

**KARAKTERISTIK BUTIR SOAL UJIAN SEMESTER DAN KEMAMPUAN SISWA SMA
DI KABUPATEN MUARO JAMBI**

¹⁾Melly Elvira, ²⁾Samsul Hadi

¹⁾Program Studi (S-2) Penelitian dan Evaluasi Pendidikan PPs Universitas Negeri Yogyakarta

²⁾Fakultas Teknik Universitas Negeri Yogyakarta

¹⁾mellyelvira.me@gmail.com, ²⁾samsul.hd@gmail.com

Abstrak

Penelitian ini bertujuan untuk mendeskripsikan: (1) karakteristik butir soal ujian akhir semester mata pelajaran kimia dan (2) kemampuan siswa kelas X SMA di Kabupaten Muaro Jambi. Penelitian ini merupakan penelitian deskriptif evaluatif dengan pendekatan survei. Subjek penelitian adalah siswa kelas X. Objek penelitian adalah perangkat soal dan respon siswa dari enam sekolah yang dipilih dengan teknik *proportional stratified random sampling*. Analisis data menggunakan lembar telaah, program *ITEMAN* pada teori tes klasik, dan program *BILOG-MG* pada teori respon butir. Hasil analisis diperoleh sebagai berikut: (1) telaah butir soal menghasilkan 33 butir berkualitas baik, analisis teori tes klasik menghasilkan 10 butir berkualitas baik, dan analisis teori respon butir menghasilkan 13 butir berkualitas baik. (2) kemampuan siswa berdasarkan teori tes klasik dan teori respon butir, sama-sama menghasilkan SMAN 3 Muaro Jambi dengan kemampuan tertinggi dan SMAN 6 Muaro Jambi dengan kemampuan terendah.

Kata Kunci: *analisis butir, teori tes klasik, teori respon butir, kemampuan siswa*

**THE CHARACTERISTICS OF FINAL EXAM TEST ITEMS AND THE ABILITY LEARNING
OF THE HIGH SCHOOLS STUDENTS IN MUARO JAMBI DISTRICT**

¹⁾Melly Elvira, ²⁾Samsul Hadi

¹⁾Program Studi (S-2) Penelitian dan Evaluasi Pendidikan PPs Universitas Negeri Yogyakarta

²⁾Fakultas Teknik Universitas Negeri Yogyakarta

¹⁾mellyelvira.me@gmail.com, ²⁾samsul.hd@gmail.com

Abstract

This study aims to describe: (1) the characteristic of the items of the final exam chemistry test and (2) the ability of the students of state high schools in Muaro Jambi District. This study was descriptive evaluative with the survey approach. The subject was first grade students. The object was the test kit and all the answer sheets from six senior high schools in Muaro Jambi District which were established using the proportional stratified random sampling technique. The data analysis was done by using sheet judgements, Iteman for Classical Test Theory (CTT) and Bilog-MG for Item Response Theory (IRT). The result of analysis obtained data as follow: (1) Examine item results 33 good items, the analysis by using the CTT results 10 good items, the analysis by using the IRT results in 13 good items. (2) the students' ability according to CTT and IRT result schools with the highest average owned by SMA Negeri 3 Muaro Jambi and the lowest average owned by SMA Negeri 6 Muaro Jambi.

Keywords: *item characteristic, classical test theory, item response theory, students' ability*

Pendahuluan

Pengukuran dan penilaian merupakan dua hal yang tidak terlepas dari sistim evaluasi. Pengukuran merupakan proses pemberian angka dengan membandingkan hasil amatan dengan kriteria yang telah ditetapkan. Hal ini sejalan dengan pendapat Olivo, Cummings, Fuentes & Saltaji (2004, p.1) yang menyatakan bahwa berdasarkan hasil pengukuran maka diperoleh informasi sebagai bahan masukan kepada penyelenggara pendidikan dalam pengambilan keputusan. Proses pengukuran memerlukan alat ukur yang baik, sebab alat ukur ini yang nantinya akan memberikan informasi mengenai gambaran seseorang yang sedang diukur. Namun dalam melakukan pengukuran, alat ukur yang digunakan haruslah valid dan reliabel agar menghasilkan data dan memberikan informasi yang akurat. Data yang dihasilkan nantinya dijadikan acuan dalam melakukan penilaian.

Penilaian dilakukan berdasarkan poin-poin hasil pengukuran. Data hasil pengukuran kemudian dianalisis melalui sebuah prosedur yang sistematis, selanjutnya diinterpretasikan untuk membuat sebuah kesimpulan. Miller, Lynn & Ground (2009, p.29) berpendapat bahwa melalui proses penilaian yang terintegritas dapat ditentukan kemampuan siswa serta memonitor perkembangan dalam mencapai tujuan pembelajaran. Penilaian diberikan sebagai acuan dalam meningkatkan kualitas pembelajaran. Salah satu cara melakukan penilaian dalam pembelajaran adalah dengan melakukan evaluasi hasil belajar siswa.

Evaluasi hasil belajar merupakan salah satu upaya yang dilakukan untuk mengukur kualitas pendidikan. Salah satu cara mengevaluasi hasil belajar yaitu melalui pelaksanaan Ujian Akhir Semester (UAS). Ujian akhir semester merupakan kegiatan yang dilakukan pendidik dalam upaya mengukur pencapaian kompetensi siswa diakhir semester. Hasil ujian yang diperoleh akan menentukan apakah mereka berhasil dalam memahami pembelajaran yang telah diberikan, mampu melanjutkan ketingkat lebih tinggi atau tetap di kelas yang sama.

Alat ukur yang digunakan dalam ujian akhir semester di antaranya menggunakan instrumen tes. Tes merupakan salah satu bentuk alat ukur yang terdiri dari butir-butir pertanyaan untuk mengukur kemampuan siswa dalam mencapai tujuan pembelajaran yang telah dilakukan. Ketepatan dalam melaksanakan peni-

laian ditentukan oleh baik buruknya perangkat tes yang digunakan oleh guru. Jika sebuah perangkat tes memiliki karakteristik yang tidak baik maka hasil penilaian menjadi tidak valid, oleh sebab itu diperlukan kemampuan guru dalam membuat instrumen penilaian yang berkualitas. Hal ini diungkapkan pula oleh Nurung (2008, p.4) dalam penelitiannya yang mengatakan bahwa pengelolaan ujian dan mutu bahan ujian yang digunakan perlu mendapat perhatian agar hasil tes dapat mencerminkan kemampuan siswa yang sebenarnya.

Tes yang baik mampu memperoleh informasi terkait kemampuan siswa, oleh karena itu dalam penulisan butir soal harus secara benar dan sesuai dengan tujuan dilaksanakan tes tersebut. Hal ini dijelaskan pula oleh Mulyana (2007, p.238) yang menyatakan bahwa tes yang baik dapat mengetahui kriteria siswa yang diterima, sehingga sekolah dapat memprediksi keberhasilan siswa yang diterima, untuk itu perlu dihindari kesalahan prediksi akibat kesalahan pengukuran.

Senada dengan yang dijelaskan oleh Boopathiraj & Chellamani (2013, p.192) yang menyimpulkan bahwa ukuran sebuah soal yang diterima bergantung pada tingkat kesulitan dan tujuan tes tersebut dibuat. Tes yang tidak baik sebaiknya dibuang atau direvisi sebelum dimasukkan kembali kedalam tes yang akan diujikan. Rumusan penilaian yang diperoleh dari hasil pengukuran berdasarkan perangkat tes yang digunakan berpengaruh pada berbagai keputusan atau kebijakan pendidikan.

Soal ujian akhir semester menggunakan bentuk tes pilihan ganda. Instrumen tes yang dibuat dalam bentuk butir soal pilihan ganda harusnya memiliki kualitas yang baik dalam tingkat kesukaran butir soal, daya beda, distraktor, validitas maupun reabilitasnya. Jika soal memiliki kualitas yang baik, maka hasil yang diperoleh dapat mengukur ketercapaian kompetensi siswa tiap akhir semester.

Soal ujian akhir semester dibuat oleh beberapa orang guru yang terlibat dalam tim Musyawarah Guru Mata Pelajaran (MGMP). Soal kemudian digunakan di seluruh SMAN di Kabupaten Muaro Jambi. Soal dibuat pada mata pelajaran kimia. Mata pelajaran kimia merupakan salah satu pelajaran yang memiliki karakteristik dan memerlukan keterampilan dalam memecahkan masalah. Hal ini sejalan dengan pernyataan Brown, Lemay & Bursten (1991, p.3) yang menjelaskan bahwa kimia mempelajari tentang zat, terutama mengenai struktur,

komposisi materi dan perubahan-perubahannya. Salah satu tujuan pembelajaran ilmu-ilmu kimia di SMA adalah agar siswa memahami konsep kimia dan keterkaitan serta penerapannya baik dalam kehidupan sehari-hari maupun dengan melibatkan teknologi. Oleh karena itu, diperlukan butir soal yang mampu mengukur kemampuan siswa sesuai dengan tujuan pembelajaran kimia yang diharapkan.

Hasil prasurvei yang dilakukan pada salah satu guru mata pelajaran kimia yang tergabung dalam MGMP, diketahui bahwa soal yang dibuat oleh MGMP belum dianalisis secara empiris. Pemilihan butir soal hanya dilihat dari materi dan penggunaan bahasa, sedangkan untuk validitas dan reabilitasnya tidak dianalisis. Selain itu, guru-guru yang terlibat dalam tim pembuat soal hanya mengambil butir soal dari beberapa sumber buku atau mengambil dari soal-soal ujian tahun sebelumnya. Soal dibuat sedikit berbeda dari soal sebelumnya dengan mengganti beberapa kata kerja operasional dengan kata yang memiliki makna sama. Soal yang berasal dari buku tes yang beredar dipasaran tidak diujicobakan sehingga karakteristik butir soal yang digunakan pada ujian akhir semester tidak diketahui.

Butir soal dikatakan baik apabila telah melalui beberapa tahapan analisis. Analisis butir soal ditelaah terlebih dahulu, kemudian dilanjutkan dengan tahap analisis data berupa uji daya beda, taraf kesukaran dan keberfungsian pengecoh pada tiap butir soal. Butir soal yang memenuhi syarat dapat disimpan untuk menjadi acuan pembuatan soal selanjutnya. Akan tetapi pada kenyataannya pengujian ini belum dilakukan, sehingga karakteristik butir soal buatan guru sering dikatakan rendah karena belum memenuhi syarat-syarat tes yang baik.

Karakteristik butir soal yang tidak diketahui membuat banyaknya kesalahan dari perangkat tes ujian yang digunakan. Kesalahan perangkat tes dapat terjadi karena adanya ketidaksesuaian antara tingkat kesukaran butir soal dengan tingkat kemampuan siswa. Soal-soal yang terlalu sukar dan terlalu mudah akan menghasilkan Standar Baku Pengukuran atau *Standard Error of Measurement* (SEM) yang besar, sehingga kemampuan siswa yang sebenarnya sulit untuk diungkapkan. Besarnya kemampuan siswa terlihat dari besarnya peluang menjawab benar dari sejumlah butir tes. Kemampuan siswa hanya terlihat dari butir yang telah dikalibrasi.

Langkah yang sebaiknya dilaksanakan untuk mengatasi terjadinya kesalahan perangkat tes adalah dengan menganalisis butir soal pada sebuah perangkat tes sebelum diujikan kepada siswa. Analisis butir soal dikerjakan melalui beberapa tahap. Tahapan pertama adalah analisis butir soal berdasarkan telaah butir yaitu lembar telaah yang diisi oleh tiga orang *expert judgement* untuk menilai dari segi konstruksi, materi dan bahasa. Selanjutnya menganalisis berdasarkan pendekatan teori tes klasik dan teori respon butir. Analisis berdasarkan teori tes klasik dilakukan dengan bantuan program *ITEMAN*. Hasil analisis akan menunjukkan daya beda, tingkat kesulitan dan keberfungsian pengecoh pada tiap butir soal yang diujikan.

Langkah selanjutnya menganalisis berdasarkan teori respon butir dengan bantuan program *BILOG-MG*. Hulin, Drasgow, & Parsons (1983, p.14) menyatakan bahwa kemungkinan peserta tes menjawab soal dipahami sebagai fungsi logistik perbedaan parameter yang dimasukkan ke dalam model. Terdapat beberapa model logistik yaitu model satu parameter (1P), dua parameter (2P), dan tiga parameter (3P). Hasil analisis terdiri dari daya beda, tingkat kesulitan dan tebakan semu. Teori respons butir menggunakan distribusi probabilitas dalam bentuk kumulatif. Proporsi menjawab benar tidak hanya menjelaskan butir tes tetapi juga menjelaskan kemampuan kelompok peserta yang dites.

Berdasarkan hasil prasurvei yang telah dilakukan, maka melalui penelitian ini diharapkan peneliti dapat mengetahui kualitas butir soal buatan guru yang digunakan dalam ujian akhir semester melalui hasil analisis telaah dan uji karakteristik butir soal untuk digunakan sebagai tolak ukur dalam melihat kemampuan siswa dalam menghadapi ujian akhir semester dan sebagai acuan dalam menulis soal-soal ujian berikutnya.

Metode Penelitian

Penelitian ini merupakan penelitian deskriptif evaluatif dengan pendekatan survei, yang mendeskripsikan karakteristik butir soal dan kemampuan siswa pada mata pelajaran kimia melalui hasil ujian akhir semester SMA negeri di Kabupaten Muaro Jambi tahun ajaran 2015/2016 berdasarkan hasil telaah *expert judgement*, hasil analisis teori tes klasik dan teori respon butir, serta mendeskripsikan kemampuan siswa masing-masing sekolah.

Penelitian ini dilaksanakan di SMAN se-Kabupaten Muaro Jambi, Provinsi Jambi. Pelaksanaannya dilakukan setelah dilaksanakan ujian akhir semester ganjil tahun pelajaran 2015/2016. Penelitian dimulai dengan prasarvei bulan Juli-Desember 2015 dan dilanjutkan dengan pengambilan data pada bulan Januari-Februari 2016.

Subjek penelitian ini adalah siswa kelas X SMAN Kabupaten Muaro Jambi yang mengikuti ujian akhir semester ganjil mata pelajaran kimia tahun ajaran 2015/2016. Objek penelitian ini adalah perangkat tes dan respon siswa berupa lembar jawaban pada ujian akhir semester ganjil yang berasal dari 6 SMAN di Kabupaten Muaro Jambi. Sebanyak 6 sekolah dipilih dari 12 sekolah yang menggunakan kurikulum KTSP. Sekolah yang menerapkan kurikulum KTSP menggunakan perangkat tes yang dibuat oleh MGMP, sedangkan sekolah yang menggunakan kurikulum 2013 membuat perangkat tes di sekolah masing-masing. Pemilihan 6 sekolah dilakukan dengan menggunakan teknik *Proportional Stratified Random Sampling* yaitu dipilih berdasarkan ranking hasil Ujian Nasional mata pelajaran kimia tahun ajaran 2014/2015. *Ranking* dipilih kembali berdasarkan nilai tertinggi, sedang dan terendah. Sekolah yang menjadi objek penelitian adalah SMAN 3 Muaro Jambi, SMAN 4 Muaro Jambi, SMAN 6 Muaro Jambi, SMAN 10 Muaro Jambi, SMAN 11 Muaro Jambi dan SMAN 12 Muaro Jambi.

Teknik pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi. Dokumen digunakan untuk memperoleh data dalam penelitian ini antara lain: Spesifikasi soal ujian akhir semester mata pelajaran kimia, lembar soal ujian akhir semester mata pelajaran kimia, kunci jawaban ujian akhir semester mata pelajaran kimia, dan lembar jawaban siswa peserta ujian akhir semester mata pelajaran kimia tahun ajaran 2015/2016 di Kabupaten Muaro Jambi.

Selain itu, pengumpulan data juga dilakukan dengan menggunakan kartu telaah. Kartu telaah berupa pedoman penulisan instrumen yang nantinya dijadikan acuan untuk menilai butir soal ujian yang ada. Kartu telaah diisi oleh beberapa ahli untuk menelaah instrumen tes secara kualitatif dari segi bahasa maupun materi. Melalui kartu telaah, instrumen tes dianalisis secara kualitatif.

Teknik analisis data dalam penelitian ini dilakukan dengan beberapa tahapan. Tahapan pertama menganalisis perangkat tes dengan

menelaah butir soal yang dilakukan dari aspek materi atau isi, aspek konstruksi, dan aspek bahasa berdasarkan kaidah-kaidah telaah butir soal yang dikeluarkan oleh Tim Pedoman Umum Pengembangan Penilaian. Telaah dilakukan secara terpisah oleh beberapa orang *expert judgement* yaitu ahli pada bidang studi yang diujikan, ahli bidang pengukuran dan ahli dalam penulisan bahasa. Masing-masing ahli menelaah butir soal dengan mengisi kartu pedoman telaah, kemudian hasil telaah dirangkum, dibuat kesimpulan dan ditentukan karakteristik perangkat tes yang memenuhi kriteria. Penelaah aspek materi dan konstruksi berjumlah tiga orang, sehingga penilaian dikategorikan baik jika dipilih oleh minimal dua orang penelaah. Lain halnya dengan penelaah aspek materi yang hanya satu orang, hasil hanya berdasarkan jawaban penelaah saja.

Selanjutnya perangkat tes dianalisis berdasarkan pendekatan teori tes klasik. Analisis butir soal dilakukan dengan menggunakan program *Item and Analysis (ITEMAN)* versi 3.00. Hasil analisis berupa data statistik mengenai tingkat kesukaran, daya beda, dan keberfungsian pengecoh. Tingkat kesukaran butir soal dihitung berdasarkan proporsi menjawab benar pada setiap butir soal. Mardapi (2012, p.128) menyatakan tentang indeks kesukaran butir soal yang dapat diterima adalah 0,30–0,80. Daya pembeda dihitung berdasarkan nilai *point biserial*, dikategorikan baik jika indeks daya beda $>0,3$ dan keberfungsian pengecoh dapat dilihat dari distribusi jawaban peserta tes. Pilihan jawaban paling sedikit dipilih oleh 5% peserta tes dan koefisien korelasi biserial pada alternatif jawaban bernilai negatif kecuali kunci jawaban. Masing-masing hasil analisis memiliki kriteria nilai dalam menentukan butir termasuk kategori baik, sedang atau bahkan tidak digunakan.

Analisis selanjutnya menggunakan teori respon butir, dilakukan dengan bantuan program *BILOG-MG* yang menghasilkan output dalam tiga fase. Fase pertama merupakan estimasi butir soal berdasarkan teori tes klasik, fase kedua merupakan estimasi butir soal berdasarkan teori respon butir dan fase ketiga merupakan estimasi kemampuan peserta tes. Model teori yang paling sering digunakan adalah model logistik. Ada tiga model logistik menurut Hambleton (1985, p.49), diantaranya: model logistik satu parameter (1PL), model logistik dua parameter (2PL), dan model logistik tiga parameter (3PL). Lebih lanjut Hambleton

(1985, p.107) menegaskan bahwa suatu butir dikatakan baik jika interval b_i berada antara -2 sampai +2 dengan syarat distribusi kemampuan peserta adalah -3 sampai dengan +3. Jika b_i kurang dari -2 mendekati +2, artinya tingkat kesulitan butir soal tinggi. Begitupun sebaliknya, jika nilai b_i mendekati -2 maka tingkat kesulitan butir soal rendah.

Kemampuan siswa dianalisis berdasarkan skor pada hasil analisis dengan teori tes klasik dan berdasarkan nilai *ability* pada hasil analisis teori respon butir. Kemampuan siswa masing-masing SMA berdasarkan hasil analisis kemudian dirata-ratakan untuk dapat dibandingkan dalam melihat kemampuan siswa antar SMA pada mata pelajaran kimia tahun pelajaran 2015/2016.

Hasil Penelitian dan Pembahasan

Hasil Penelitian

Karakteristik Butir Soal UAS Kimia Berdasarkan Kriteria Telaah Butir Soal

Butir soal ditelaah oleh 4 orang ahli dengan menggunakan lembar telaah butir. Lembar telaah berisi pernyataan yang terbagi dalam tiga aspek yaitu bahasa, konstruksi dan materi. Berdasarkan hasil telaah butir sebagian besar butir soal dikategorikan sebagai butir yang baik, dan sebagian lainnya termasuk dalam kategori cukup baik serta memerlukan revisi.

Butir soal yang berada pada kategori kurang baik masih bisa diterima untuk dijadikan butir soal ujian semester meski diperlukan revisi dari berbagai aspek. Lain halnya dengan butir yang termasuk dalam kategori tidak baik, butir akan digugurkan dan diganti dengan butir yang sudah sesuai dengan kriteria. Berdasarkan hasil analisis telaah butir diperoleh butir yang tidak baik sebanyak 7 butir soal yang terdiri dari butir nomor 5, 19, 13, 14, 21, 27 dan 34. Jika dipersentasekan jumlah butir yang tidak baik, maka akan diperoleh sekitar 17,5% butir dikategorikan tidak baik. Hasil rangkuman karakteristik butir berdasarkan analisis teoritis dijelaskan pada Tabel 1.

Selain menggunakan lembar telaah, beberapa orang guru dimintai untuk memberikan skor terhadap butir soal guna menghitung validitas isi. Validitas isi bertujuan untuk memastikan apakah butir soal sudah sesuai dan relevan dengan kisi-kisi soal, tujuan pembelajaran dan indikator pembelajaran yang ingin capai. Butir soal diperiksa dan dicocokkan de-

ngan perangkat pembelajaran lainnya untuk melihat apakah butir soal sudah mencerminkan domain indikator yang sudah ditetapkan. Mencocokkan butir dengan kisi-kisi, tujuan pembelajaran dan indikator pembelajaran dilakukan oleh *rater* yang merupakan guru-guru kimia SMAN di Kota Yogyakarta. Pencocokan butir yang dilakukan *rater* dinilai dalam bentuk skor 1 – 4. Skor tertinggi adalah 4 sedangkan skor terendah adalah 1.

Tabel 1. Rangkuman Telaah Butir Soal

Aspek	Telaah Kualitatif					
	Baik		Kurang Baik		Tidak Baik	
	Jml	%	Jml	%	Jml	%
Materi	33	82,5	6	15	1	2,5
Konstruksi	18	45	18	45	4	10
Bahasa	38	95	-	-	2	5

Berdasarkan hasil perhitungan dapat disimpulkan bahwa butir yang memiliki skor rendah dan indeks validitas rendah sebanyak 4 butir yaitu butir 2 (0,47), 19 (0,60), 27 (0,40) dan 30 (0,60). Butir dikatakan baik jika memiliki indeks validitas 0,67 atau lebih dari 0,67. Sehingga dapat disimpulkan bahwa butir memiliki validitas baik berdasarkan skor yang diberikan oleh *rater* berjumlah 36 butir atau sekitar 90% butir dikatakan valid.

Karakteristik Butir Soal UAS Kimia Berdasarkan Teori Tes Klasik

Menganalisis butir soal berdasarkan pemdekatan teori tes klasik dilakukan dengan bantuan program *ITEMAN* versi 3.00 yang akan menghasilkan karakteristik butir soal yang meliputi tingkat kesukaran, daya beda, dan efektivitas distraktor. Masing masing hasil memiliki kriteria. Sebanyak 40 butir soal ujian akhir semester kimia tahun pelajaran 2015/2016 dengan respon sebanyak 758 siswa di Kabupaten Muaro Jambi yang dianalisis berdasarkan kriteria tingkat kesulitan menghasilkan data yang disajikan pada Tabel 2.

Berdasarkan Tabel 2, diketahui bahwa butir soal yang memiliki tingkat kesulitan sedang sebanyak 26 butir soal atau sekitar 65% dari total butir soal yang dianalisis. Sebanyak 13 butir soal termasuk dalam kategori soal dengan tingkat kesulitan tinggi atau sekitar 32,5%, sedangkan butir soal dengan tingkat kesulitan mudah hanya terdapat satu soal saja.

Tabel 2. Tingkat Kesulitan Butir Soal

Kategori	Jumlah Butir	Persentase (%)
$p > 0,8$ (Mudah)	1	2,5
$0,3 \leq p \leq 0,8$ (Sedang)	26	65
$p < 0,3$ (Sulit)	13	32,5
Jumlah	40	100

Selanjutnya hasil analisis butir soal berdasarkan kriteria daya beda menghasilkan data yang disajikan pada Tabel 3.

Tabel 3. Daya Beda Butir Soal

Kategori	Jumlah	Persentase (%)
$> 0,30$ (Baik)	22	55
$< 0,30$ (Tidak Baik)	18	45
Jumlah	40	100

Daya beda butir soal berdasarkan hasil analisis menghasilkan butir soal dengan kategori baik sebanyak 22 butir soal atau sekitar 55% dan butir soal dengan kategori tidak baik sebanyak 18 butir atau sekitar 45% dari total butir soal yang dianalisis. Hal ini dapat dikatakan bahwa butir soal dengan daya beda baik berjumlah setengah dari butir soal yang ada.

Setiap pertanyaan memiliki distraktor berupa pilihan jawaban yang terdiri dari kunci jawaban dan pengecoh. Keefektifan distraktor bertujuan untuk melihat seberapa banyak siswa dalam memilih tiap-tiap pilihan jawaban yang telah disediakan. Adapun hasil analisis seperti disajikan pada Tabel 4.

Tabel 4. Efektivitas Distraktor Butir Soal

Kategori	Jumlah	Persentase (%)
$> 5\%$ (Baik)	30	75
$< 5\%$ (Tidak Baik)	10	25
Jumlah	40	100

Distraktor yang baik jika memiliki nilai lebih dari 0,05 pada setiap pilihan jawaban artinya lebih dari 5% siswa memilih pilihan jawaban yang benar dan pengecoh yang telah disediakan. Sebanyak 30 butir soal (75%) memiliki distraktor yang baik sedangkan 10 butir soal lainnya (25%) memiliki distraktor yang tidak baik yaitu dipilih kurang dari 5% siswa yang menjawab soal. Terdapat 2 butir soal yang me-

miliki peringatan “*check the key*”, yaitu soal dengan nomor 39 dan 40. Hal ini disebabkan oleh siswa tidak memilih kunci jawaban sebagai jawaban mereka, namun cenderung memilih pilihan jawaban pengecoh.

Karakteristik Butir Soal UAS Kimia Berdasarkan Teori Respon Butir

Sebelum dilakukan analisis butir, terlebih dahulu dilakukan uji prasyarat yaitu menguji terpenuhi atau tidaknya tahapan uji prasyarat yang mendasari teori respon butir. Uji prasyarat yang dilakukan terdiri dari uji asumsi, independensi lokal dan invariansi parameter.

Uji Prasyarat Teori Respon Butir

Persyaratan unidimensi dapat dikatakan terpenuhi atau tidak, dibuktikan melalui metode analisis faktor yang bertujuan untuk memperlihatkan pada kelompok faktor mana butir tes itu berada. Setiap faktor hanya menunjukkan suatu dimensi indikator tes. Setiap dimensi indikator terhimpun dalam satu faktor yang melibatkan beberapa butir tes yang diperlukan. Sebelumnya diperlukan uji kelayakan dengan berdasarkan uji Kaiser-Meyer-Olki (KMO) dan uji *Bartlett's* seperti pada hasil berikut ini:

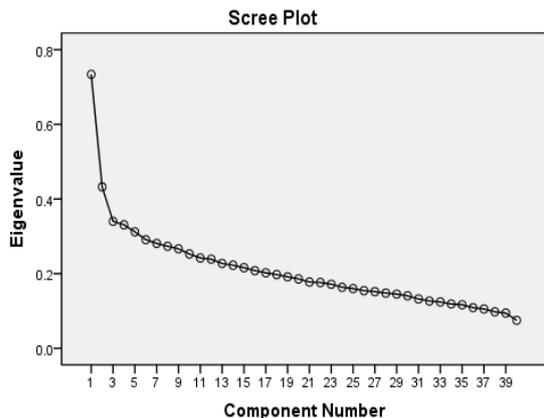
Tabel 5. Nilai KMO and Bartlett's Test^a

KMO and Bartlett's Test ^a		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.665
	Approx. Chi-Square	2530.335
Bartlett's Test	Df	780
of Sphericity	Sig.	.000

a. Based on correlations

Hasil menunjukkan bahwa nilai KMO sebesar 0,665. Nilai $0,665 > 0,5$ menunjukkan bahwa ukuran sampel untuk perangkat ini telah sesuai. Hasil Uji *bartlett's* bertujuan untuk menguji normalitas data yang menunjukkan bahwa nilai sig kurang dari 5%, artinya butir soal ujian semester kimia merupakan butir-butir yang saling independen.

Hasil analisis dengan menggunakan analisis faktor diperoleh nilai eigen dengan satu komponen domain yang mampu menunjukkan bahwa perangkat tes memenuhi uji asumsi unidimensi. Berdasarkan nilai eigen yang diperoleh, dapat pula disajikan dengan *scree plot* pada Gambar 1.



Gambar 1. *Scree Plot* Uji Asumsi

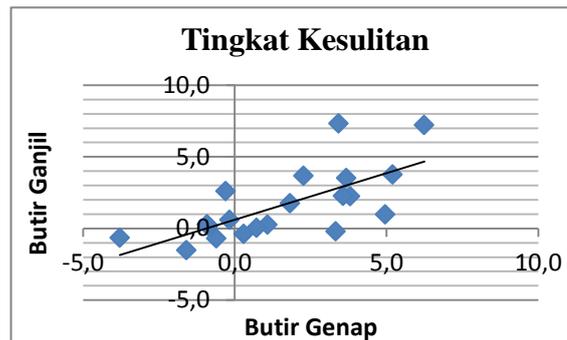
Selanjutnya dilakukan uji indepedensi lokal bertujuan untuk melihat apakah kemampuan peserta dalam suatu sub populasi yang sama independen terhadap butir. Pengujian inde-pendensi lokal dapat dilakukan dengan memeriksa nilai matriks variansi-kovariansi dari data kemampuan yang telah dibuat dalam sub-sub populasi. Kemampuan siswa terlebih dahulu dikelompokkan dan diurutkan berdasarkan nilai tertinggi hingga terendah siswa. Kemampuan siswa yang digunakan adalah kemampuan siswa berdasarkan model yang paling cocok yaitu model logistik 2 PL.

Hasil yang diperoleh menunjukkan bahwa menunjukkan bahwa nilai elemen di luar diagonal mendekati 0, artinya bahwa perangkat tes ujian semester kimia memenuhi uji asumsi indepedensi lokal.

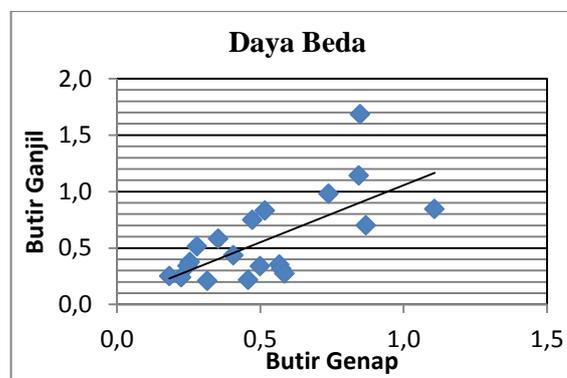
Langkah terakhir yang dilakukan adalah uji invariansi parameter. Invaransi parameter terbagi menjadi dua bagian yaitu invariansi parameter butir dan invariansi parameter siswa. Uji invariansi parameter butir bertujuan untuk mengetahui apakah karakteristik butir akan berubah meski dikerjakan oleh siswa yang berbeda kemampuannya, sedangkan uji invariansi parameter siswa bertujuan untuk mengamati apakah kemampuan siswa berubah jika diberikan soal dengan tingkat kesulitan dan daya beda yang berbeda. Soal dibagi menjadi dua bagian berdasarkan ganjil dan genap, kemudian diukur invariansi parameternya seperti disajikan pada Gambar 2.

Berdasarkan Gambar 2 dan 3 dapat dilihat bahwa banyak titik pada diagram tidak berada pada garis diagonal, hal ini menunjukkan bahwa adanya indikasi invariansi parameter butir. Artinya karakteristik akan berubah jika dikerjakan oleh siswa dengan kemampuan yang berbeda. Hasil uji invariansi parameter kemam-

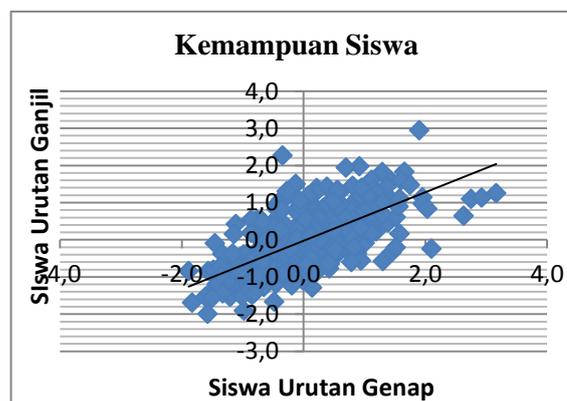
puan siswa yang diperoleh dari kemampuan siswa ganjil dan genap disajikan pada Gambar 4.



Gambar 2. Parameter Butir Tingkat Kesulitan Uji Invariansi Parameter



Gambar 3. Parameter Butir Daya Bada Uji Invariansi Parameter



Gambar 4. Parameter Kemampuan Siswa Invariansi Parameter

Gambar 4 menunjukkan bahwa titik pada diagram relatif dekat pada satu garis diagonal, artinya uji asumsi invariansi parameter siswa terpenuhi.

Hasil Analisis Butir

Butir soal dianalisis dengan bantuan program *BILOG_MG* dengan model logistik 2

parameter yang menghasilkan tingkat kesukaran dan daya beda butir serta kesesuaian dengan model. Kategori hasil analisis butir dapat dilihat pada Tabel 6.

Tabel 6. Analisis Butir Model 2PL

Kategori	Daya Beda	
	Jumlah	Persentase (%)
Baik	40	100
Tidak Baik	0	0
Total	40	100

Tabel 7. Analisis Butir Model 2PL

Kategori	Tingkat Kesulitan	
	Jumlah	Persentase (%)
Baik	22	55
Tidak Baik	18	45
Total	40	100

Tabel 8. Analisis Butir Model 2PL

Kategori	Fit Model	
	Jumlah	Persentase (%)
Fit	29	72,5
Tidak Fit	11	27,5
Total	40	100

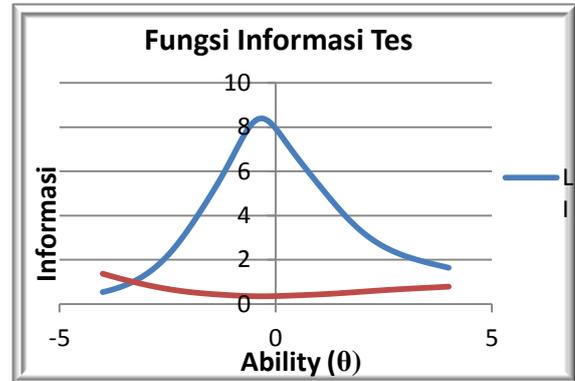
Berdasarkan Tabel 6, diketahui bahwa dengan menggunakan model logistik 2 PL seluruh butir dikategorikan baik untuk daya beda. Lain halnya dengan tingkat kesukaran pada Tabel 7 yang memiliki kategori baik hanya berjumlah 22 butir sedangkan 18 butir lainnya dikategorikan tidak baik. Kesesuaian butir dengan model dapat dilihat pada Tabel 8 diperoleh sebanyak 29 butir yang fit model atau sekitar 72,5%.

Fungsi Informasi

Fungsi informasi dianalisis berdasarkan model logistik 2 parameter. Berikut disajikan grafik yang menggambarkan hasil analisis fungsi informasi pada Gambar 5.

Berdasarkan grafik diketahui bahwa dalam mengukur kemampuan siswa, fungsi informasi perangkat tes berada pada kemampuan siswa sedang hingga kemampuan rendah. Artinya informasi yang diperoleh dari perangkat tes ini akan maksimal jika diberikan pada siswa yang berkemampuan sedang hingga kemampuan rendah yang ditunjukkan pada puncak grafik

(8,1576) berada pada kemampuan -0.1. Fungsi informasi memiliki nilai SEM terkecil sebesar 0,3501 tepat pada saat fungsi informasi terbesar diberikan. Besar kecilnya kesalahan baku pengukuran dipengaruhi oleh besar kecilnya nilai fungsi informasi suatu tes.

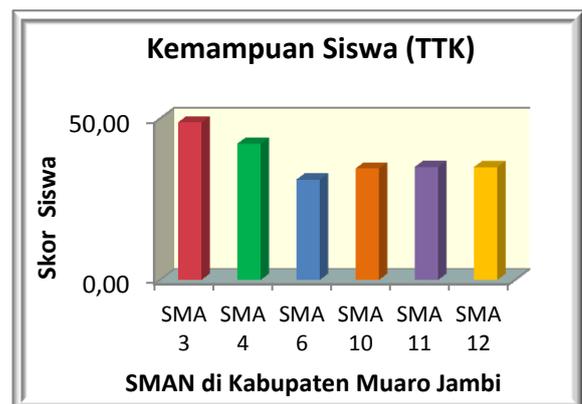


Gambar 5. Grafik Fungsi Informasi Tes

Kemampuan Siswa Berdasarkan Hasil Ujian Akhir Sekolah

Analisis dengan Teori Tes Klasik

Kemampuan siswa berdasarkan dengan teori tes klasik dilihat dari hasil analisis dengan menggunakan program *ITEMAN*. Hasil analisis berupa skor yang diperoleh siswa berdasarkan skor hasil ujian tiap sekolah kemudian dirata-ratakan dan dibandingkan kemampuannya. Berikut grafik yang menjelaskan perbandingan kemampuan siswa SMA yang menjadi objek penelitian di Kabupaten Muaro Jambi:



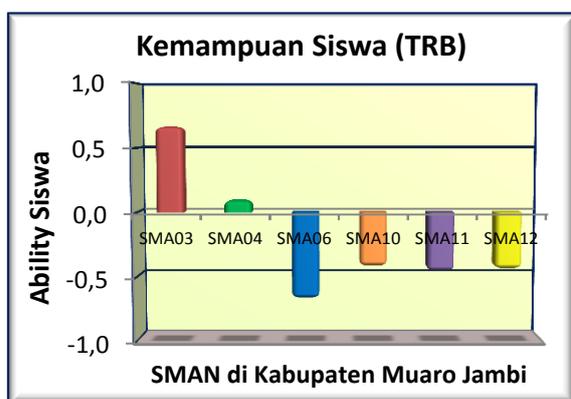
Gambar 6. Grafik Kemampuan Siswa Berdasarkan Teori Tes Klasik

Kemampuan siswa yang paling tinggi dimiliki oleh siswa SMAN 3 Muaro Jambi (48,761), kemudian berturut-turut dimiliki oleh SMAN 4 Muaro Jambi (42,125), SMAN 11 Muaro Jambi (34,949), SMAN 12 Muaro Jambi

(34,9), SMAN 10 Muaro Jambi (34,479) dan SMAN 6 Muaro Jambi (31,029). SMAN 6 menempati kemampuan paling rendah berdasarkan skor yang mereka hasilkan.

Analisis dengan Teori Respon Butir

Kemampuan siswa yang dianalisis berdasarkan hasil analisis *BILOG-MG* dapat diamati pada fase 3. Terlihat masing masing kemampuan siswa dan jumlah butir soal yang telah dikerjakan. Analisis butir dengan program *BILOG-MG* dilakukan dengan berdasarkan model logistik 2PL. Kemampuan siswa pada tiap sekolah nantinya akan dirata-ratakan dan dibandingkan bagaimana peringkat kemampuan siswa berdasarkan hasil ujian semester yang telah dilakukan. Rata-rata kemampuan siswa ditunjukkan pada Gambar 7.



Gambar 7. Grafik Kemampuan Siswa Berdasarkan Teori Respon Butir

Berdasarkan grafik terlihat bahwa rata-rata kemampuan siswa tertinggi dimiliki oleh SMAN 3 Muaro Jambi sebesar 0,667, kemudian SMAN 4 Muaro Jambi sebesar 0,096, SMAN 10 Muaro Jambi sebesar -0,422, SMAN 11 Muaro Jambi sebesar -0,461 SMAN 12 Muaro Jambi sebesar -0,443, dan SMAN 6 Muaro Jambi sebesar -0,676 menempati urutan terendah kemampuan siswa.

Pembahasan

Karakteristik Butir Soal UAS Kimia Berdasarkan Kriteria Telaah Butir Soal

Pada bagian aspek materi terdapat 4 butir soal yang perlu direvisi karena butir soal masih belum sesuai dengan indicator, soal masih memiliki pilihan jawabannya dianggap kurang homogen dan logis.

Aspek konstruksi menghasilkan butir yang kurang sesuai pada kriteria pokok soal

yang belum dirumuskan dengan jelas dan tegas. Masih terdapat butir soal yang mengandung pernyataan yang bersifat negatif ganda pada butir pertama. Selain itu ada beberapa butir yang memiliki panjang rumusan jawaban tidak sama yaitu pada butir 1, 9, 15, 17 dan 34, sedangkan butir 21 dianggap tidak memenuhi kriteria ini karena panjang jawaban pada soal tidak sama dan memiliki pilihan jawaban yang ganda. Beberapa pilihan jawaban tidak berurut dari yang terkecil hingga yang terbesar namun terbalik dari yang terbesar hingga terkecil seperti pada butir nomor 25 dan 36.

Selanjutnya bagian akhir telaah butir dari aspek bahasa pada menjelaskan bahwa butir 13 dan 34 perlu direvisi karna tidak tidak begitu jelas antara subjek, predikat dan objek dalam sebuah kalimat. Selain itu kata "diatas" tidak berjarak antara kata depan, yang seharusnya adalah "di atas". Butir soal yang baik salah satunya memiliki validitas butir yang baik pula. Validitas yang diujikan pada perangkat tes adalah validitas isi yang bertujuan untuk menilai kecocokan butir terhadap kisi-kisi soal. Butir yang tidak valid yaitu butir soal belum mengukur apa yang seharusnya diukur berjumlah 4 butir soal yaitu butir ke 2, 19, 27 dan 30. Butir dikatakan memiliki validitas isi yang baik jika memiliki indeks validitas 0,67 atau lebih dari 0,67

Karakteristik Butir Soal UAS Kimia Berdasarkan Teori Tes Klasik

Hasil analisis yang diperoleh butir dengan karakteristik baik hanya 10 butir saja. Tingkat kesulitan butir soal cukup tinggi, sehingga dapat dikategorikan sebagai soal sulit untuk siswa yang mengerjakannya. Soal yang dianggap sulit bagi siswa, menghasilkan jawaban yang tidak beragam, sehingga tidak mampu membedakan kemampuan siswa.

Minimnya butir soal yang baik dapat disebabkan oleh banyak faktor. Salah satunya adalah kualitas perangkat soal yang rendah. Tingkat kesukaran butir tidak menyebar secara proposional sehingga siswa yang kurang pandai tidak mampu menjawab soal yang diujikan.

Selain itu, terdapat banyak soal yang memiliki distraktor tidak efektif. Hal ini dapat disebabkan oleh banyaknya siswa yang berkemampuan tinggi lebih memilih jawaban jawaban pengecoh atau sama sekali tidak dipilih oleh siswa. distraktor yang baik adalah distraktor yang dipilih oleh sedikit siswa yang berkemampuan rendah, namun jumlah siswanya

signifikan. Soal yang terlalu sulit dapat juga menyebabkan distraktor tidak berfungsi atau terdapatnya kesalahpahaman konsep pada siswa.

Butir soal nomor 39 perlu ditinjau dan direvisi kembali karena pada hasil analisis tampak adanya peringatan *CHECK THE KEY C was specified, D works better* yang menunjukkan bahwa kunci jawaban C kurang tepat dan alternatif jawaban D tampak berfungsi lebih baik. Artinya butir C yang merupakan kunci jawaban soal justru dipilih oleh sedikit siswa. Siswa justru cenderung memilih alternatif jawaban A yang bukan kunci jawaban tapi bagian dari pengecoh. Tampak pula pada pilihan jawaban A yang memiliki kode tanda tanya (?) pada option jawaban. Begitu pula pada butir soal nomor 40 yang memiliki peringatan *CHECK THE KEY D was specified, C works better* yang menunjukkan bahwa kunci jawaban D kurang tepat dan alternatif jawaban C tampak berfungsi lebih baik. Sebaiknya dilakukan revisi pada butir soal ke 39 dan 40 yaitu dengan memeriksa kembali kunci jawaban. Apabila kunci jawaban ternyata salah lakukan revisi dan kembali analisis soal, namun apabila ternyata kunci jawaban sudah benar maka kemungkinan kesalahan terletak pada kesalahan penguasaan konsep oleh siswa.

Karakteristik Butir Soal UAS Kimia Berdasarkan Teori Respon Butir

Hasil analisis melalui pendekatan teori respon butir menghasilkan karakteristik butir baik sebanyak 13 butir. Tingkat kesulitan butir soal yang baik sebanyak 22 butir dengan daya beda butir baik secara keseluruhan. Namun butir soal yang baik berdasarkan tingkat kesulitan dan daya beda beberapa tidak fit dengan model, sehingga tidak dapat dikategorikan sebagai butir yang baik. Butir baik jika memiliki tingkat kesulitan, daya beda dan kecocokan model yang sesuai dengan kriteria.

Besarnya kemampuan siswa dalam menjawab soal ditunjukkan pada nilai fungsi informasi. Hasil perhitungan menunjukkan bahwa perangkat tes ujian semester kimia di Kabupaten Muaro Jambi mampu dikerjakan oleh siswa yang memiliki kemampuan sedang ke rendah. Hal ini ditunjukkan pada grafik yang telah di sajikan sebelumnya. Fungsi informasi dengan SEM memiliki hubungan yang berbanding terbalik *kuadratik*. Besar kecilnya nilai fungsi informasi suatu tes dipengaruhi oleh

kesalahan baku pengukuran. Semakin besar fungsi informasi maka SEM akan semakin kecil, terlihat pada grafik bahwa nilai SEM terendah berada pada saat perangkat tes memberikan fungsi informasi yang maksimal.

Hasil analisis dengan bantuan program *BILOG-MG* dan *ITEMAN* menghasilkan perbedaan jumlah butir yang memenuhi kriteria baik. Karakteristik butir soal berdasarkan hasil analisis program *ITEMAN* menunjukkan bahwa butir memiliki tingkat kesulitan cukup baik namun memiliki butir soal dengan daya beda yang rendah cukup banyak. Pengecoh berfungsi cukup baik pada soal dengan tingkat kesulitan cukup tinggi, sedangkan untuk soal yang memiliki tingkat kesulitan yang rendah pengecoh belum berfungsi dengan baik. Jika mengikuti kriteria yang ada, maka terdapat 10 butir soal yang memiliki karakteristik baik yaitu butir soal ke 2, 7, 9, 11, 12, 13, 22, 23, 24 dan 25.

Kemampuan Siswa Berdasarkan Hasil Ujian Akhir Sekolah

Berdasarkan hasil analisis, diketahui bahwa rata-rata kemampuan siswa tertinggi diraih oleh siswa SMAN 3 Muaro Jambi dan rata-rata kemampuan terendah diperoleh siswa SMAN 6 Muaro Jambi. Hasil yang sama diperoleh ketika kemampuan siswa dihitung berdasarkan hasil analisis teori respon butir. Siswa dengan rata-rata kemampuan tinggi dicapai oleh SMAN 3 Muaro Jambi, sedangkan siswa dengan rata-rata kemampuan kategori rendah dimiliki oleh SMAN 6 Muaro Jambi.

Terjadi perbedaan pada urutan ketiga, keempat dan kelima yaitu SMAN 12 Muaro Jambi menempati urutan ketiga berdasarkan teori tes klasik, namun berada di urutan ke empat ketika dianalisis dengan menggunakan teori respon butir. Hal ini karena karena teori respon butir menganalisis berdasarkan butir soal yang tidak terikat dengan sampel. Peluang menjawab benar siswa akan dihubungkan dengan parameter butir dan parameter peserta tes. Lain halnya dengan teori tes klasik yang mengestimasi kemampuan peserta berdasarkan soal yang diberikan. Jika siswa berkemampuan rendah mengerjakan soal dengan dengan tingkat kesulitan diatas kemampuannya, maka soal tidak akan dapat dikerjakan, sehingga hasil analisis hanya berdasarkan skor yang dihasilkan.

Rendahnya kemampuan siswa pada hasil ujian semester disebabkan oleh banyak faktor, salah satunya adalah rendahnya kualitas

tes yang dibuat oleh guru. Tes yang baik mampu memperoleh informasi terkait kemampuan siswa, oleh karena itu dalam penulisan butir soal harus secara benar dan sesuai dengan tujuan dilaksanakan tes tersebut. Ukuran sebuah soal yang diterima bergantung pada tingkat kesulitan dan tujuan tes tersebut dibuat. Tes yang tidak baik sebaiknya dibuang atau direvisi sebelum dimasukkan kembali kedalam tes yang akan diujikan. Soal yang baik seyogyanya didukung oleh kemampuan siswa yang baik pula guna meningkatkan kualitas pendidikan.

Simpulan dan Saran

Simpulan

Karakteristik butir berdasarkan kriteria telaah ahli menunjukkan bahwa soal ujian semester kimia SMA di Kabupaten Muaro Jambi memiliki butir soal yang baik dari aspek materi, aspek konstruksi dan aspek bahasa sebanyak 33 butir soal atau sebesar 82,5% dari total butir yang ditelaah. Karakteristik butir soal yang ditinjau berdasarkan pendekatan teori tes klasik menghasilkan butir memenuhi kriteria sebanyak 10 butir soal (25%). Karakteristik butir soal yang ditinjau berdasarkan pendekatan teori respon butir menghasilkan butir sebanyak 13 butir soal (32,5%). Hasil analisis dengan menggunakan teori tes klasik menghasilkan butir tes lebih sedikit karena butir baik menurut analisis teori respon butir memiliki distraktor yang tidak baik pada hasil analisis teori tes klasik.

Kemampuan siswa pada ujian akhir semester mata pelajaran kimia di SMA se-Kabupaten Muaro Jambi berdasarkan pendekatan teori tes klasik dan teori respon butir sama sama menunjukkan bahwa rata-rata kemampuan siswa SMAN 3 Muaro Jambi berada pada urutan tertinggi, sedangkan urutan terendah ditempati oleh SMAN 6 Muaro Jambi.

Saran

Mengingat banyaknya saran yang diberikan oleh penelaah terhadap butir soal yang ada. Tim penyusun soal sebaiknya melakukan telaah butir soal terlebih dahulu sebelum soal digunakan agar butir soal memiliki kualitas baik secara materi, konstruksi dan bahasa. Kemudian setelah soal selesai digunakan, perlu dilakukan analisis dengan bantuan program untuk mengamati bagaimana karakteristik butir

soal yang ada meliputi daya beda, tingkat kesulitan dan pengecoh.

Daftar Pustaka

- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, 2, 189-193.
- Brown, T. L., Lemay, E. H. Jr., & Bursten, B., E. (1991). *Chemistry the central science*. Englewood cliffs. New York: Prentice Hall, Inc.
- Hambleton., K.R., & Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston: Kluwer Nijhoff Publishing.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory application to psychological measurement*. Washington, D.C: Dow Jones-Irwin.
- Mardapi, Djemari. (2012). *Pengukuran penilaian & evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching*. New Jersey: Pearson Education Inc.
- Mulyana, Kabul. (2007). Karakteristik soal tes masuk SMP negeri di kabupaten bantul. *Jurnal Penelitian dan Evaluasi Pendidikan*, 10(1), 235-248. Retrieved from <http://journal.uny.ac.id/index.php/jpep/article/view/1985>
- Nurung, Muh. (2008). *Kualitas tes ujian akhir sekolah berstandar nasional (UASBN) IPA SD tahun pelajaran 2007/2008 di kota Kendari*. Tesis magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
- Olivo, S. A., Cumming, G. G., Fuentes, J., Saltaji, H. (2014). Identifying items to assess methodological quality in physical therapy trials: a factor analysis. *Journal Physical Therapy*, 94, 1227-1284.