

ANALISIS SKALA KEMAMPUAN MAHASISWA KIMIA MELALUI TES ENTRY BEHAVIOUR METODE COMPUTERIZE ASSISTED TEST (CAT) MENGGUNAKAN ITEM RESPONDS THEORY (IRT)

Y. Manoppo ^{1*}, J. B. Manuhuttu¹

¹Departement of Chemistry-FKIP, Pattimura University Ambon

*molucanano@yahoo.com

ABSTRACT

This study aims to analyze scales abilities through tests chemistry student entry behaviour method computerize assisted test (CAT) used items response theory (IRT) with logistics 1-parameter. The subject of research is the whole answer sheet test participants entry behaviour student course of study chemical years education 2019. The research is quantitative research adopting ex-post facto. The analysis of the theory test approach with classic show 79,5 % grains have difficulty rating of grains serves both, 57 % grains of resources is the difference not qualified, and 72 % grains of having the rapscallion serves both with the reliability test 0,832. An analysis by approach response theory grains show there are 11 (27,5 %) grains of about suitable (fit) with, model the function of the maximum information 5,48 on $\theta = -0,21$, and the size of a Standard of Error measuring (SEM) 0,21.

Keywords: *Entry Behaviour, Classical Test Theory, Item Respons Theory, Rasch.*

ABSTRAK

Penelitian ini bertujuan untuk menganalisis Skala Kemampuan Mahasiswa Kimia Melalui Tes Entry Behaviour Metode *Computerize Assisted Test (CAT)* Menggunakan Item Respons Theory (IRT) Model Logistik 1-Parameter. Subjek penelitian adalah seluruh lembar jawaban peserta Tes *Entry Behaviour* Mahasiswa Program Studi Pendidikan Kimia Tahun 2019. Penelitian ini merupakan penelitian kuantitatif menggunakan pendekatan *ex-post facto*. Hasil analisis dengan pendekatan teori tes klasik menunjukkan 79,5% butir memiliki tingkat kesulitan butir berfungsi baik, 57% butir daya bedanya belum memenuhi syarat, dan 72% butir memiliki pengecoh berfungsi baik dengan indeks reliabilitas tes 0,832. Analisis dengan pendekatan teori respons butir menunjukkan terdapat 11 (27,5%) butir soal cocok (*fit*) dengan model, fungsi informasi maksimum 5,48 pada $\theta = -0,21$, dan besarnya kesalahan pengukuran (*SEM*) 0,21.

Kata Kunci: *Entry Behaviour, Classical Test Theory, Item Respons Theory, Rasch.*

PENDAHULUAN

Mutu pendidikan dapat dilihat dari hasil evaluasi terhadap prestasi belajar siswa dan evaluasi terhadap prestasi belajar siswa sebagian besar bersumber pada hasil pengukuran. Berdasarkan hasil pengukuran dapat diperoleh informasi yang bermakna sebagai masukan kepada pihak-pihak yang terkait dengan pendidikan, sebagai dasar pengambilan keputusan atau sebagai dasar penilaian dalam rangka menentukan kebijakan bagi guru atau lembaga penyelenggara pendidikan terhadap peserta didik.

Untuk menilai mutu setiap butir soal yang pada akhirnya untuk mengetahui seperangkat soal dalam berbagai aspek, perlu dilakukan analisis butir soal. Tujuan utama analisis butir soal adalah untuk mendapatkan informasi tentang karakteristik setiap butir soal, baik melalui telaah butir soal maupun

analisis empiris. Hasil analisis tersebut dapat digunakan untuk mengetahui mutu soal, mutu belajar peserta didik, dan sebagai salah satu indikasi keberhasilan lembaga pendidikan.

Konsep Paradigma pengembangan butir-butir tes berdasarkan acuan patokan digunakan untuk mengukur sejauh mana peserta didik telah mencapai tujuan instruksional. Hal ini dapat dilakukan dengan cara membandingkan penampilan peserta didik dalam pengujian dengan patokan yang telah ditentukan sebelumnya. Tes acuan patokan disebut juga tes acuan tujuan dalam pengukuran proses pembelajaran yang berpusat pada peserta didik (*learned-centered*) adalah penilaian yang berpusat pada peserta didik (*learner-centered assessment*). Definisi *learner-centered assessment* sejajar dengan definisi tradisional test acuan patokan, sebagai element utama dari pembelajaran yang didesain secara sistematis. Bagi seorang perancang pembelajaran harus mengembangkan butir tes acuan patokan, karena hasil tes pengukuran tersebut berguna untuk: (i). mendiagnosis dan menempatkan dalam kurikulum; (ii). Memeriksa hasil belajar dan kesalahan pengertian sehingga dapat diberikan pembelajaran remedial sebelum pembelajaran dilanjutkan; (iii). menjadi dokumen kemajuan belajar.

Dalam mengembangkan butir-butir tes acuan patokan, Dick and Carrey (1985) merekomendasikan empat macam tes acuan patokan antara lain: (1) *Entry behavior* test yang diberikan kepada peserta didik sebelum memulai pembelajaran. Tes ini berguna untuk mengukur keterampilan syarat atau keterampilan yang harus sudah dikuasai sebelum pembelajaran dimulai. Keterampilan syarat akan muncul di bawah garis *entry behavior*. (2) *Pre-test* yang dilakukan pada awal pembelajaran untuk mengetahui apakah peserta didik sudah menguasai beberapa atau semua keterampilan yang akan diajarkan. Tujuannya adalah untuk efisiensi. Jika semua keterampilan sudah dikuasai maka tidak perlu ada pembelajaran. Namun jika hanya sebagian materi yang sudah dikuasai maka data tes ini memungkinkan desainer untuk lebih efisien. Mungkin hanya review atau pengingat yang dibutuhkan. Bila program tersebut merupakan sesuatu yang baru, maka tes inipun dapat ditiadakan. Maksud dari pretes ini bukanlah untuk menentukan nilai akhir tetapi lebih mengenal profil anak didik berkenaan. Biasanya pretest dan *entry behavior* test dijadikan satu.

Hasil dari tes *entry behavior* dapat digunakan untuk mengetahui apakah peserta didik siap memulai pembelajaran, sedangkan dari hasil pretest dapat diutuskan apakah pembelajaran akan menjadi terlalu mudah untuk peserta didik. (3) *Practice test* yang diberikan selama peserta didik sedang dalam proses belajar (Uno, 2007: 28, tes tersebut tes sisipan). Tes ini berfungsi untuk melihat apakah peserta didik memang telah dapat menangkap apa yang sedang dibicarakan dan juga untuk membuat peserta didik lebih aktif berpartisipasi selama pembelajaran. Tes ini memungkinkan peserta didik untuk menampilkan pengetahuan dan keterampilan baru dan untuk refleksi diri sampai level berapa keterampilan dan pengetahuan mereka. Tes ini berisi keterampilan yang lebih sedikit dan lebih fokus pada materi setiap. Hasil tes ini digunakan instruktur untuk memberikan feedback dan untuk memonitor pembelajaran. (4) *Post-test* yang diparalel dengan *pre-test*. Sama dengan *pre-test*, *post-test* mengukur tujuan pembelajaran. *Post-test* harus menilai semua objektif dan terutama fokus pada objektif terakhir. Namun jika waktu tidak memungkinkan, maka hanya tujuan akhir dan keterampilan penting saja yang diujikan. *Post-test* mungkin digunakan untuk menilai performa peserta didik dan untuk memberi kredit karena telah menyelesaikan program. Tujuan yang terutama dari tes ini adalah agar pendidik dapat mengidentifikasi area pembelajaran yang tidak bisa dilakukan dengan baik. Jika peserta didik gagal dalam tes, pendidik harus dapat teridentifikasi dalam proses pembelajaran yang mana tidak dimengerti oleh peserta didik. Tes ini merupakan tes acuan patokan yang mencakup pengukuran semua tujuan intruksional khusus yang ada terutama tujuan intruksional yang bersifat terminal. Dengan tes ini dapat diketahui bagian-bagian mana diantara pembelajaran yang belum dicapai.

Terdapat beberapa saran yang dapat membantu dalam menentukan berapa banyak tes item pilihan yang diperlukan. Jika tes item memerlukan sebuah format respon yang memungkinkan peserta didik dapat menebak jawaban dengan benar, maka dapat memasukkan beberapa tes item paralel

untuk tujuan yang sama jika kemungkinan menebak jawaban yang benar kecil kemungkinan, dapat diputuskan satu atau dua item untuk menentukan kemampuan peserta didik.

Tiap-tiap program studi yang ada di FKIP Unpatti memiliki karakteristik bidang studi yang berbeda-beda, sehingga menghasilkan perbedaan kemampuan dan prestasi belajar pada mahasiswa. Untuk itu perlu dilakukan penelitian terhadap validity differensial. Hal ini bertujuan agar pemilihan program studi yang akan diminati mahasiswa memiliki kesesuaian bidang dengan kemampuannya. Dengan mengetahui kemampuan setiap mahasiswa, maka setiap mahasiswa dapat diseleksi untuk dapat menempati kelas Unggulan.

Karakteristik soal tes tes entry behaviour tidak hanya sebatas pada deteksi kemampuan setiap peserta tes, tetapi kedepan estimasi kemampuan dapat diintegrasikan pada penyusunan soal berbasis online yang lebih dikenal dengan CAT (Computerize Adaptif Test). CAT memformulasikan estimasi kemampuan peserta tes disamping menggunakan skor jawaban benar yang diperoleh oleh masing-masing peserta tes.

METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif menggunakan pendekatan ex-post facto. Pendekatan ini dapat diartikan sebagai pendekatan yang mencoba mengungkap dampak dari suatu perlakuan dan menelusuri faktor-faktor penyebabnya. Oleh karena itu dalam penelitian ini tidak dilakukan perlakuan apapun terhadap variabel penelitian. Data utama yang digunakan dalam penelitian ini adalah data set berupa jawaban/respons (lembar jawaban komputer/LJK) peserta Tes Entry Behaviour Mahasiswa Program Studi Pendidikan Kimia Tahun 2019.

Objek penelitian adalah Lembar Respon Jawaban Mahasiswa Baru calon guru kimia di Program Studi Pendidikan Kimia FKIP Unpatti. Dalam penelitian ini tidak dilakukan perlakuan apapun terhadap variabel penelitian. Data utama yang digunakan dalam penelitian ini adalah data set berupa jawaban/respons (lembar jawaban tes) peserta Tes Entry behaviour 2019 sebanyak 49 peserta.

Prosedur Penelitian

Menganalisis butir tes dengan perangkat lunak Iteman versi 3.0 untuk melihat karakteristik butir tes yang meliputi tingkat kesukaran butir, daya pembeda butir, dan statistik sebaran jawaban. Selain itu ditentukan juga kehandalan/reliabilitas tes, kesalahan pengukuran (SE) dan distribusi skor.

Menganalisis butir dengan pendekatan teori respons butir menggunakan program Bilog-MG 3.0. Dari analisis ini diperoleh parameter butir, parameter peserta, fungsi informasi tes, dan grafik informasi tes. Dari analisis ini diperoleh besarnya Fungsi Informasi maksimum tiap butir, dan skala kemampuan θ yang diambil antara -4,0 dan 4,0 dengan interval 0,25.

Menaksir kecocokan data dengan model dan parameter butir dengan program Bilog-MG 3.0. Kecocokan butir dengan model dapat dilihat dari harga p yang muncul dari hasil analisis dengan program Bilog-MG. Untuk butir yang memiliki harga p lebih kecil dari 0,01 dianggap tidak cocok dengan model logistik 1-parameter (Rasch Model). Data yang digunakan dalam penelitian ini adalah respon peserta tes Entry Behaviour Program Studi Pendidikan Kimia FKIP Unpatti Tahun 2017.

Teknik Analisis Data

1. Analisis Karakteristik Butir Soal

Analisis terhadap butir soal secara kuantitatif (empiris) menggunakan pendekatan teori tes klasik dan teori respon butir (IRT).

2. Analisis Empiris dengan Teori Tes Klasik

1) Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir dianalisis dengan menghitung proporsi menjawab benar (p) sebuah butir soal. Apabila indeks $p < 0,30$ maka butir soal termasuk sukar, apabila $0,30 \leq p \leq 0,70$ maka butir soal memiliki tingkat kesukaran yang sedang, jika $p > 0,70$ maka soal termasuk soal yang mudah.

2) Daya Pembeda Butir Soal

Daya beda butir soal (item) adalah korelasi antara skor butir soal dengan skor total, dihitung dengan rumus korelasi point biserial. Besarnya daya beda butir soal untuk menyatakan bahwa butir soal baik adalah minimal 0,3. Sedangkan item soal dengan point biserial dibawah 0,3 termasuk item yang kurang baik.

3) Keberfungsian Pengecoh

Untuk mengetahui suatu pengecoh berfungsi secara baik dapat dilihat dari distribusi jawaban. Suatu pengecoh dapat dikatakan berfungsi baik jika paling sedikit pilihan jawaban dipilih oleh 5% peserta tes. Beberapa literatur pada penelitian yang lain juga menyebutkan 2%, tetapi sebenarnya jika ada yang menjawab saja sudah bisa dikategorikan bahwa pengecohnya (distractor) sudah berfungsi dengan baik.

4) Indeks Reliabilitas.

Nilai indeks reliabilitas tes dapat dilihat pada nilai koefisien alpha. Semakin besar nilai alpha menunjukkan semakin reliabel tes tersebut dan semakin kecil tingkat kesalahan pengukuran. Standar yang digunakan dalam penelitian ini adalah yang mensyaratkan instrumen yang reliabel haruslah memiliki koefisien reliabilitas minimal 0,7.

3. Analisis Empiris Butir dengan Teori Respon Butir

Analisis dengan program Bilog-MG menghasilkan output dalam bentuk tiga fase. Fase pertama merupakan estimasi butir berdasar teori klasik, fase kedua estimasi parameter butir berdasar Teori Respons Butir dan fase ketiga estimasi kemampuan peserta.

Pada fase pertama diperoleh info tentang banyaknya peserta tes yang menjawab benar, proporsi peluang menjawab benar dibagi peluang menjawab salah serta koefisien korelasi biserial. butir yang memiliki nilai koefisien biserial negatif dapat mengganggu proses analisis selanjutnya, sehingga butir tersebut tidak diikutkan dalam analisis berikutnya.

Fase kedua, estimasi parameter teori respon butir. Pada fase ini diperoleh tentang informasi parameter butir sesuai dengan model teori tespon butir yang digunakan. Selanjutnya untuk model 1-PL didapatkan estimasi tingkat kesukaran, model 2-PL diperoleh estimasi parameter berupa tingkat kesukaran dan daya beda, sedangkan model 3-PL didapatkan informasi tentang tingkat kesukaran, daya beda, dan tebakan semu. Selain parameter butir, pada fase kedua juga dihasilkan statistik kecocokan suatu butir dengan model atau goodness of fit statistic. Model yang digunakan untuk estimasi parameter adalah model logistik yang banyak menerima butir yang cocok.

Secara empiris, kualitas butir ditelaah berdasarkan kecocokan data dengan model dan nilai parameter butir. Kecocokan suatu butir dengan model dapat dilihat dari nilai chi kuadrat butir dibandingkan dengan harga kritik distribusi chi kuadrat sesuai dengan dk butir yang bersangkutan pada taraf signifikansi α . Butir dikatakan tidak cocok model jika nilai butir lebih besar dari harga distribusi pada nilai kritisnya, sebaliknya butir cocok dengan model jika nilai item lebih kecil atau sama dengan nilai distribusi. Atau dikatakan cocok model jika probabilitas lebih besar dari 0,01.

Berdasarkan perbandingan ketiga model parameter logistik untuk uji kecocokan model, maka model 1-PL pada umumnya lebih banyak menghasilkan butir yang fit (cocok). Sehingga untuk keseragaman model analisis, estimasi parameter butir untuk seluruh materi uji dianalisis menggunakan model 1-PL.

Fase ketiga didapatkan estimasi parameter kemampuan peserta dan fungsi informasi tes. Pada penelitian ini informasi yang digunakan hanya fungsi informasi tes, sehingga estimasi parameter kemampuan peserta tidak dianalisis. Besarnya fungsi informasi tes dihitung menggunakan program Excel.

Berdasarkan kriteria di atas, untuk menentukan kualitas butir yang baik dengan pendekatan teori respon butir didasarkan pada kriteria: (1) butir cocok model, dan (2) tingkat kesulitan -2 hingga $+2$ (Hambleton & Swaminathan, 1985 : 36).

HASIL PENELITIAN

1. Karakteristik Butir Tes Klasik dengan Program Iteman

a. Statistik Butir Soal

Pada tampilan hasil statistik butir soal dapat diketahui hasilnya untuk soal yang dikategorikan sukar 17 butir, kategori sedang 23 butir, dan tidak terdapat soal dalam kategori mudah. Soal yang memiliki daya beda yang baik 22 butir, kurang baik 18 butir sedangkan untuk pengecoh 36 butir yang berfungsi baik.

b. Statistik Tes/Skala

Berdasarkan statistik hasil analisis program Microcat Iteman versi 3.0 terhadap 30 butir soal Kimia Tes Entry Behaviour Tahun 2013 yang direspon oleh peserta ujian sebanyak 85 orang pada Tabel 1 dapat bahwa peserta rata-rata dapat menjawab dengan benar sebesar 13 butir soal (mean = 12,518), berarti rata-rata kurang dari setengah (32,5%) dari jumlah butir soal dapat dijawab dengan benar oleh peserta ujian, skor tertinggi 29 dan terendah 3. Nilai rerata (12,518) yang berdekatan dengan nilai median (12,00). Rata-rata tingkat kesukaran butir soal pada umumnya sedang (mean $p = 0,313$), cenderung banyak pada kisaran indeks 0,30 sampai 0,70. Butir soal cukup mampu membedakan kelompok siswa berkemampuan tinggi dengan kelompok siswa berkemampuan rendah, hal tersebut dinyatakan oleh rerata indeks daya pembeda 0,269.

Indeks reliabilitas soal 0,679 dapat dinyatakan baik, artinya 68% perbedaan skor yang terdapat pada peserta ujian adalah kemampuan mereka yang sebenarnya, sedangkan 32% perbedaan skor yang ada merupakan kesalahan pengukuran. Menurut Linn (Djemari Mardapi, 1999: 14), kesalahan pengukuran bisa disebabkan oleh variasi acak dalam diri manusia, faktor lingkungan, subjektivitas pengukur, dan alat ukur. Kesalahan baku pengukuran pada perangkat soal hasil analisis secara klasik adalah sebesar 2,736.

Tabel 1. Karakteristik Tes Entry Behaviour dengan Program Iteman

Karakteristik	Nilai
<i>N of Item</i>	40
<i>N of Examinees</i>	85
<i>Mean</i>	12,518
<i>Variance</i>	23,309
<i>Std. Dev.</i>	4,828
<i>Skew</i>	0,900
<i>Kurtosis</i>	1,919
<i>Minimum</i>	3,000
<i>Maximum</i>	29,000
<i>Median</i>	12,000
<i>Alpha</i>	0,679
<i>SEM</i>	2,736
<i>Mean P</i>	0,313
<i>Mean Item-Tot.</i>	0,269
<i>Mean Biserial</i>	0,363

Berdasarkan Tabel 1, rerata tingkat kesulitan dan daya pembeda soal yaitu 0,313 dan 0,269 artinya soal tes ini rata-rata dijawab dengan benar butirnya sebesar 31% dan soal dapat membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah sebesar 33,1%. Hasil analisis soal Entry Behaviour Kimia menunjukkan bahwa tes tersebut memiliki reliabilitas yang cukup baik yaitu 0,679. Kesalahan pengukuran dapat disebabkan oleh beberapa faktor, seperti lingkungan ketika ujian yang kurang mendukung, kondisi psikologis maupun biologis peserta tes atau kurang menguasai materi yang ada.

Berdasarkan analisis program Iteman, soal Entry Behaviour Kimia memiliki kualitas sedang pada tingkat kesulitan, karena 23 (57,5%) butir soal memenuhi kriteria dan 17 (42,5%) butir soal yang tidak memenuhi kriteria. Hal ini didukung juga dengan 90% pengecoh bekerja dengan baik meskipun dari kategori daya beda hanya 55%. Hasil temuan dari 40 butir soal, terdapat 8 butir soal dengan tanda "check the key" dari print out program Iteman, yaitu butir soal nomor 3, 7, 13, 27, 33, 36 dan 39. Tanda tersebut merupakan suatu peringatan bagi pembuat soal untuk melihat kembali pada pilihan jawaban apakah sudah tepat atau belum Selanjutnya dari 40 butir soal Entry Behaviour yang diujikan ditemukan 12 butir soal yang perlu diperiksa kembali sebelum digunakan. Dengan demikian, berdasarkan kriteria yang telah ditetapkan dalam penentuan penerimaan butir soal, maka butir soal yang tidak sesuai dengan teori tes klasik adalah sebanyak 17 (42,5%) butir perlu direvisi dan 4 (10%) butir perlu diganti.

2. Karakteristik Karakteristik Butir Soal berdasarkan Pendekatan IRT

Pada analisis fase pertama, nilai biserial digunakan untuk menentukan apakah suatu butir diikuti pada analisis berikutnya atau tidak. Fase pertama hasil output Bilog-MG menunjukkan bahwa dari 40 butir yang dianalisis terdapat 29 butir soal yang tidak dapat dilanjutkan pada fase kedua. Hal ini untuk mengantisipasi munculnya nilai negatif yang tidak signifikan sehingga akan menyebabkan keputusan bias. Dalam konteks ujian yang beracuan kriteria (criterion referenced test (CRT)) butir soal dengan tingkat kesulitan mudah dapat digunakan.

Butir soal yang tidak diikuti dalam fase pertama dapat dilihat pada Tabel 2 berikut.

Tabel 2. Butir Hasil Analisis Bilog Fase Pertama yang Tidak Dapat Dilanjutkan ke Fase Kedua

Butir	<i>a</i>	<i>b</i>	Kecocokan model	Ket
1	0,468	-1,910	Cocok	Baik
2	0,492	-4,105	TC	TB
4	0,727	-2,689	Cocok	TB
5	0,716	-2,474	Cocok	TB
6	0,572	-2,863	Cocok	TB
7	0,587	-3,432	TC	TB
8	1,174	-1,596	Cocok	Baik
10	0,808	-1,990	Cocok	Baik
12	0,805	-1,040	Cocok	Baik
13	0,792	-1,628	Cocok	Baik
15	0,409	-3,551	TC	TB
16	0,763	-1,912	Cocok	Baik
17	0,978	-1,694	Cocok	Baik
20	0,887	-1,947	Cocok	Baik
23	0,764	-1,896	Cocok	Baik
24	1,279	-1,250	TC	TB
25	0,592	-1,883	Cocok	Baik
29	1,417	-1,749	Cocok	Baik
30	1,035	-1,893	Cocok	Baik
31	1,280	-1,273	Cocok	Baik
32	0,507	-2,387	Cocok	TB
34	1,056	-1,827	Cocok	Baik
35	0,548	-2,146	Cocok	TB
37	0,703	-2,458	TC	TB

Keterangan : TC=Tidak Cocok, TB=Tidak Baik

22	0,000	Tidak Dilanjutkan
26	0,124	Tidak Dilanjutkan
27	0,051	Tidak Dilanjutkan
28	0,000	Tidak Dilanjutkan
29	0,250	Tidak Dilanjutkan
31	0,265	Tidak Dilanjutkan
32	-0,160	Tidak Dilanjutkan
34	0,000	Tidak Dilanjutkan
35	-0,019	Tidak Dilanjutkan
36	-0,028	Tidak Dilanjutkan
37	0,000	Tidak Dilanjutkan
38	-0,325	Tidak Dilanjutkan
39	0,000	Tidak Dilanjutkan
40	0,000	Tidak Dilanjutkan

Suatu butir dengan biserial kurang dari 0,3 tidak diikutkan pada analisis berikutnya, karena akan mengganggu proses analisis yakni munculnya nilai negatif yang tidak signifikan sehingga akan menimbulkan keputusan bias. Keputusan bias yang terjadi akan cenderung menolak butir pada sampel yang besar dan menerima butir pada sampel yang kecil. Menarik jika diperhatikan butir dengan biserial pada output tersebut yang memiliki nilai negatif dimana setelah dikorelasikan dengan bentuk butir pertanyaan, ternyata opsi jawaban yang diberikan memberikan sedikit kebingungan bagi siswa dalam menjawab oleh karna terlalu sulit dibedakan butir jawabannya.

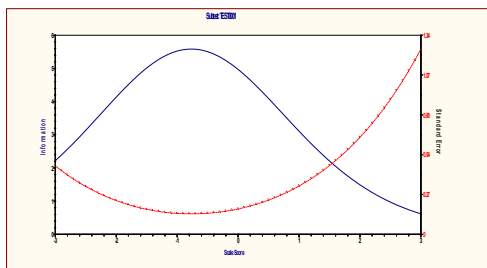
Pada fase kedua hasil analisis model teori respons-butir dapat dikemukakan bahwa dari 11 butir soal yang dianalisis terdapat 10 butir soal yang baik atau 35 % dan 10 butir soal yang tidak baik atau 25%. Butir soal yang baik adalah butir soal dengan tingkat kesukaran soal (*bi*) terletak antara (-2, 2). Hasil analisis estimasi parameter butir soal Entry Behaviour Kimia menggunakan model 1-PL. Hasilnya, jumlah butir soal kimia yang layak dianalisis menggunakan IRT adalah sebanyak 11 butir.

Rerata tingkat kesukaran soal kimia ditunjukkan oleh indeks $-0,41$ yang cenderung mengarah ke soal dengan kategori yang mudah.

Hasil estimasi parameter tingkat kesukaran menunjukkan terdapat 15 butir yang berada pada interval -2 sampai 2 , yang berarti memiliki tingkat kesukaran yang baik. Sebanyak 9 butir soal memiliki indeks kesukaran dibawah $-2,0$, sehingga dapat dikatakan sebagai butir yang mudah.

Hasil estimasi parameter di atas dapat dijadikan informasi untuk menentukan karakteristik butir yang baik. Butir yang baik hendaknya memenuhi 3 persyaratan yaitu, butir cocok model dan tingkat kesulitan -2 sampai $+2$. Berdasarkan kriteria tersebut maka dapat dikatakan bahwa untuk memilih soal yang baik terdapat 9 butir atau 35% butir soal yang memenuhi persyaratan sebagai butir yang baik. Butir soal tersebut adalah butir nomor 3, 18, 19, 23, 24, 31, 33, 35, dan 38

Estimasi parameter butir dengan model 1 parameter logistik pada soal Entry Behaviour Kimia, menghasilkan nilai informasi tes sebesar $5,28$ dengan tingkat kesalahan pengukuran $0,21$. Nilai tersebut tercapai jika parameter kemampuan mahasiswa (θ) sebesar $-0,41$. Perhitungan fungsi informasi tes dihitung dengan bantuan program Excel 2007 for windows. Grafik nilai fungsi informasi tes dengan parameter kemampuan peserta yang sesuai disajikan dalam Gambar 1 berikut.



Gambar 1. Grafik Nilai Fungsi Informasi, Sekor Kemampuan dan Kesalahan Baku

Kesalahan pengukuran perangkat soal Entry Behaviour Kimia berdasarkan hasil analisis menunjukkan bahwa besarnya kesalahan pengukuran sebesar $0,21$. Kesalahan pengukuran dipergunakan untuk memahami kesalahan yang bersifat acak yang mempengaruhi skor peserta yang sebenarnya dalam pelaksanaan tes. Hasil perhitungan kesalahan pengukuran, diperoleh kesalahan pengukuran sebesar $0,21$ dan skor maksimum yang dapat dicapai oleh peserta adalah 29 , maka skor peserta Entry Behaviour Kimia sebenarnya berkisar pada $X_i \pm 0,21$.

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan pada soal Entry Behaviour Kimia dapat disimpulkan sebagai berikut.

1. Berdasarkan pendekatan teori tes klasik terdapat 79,5% butir memiliki tingkat kesulitan butir berfungsi baik, 57% butir daya bedanya belum memenuhi syarat dan 72% butir pengecoh (distractor) berfungsi baik dengan indeks reliabilitas tes $0,679$.
2. Berdasarkan pendekatan teori respons butir terdapat 11 (27,5%) butir soal cocok (fit) dengan model, fungsi informasi maksimum $5,48$ pada $\theta = -0,21$, dan besarnya kesalahan pengukuran (SEM) $0,21$.

DAFTAR PUSTAKA

- Allen, M. J. & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole Publishing Company.
- Badrun Kartowagiran, (2006). Teori klasik dan teori respon butir. Bahan Kuliah. Yogyakarta: Pasca Sarjana Universitas Negeri Yogyakarta
- Brennan, R. L. (Ed.) (2006). Educational measurement (4thed.). Portsmouth: Greenwood Publishing Group
- DeMars, C. E. (2010). Understanding measurement: item response theory. New York: Oxford University Press.
- Hambleton, R.K. & Swaminathan, H. (1985). Item response theory. Boston, MA: Kluwer Inc.
- Hullin, C. L., et al. (1983). Item response theory: Application to psychological measurement. Homewood, IL: Dow Jones-Irwin.
- Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). Fundamental of item response theory. Newbury Park, CA : Sage Publication Inc
- Guilford, JP (1936). Psychometric methods. New York: Mc Graw Hill.
- Mehrens, W.A. & Lehmann, I.J. (1973). Measurement and evaluation in education and psychology. New York: Hold, Rinehart and Wiston, Inc.
- Nitko, A. J., & Brookhart, S. M. (2007). Educational assessment of students, 5th Edition. Upper Saddle River, NJ: Pearson Education, Inc.
- Thorndike, R. M., & Thorndike, C. T. (2010). Measurement and evaluation in psychology and education (8th ed.). Pearson: Upper Saddle River, NJ.
- Mislevy, R.J., & Bock, R. D. (1990). BILOG-3 (3nded.): Item analysis and test scoring with binary logistic models. Mooresville, IN: Scientific Software, Inc.
- Putri, R. I, Dkk. (2015). Keefektifan Strategi React Ditinjau Dari Prestasi Belajar, Kemampuan Penyelesaian Masalah, Koneksi Matematis, Self Efficacy, Jurnal Riset Pendidikan Matematika, Vol 2, No 2, 262-272
- Kubiszyn, T., & Borich, G. (2010). Educational testing & measurement: classroom application and practice (9th ed.). New York: Jhon Wiley and Sons.