRM) adalah sebagai berikut:

Tabel 1. Pedoman Penskoran dengan *Graded Respon Model* dengan 4 kategori

No	Aspek Penilaian	Skor
1	Siswa tidak dapat menyelesaikan langkah atau tahapan pertama dengan benar	0
2	Siswa dapat menyelesaikan langkah atau tahapan pertama dengan benar, tetapi tidak dapat menyelesaikan tahapan kedua	1
3	Siswa dapat menyelesaikan dua langkah atau tahapan pertama dan kedua dengan benar, tetapi tidak dapat menyelesaikan tahap ketiga	2
4	Siswa dapat menyelesaikan keseluruhan tahapan atau langkah dengan benar	3

Dengan menggunakan pedoman penskoran tersebut, instrumen penilaian dalam bentuk *testlet* yang dikembangkan akan dapat digunakan untuk mengukur

kemampuan berpikir tingkat tinggi siswa, sejauh mana tahapan yang dapat diselesaikan siswa. Salah satu contoh butir soal dari instrumen penilaian bentuk testlet adalah sebagai berikut:

Stem:

Pada reaksi redoks, atom atau ion ada yang mengalami reaksi oksidasi dan ada yang mengalami reduksi. Bagian reduksi dan oksidasi masing-masing disebut setengah reaksi. Diketahui persamaan reaksi redoks:



Butir Soal:

- Bilangan oksidasi Mn berubah dari
 - +7 menjadi +4
 - +4 menjadi +7
 - 0 menjadi -1
 - 1 menjadi 0
 - +7 menjadi +2
- Pada reaksi di atas setiap mol MnO_4^- menangkap elektron sebanyak
 - 1 mol
 - 2 mol
 - 3 mol
 - 5 mol
 - 7 mol
- Persamaan reaksi setara dari persamaan tersebut yaitu
 - $6\text{Cl}^- + 2\text{MnO}_4^- + 4\text{H}_2\text{O} \longrightarrow \text{Cl}_2 + 2\text{MnO}_2 + 8\text{OH}^-$
 - $6\text{Cl}^- + 2\text{MnO}_4^- + 4\text{H}_2\text{O} \longrightarrow \text{Cl}_2 + 2\text{MnO}_2 + 8\text{OH}^-$
 - $3\text{Cl}^- + \text{MnO}_4^- + 2\text{H}_2\text{O} \longrightarrow \text{Cl}_2 + \text{MnO}_2 + 4\text{OH}^-$
 - $3\text{Cl}^- + 2\text{MnO}_4^- + 2\text{H}_2\text{O} \longrightarrow 3\text{Cl}_2 + 2\text{MnO}_2 + 4\text{OH}^-$
 - $3\text{Cl}^- + 3\text{MnO}_4^- + 2\text{H}_2\text{O} \longrightarrow \text{Cl}_2 + \text{MnO}_2 + 4\text{OH}^-$

Instrumen penilaian bentuk testlet akan memudahkan guru untuk mengetahui kemampuan siswa dalam menyelesaikan setiap tahapan soal. Tahapan-tahapan soal akan menunjukkan seberapa kemampuan berpikir tingkat tinggi siswa, yaitu kemampuan menggunakan keterampilan berpikir analisis, evaluatif, kreatif, kritis, dan bernalar logis.

Berdasarkan hasil angket analisis kebutuhan siswa dapat diketahui bahwa

66,67% siswa dari 120 siswa kelas XI Kimia Analisis SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, dan SMK SMAK Makassar yang tahun 2016 mendapatkan materi elektrokimia, menyatakan materi elektrokimia merupakan salah satu materi yang membutuhkan kemampuan berpikir tingkat tinggi karena pada materi ini terdapat beberapa topik (misalnya elektrolisis) di mana siswa membutuhkan menganalisis reaksi yang terjadi di anoda dan katoda, menganalisis hasil reaksi yang terjadi, dan mengevaluasi persamaan reaksi yang terjadi di anoda dan katoda. Empat puluh siswa lainnya mempunyai jawaban yang berbeda; 16,66% siswa memilih materi larutan penyangga, 15% siswa memilih materi hidrolisis garam, dan 1,67% siswa memilih materi stoikiometri yang membutuhkan kemampuan berpikir tingkat tinggi. Materi elektrokimia pada bidang keahlian atau jurusan Kimia Analisis bagi siswa SMK, merupakan materi teori pada mata pelajaran Kimia Fisika. Berdasarkan angket dan diskusi dengan guru mata pelajaran kimia fisika, kesulitan siswa dalam berpikir tingkat tinggi untuk materi elektrokimia ini akan sangat terlihat pada saat ujian semester. Kesulitan siswa terlihat pada saat siswa menentukan hasil reaksi elektrolisis dan jumlah endapan atau gas yang diperoleh dari reaksi elektrolisis, yang mana reaksi elektrolisis termasuk dalam materi elektrokimia. Keempat guru yang mengisi angket menyatakan bahwa berpikir tingkat tinggi perlu dilatihkan kepada siswa, dan perlu dikembangkan instrumen penilaian yang dapat mengukur kemampuan berpikir tingkat tinggi siswa. Siswa pun mengungkapkan hal yang sama, bahwa perlu dilakukan pengembangan soal untuk mengukur kemampuan berpikir tingkat tinggi.

Berdasarkan hasil analisis pendahuluan yang telah dilakukan terhadap guru dan siswa, maka tujuan

pengembangan ini adalah untuk mengetahui:

1. Kualitas produk pengembangan yang berupa instrumen penilaian *testlet* untuk mengukur kemampuan berpikir tingkat tinggi pada materi elektrokimia, yang ditinjau dari: 1) validitas: isi, bahasa, dan konstruk; 2) kualitas butir soal yang meliputi; reliabilitas, daya pembeda, tingkat kesukaran, dan indeks pengecoh dari butir soal.
2. Gambaran instrumen penilaian *testlet* dalam menilai atau mengukur kemampuan berpikir tingkat tinggi pada materi elektrokimia di SMK.

Metode Penelitian

Model pengembangan yang digunakan dalam pengembangan instrumen penilaian *testlet* untuk mengukur kemampuan berpikir tingkat tinggi ini adalah *research and development* atau penelitian pengembangan. Produk yang dikembangkan berupa instrumen penilaian *testlet* untuk mengukur kemampuan berpikir tingkat tinggi pada materi elektrokimia.

Langkah-langkah pada penelitian ini merupakan langkah penelitian dan pengembangan yang dikembangkan oleh Borg dan Gall (2007). Kesepuluh langkah tersebut adalah: 1) Penelitian dan pengumpulan data (*research and information collecting*), 2) Perencanaan (*planning*), 3) Pengembangan draf awal (*develop preliminary from product*), 4) Uji lapangan awal (*preliminary field testing*), 5) Revisi hasil uji coba (*main product revision*), 6) Uji lapangan utama (*main field testing*), 7) Penyempurnaan produk hasil uji coba lapangan (*operating product revision*), 8) Uji lapangan operasional (*operational field testing*), 9) Penyempurnaan dan produk akhir (*final product revision*), 10) Desiminasi dan implementasi (*dissemination and implementation*).

Dalam penelitian dan pengembangan instrumen penilaian *testlet* untuk mengukur kemampuan berpikir tingkat tinggi ini hanya sampai langkah ke tujuh dari langkah penelitian dan pengembangan Borg dan Gall yaitu pada langkah penyempurnaan produk hasil uji coba lapangan. Langkah ke delapan sampai sepuluh tidak dilakukan karena pada langkah ini membutuhkan waktu yang lama dan biaya yang cukup besar.

Sampel pada penelitian ini adalah siswa SMK dengan bidang keahlian kimia analisis yang berasal dari SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, dan SMK SMAK Makassar. Keempat sekolah ini dinaungi Kementerian Perindustrian Republik Indonesia, yang terletak di Kota Yogyakarta, Bogor, Padang, dan Makassar. Sekolah-sekolah ini dipilih untuk ketersebaran data, di mana 2 sekolah terdapat di Pulau Jawa, 1 sekolah di Pulau Sumatera dan 1 Sekolah di Pulau Sulawesi.

Instrumen yang digunakan dalam penelitian pengembangan ini yaitu angket analisis kebutuhan guru dan siswa, kisi-kisi instrumen penilaian *testlet* yang dikembangkan, lembar validasi instrumen penilaian *testlet* yang dikembangkan, dan angket uji coba lapangan awal. Pengolahan data dalam penelitian ini dilakukan dengan menggunakan analisis kuantitatif, meliputi validitas instrumen penilaian *testlet* yang dikembangkan, dari segi isi, bahasa, dan konstruk; analisis butir soal, dan analisis kelayakan instrumen penilaian *testlet* yang dikembangkan dalam mengukur kemampuan berpikir tingkat tinggi siswa. Metode pengumpulan data dalam penelitian ini adalah dengan teknik angket untuk mengetahui kebutuhan guru dan siswa, untuk mengetahui keterbacaan instrumen penilaian *testlet* yang dikembangkan, dan pengujian instrumen penilaian *testlet* yang dikembangkan kepada siswa pada uji lapangan utama.

Pada tahap pengembangan draf awal instrumen penilaian *testlet* untuk mengukur kemampuan berpikir tingkat tinggi diperbaiki atau direvisi berdasarkan saran atau masukan dari para ahli. Sebelum diujicobakan instrumen penilaian *testlet* yang dikembangkan divalidasi oleh ahli materi, ahli evaluasi, ahli pembelajaran, praktisi pendidikan, dan rekan sejawat dengan menggunakan formula Aiken. Kriteria yang digunakan adalah jika $V \geq 0,76$ maka butir soal dinyatakan valid dan dapat dilanjutkan analisisnya.

Hasil Penelitian dan Pembahasan

Data hasil penilaian validator terhadap kelayakan instrumen penilaian *testlet* dapat disimpulkan bahwa semua aspek dalam instrumen penilaian *testlet* tersebut telah dinyatakan valid atau telah layak digunakan dengan revisi. Tujuan dari diskusi ini adalah untuk menyepakati indikator kemampuan berpikir tingkat tinggi yang digunakan dalam pengembangan instrumen penilaian *testlet*, menjangkau informasi, dan sebagai tahap awal validasi instrumen yang dikembangkan. Tahap ini merupakan tahap mengembangkan bentuk awal produk (*develop of the preliminary form of the product*).

Pengujian kualitas soal dari segi isi atau validasi isi ini dilakukan oleh 7 orang validator. Diskusi ini menghasilkan masukan dan saran terhadap perbaikan soal yang telah dikembangkan. Sebanyak 36 butir soal yang diuji validitasnya menggunakan formula Aiken menunjukkan hasil lebih dari 0,76; artinya semua soal yang dikembangkan dapat dikatakan valid secara isi. Penelitian Shidiq (2016: 87) dapat diketahui bahwa 30 butir soal instrumen penilaian *testlet* yang dikembangkan pada materi hidrolisis garam untuk mengukur keterampilan proses sains dinyatakan layak dan memenuhi kriteria sebagai soal yang baik

dari hasil validitas isi menggunakan formula Aiken. Validnya instrumen yang dikembangkan secara isi ini merupakan salah satu indikator kualitas instrumen yang dikembangkan. Semakin banyak validator yang memberikan masukan dan saran diharapkan dapat menambah kualitas produk instrumen yang dihasilkan.

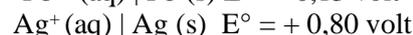
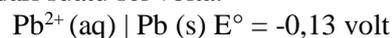
Kualitas instrumen yang dikembangkan dapat ditentukan dengan cara pengujian empiris (empirical analysis). Pengujian empiris pada pengembangan instrumen penilaian ini dilakukan dengan analisis butir soal, yang terdiri dari pengujian reliabilitas instrumen, daya pembeda, tingkat kesukaran dan kualitas pengecoh jawaban soal. Pengujian produk instrumen *testlet* dilakukan sebanyak 2 kali dengan responden yang berbeda setiap ujinya, tahap pertama yaitu uji coba dan revisi produk awal (*preliminary field and product revision*), tahap ini dilakukan dengan memberikan produk awal instrumen penilaian *testlet* pada 16 siswa yaitu 4 siswa secara acak di setiap sekolah (SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, dan SMK SMAK Makassar). Setelah siswa mendapatkan instrumen penilaian *testlet* untuk mengukur kemampuan berpikir tingkat tinggi, kemudian siswa diberikan angket untuk mengetahui respon dan tanggapan mereka terhadap instrumen penilaian *testlet*.

Tahap ini juga bertujuan untuk menguji keterbacaan instrumen, sehingga diharapkan saat memasuki tahap berikutnya instrumen sudah dapat dibaca dan dipergunakan secara jelas. Hasil angket pada tahap uji coba ini menjelaskan bahwa instrumen awal yang diberikan masih terdapat kekeliruan pada soal nomor 10. Hasil ini menjadi dasar untuk merevisi instrumen awal. Tanggapan dan saran siswa secara umum adalah soal nomor 10, kurang lengkap sehingga membingungkan untuk dikerjakan, namun secara umum

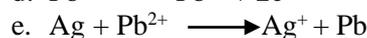
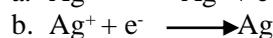
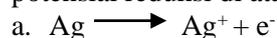
instrumen yang diberikan dapat digunakan sebagai instrumen untuk mengukur kemampuan berpikir tingkat tinggi.

Perubahan pada soal nomor 10 adalah;

Stem: Diketahui data potensial reduksi dari suatu sel volta:

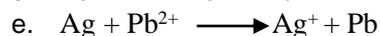
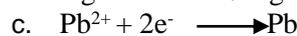
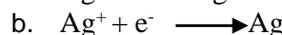
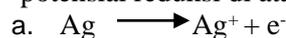


10. Persamaan reaksi reduksi dari data potensial reduksi di atas yaitu



menjadi,

10. Persamaan reaksi reduksi dari suatu sel volta yang dapat terjadi dari data potensial reduksi di atas yaitu



Uji coba selanjutnya adalah uji lapangan utama dan revisi (*main field test and product revision*). Pada tahap ini instrumen yang telah mengalami revisi dari tahap sebelumnya diberikan pada satu kelas (diambil 20 siswa) dari masing-masing sekolah dengan jumlah total 80 orang siswa. Pada tahap ini jawaban dari siswa setelah mengerjakan *testlet* elektrokimia, dianalisis untuk mengetahui kualitas butir soal yang diberikan. Analisis yang dilakukan meliputi reliabilitas, daya pembeda, tingkat kesukaran dan indeks pengecoh. Selain itu, pada tahap ini juga dianalisis kemampuan berpikir tingkat tinggi yang dimiliki oleh siswa. Hasil dari analisis ini digunakan sebagai dasar untuk merevisi produk. Analisis dilakukan secara manual, menggunakan *software microsoft excel*. Reliabilitas butir soal pada uji lapangan utama ini adalah 0,889

yang termasuk dalam kategori sangat tinggi.

Daya pembeda item adalah seberapa baik suatu item dapat membedakan peserta tes dari konstruk yang diukur. Daya pembeda suatu butir soal berfungsi untuk menentukan dapat tidaknya suatu soal membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. Daya pembeda dapat diketahui dengan melihat besar kecilnya angka indeks diskriminasi item.

Pada uji lapangan utama ini, diperoleh daya pembeda 30,56% butir soal termasuk dalam kategori cukup, 38,88% butir soal termasuk kategori baik, dan 30,56% soal termasuk kategori baik sekali. Suatu butir soal tidak dapat membedakan kemampuan siswa dengan baik apabila pengecoh tidak berfungsi dengan baik atau soal terlalu mudah.

Tingkat kesukaran item menyatakan proporsi banyaknya peserta tes yang menjawab dengan benar item tersebut terhadap seluruh peserta tes. Angka yang dapat memberikan petunjuk tingkat kesukaran item tes disebut dengan istilah *difficulty index* yang dilambangkan dengan huruf P yang berarti *proportion* (proporsi). Semakin tinggi nilai P, maka semakin mudah suatu butir soal dan semakin rendah nilai P, maka semakin sukar butir soal tersebut. Pada uji coba lapangan utama, kriteria soal sukar sebanyak 13,88%, soal sedang sebanyak 63,89%, soal mudah sebanyak 16,67% dan soal sangat mudah sebanyak 5,56%. Semakin banyak butir soal dalam kategori sangat mudah, maka daya pembeda butir soal tersebut akan turun, karena baik siswa yang memiliki kemampuan tinggi maupun

rendah, akan menjawab dengan benar pada soal dengan kategori sangat mudah. Hal ini mengakibatkan antara siswa yang memiliki kemampuan tinggi dan rendah sulit untuk dibedakan pada butir soal tersebut.

Analisis butir soal yang berikutnya adalah indeks pengecoh. Indeks pengecoh atau distraktor dapat dikatakan telah menjalankan fungsinya dengan baik apabila indeks pengecoh tersebut telah memiliki daya rangsang atau daya tarik sedemikian rupa sehingga peserta tes (khususnya kemampuan rendah) merasa bimbang dan ragu-ragu sehingga pada akhirnya mereka memilih indeks pengecoh sebagai jawaban betul. Pengecoh dikatakan telah berfungsi dengan baik apabila dipilih lebih dari 5% jumlah peserta tes.

Uji lapangan utama dilakukan dengan memberikan instrumen penilaian *testlet* kepada 80 siswa dari 4 sekolah representatif (masing-masing 20 siswa, dari satu kelas). Data dari uji coba ini berupa analisis butir soal. Revisi pada tahap ini didasarkan pada dua hal tersebut. Data analisis butir soal yang memberikan hasil kurang baik akan direvisi. Tabel 2 menunjukkan beberapa *option* yang memiliki nilai indeks pengecoh 0,00. Hal ini menunjukkan bahwa *option* jawaban tersebut tidak dipilih sama sekali oleh seluruh peserta tes dan menunjukkan bahwa pengecoh tersebut tidak berfungsi dengan baik. Untuk tahap pengujian berikutnya pengecoh yang memiliki nilai 0,00 harus direvisi.

Tabel 2. Data Butir Soal Berkualitas Jelek

No Soal	Daya Pembeda	Tingkat Kesukaran	Indeks Pengecoh (%)				
			A	B	C	D	E
1	0,230	0,925	92,5	5	2,5	0,00	0,00
2	0,282	0,913	3,75	91,25	3,75	1,25	0,00
4	0,403	0,850	85	8,75	2,5	0,00	3,75
6	0,468	0,825	2,5	82,5	10	5	0,00
7	0,247	0,888	88,75	8,75	1,25	1,25	0,00

No Soal	Daya Pembeda	Tingkat Kesukaran	Indeks Pengecoh (%)				
			A	B	C	D	E
8	0,413	0,313	65	31,25	0,00	3,75	0,00
9	0,307	0,838	83,75	2,5	8,75	0,00	5
14	0,629	0,300	30	8,75	1,25	60	0,00
15	0,222	0,450	30	45	6,25	18,75	0,00
18	0,645	0,400	0,00	40	20	28,75	11,25
32	0,574	0,413	41,25	33,75	0,00	17,5	7,5
34	0,450	0,338	32,5	0,00	1,25	33,75	32,5

Tabel 2 menunjukkan bahwa ada enam nomor, yaitu nomor 1, 2, 4, 6, 7, dan 9 pengecoh D atau E tidak ada pemilih sama sekali. Pada beberapa soal yang tergolong sangat mudah dan memiliki kunci jawaban A atau B, dimungkinkan siswa tidak lagi melihat *option* yang lain, sehingga pengecoh tidak diperhatikan atau tidak dibaca lagi oleh siswa. Alternatif solusi untuk fenomena seperti ini adalah kunci jawaban (apabila bukan angka) dapat diletakkan di *option* terakhir atau *option* E.

Pada umumnya, instrumen penilaian hanya berfungsi untuk mengukur kemampuan kognitif peserta tes, tetapi pengembangan instrumen penilaian *testlet* berdasarkan indikator kognitif yang disesuaikan dengan indikator kemampuan berpikir tingkat tinggi. Hal ini memungkinkan instrumen penilaian *testlet* yang dikembangkan dapat mengukur kemampuan berpikir tingkat tinggi yang dimiliki siswa. Berbeda dengan soal pilihan ganda pada umumnya, soal pilihan ganda *testlet* ini memiliki pedoman penskoran yang tidak biasa. Hal ini dimaksudkan agar siswa dapat lebih mengetahui atau memperhatikan proses penemuan jawaban pada soal pilihan ganda.

Instrumen *testlet* yang dikembangkan terdiri dari 12 stem utama, setiap stem memiliki 3 soal pendukung

yang merupakan soal yang hirarkis, artinya siswa tidak akan dapat menjawab soal nomor 3 dengan benar tanpa melalui soal nomor 2, dan siswa tidak akan dapat menjawab soal nomor 2 dengan benar tanpa melalui soal nomor 1. Soal yang hirarkis seperti ini memungkinkan untuk diterapkannya model GRM. Penerapan model GRM pada soal *testlet* ini adalah guru atau pengguna soal dapat mengetahui sejauh mana proses pengetahuan yang dimiliki siswa, yang pada penelitian ini adalah kemampuan berpikir tingkat tinggi siswa. Setiap soal dibuat berjenjang, soal pertama dari setiap stem merupakan soal konsep dasar yang seharusnya mudah dan dipahami siswa. Dua soal berikutnya merupakan soal pengembangan dari konsep utama untuk melatih kemampuan berpikir siswa. Hal ini sesuai dengan pendapat Susongko (2010: 287); Yamtinah, Haryono, dan Martini (2014: 9) yang menyatakan bahwa skor yang didapatkan penempuh atau peserta tes pada item politomus model GRM merupakan penjumlahan skor dari setiap langkah pengerjaan soal. Distribusi soal berdasarkan indikator kemampuan berpikir tingkat tinggi siswa disajikan pada tabel 3. Hasil analisis kemampuan berpikir tingkat tinggi siswa pada uji coba lapangan utama disajikan pada tabel 4.

Tabel 3. Distribusi Soal Berdasarkan Indikator Kemampuan Berpikir Tingkat Tinggi

Indikator	Nomor Soal
Menganalisis	1, 4, 5, 6, 7, 8, 9, 15, 22, 23, 24, 34
Mengevaluasi	2, 3, 19, 20, 31, 35
Mengkreasi	18, 25, 26, 27, 30
Berpikir Kritis	10, 11, 12, 13, 14, 16, 17, 28
Bernalar Logis	21, 29, 32, 33, 36

Tabel 4. Ketuntasan Indikator Kemampuan Berpikir Tingkat Tinggi pada Uji Lapangan Utama

Indikator	Persen Ketuntasan (%)				Total
	SMK SMTI Yogyakarta	SMK SMAK Bogor	SMK SMAK Padang	SMK SMAK Makassar	
Menganalisis	49,17	75,83	42,08	77,5	61,15
Mengevaluasi	48,33	65	30,83	65,83	52,5
Mengkreasi	21	57	0	58	34,25
Berpikir Kritis	23,13	77,5	12,5	56,88	42,5
Bernalar Logis	50	53	0	46	37,25

Tabel 4 menunjukkan persentase pada indikator menganalisis mendapatkan persentase tertinggi di 3 sekolah, kecuali SMK SMTI Yogyakarta. Hal ini membuktikan bahwa kemampuan siswa dalam menganalisis sebuah kasus (soal elektrokimia) yaitu kemampuan dalam menguraikan informasi ke dalam bagian-bagiannya dan menjelaskan alasan yang digunakan untuk menghubungkan bagian satu dengan yang lainnya sudah cukup baik. Pada instrumen yang dikembangkan ini, berpikir analisis atau menganalisis tampak pada kemampuan siswa dalam menentukan bilangan oksidasi, menuliskan persamaan reaksi dari suatu kasus, menyetarakan suatu persamaan reaksi menggunakan perubahan bilangan oksidasi, dan menghitung arus serta massa yang mengendap pada suatu sel elektrolisis.

Persentase terendah dari indikator kemampuan berpikir tingkat tinggi pada indikator bernalar logis pada 3 sekolah, yaitu SMAK Bogor, SMAK Padang, dan SMAK Makassar. Tetapi untuk siswa di SMTI Yogyakarta, persentase pada indikator bernalar logis merupakan persentase tertinggi. Kemampuan bernalar logis merupakan kemampuan dalam menyatakan suatu fakta, suatu kebenaran sesuai dengan penalaran. Kemampuan bernalar logis pada soal yang dikembangkan ini tampak pada kasus reaksi pendesakan, *electroplating*, dan proses korosi besi.

Kemampuan mengevaluasi adalah kemampuan dalam mempertimbangkan dan menilai materi terhadap suatu kriteria.

Pada instrumen yang dikembangkan ini, kemampuan mengevaluasi tampak pada kasus menentukan zat pengoksidasi atau pereduksi, urutan kenaikan daya oksidasi atau reduksi dari beberapa logam dan menyimpulkan pernyataan yang benar atau tidak benar pada proses *electroplating*.

Kemampuan berpikir kritis adalah kemampuan dalam berpikir dengan penuh alasan yang masuk akal dan reflektif yang fokus dalam menentukan keputusan. Pada instrumen yang dikembangkan ini, kemampuan mengevaluasi tampak pada kasus penentuan notasi sel volta dari data potensial reduksi dan penentuan reaksi yang berlangsung di katoda atau anoda.

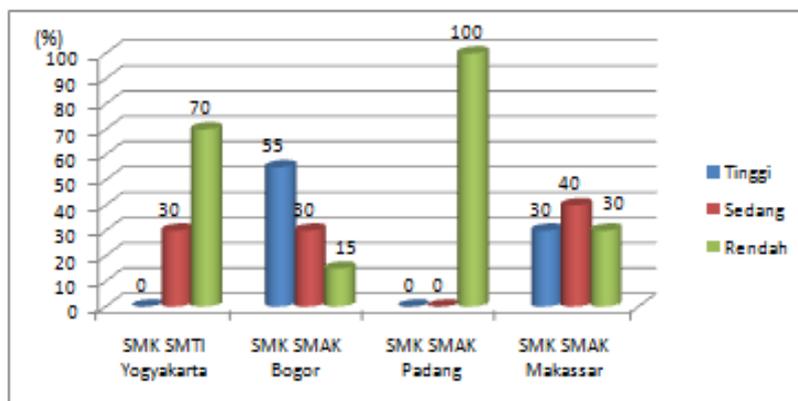
Kemampuan mengkreasi adalah kemampuan dalam menempatkan sesuatu yang tidak sama menjadi satu kesatuan yang baru atau mengorganisasikan kembali sesuatu yang telah ada menjadi sesuatu yang baru. Pada instrumen yang dikembangkan ini, kemampuan mengkreasi tampak pada kasus penentuan volume hasil reduksi atau oksidasi yang terintegrasi dengan hipotesis Avogadro dan penentuan pH setelah elektrolisis.

Dari data pada tabel 4.11 dapat dilihat bahwa kecenderungan siswa di SMK SMTI Yogyakarta menggunakan kemampuan menganalisis, mengevaluasi, dan bernalar logis, tetapi untuk kemampuan mengkreasi dan berpikir kritis masih sangat kurang. Siswa di SMK SMAK Bogor dan SMK SMAK Makassar sudah menggunakan kemampuan berpikir tingkat tingginya, hanya perlu dilatihkan agar persentase ketuntasan siswa dalam kemampuan berpikir tinggi dapat optimal. Siswa SMK SMAK Padang hanya

menggunakan kemampuan menganalisis, mengevaluasi, dan berpikir kritis; untuk kemampuan mengkreasi dan bernalar logis belum digunakan secara optimal.

Hasil pengujian ini dapat digunakan untuk pengkategorian siswa berdasarkan kemampuan berpikir tingkat

tinggi yang dimiliki. Pada penelitian ini pengkategorian skor siswa dibagi menjadi 3 yaitu tinggi, sedang, dan rendah untuk kemampuan berpikir tingkat tinggi yang dimiliki. Data perbandingan skor siswa setiap sekolah disajikan pada Gambar 1.

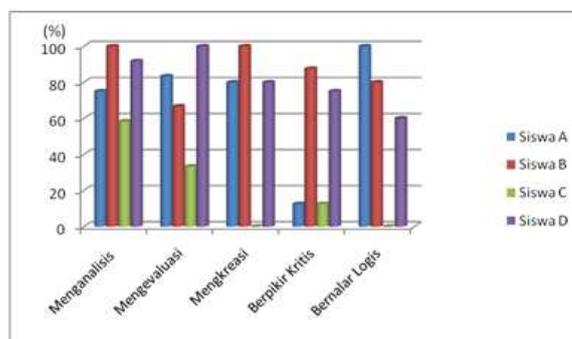


Gambar 1. Perbandingan Skor Siswa pada Uji Lapangan Utama

Pada penelitian ini, kemampuan instrumen penilaian *testlet* dalam mengukur kemampuan berpikir tingkat tinggi pada materi elektrokimia, dianalisis berdasarkan persentase skor capaian siswa pada masing-masing indikator kemampuan berpikir tingkat tinggi siswa. Pada tabel 5, disajikan persen ketuntasan siswa pada masing-masing indikator kemampuan berpikir tingkat tinggi, *sample* 4 siswa yang masing-masing 1 siswa dari masing-masing sekolah.

Tabel 5. Ketuntasan Indikator Kemampuan Berpikir Tingkat Tinggi Siswa (*sample*) Pada Uji Lapangan Utama

Indikator	Ketuntasan (%)			
	Siswa A (SMK SMTI Yogyakarta)	Siswa B (SMK SMAK Bogor)	Siswa C (SMK SMAK Padang)	Siswa D (SMK SMAK Makassar)
Menganalisis	75,00	100,00	58,33	91,67
Mengevaluasi	83,33	66,67	33,33	100,00
Mengkreasi	80,00	100,00	0,00	80,00
Berpikir Kritis	12,50	87,50	12,50	75,00
Bemalar Logis	100,00	80,00	0,00	60,00



Gambar 2. Perbandingan Skor Siswa pada Uji Lapangan Utama

Pada tabel 5 dan gambar 2 dapat diketahui bahwa siswa A perlu peningkatan pada kemampuan berpikir kritis, karena persentasenya paling kecil dibandingkan kemampuan menganalisis, mengevaluasi, mengkreasi, dan bernalar logis, sedangkan siswa B perlu peningkatan pada kemampuan mengevaluasi. Siswa C perlu peningkatan pada kelima indikator kemampuan berpikir tingkat tinggi, yaitu kemampuan menganalisis, mengevaluasi, mengkreasi, berpikir kritis, dan bernalar logis. Siswa D, perlu peningkatan pada

kemampuan bernalar logis. Siswa A menonjol pada kemampuan bernalar logis, sedangkan siswa B menonjol pada kemampuan menganalisis dan mengkreasi, serta siswa D menonjol pada kemampuan mengevaluasi.

Kesimpulan dan Rekomendasi

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa:

1. Instrumen penelitian *testlet* yang dikembangkan dinyatakan layak dan memenuhi kriteria sebagai suatu soal yang baik dengan hasil validitas isi menggunakan formula Aiken didapatkan seluruh soal yang dikembangkan memiliki angka validitas lebih dari 0,78 yang dikatakan valid, memiliki reliabilitas tes pada uji lapangan utama 0,889 yang tergolong sangat tinggi, memiliki daya pembeda dengan persentase 30,56% cukup, 38,88% baik, dan 30,56% baik sekali, serta memiliki tingkat kesukaran 13,88% sukar, 63,89% sedang, 16,67% mudah, dan 5,56% sangat mudah.
2. Instrumen *testlet* untuk mengukur kemampuan berpikir tingkat tinggi pada materi elektrokimia yang dikembangkan mampu mengukur kemampuan berpikir tingkat tinggi siswa dengan persentase kemampuan siswa dalam menganalisis sebesar 61,15%; mengevaluasi 52,5%; mengkreasi 34,25%; berpikir kritis 42,5%; dan bernalar logis 37,25% pada uji coba lapangan utama.

Rekomendasi

Kepada guru: 1) Instrumen penilaian *testlet* untuk mengukur kemampuan berpikir tingkat tinggi pada materi elektrokimia ini digunakan untuk kelas XI Kimia Analisis, sebagai tes formatif.

Kepada peneliti lain: 1) Hasil penelitian dan pengembangan dapat digunakan sebagai acuan untuk penelitian berikutnya yang sejenis dengan materi yang berbeda. 2) Prosedur penelitian dan pengembangan dilanjutkan pada langkah ke-8 sampai ke-10 Borg dan Gall yaitu uji coba pelaksanaan lapangan, revisi akhir, diseminasi dan implementasi. 3) Dibutuhkan kriteria yang pasti dalam pengelompokan item soal ke dalam indikator kemampuan berpikir tingkat tinggi yang telah dibuat.

Daftar Pustaka

- Aiken, L. (1985). Three Coefficient for Analyzing The Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45, 131-142.
- Borg, W.R., & Gall, M.D. (2007). *Educational research*. New York: Longman, Inc
- Fischer C., Bol L., & Pribesh S. (2011). An Investigation og Higher Order Thinking Skills in Smaller Learning Community Social Studies Classrooms. *American Secondary Education Spring Journals*. 39:2.
- King F.J., Goodson L., & Rohani F. (2004). Higher Order Thinking Skills, Definition, Teaching Strategies, Assessment.
- Susongko, P. (2010). Testlet Sebagai Alternatif Pengembangan Bentuk Tes Prestasi Belajar pada Pembelajaran Kimia di Sekolah. *Jurnal Penelitian dan Evaluasi Pendidikan*, 14(2) 269-288.
- Gani T., Auliah A., Faika S., (2011). Penguasaan Pengetahuan Deklaratif dan Kemampuan Berpikir Tingkat Tinggi Mahasiswa Prodi. Pendidikan Kimia. *J. Chemica*. 12:1-9.
- Shidiq, Ari Syahidul. (2016). Pengembangan Instrumen Penilaian Testlet untuk Mengukur Keterampilan Proses Sains pada Materi Hidrolisis Garam

untuk Siswa Kelas XI SMA/MA.
Tesis S2 Fakultas Keguruan dan Ilmu Pendidikan Universitas Sebelas Maret. Surakarta. (Unpublished).

Wahyuni, I.T., Yamtinah S., & Utami B. (2015). Pengembangan Instrumen Pendeteksi Kesulitan Belajar Kimia Kelas X Menggunakan Model Testlet. *Jurnal Pendidikan Kimia*, 4(4) 222-231.

Wainer, H, Sireci, S. G & Thissen, D. (1991). Differential Testlets Functioning: Definition and Detecting. *Journal of Education Measurement*, 24, 185-201.

Yamtinah S., Haryono, & Martini, K.S. (2014). Profil Individu Peserta Didik Pelengkap Tes Jenis Testlet Sebagai Alternatif Pendeteksi Kesulitan Belajar Kimia. *Jurnal Profesi Pendidik*, 1(1) 1-10.