

KARAKTERISTIK KEMAMPUAN SISWA MELALUI UJIAN SEKOLAH MATA PELAJARAN KIMIA PADA SMA DI KECAMATAN TELUK AMBON BAGUALA MENGGUNAKAN *CLASSICAL TEST THEORY (CTT)* DAN *ITEM RESPONS THEORY (IRT)* MODEL RASCH

¹Yance Manoppo, ¹Leny Latuny

¹ Program Studi Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan - Universitas Pattimura, Ambon, Indonesia

[*molucanano@yahoo.com](mailto:molucanano@yahoo.com)

Received: 09 October 2020 / Accepted: 13 January 2021 / Published: 19 January 2021

ABSTRACT

This study aims to analyze scales abilities through final exam chemistry at high school in Teluk Ambon Baguala district used items response theory (IRT) with logistics 1-parameter. The subject of the research is the answer sheets of all participants. The research is quantitative research by adopting ex-post facto approach. The analysis of the theory test approach with classic shows that 80% grains have difficulty rating of grains serves both, 100 % grains of resources have not met the requirements, and 100 % grains of having the rapscallion serves both with the reliability test 0,897. An analysis by response theory approach grains show there are 17 (85 %) grains of about suitable (fit) with, model the function of the maximum information 4,12 on $\theta = -0,41$, and the size of a Standard of Error measuring (SEM) 0,43.

Keywords: Final Exam, Classical Test Theory, Item Respons Theory, Rasch Model.

ABSTRAK

Penelitian ini bertujuan untuk mengetahui Karakteristik Kemampuan Siswa Melalui Ujian Sekolah Mata Pelajaran Kimia Pada SMA Di Kecamatan Teluk Ambon Baguala Menggunakan *Classical Test Theory (CTT)* Dan *Item Respons Theory (IRT)* Model Rasch. Subjek penelitian adalah seluruh lembar jawaban peserta Ujian Sekolah Mata Pelajaran Kimia. Penelitian ini merupakan penelitian kuantitatif menggunakan pendekatan *ex-post facto*. Hasil analisis dengan pendekatan teori tes klasik menunjukkan 80% butir memiliki tingkat kesulitan butir berfungsi baik, 100% butir daya bedanya belum memenuhi syarat, dan 100% butir memiliki pengecoh berfungsi baik dengan indeks reliabilitas tes 0,897. Analisis dengan pendekatan teori respons butir Model Rasch menunjukkan terdapat 17 (85%) butir soal cocok (*fit*) dengan model, fungsi informasi maksimum 4,12 pada $\theta = -0,41$, dan besarnya kesalahan pengukuran (*SEM*) 0,43.

Kata kunci: Ujian Sekolah, *Classical Test Theory*, *Item Respons Theory*, *Model Rasch*.

PENDAHULUAN

Evaluasi belajar dalam sistem pendidikan merupakan kegiatan yang biasa dilakukan oleh pihak sekolah terhadap para peserta didik, dengan segala tujuan dan fungsinya. Sejak sekolah dasar, para peserta didik sudah mengenal istilah ujian sekolah, di Indonesia evaluasi belajar dibagi menjadi beberapa tahap, tahapan dibagi mengikuti pola satu tahun pelajaran dalam kalender pendidikan. Seperti ujian semester satu, ujian semester dua, dan ada juga ujian tengah semester yang dilaksanakan dipertengahan semester. Selain itu, siswa yang duduk di kelas IX dan kelas XII diharuskan mengikuti ujian nasional (UN) atau Ujian Sekolah Berbasis Nasional (USBN) agar dapat dinyatakan lulus.

USBN dikembalikan pada esensinya, yaitu asesmen akhir jenjang dilakukan guru dan sekolah. Kelulusan siswa pada akhir jenjang memang merupakan wewenang sekolah yang didasarkan pada penilaian oleh guru. Hal ini sesuai dengan UU Sisdiknas dan juga prinsip pendidikan bahwa yang paling memahami siswa adalah guru. Selain itu, asesmen akhir jenjang

oleh sekolah memungkinkan penilaian yang lebih komprehensif, tidak hanya didasarkan pada tes tertulis pada akhir tahun. Hal ini juga mendorong sekolah mengintensifkan dan memperluas pelibatan guru dalam semua tingkat dalam proses asesmen.

Mendukung kebijakan Menteri Pendidikan dan Kebudayaan melalui program “Merdeka Belajar”, gantinya adalah ujian yang dikelola tiap-tiap sekolah. Ujian tersebut dapat dilaksanakan dalam beragam bentuk asesmen sesuai dengan kompetensi yang diukur. Dari sisi bentuk ujian, guru boleh dan diharapkan menggunakan beragam bentuk asesmen. Hal ini bisa berupa tes tertulis seperti saat ini. Namun guru juga disarankan menggunakan asesmen bentuk lain seperti penugasan, portofolio siswa, dan project kolaboratif. Dari sisi waktu pelaksanaan, asesmen yang menjadi bagian dari ujian ini tidak selalu harus dilakukan di penghujung tahun ajaran sebagaimana ujian konvensional selama ini. Misalnya, nilai ujian akhir jenjang bisa didasarkan pada penilaian portofolio dan penugasan yang dilakukan sejak semester ganjil. Kedua perubahan ini memungkinkan kompetensi siswa dinilai secara lebih komprehensif. Perubahan ini juga memungkinkan penilaian yang lebih terdiferensiasi, sesuai dengan kebutuhan individual siswa.

Guru menjadi lebih merdeka dalam mengajar dan melakukan asesmen siswa. Guru dapat melakukan asesmen yang lebih sesuai untuk kebutuhan siswa dan situasi kelas/sekolahnya. Hal ini juga mendorong guru untuk terus mengembangkan kompetensi profesionalnya, terutama terkait asesmen siswa. Dengan perubahan kebijakan ini, sekolah perlu mendukung praktik asesmen yang baik, yakni asesmen yang berdampak positif pada proses dan hasil belajar siswa. Hal ini bisa dilakukan dengan memfasilitasi guru untuk berkolaborasi mengenai strategi asesmen yang tepat bagi siswa dan kondisi sekolah masing-masing.

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.

Karakteristik soal tes Ujian Sekolah 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.

Berdasarkan uraian di atas untuk menguji mutu setiap butir soal Kimia tes Ujian Sekolah, diperlukan analisis yang dapat memberikan informasi tentang karakteristik setiap butir soal yang nantinya menggambarkan skala kemampuan responden dengan menggunakan teori tes klasik dan teori respon butir.

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 Ujian Sekolah Siswa SMA N 4 Ambon.

Objek penelitian adalah Lembar Respon Jawaban Siswa. 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 Ujian Sekolah 2020 sebanyak 217 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 Ujian Sekolah SMA N 4 Ambon 2020.

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).

a. 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.

b. 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 Rasch. 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 Rasch pada umumnya lebih banyak menghasilkan butir yang fit (cocok). Sehingga untuk keseragaman model analisis, estimasi parameter butir untuk seluruh materi uji dianalisis menggunakan model Rasch.

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 Ujian Sekolah Mata Pelajaran Kimia SMA N 4 Ambon yang direspon oleh peserta ujian sebanyak 217 orang pada Tabel 1 dapat bahwa peserta rata-rata dapat menjawab dengan benar sebesar 12 butir soal (mean = 12,207), berarti rata-rata kurang dari setengah (32,5%) dari jumlah butir soal dapat dijawab dengan benar oleh peserta ujian, skor tertinggi 20 dan terendah 1. Nilai rerata (12,207) yang berdekatan dengan nilai median (12,00). Rata-rata tingkat kesukaran butir soal pada umumnya sedang (mean p = 0,610), 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 0,582.

Indeks reliabilitas soal 0,897 dapat dinyatakan baik, artinya 89,7% perbedaan skor yang terdapat pada peserta ujian adalah kemampuan mereka yang sebenarnya, sedangkan 10,3% 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 1,780.

Tabel 1. Karakteristik Ujian Sekolah Kimia dengan Program Iteman

N of Items	20
N of Examinees	217
Mean	12.207
Variance	30.754
Std. Dev.	5.546
Skew	-0.174
Kurtosis	-1.334
Minimum	1.000
Maximum	20.000
Median	12.000
Alpha	0.897
SEM	1.780
Mean P	0.610
Mean Item-Tot.	0.582
Mean Biserial	0.752

Berdasarkan Tabel 1, rerata tingkat kesulitan dan daya pembeda soal yaitu 0,610 dan 0,582 artinya soal tes ini rata-rata dijawab dengan benar butirnya sebesar 61% dan soal dapat membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah sebesar 61%. Hasil analisis soal Ujian Sekolah Kimia menunjukkan bahwa tes tersebut memiliki reliabilitas yang cukup baik yaitu 0,897. 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 pogram Iteman, Soal Ujian Sekolah Kimia memiliki kualitas sedang sebanyak 16 (80%) butir soal memenuhi kriteria dan 4 (20%) butir soal yang tidak memenuhi kriteria. Hal ini didukung juga dengan 100% pengecoh bekerja dengan baik.

Selanjutnya dari 20 butir soal Ujian Akhir yang diujikan ditemukan 4 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 4 (20%) butir perlu direvisi.

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 20 butir yang dianalisis terdapat 2 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 diikutkan dalam fase pertama dapat dilihat pada **Tabel 2** berikut.

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

ITEM STATISTICS FOR SUBTEST TEST0001							
ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT	ITEM*TEST CORRELATION	
						PEARSON	BISERIAL
1	ITEM0001	217.0	0.0	0.0	99.99	0.000	0.000
2	ITEM0002	0.0	0.0	0.0	99.99	0.000	0.000
3	ITEM0003	217.0	135.0	62.2	-0.50	0.606	0.773
4	ITEM0004	217.0	159.0	73.3	-1.01	0.442	0.594
5	ITEM0005	217.0	182.0	83.9	-1.65	0.523	0.787
6	ITEM0006	217.0	92.0	42.4	0.31	0.554	0.699
7	ITEM0007	217.0	146.0	67.3	-0.72	0.565	0.735
8	ITEM0008	217.0	103.0	47.5	0.10	0.602	0.755
9	ITEM0009	217.0	98.0	45.2	0.19	0.357	0.449
10	ITEM0010	217.0	147.0	67.7	-0.74	0.533	0.694
11	ITEM0011	217.0	158.0	72.8	-0.99	0.423	0.567
12	ITEM0012	217.0	120.0	55.3	-0.21	0.410	0.515
13	ITEM0013	217.0	125.0	57.6	-0.31	0.613	0.774
14	ITEM0014	217.0	141.0	65.0	-0.62	0.653	0.841
15	ITEM0015	217.0	121.0	55.8	-0.23	0.435	0.547
16	ITEM0016	217.0	116.0	53.5	-0.14	0.541	0.679
17	ITEM0017	217.0	123.0	56.7	-0.27	0.565	0.712
18	ITEM0018	217.0	146.0	67.3	-0.72	0.567	0.737
19	ITEM0019	217.0	139.0	64.1	-0.58	0.526	0.675
20	ITEM0020	217.0	154.0	71.0	-0.89	0.489	0.648

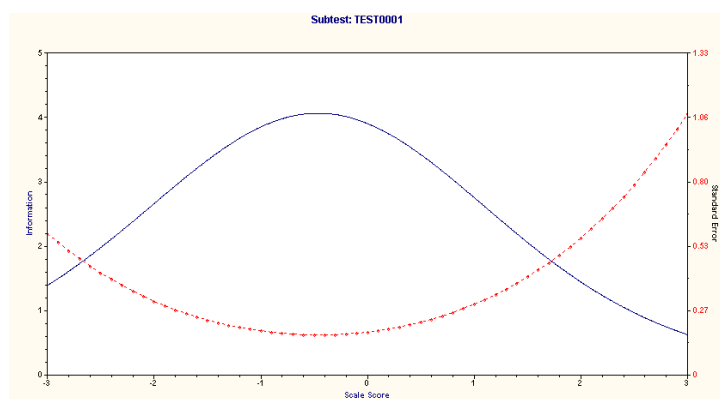
Suatu butir dengan biserial kurang dari 0,3 tidak diikutkan pada analisis berikutnya, karena akan mengganggu proses analisis yakni munculnya nilai negatif yang tidak signifikan sehingga akan menimbulkan keputusan bias. Keputusan bias yang terjadi akan cenderung menolak butir pada sampel yang besar dan 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 18 butir soal yang dianalisis terdapat 17 butir soal yang baik atau 94 % dan 1 butir soal yang tidak baik atau 6%. Butir soal yang baik adalah butir soal dengan tingkat kesukaran soal (bi) terletak antara (-2, 2). Hasil analisis estimasi parameter butir soal Ujian Sekolah Kimia menggunakan model Rach. Hasilnya, jumlah butir soal kimia yang layak dianalisis menggunakan IRT adalah sebanyak 18 butir.

Hasil estimasi parameter tingkat kesukaran menunjukkan terdapat 17 butir yang berada pada interval -2 sampai 2, yang berarti memiliki tingkat kesukaran yang baik. Sebanyak 1 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 17 butir atau 94% butir soal yang memenuhi persyaratan sebagai butir yang baik.

Estimasi parameter butir dengan model Rasch pada soal Ujian Sekolah Kimia, menghasilkan nilai informasi tes sebesar 4,12 dengan tingkat kesalahan pengukuran -0,43. Nilai tersebut tercapai jika parameter kemampuan mahasiswa(θ) sebesar -0,41. Perhitungan fungsi informasi tes dihitung dengan bantuan program Excel 2019 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 Ujian Akhir Kimia berdasarkan hasil analisis menunjukkan bahwa besarnya kesalahan pengukuran sebesar 0,78. 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 0,78 dan skor maksimum yang dapat dicapai oleh peserta adalah 20.

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan pada soal Ujian Akhir Kimia dapat disimpulkan sebagai berikut.

1. Berdasarkan pendekatan teori tes klasik terdapat 80% butir memiliki tingkat kesulitan butir berfungsi baik, 100% butir daya bedanya memenuhi syarat dan 100% butir pengecoh (distractor) berfungsi baik dengan indeks reliabilitas tes 0,897.
2. Berdasarkan pendekatan teori respons butir terdapat 17 (85%) butir soal cocok (fit) dengan model, fungsi informasi maksimum 4,12 pada $\theta = -0,41$, dan besarnya kesalahan pengukuran (SEM) 0,43.

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: clashroom application and practice* (9th ed.). New York: Jhon Wiley and Sons.