



## Analisis uas biologi kelas x dengan teori tes klasik dan item response theory (rasch model)

Sri Handayani<sup>a, 1,\*</sup>

<sup>a</sup> SMK Negeri 1 Lokop, Kabupaten Aceh Timur, Aceh, 24392, Indonesia.

<sup>1</sup> [eshada20@gmail.com](mailto:eshada20@gmail.com) \*

\* Corresponding author.

### INFORMASI ARTIKEL

#### Lini Masa Artikel

Draft diterima : 2022-07-11  
 Revisi diterima : 2022-09-17  
 Diterbitkan : 2022-10-04

#### Kata Kunci

Analisis soal;  
 Tes klasik;  
 Model RASCH.

#### ABSTRAK

Analisis soal merupakan suatu proses untuk menguji kualitas setiap item soal. Analisis soal UAS Biologi Kelas X ini bertujuan untuk menguji kualitas soal ujian akhir semester genap kelas X mata pelajaran biologi pada tahun pelajaran 2019/2020. Analisis ini menggunakan teori tes klasik yang dilakukan dengan program Microsoft Excel dan item response theory dengan menggunakan model Rasch. Kualitas item soal yang diuji meliputi tingkat kesukaran, daya pembeda soal, efektifitas distractor, validitas dan reliabilitas. Soal berjumlah 40 butir pilihan ganda yang diujikan kepada 41 siswa. Terdapat perbedaan hasil analisis dengan CTT dan Model Rasch. Hasil analisis dengan CTT menunjukkan tingkat kesukaran terdapat dua kategori; daya pembeda 5 kategori; efektifitas distraktor yang berfungsi 72,5% dan yang tidak berfungsi 27,5%; validitas butir soal yang valid 65% dan tidak valid 35%; reliabilitas soal secara umum adalah reliabel dengan kategori sangat tinggi (nilai  $r_{11}=0,855$ ). Analisis dengan model Rasch menunjukkan tingkat kesukaran terdapat tiga kategori; daya pembeda 3 kategori; efektifitas distraktor yang berfungsi 85% dan yang tidak berfungsi 15%; validitas butir soal yang valid dan tidak valid seimbang yaitu 50%; reliabilitas item sebesar 0,79 (cukup) dan reliabilitas person 0,82 (bagus).

#### ABSTRACT

#### Class X Biology Exam analysis with classical test theory and item response theory (rasch model)

Problem analysis is a process to test the quality of each item of the question. The analysis aims to test the quality of the final exam questions. This analysis uses classical test theory which is carried out with the Microsoft Excel program and item response theory using the Rasch model. The quality of the items tested included the level of difficulty, discriminatory power of questions, distractor effectiveness, validity and reliability. The questions totaled 40 multiple choice items which were tested on 41 students. There are differences in the results of the analysis with the CTT and the Rasch model. The results of the CTT analysis show that there are two categories of difficulty levels; distinguishing power of 5 categories; the effectiveness of the distractors that functioned 72.5% and those that did not functioned 27.5%; the validity of the items that are valid 65% and invalid 35%; The reliability of the questions in general is reliable with a very high category ( $r_{11} = 0.855$ ). The analysis using the Rasch model shows that there are three categories of difficulty levels; distinguishing power of 3 categories; the effectiveness of the distractors that work 85% and those that don't work 15%; the validity of the valid and invalid items is balanced, namely 50%; item reliability is 0.79 (enough) and person reliability is 0.82 (good).

#### Cara Sitasi Artikel Ini (APA Style):

Handayani, S. (2022). Analisis uas biologi kelas x dengan teori tes klasik dan item response theory (rasch model). *Bio-Pedagogi*. 11(2), 76-90. <https://dx.doi.org/10.20961/bio-pedagogi.v11i2.63177>

Artikel ini berakses bebas dibawah lisensi [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).



---

## PENDAHULUAN

Dunia pendidikan pasti selalu berkaitan dengan evaluasi/penilaian, baik evaluasi proses, evaluasi hasil maupun evaluasi program pendidikan. Evaluasi proses belajar menyangkut kegiatan guru dalam pembelajaran, interaksi guru dan peserta didik maupun keterlaksanaan proses pembelajaran tersebut. Evaluasi hasil belajar menyangkut bagaimana hasil belajar siswa pada ranah kognitif, psikomotorik dan afektif baik jangka pendek maupun jangka panjang. Evaluasi program pendidikan berarti cakupannya lebih luas lagi yaitu mencakup seluruh komponen dalam pendidikan, misalnya tujuan, sasaran dan isi program. Evaluasi atau penilaian merupakan bagian dari pelaksanaan pendidikan dan secara keseluruhan tidak dapat dipisahkan dari kegiatan mengajar. Evaluasi proses bertujuan menilai keefektifan dan efisiensi kegiatan pembelajaran sebagai bahan untuk perbaikan dan penyempurnaan program pelaksanaannya (Sitorus et al., 2017)

Ujian Akhir Semester (UAS) merupakan salah satu alat evaluasi hasil belajar siswa, yang digunakan guru untuk mengetahui tingkat pencapaian kompetensi siswa di akhir pembelajaran selama satu semester. Ada berbagai macam bentuk soal yang dapat digunakan untuk mengukur ketercapaian kompetensi siswa, misalnya soal pilihan ganda, uraian panjang, maupun jawaban singkat. Biasanya bentuk soal pilihan ganda adalah yang paling sering digunakan oleh guru. Hal ini berdasarkan pertimbangan bahwa soal pilihan ganda lebih mudah dalam penentuan skornya sehingga memudahkan saat mengoreksi/memeriksa jawaban siswa dan tentu saja prosesnya akan lebih cepat dibandingkan soal yang berbentuk uraian. Apalagi sekarang hal ini lebih dimudahkan lagi dengan adanya program atau aplikasi komputer. Selain itu, cakupan materi yang dapat dievaluasi dengan menggunakan pilihan ganda juga lebih luas dan banyak daripada soal uraian karena biasanya jumlah soal pilihan ganda lebih banyak daripada soal uraian untuk jangka waktu pengerjaan yang sama.

Analisis butir soal merupakan suatu proses untuk menguji kualitas setiap item soal. Analisis butir soal adalah suatu langkah atau cara yang sistematis (Amalia & Widayati, (2012). Hal ini karena hasil dari analisis setiap butir soal tersebut bisa menunjukkan informasi yang bersifat spesifik. Selain itu, analisis butir soal juga untuk menentukan soal yang cacat atau tidak berfungsi dalam penggunaannya (Kurniawan et al., 2017). Uji kriteria instrumen penilaian hasil belajar melalui analisis butir soal penting dilaksanakan untuk mengetahui baik tidaknya butir-butir soal yang diujikan untuk mengukur kemampuan siswa (Yustika et al., 2014). Analisis soal bertujuan untuk menguji atau mengetahui kualitas dari butir soal. Hal ini meliputi kegiatan menelaah maupun mengkaji setiap item soal agar diperoleh soal yang bermutu atau memiliki kualitas bagus. Selain itu juga membantu dalam merevisi soal ataupun membuang soal yang tidak baik. Analisis soal juga dapat untuk mengetahui kemampuan siswa (analisis siswa). Dari analisis butir soal ini dapat diidentifikasi butir soal mana yang baik dan tidak baik serta butir soal mana yang dapat masuk ke dalam bank soal, direvisi, atau dibuang (Oktanin, 2015)

Soal dikatakan bermutu jika soal yang telah dibuat dapat memberikan informasi dengan tepat sesuai dengan tujuannya, antara lain mengetahui siswa yang memiliki kemampuan tinggi atau rendah terhadap materi yang diujikan, mengetahui kualitas soal terkait tingkat kesukaran soal, valid atau tidak serta daya pembedanya, kira-kira apakah soal tersebut dapat membedakan antara siswa yang memiliki kemampuan tinggi dan rendah atau tidak. Oleh karena itu, analisis butir soal penting dalam memperbaiki proses belajar mengajar. Hal ini karena melalui soal yang telah diujikan kepada siswa, guru dapat juga menganalisis materi pelajaran apa yang susah maupun yang mudah dipahami oleh siswa.

Analisis butir soal dapat dilakukan secara klasik dan modern. Teori tes klasik dirasa memiliki beberapa kelemahan, misalnya tingkat kesukaran serta daya beda butir soal tergantung pada kelompok peserta yang mengerjakan sehingga para ahli pengukuran berusaha mencari alternatif sebagai upaya mengatasi kelemahan-kelemahan yang ada pada teori klasik. Ciri penting model

Rasch adalah tidak mengandung parameter diskriminasi dan parameter terkaan. Pada model ini, diasumsikan bahwa kesukaran butir merupakan satu-satunya karakteristik butir yang mempengaruhi kinerja tes (Alfarisa & Purnama, (2019)

Teori tes klasik (*Classical Test Theory/CTT*) merupakan metode analisis soal yang sering digunakan. CTT meliputi dua komponen yaitu tingkat kesukaran dan daya beda. CTT mengalami perkembangan yaitu *Item Response Theory (IRT)*. IRT dapat merepresentasikan kemampuan setiap individu pada berbagai tingkat kemampuan siswa yang mengikuti ujian tersebut (Fernanda & Hidayah, 2020).

Kebanyakan guru maupun calon guru telah memiliki kemampuan yang memadai tentang bagaimana merencanakan proses pembelajaran (RPP), baik merumuskan tujuan pembelajaran sesuai indikator, menentukan bahan pelajaran secara terperinci, memilih dan menentukan metode pembelajaran sesuai materi, serta menyiapkan alat dan media pembelajaran. Namun, masih banyak yang belum memahami dalam menentukan penilaian hasil belajar bahkan bagaimana menganalisis hasil belajar tersebut. Padahal hal ini penting sebagai bahan evaluasi terkait hasil belajar siswa dalam mengukur ketercapaian kompetensi dan refleksi bagi guru untuk pelaksanaan pembelajaran selanjutnya. Oleh karena itu, kemampuan dan ketrampilan melakukan evaluasi dan menganalisisnya harus dimiliki oleh setiap guru maupun calon guru. Menguasai kemampuan ini tidaklah mudah, tetapi memerlukan latihan dan pengalaman lapangan yang memadai. Kemampuan dalam bidang evaluasi hasil belajar bukan hanya bermanfaat dalam proses belajar mengajar, tetapi juga bermanfaat dalam rangka penelitian ilmiah, yaitu tentang bagaimana cara membuat alat ukur yang valid dan reliabel.

Kajian ini membahas tentang beberapa hal terkait analisis butir soal dari instrumen soal UAS Biologi kelas X SMK yang telah dibuat agar bisa dilihat kualitasnya yaitu meliputi kriteria apakah instrumen ini baik atau tidak baik, yang terdiri dari validitas butir soal, reliabilitas instrumen, tingkat kesukaran soal, daya pembeda, dan efektifitas distraktor.

## METODE

Metode analisis soal yang digunakan dalam kajian ini adalah teori tes klasik/classical test theory (CTT) melalui aplikasi Microsoft excel dan item response theory dengan model rasch melalui aplikasi Winstep. Metode CTT dipilih karena dianggap lebih mudah sedangkan pemilihan model rasch karena model ini memungkinkan analisis data secara lebih mendalam, bahkan pada tingkat per siswa maupun per item soal. Melalui hasil pengukuran dengan model rasch, kesimpulan yang lebih akurat dan bermakna dapat dibuat atas data yang terkumpul (Sumintono, 2018). Instrumen tes merupakan soal ujian akhir sekolah (UAS) semester genap Kelas X pada mata pelajaran Biologi. Instrumen terdiri dari 40 butir soal pilihan ganda yang diujikan kepada 41 siswa kelas X di salah satu SMK Negeri di Aceh Timur.

## HASIL DAN PEMBAHASAN

Hasil analisis UAS Biologi Kelas X, yang meliputi tingkat kesukaran (*difficulty indexes*), daya pembeda, efektifitas distraktor, validitas dan reliabilitas menggunakan metode CTT disajikan dalam Lampiran 1.

Semakin kecil nilai DI maka soal semakin sulit. Sedangkan nilai DP, semakin kecil maka semakin rendah daya pembeda soal tersebut. Sebagian besar soal sudah memiliki efek distraktor, tetapi pada soal nomor 32 ada dua opsi distractor yang tidak dipilih oleh siswa yaitu A dan E (0%). Namun, opsi yang menjadi kunci jawaban (opsi D) pun memiliki persentasi paling tinggi yang dijawab siswa yaitu sebesar 49%. Soal ini memiliki nilai ID termasuk kategori sedang (ID=0,4878) dan nilai DP tergolong sangat tinggi (DP= 0,7272) yang berarti soal ini dapat membedakan antara siswa yang memiliki kemampuan tinggi dan siswa yang kemampuannya rendah. Namun secara umum soal ini termasuk soal yang valid dan dapat mengukur apa yang seharusnya diukur.

Hasil menunjukkan bahwa soal dengan nilai DI rendah (sulit) ataupun DP rendah maka tergolong soal yang tidak valid, yaitu misalnya pada soal nomor 3, 4, 6, 11, 15, 20, 25, 30, 31, 35, 36 dan 37. Namun ternyata ada juga soal dengan ID dan DP yang termasuk kategori sedang namun termasuk soal yang tidak valid, misalnya pada soal nomor 16 dan 23. Berikut ini disajikan ringkasan kualitas butir soal UAS Biologi kelas X dengan menggunakan CTT.

Tabel 1. Ringkasan Persentase Kualitas Butir Soal

No.	Indikator							
	DI	%	DP	%	Distraktor	%	Validitas	%
1	Mudah	0	Sangat Rendah	5	Berfungsi ( $\geq 5\%$ )	72,5	Valid	65
2	Sedang	57,5	Rendah	22,5	Tidak Berfungsi ( $< 5\%$ )	27,5	Tidak Valid	35
3	Sulit	42,5	Sedang	20				
4			Tinggi	30				
5			Sangat Tinggi	22,5				

Tabel 1 menunjukkan bahwa persentase tingkat kesukaran (DI) paling tinggi adalah pada soal dengan kategori sedang yaitu sebesar 57,5%. Persentase tertinggi pada indikator daya pembeda (DP) adalah pada soal dengan kategori daya pembedanya tinggi yaitu sebesar 30%. Indikator keefektifan distractor yaitu 72,5% opsi pengecoh adalah berfungsi, artinya distractor/pengecoh tersebut dipilih  $\geq 5\%$  oleh siswa-siswa peserta ujian. Validitas soal menunjukkan 65% butir soal adalah valid. Secara keseluruhan instrumen soal UAS ini adalah reliabel yaitu dengan nilai *Cronbach Alpha* ( $r_{11}$ ) = 0,855 dan termasuk kategori sangat tinggi.

Menurut (Rahayu et al., n.d.), tingkat kesukaran suatu butir soal adalah proporsi antara siswa yang dapat menjawab benar soal tersebut dengan jumlah seluruh siswa yang mengikuti tes/ujian. Menurut (Erfan et al., 2020), soal yang berkualitas baik adalah soal yang tidak terlalu mudah namun juga tidak terlalu sulit. Dengan demikian soal yang baik adalah soal yang memiliki tingkat kesukaran dengan kategori sedang. Tingkat kesukaran item soal untuk pilihan ganda dihitung dengan rumus:

$$p = \frac{\text{jumlah siswa yang menjawab benar}}{\text{jumlah siswa yang mengikuti tes}} \quad (\text{Nitko, 2011})$$

Tabel 2. Kategori Tingkat Kesukaran Soal

Koefisien	Kategori
$p < 0,3$	Sukar
$0,3 \leq p \leq 0,7$	Sedang
$p > 0,7$	Mudah

(Sudjana, 2014)

Hasil analisis menunjukkan bahwa tingkat kesukaran soal tidak terdapat yang berkategori mudah. Soal yang berkategori sedang dan sulit hampir seimbang. Berarti instrumen ini belum memenuhi proporsi perbandingan yang baik menurut Sudjana (2014), proporsi perbandingan 3-5-2 atau 30% butir soal dengan kategori mudah, 50% butir soal dengan kategori sedang, dan 20% butir soal dengan kategori sukar.

Analisis daya pembeda berarti proses pengkajian setiap butir soal dengan tujuan untuk mengetahui kemampuan soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dengan peserta didik yang memiliki kemampuan rendah (Septiana, 2016). Analisis daya pembeda bertujuan untuk mengetahui kemampuan soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dengan peserta didik yang memiliki kemampuan rendah. Daya pembeda berfungsi untuk membedakan antara siswa yang berkemampuan tinggi dan siswa berkemampuan rendah. Daya pembeda butir soal dapat diketahui dengan seberapa besar indeks diskriminasi. Daya pembeda berfungsi untuk mendeteksi perbedaan individual di antara para peserta tes sekecil-kecilnya (Iskandar & Rizal, 2018)

Hasil analisis juga menunjukkan bahwa proporsi soal yang memiliki daya pembeda sangat rendah, rendah, sedang, tinggi dan sangat tinggi tidak merata. Namun soal dengan kategori sedang, tinggi dan sangat tinggi menunjukkan proporsi yang seimbang. Hal ini sesuai pendapat Warju, Ariyanto, Soeryanto, & Trisna. (2020) bahwa soal dapat dikatakan telah memenuhi kriteria apabila terletak pada kategori sedang, tinggi dan sangat tinggi. Sedangkan soal dikatakan tidak memenuhi kriteria apabila masuk kategori rendah dan sangat rendah. Soal yang tidak memenuhi kriteria berarti soal tersebut belum mampu membedakan siswa yang memiliki kemampuan tinggi dengan siswa yang memiliki kemampuan rendah.

Suatu distraktor (pengecoh) dikatakan efektif jika dipilih oleh beberapa peserta tes atau minimal dipilih oleh 5% peserta tes. Keefektifan distraktor dikatakan berfungsi jika opsi pengecoh tersebut banyak dipilih oleh peserta tes dari kelompok bawah. Namun jika banyak kelompok atas yang memilih opsi pengecoh tersebut maka dikatakan keefektifan pengecoh tidak berfungsi (Arifin, 2017a)

Analisa juga menunjukkan bahwa terdapat pengecoh yang tidak berfungsi atau keefektifannya kurang signifikan misalnya opsi D nomor 9, opsi A dan E pada nomor 32. Keseluruhan ada 11 butir soal (27,5%) yang distraktornya tidak berfungsi yaitu nomor 9, 10, 11, 12, 23, 26, 27, 28, 32, 39 dan 40. Pada butir-butir soal ini terdapat opsi jawaban yang dijawab siswa kurang dari 5%. Hal ini menunjukkan bahwa pengecoh terlalu mencolok perbedaannya bila dibandingkan pilihan jawaban yang lainnya sehingga kurang menimbulkan ketertarikan siswa untuk memilih pengecoh tersebut.

Validitas menunjukkan seberapa tepat dan cermat suatu butir soal dapat mengukur apa yang akan diukur, kaitannya dengan pembelajaran tentunya hal ini adalah hasil belajar siswa. Soal dikatakan valid jika mampu mengukur obyek tersebut sesuai kriteria yang telah ditentukan. Validitas soal dikatakan tinggi jika soal tersebut dapat melakukan fungsi ukurnya artinya memberikan hasil pengukuran yang tepat sesuai maksud atau tujuan dilakukannya pengukuran. Jika validitas tes rendah akan menghasilkan data yang tidak relevan dengan tujuan pengukuran (Amalia & Widayati, 2012). Validitas butir soal dilihat berdasarkan perhitungan korelasi product moment. Soal dikatakan valid jika koefisien korelasi  $r_{xy} > r_{tabel}$  dan jika  $r_{xy} \leq r_{tabel}$  maka soal dapat dianggap tidak valid (Erfan et al., 2020). Terdapat 12 soal yang tidak valid (35%) dan sisanya (65 %) soal termasuk valid.

Reliabilitas adalah tingkat kekonsistenan dari suatu instrumen. Reliabilitas tes menyatakan bahwa suatu tes sudah teliti dan dapat dipercaya sesuai dengan kriteria yang telah ditentukan. Apabila suatu tes selalu memberikan hasil yang sama meskipun diteskan pada kelompok yang sama dan pada waktu atau kesempatan yang berbeda maka tes atau instrument tersebut dikatakan reliabel (Arifin, 2017). Analisis reliabilitas dilakukan terhadap sejumlah butir soal tes secara keseluruhan, namun analisis validitas dilakukan pada masing-masing butir soal tes (Son, 2019). Hasil analisis menunjukkan bahwa soal UAS Biologi Kelas X adalah reliabel.



Dari *Wright Map* juga terlihat bahwa soal nomor 21 dan 37 memiliki bobot yang paling tinggi karena tidak ada satu siswa pun yang mampu menjawab soal tersebut. Soal nomor 20, 25, 35 dan 13 juga demikian, tidak ada siswa yang mampu menjawabnya. Soal-soal tersebut memiliki nilai logit tertinggi sehingga tidak ada siswa yang mampu menjawabnya. Oleh karena itu, soal-soal tersebut dikenal sebagai *item free person*. Nilai logit soal lebih tinggi daripada nilai logit siswa. Siswa yang memiliki kemampuan paling tinggi atau siswa yang dianggap paling pandai yaitu siswa dengan kode 17P (artinya siswa dengan nomor 17 dan berjenis kelamin perempuan) mampu menjawab soal nomor 36 dan 7 serta soal lainnya yang memiliki nilai logit dibawahnya misalnya juga soal nomor 15, 11, 38, 6, 28 dan 29. Siswa 10L dan 19P memiliki nilai logit yang sama dan mampu menjawab soal dengan logit yang setara juga yaitu soal nomor 11, 38 dan 6 serta mampu menjawab soal pada nilai logit di bawahnya yaitu soal nomor 28 dan 29. Namun, kedua siswa tersebut tidak akan mampu menjawab soal nomor 15 karena memiliki nilai logit yang lebih tinggi sehingga soal tersebut dianggap soal sulit bagi kedua siswa tersebut, apalagi soal nomor 36 dan 7.

Soal yang paling baik adalah soal yang berada pada posisi mistar logit M. Pada instrumen soal UAS Biologi Kelas X ini adalah soal nomor 27 dan 39. *Wright Map* juga menunjukkan bahwa banyak siswa yang memiliki nilai logit di bawah nilai logit soal. Hal ini berarti siswa-siswa tersebut memiliki kemampuan yang lebih rendah daripada bobot soal. Berdasarkan terlihat bahwa soal yang memiliki bobot paling rendah adalah soal nomor 10, namun ternyata soal ini hanya mampu dijawab siswa 04L, 21P dan 38P. Siswa 24P, 27L, 34P, 37P, 09P, 16P, 33L, 35P, 41L, 05L, 07P, 01L, 28L, dan 23 P tidak mampu menjawab soal nomor 10 meskipun soal tersebut memiliki bobot yang paling rendah.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.		INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXPS%		
21	3	41	2.00	.62	.80	-.31	.35	-.91	.46	.23	92.7	92.6	521	
37	3	41	2.00	.62	1.02	-.21	1.41	-.71	.16	.23	92.7	92.6	537	
20	5	41	1.39	.50	.97	.01	1.06	.31	.28	.29	87.8	87.7	520	
25	5	41	1.39	.50	1.15	-.51	1.45	.91	.11	.29	87.8	87.7	525	
35	5	41	1.39	.50	1.02	-.22	1.40	2.01	.14	.29	87.8	87.7	535	
13	6	41	1.16	.47	.85	-.51	.71	-.51	.46	.31	85.4	85.3	513	
7	7	41	.95	.44	.83	-.61	.55	-1.11	.53	.33	85.4	82.9	57	
36	7	41	.95	.44	1.19	.81	1.43	1.01	.11	.33	80.5	82.9	536	
15	8	41	.77	.42	1.21	-.91	1.64	1.51	.08	.35	78.0	80.6	515	
6	9	41	.60	.41	1.10	-.51	1.10	.41	.26	.36	75.6	78.4	56	
11	9	41	.60	.41	1.22	1.11	1.44	1.21	.11	.36	80.5	78.4	511	
38	9	41	.60	.41	.91	-.41	.71	-.81	.48	.36	80.5	78.4	538	
28	10	41	.44	.39	1.01	-.11	1.16	.61	.33	.37	75.6	76.5	528	
29	10	41	.44	.39	1.03	-.21	1.22	.81	.31	.37	80.5	76.5	529	
30	11	41	.29	.38	1.45	2.21	1.57	1.81	-.05	.39	68.3	74.9	530	
4	12	41	.14	.37	1.24	1.41	1.36	1.31	.14	.39	70.7	73.7	54	
12	12	41	.14	.37	.90	-.51	.87	-.41	.49	.39	75.6	73.7	512	
27	13	41	.01	.37	.96	-.21	1.02	.11	.42	.40	80.5	72.8	527	
39	13	41	.01	.37	.81	-1.21	.70	-1.31	.60	.40	80.5	72.8	539	
2	14	41	-.13	.36	.79	-1.41	.75	-1.21	.60	.41	87.8	72.1	52	
3	14	41	-.13	.36	1.18	1.11	1.35	1.51	.21	.41	63.4	72.1	53	
22	14	41	-.13	.36	.81	-1.21	.71	-1.41	.60	.41	78.0	72.1	522	
33	15	41	-.25	.36	1.08	-.51	1.07	.41	.34	.41	70.7	71.5	533	
5	17	41	-.50	.35	1.04	-.31	1.00	.11	.39	.42	65.9	70.4	55	
34	17	41	-.50	.35	.73	-2.01	.72	-1.71	.67	.42	85.4	70.4	534	
23	18	41	-.62	.35	1.16	1.11	1.23	1.31	.26	.42	65.9	69.9	523	
24	18	41	-.62	.35	1.11	.81	1.21	1.21	.30	.42	70.7	69.9	524	
17	19	41	-.74	.35	.79	-1.61	.81	-1.21	.61	.42	82.9	69.4	517	
19	19	41	-.74	.35	.88	-.91	.87	-.81	.53	.42	73.2	69.4	519	
26	19	41	-.74	.35	.81	-1.41	.83	-1.01	.59	.42	82.9	69.4	526	
40	19	41	-.74	.35	.87	-.91	.84	-1.01	.54	.42	78.0	69.4	540	
14	20	41	-.86	.34	1.02	-.21	.99	.01	.41	.42	63.4	68.9	514	
16	20	41	-.86	.34	1.17	1.21	1.23	1.41	.25	.42	58.5	68.9	516	
32	20	41	-.86	.34	.85	-1.11	.82	-1.11	.57	.42	73.2	68.9	532	
1	21	41	-.98	.34	.79	-1.61	.74	-1.71	.62	.42	78.0	68.5	51	
18	21	41	-.98	.34	.91	-.61	.88	-.71	.51	.42	73.2	68.5	518	
31	21	41	-.98	.34	1.51	3.41	1.56	3.01	-.06	.42	43.9	68.5	531	
8	22	41	-1.10	.34	.83	-1.31	.89	-.61	.56	.42	78.0	68.3	58	
9	24	41	-1.33	.35	.87	-1.01	.87	-.71	.53	.41	80.5	68.2	59	
10	25	41	-1.46	.35	.76	-1.91	.66	-1.91	.65	.40	80.5	68.4	510	
MEAN	13.9	41.0	.00	.39	.99	-.11	1.06	.01			77.0	75.0		
S.D.	6.1	.0	.91	.07	.19	1.21	.37	1.21			9.5	7.2		

Gambar 2. Tingkat Kesukaran dengan Model Rasch

Hasil analisis *Output Measure Order* (Gambar 2) menunjukkan bahwa soal dengan bobot tertinggi yaitu soal nomor 21 dan 37 memiliki nilai *measure* 2,00 dan soal nomor 20, 25 dan 35 dengan bobot soal lebih rendah nilai *measure*-nya adalah 1,39 serta soal nomor 13 memiliki nilai *measure* 1,16. Soal-soal tersebut tidak mampu dijawab oleh satu orang siswa pun. Soal yang mampu dijawab oleh siswa yang dianggap memiliki abilitas tinggi adalah soal nomor 7 dan 36, memiliki nilai *measure* 0,95. Dilihat model S.E (standar error), bahwa soal nomor yang memiliki nilai S.E lebih dari 0,5 atau mendekati 0,5 maka dikatakan sebagai soal yang bagus karena memiliki

soal tersebut memiliki ketelitian yang baik. Pada tabel tersebut bisa dikatakan bahwa soal nomor 21, 37, 20, 25, 35 dan 13 tergolong soal yang bagus. Nomor 21 dan 37 model S.E-nya memiliki nilai 0,62, nomor 20, 25 dan 35 adalah 0,50 dan nomor 13 adalah 0,47. Namun, justru soal-soal ini tidak mampu dijawab oleh siswa bahkan oleh siswa yang dianggap paling pandai sekalipun. Siswa yang memiliki kemampuan tinggi hanya mampu mengerjakan soal dengan nilai measure dan model S.E yang lebih rendah daripada itu.

Tingkat kesukaran setiap butir soal dapat dianalisis dengan melihat *measure logit*. Tingkat kesukaran butir soal dikategorikan sulit, sedang dan mudah berdasarkan nilai standar deviasi (SD) dan *mean measure*. Standar deviasi dari *mean measure* adalah 0,91 sehingga tingkat kesukaran butir soal UAS Biologi Kelas X dapat dikelompokkan sebagai berikut.

Tabel 3. Ringkasan Tingkat Kesukaran Butir Soal UAS Biologi dengan Model Rasch

Nilai Measure Logit	Kategori	Nomor Soal	Jumlah Butir Soal	Persentase (%)
<i>Measure logit</i> > 0,91	Sulit	7, 13, 20, 21, 25, 35, 36, 37	8	20
-0,91 < <i>Measure Logit</i> < 0,91	Sedang	2, 3, 4, 5, 6, 7, 11, 12, 14, 16, 17, 19, 22, 23, 24, 26, 27, 28, 29, 30, 32, 33, 34, 38, 39, 40	26	65
<i>Measure Logit</i> < -0,91	Mudah	1, 8, 9, 10, 18, 31	6	15

Soal dengan kategori sulit sebanyak 20% yaitu meliputi nomor 7, 13, 20, 21, 25, 35, 36 dan 37. Soal dengan kategori sedang sebanyak 65% dan soal dengan kategori mudah sebanyak 15%. Perbandingan persentase bobot soal berdasarkan SD ini lebih mendekati kriteria perbandingan yang baik menurut Sudjana (2014) yaitu proporsi perbandingan soal mudah, sedang dan sulit adalah 3-5-2. Apabila dibandingkan dengan analisis hasil CTT ternyata terdapat perbedaan hasil analisis pada tingkat kesukaran butir soal UAS Biologi Kelas X. Pada analisis dengan menggunakan CTT tidak terdapat soal dengan kategori mudah sedangkan pada analisis dengan Rasch model terdapat 6 butir soal (15%) yang termasuk kategori mudah. Hal ini karena adanya perbedaan kriteria penentuan batas suatu soal dikategorikan sulit, sedang dan mudah. Pada CTT, kategori soal sedang adalah pada rentang antara 0,3 sampai 0,7 sedangkan pada Rasch model adalah berdasarkan SD yaitu soal kategori sedang pada rentang antara -0,91 sampai 0,91.

```

SUMMARY OF 40 MEASURED Item
-----
      TOTAL      MODEL      INFIT      OUTFIT
      SCORE      COUNT      MEASURE      ERROR      MNSQ      ZSTD      MNSQ      ZSTD
-----
MEAN    13.9      41.0          .00      .39          .99      -.1      1.06      .0
S.D.     6.1         .0          .91      .07          .19      1.2      .37      1.2
MAX.    25.0      41.0          2.00      .62          1.51      3.4      2.40      3.0
MIN.     3.0      41.0          -1.46     .34          .73      -2.0     .35     -1.9
-----
REAL RMSE .41 TRUE SD .81 SEPARATION 1.96 Item RELIABILITY .79
MODEL RMSE .40 TRUE SD .81 SEPARATION 2.05 Item RELIABILITY .81
S.E. OF Item MEAN = .15
-----
UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -.99
1640 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 1667.68 with 1560 d.f. p=.0289
Global Root-Mean-Square Residual (excluding extreme scores): .4086
Capped Binomial Deviance = .2288 for 1640.0 dichotomous observations
    
```

Gambar 3. Separation Item

Asumsi pada pemodelan Rasch bahwa daya pembeda pada seluruh butir soal memiliki taraf yang sama. Pengelompokan butir soal menggunakan rumus persamaan strata (H):

$$H = \frac{[(4 \times \text{SEPARATION}) + 1]}{3}$$

(Erfan, 2020)

Nilai *item separation* adalah 1,96. Oleh karena itu dapat dihitung bahwa nilai H adalah 2,95 (dibulatkan menjadi 3). Hal ini berarti butir soal dikelompokkan menjadi 3 (rendah, sedang, tinggi).

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MEASURE ORDE									
ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE ABILITY	S.E. MEAN	OUTF WNSQ	PTPIEA CORR.	Item
21	A	0	14	34	-1.14	.22	.7	-.17	S21
	E	0	16	39	-1.05	.23	.9	-.12	
	B	0	5	12	-1.02	.42	1.0	-.04	
	D	0	3	7	-.45	.67	1.5	.13	
	C	1	3	7	.67	.16	.3	.46	
37	A	0	8	20	-1.23	.36	.8	-.16	S37
	E	0	9	22	-1.01	.23	.8	-.06	
	D	0	18	44	-.97	.24	1.1	-.05	
	C	0	3	7	.06	.22	1.9	.28	
	B	1	3	7	-.37*	.72	1.5	.16	
20	A	0	1	2	-1.79	.3	.15	-.15	S20
	E	0	10	24	-1.34	.34	.8	-.26	
	B	0	5	12	-1.03	.29	.7	-.05	
	D	0	11	27	-1.00	.25	.9	-.06	
	C	0	9	22	-.54	.32	1.5	.20	
	D	1	5	12	-.19	.49	1.1	.28	
25	C	0	11	27	-1.37	.28	.7	-.29	S25
	D	0	5	12	-1.03	.48	1.0	-.05	
	B	0	6	15	-.79	.48	1.5	.05	
	A	0	14	34	-.65	.24	1.3	.19	
	E	1	5	12	-.64	.41	1.5	.11	
35	E	0	14	34	-1.15	.22	.8	-.18	S35
	C	0	7	17	-.93	.40	1.2	-.01	
	A	0	11	27	-.81	.29	1.2	.06	
	D	0	4	10	-.71	.34	1.1	.07	
	B	1	5	12	-.56	.70	2.6	.14	
13	D	0	4	10	-1.94	.26	.3	-.36	S13
	B	0	1	2	-1.79	.3	.15	-.15	
	A	0	6	15	-1.30	.47	.8	-.17	
	E	0	15	37	-.96	.19	.9	-.04	
	C	0	9	22	-.70	.29	1.2	-.11	
	E	1	6	15	.15	.34	.7	.46	
7	B	0	8	20	-1.57	.30	.6	-.34	S7
	A	0	20	49	-1.06	.20	.9	-.16	
	D	0	2	5	-.99	.61	.8	-.02	
	E	0	4	10	-.71	.49	1.3	.07	
	C	1	7	17	.21	.19	.5	.53	
36	A	0	11	27	-1.28	.30	.8	-.24	S36
	B	0	6	15	-1.13	.40	1.0	-.10	
	D	0	12	29	-.71	.27	1.4	.13	
	E	0	5	12	-.60	.46	1.5	.12	
	C	1	7	17	-.69*	.36	1.5	.11	
15	D	0	4	10	-1.79	.35	.4	-.30	S15
	C	0	21	51	-.94	.20	1.1	-.04	
	E	0	6	15	-.61	.28	1.2	.13	
	B	0	2	5	-.31	1.29	2.7	.14	
	A	1	8	20	-.75*	.39	1.8	.08	
6	C	0	3	7	-1.12	.28	.7	-.06	S6
	E	0	7	17	-1.09	.33	.9	-.09	
	B	0	13	32	-1.07	.28	1.1	-.12	
	A	0	9	22	-.93	.39	1.3	-.01	
	D	1	9	22	-.43	.29	1.1	.26	
11	E	0	1	2	-1.61	.22	.4	-.12	S11
	D	0	9	22	-1.02	.36	1.1	-.06	
	B	0	1	2	-1.01	.7	.02	-.02	
	A	0	15	37	-.97	.22	1.0	-.05	
	B	0	6	15	-.73	.54	1.6	.07	
	C	1	9	22	-.72	.34	1.5	.11	
38	C	0	3	7	-2.27	.30	.2	-.40	S38
	E	0	1	2	-1.61	.4	.12	-.12	
	D	0	12	29	-1.05	.25	1.0	-.09	
	A	0	14	34	-1.04	.24	1.0	-.10	
	D	0	2	5	-.61	.83	1.4	.07	
	B	1	9	22	-.05	.22	.7	.48	
28	B	0	10	24	-1.34	.24	.7	-.26	S28
	C	0	13	32	-1.29	.25	.8	-.27	
	D	0	7	17	-.46	.31	1.7	.21	
	E	0	1	2	-.38	1.4	.09		
	A	1	10	24	-.35	.34	1.2	.33	
29	B	0	1	2	-1.61	.4	.12	-.12	S29
	E	0	5	12	-1.34	.37	.7	-.17	
	C	0	8	20	-1.22	.27	.8	-.16	
	A	0	6	15	-1.08	.51	1.1	-.08	
	D	0	11	27	-.80	.26	1.2	.07	
	E	1	10	24	-.38	.34	1.3	.31	
30	A	0	12	29	-1.08	.24	1.0	-.11	S30
	B	0	8	20	-.96	.42	1.5	-.03	
	E	0	8	20	-.78	.43	1.5	.07	
	D	0	2	5	-.22	.36	2.7	.27	
	C	1	11	27	-.99*	.24	1.7	-.05	
4	D	0	5	12	-2.00	.23	.3	-.43	S4
	B	0	2	5	-1.79	.00	.3	-.21	
	A	0	4	10	-1.38	.56	.9	-.16	
	E	0	18	44	-.54	.18	1.5	.34	
	C	1	12	29	-.69*	.30	1.4	.14	
12	E	0	8	20	-1.77	.24	.4	-.44	S12
	B	0	2	5	-1.70	.09	.4	-.19	
	D	0	1	2	-1.61	.4	.12	-.12	
	A	0	18	44	-.88	.20	1.2	.02	
	C	1	12	29	-.18	.24	.9	.49	
27	B	0	1	2	-2.21	.2	.22	-.22	S27
	E	0	3	7	-2.07	.37	.3	-.24	
	C	0	6	15	-1.38	.25	.6	-.20	
	D	0	18	44	-.91	.19	1.2	-.01	
	A	1	13	32	-.31	.28	1.1	.42	
39	B	0	5	12	-1.85	.25	.4	-.37	S39
	D	0	1	2	-1.79	.4	.15	-.15	
	C	0	5	12	-1.55	.35	.6	-.25	
	A	0	17	41	-1.03	.21	1.1	-.11	
	E	1	13	32	-.07	.19	.6	.60	
26	C	0	5	12	-1.49	.13	.6	-.23	S26
	A	0	12	29	-1.44	.29	1.1	-.36	
	E	0	1	2	-1.44	.6	.09	-.09	
	D	0	4	10	-1.32	.22	.7	-.14	
	B	1	19	46	-.30	.19	.8	.59	
40	B	0	9	22	-1.52	.28	.8	-.34	S40
	C	0	6	15	-1.30	.40	1.0	-.17	
	E	0	6	15	-1.30	.33	.9	-.17	
	D	0	1	2	-1.29	.7	.06	-.06	
	A	1	19	46	-.35	.19	.8	.54	
14	B	0	7	17	-1.52	.28	.7	-.29	S14
	E	0	4	10	-1.39	.72	1.3	-.17	
	D	0	4	10	-1.33	.27	.7	-.15	
	A	0	6	15	-.93	.43	1.6	-.01	
	C	1	20	49	-.50	.19	.9	.41	
16	D	0	5	12	-1.41	.55	1.2	-.20	S16
	E	0	3	7	-1.12	.44	1.0	-.06	
	C	0	8	20	-1.11	.42	1.6	-.10	
	B	0	5	12	-.94	.46	1.5	-.01	
	A	1	20	49	-.66	.19	1.1	.25	
32	C	0	3	7	-1.92	.07	.4	-.30	S32
	B	0	18	44	-1.36	.21	1.0	-.42	
	D	1	20	49	-.35	.17	.8	.57	
1	A	0	9	22	-1.77	.22	.5	-.48	S1
	C	0	5	12	-1.45	.12	.6	-.21	
	B	0	2	5	-1.18	.80	1.0	-.07	
	E	0	4	10	-1.18	.56	1.2	-.09	
	D	1	21	51	-.33	.18	.7	.62	
18	B	0	2	5	-2.13	.35	.3	-.29	S18
	A	0	2	5	-2.12	.68	.4	-.29	
	E	0	1	2	-1.61	.5	.12	-.12	
	C	0	12	29	-1.24	.18	.9	-.23	
	D	0	3	7	-1.01	.60	1.3	-.03	
	E	1	21	51	-.43	.20	.9	.51	
31	E	0	5	12	-1.27	.33	.9	-.14	S31
	D	0	3	7	-.96	.66	1.5	-.02	
	B	0	6	15	-.82	.53	1.9	.04	
	C	0	6	15	-.46	.31	1.9	.19	
	A	1	21	51	-.96*	.21	1.5	.06	
8	B	0	7	17	-1.81	.22	.5	-.43	S8
	A	0	4	10	-1.76	.17	.5	-.29	
	E	0	5	12	-1.31	.17	.7	-.16	
	D	0	3	7	-.65	.84	2.7	.08	
	C	1	22	54	-.41	.18	.9	.56	
9	B	0	4	10	-2.08	.33	.4	-.40	S9
	A	0	1	2	-1.61	.0	.10	-.10	
	E	0	7	17	-1.54	.14	.6	-.30	
	C	0	5	12	-.98	.53	1.7	-.03	
	E	1	24	59	-.48	.17	.8	.53	
10	A	0	3	7	-2.08	.21	.4	-.35	S10
	B	0							

Keefektifan distraktor juga dilihat berdasarkan *average ability* setiap opsi jawaban pada butir soal tersebut. Jika terdapat peningkatan *average ability* berarti distraktor berfungsi. Misalnya pada soal nomor 21, *average ability* opsi A adalah -1,14, opsi E adalah -1,05, opsi B adalah -1,03, opsi D adalah -0,45 dan opsi C (kunci jawaban) adalah 0,67. Berturut-turut terdapat peningkatan *average ability*, berarti keefektifan distraktor berfungsi. Pada soal nomor 37, *average ability* opsi jawaban adalah -1,23; -1,01; -0,97; 0,06 dan -0,37. Ternyata tidak terjadi peningkatan *average ability*, berarti keefektifan distraktor pada soal nomor 37 tidak berfungsi (ditandai dengan bintang).

Tabel 4. Ringkasan Keefektifan Distraktor Butir Soal UAS Biologi Kelas X dengan Model Rasch

Efektifitas Distraktor	Nomor Soal	Jumlah Butir Soal	Persentase (%)
Berfungsi	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 32, 33, 34, 35, 38, 39, 40	34	85
Tidak Berfungsi	4, 15, 30, 31, 36, 37	6	15

Item STATISTICS: MEASURE ORDER

ENTRY	TOTAL	TOTAL	MEASURE	S.E.	MODEL	SWIFT	OUTFIT	PT-MEAS	EXP.	EXACT MATCH	Item		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OSIS	EXPL	
21	3	41	2.00	.62	1.00	-.35	-.91	-.46	.23	92.7	92.6	521	
37	3	41	2.00	.62	1.00	-.43	-.71	-.18	.23	92.7	92.6	527	
20	5	41	1.39	.50	1.07	-1.06	-.31	.28	-.29	87.8	87.7	520	
25	5	41	1.39	.50	1.15	-1.45	-.91	-.11	.29	87.8	87.7	525	
15	6	41	1.39	.50	1.02	-1.08	-.20	.14	.28	87.8	87.7	535	
13	6	41	1.16	.47	1.85	-1.71	-.51	-.46	.31	85.4	85.3	513	
7	7	41	.95	.44	1.83	-1.58	-1.11	-.51	.33	85.4	85.3	517	
36	7	41	.95	.44	1.19	-1.43	-1.01	-.13	.33	85.5	82.9	536	
15	8	41	.77	.42	1.21	-1.64	-1.51	-.88	.35	78.0	88.6	515	
6	9	41	.60	.41	1.18	-1.19	-.41	-.26	.36	75.6	78.4	516	
11	9	41	.60	.41	1.22	-1.44	-1.21	-.11	.36	88.5	78.4	511	
38	9	41	.60	.41	.93	-.73	-.81	-.48	.36	88.5	78.4	538	
28	10	41	.44	.39	1.01	-1.16	-.61	-.33	.37	75.6	76.5	528	
29	10	41	.44	.39	1.03	-1.22	-.81	-.31	.37	88.5	76.5	529	
30	11	41	.29	.38	1.45	-2.17	-1.81	-.60	.39	68.3	74.9	530	
4	12	41	.14	.37	1.24	-1.36	-1.31	-.14	.39	79.7	73.7	54	
12	12	41	.14	.37	.98	-.87	-.41	-.49	.39	75.6	73.7	512	
27	13	41	.81	.37	.96	-1.02	-.11	-.42	.48	88.5	72.8	527	
39	13	41	.81	.37	.81	-1.11	-.70	-1.31	.48	88.5	72.8	539	
2	14	41	-.13	.36	.79	-1.41	-.75	-1.21	.48	87.8	72.1	52	
3	14	41	-.13	.36	1.18	-1.35	-1.51	-.21	.41	63.4	72.1	53	
22	14	41	-.13	.36	.83	-.73	-1.41	-.60	.41	78.0	72.1	522	
33	15	41	-.25	.36	1.08	-1.07	-.41	-.34	.41	79.7	71.5	533	
5	17	41	-.68	.35	1.04	-1.00	-.11	-.39	.41	65.9	79.4	51	
34	17	41	-.50	.35	.73	-1.72	-1.71	-.67	.42	85.4	79.4	534	
23	18	41	-.62	.35	1.16	-1.23	-1.31	-.26	.42	65.9	69.9	523	
24	18	41	-.62	.35	1.12	-1.23	-1.21	-.30	.42	79.7	69.9	524	
17	19	41	-.74	.35	.79	-1.61	-.81	-1.21	.61	82.9	69.4	517	
19	19	41	-.74	.35	.86	-.87	-.81	-.53	.61	73.2	69.4	519	
26	19	41	-.74	.35	.81	-1.41	-.83	-1.01	.59	82.9	69.4	526	
40	19	41	-.74	.35	.87	-.84	-1.01	-.84	.61	78.0	69.4	540	
14	20	41	-.86	.34	1.02	-.99	-.81	-.41	.62	63.4	68.5	514	
16	20	41	-.86	.34	1.17	-1.23	-1.41	-.25	.62	58.5	68.9	516	
12	20	41	-.86	.34	.95	-1.82	-1.51	-.57	.62	73.2	68.9	512	
1	21	41	-.98	.34	.79	-1.61	-.74	-1.71	.62	42.7	78.0	68.5	51
18	21	41	-.98	.34	.92	-.88	-.71	-.51	.62	73.2	68.5	518	
31	21	41	-.98	.34	1.51	-3.16	-3.01	-.86	.62	43.9	68.5	531	
8	22	41	-1.10	.34	.83	-1.11	-.89	-.61	.56	42.7	78.0	68.3	58
9	24	41	-1.33	.35	.87	-1.61	-.87	-.71	.53	68.5	68.2	59	
10	25	41	-1.46	.35	.76	-1.15	-.66	-1.91	.65	68.5	68.4	510	
MEAN	13.9	41.0	.00	.39	.99	-1.00	.01			77.0	75.0		
S.D.	6.1	.0	.91	.07	.19	1.17	1.21			9.5	7.2		

Gambar 5. Validitas Butir Soal

Analisis butir soal dengan model Rasch dapat dilihat dengan tiga kriteria yaitu nilai *outfit MNSQ*, *outfit ZSTD* dan *Point Measure Correlation (Pt-measure Corr)*. Suatu butir soal dikatakan valid jika  $0,5 < outfit\ MNSQ < 1,5$ ;  $-2,0 < outfit\ ZSTD < +2,0$  dan  $0,4 < Point\ Measure\ Correlation\ (Pt-measure\ Corr) < 0,85$ .

Tabel 5. Ringkasan Validitas Butir Soal UAS Biologi Kelas X dengan Model Rasch

Validitas	Nomor Soal	Jumlah Butir Soal	Persentase (%)
Valid	1,2, 7, 8, 9, 10, 12, 13, 16, 17, 18, 19, 22, 26, 27, 32, 34, 38, 39, 40	20	50
Tidak Valid	3, 4, 5, 6, 11, 14, 15, 20, 21, 23, 24, 25, 28, 29, 30, 31, 33, 35, 36, 37	20	50

Analisis menunjukkan bahwa 20 butir soal (50%) valid dan 20 butir soal lainnya (50%) tidak valid. Jika dibandingkan dengan hasil analisis dengan menggunakan CTT, bahwa 65% soal valid dan 35% soal tidak valid menunjukkan adanya perbedaan. Hal ini karena kriteria kevalidan butir soal dengan menggunakan model Rasch lebih detail yaitu meliputi tiga kriteria (*outfit MNSQ*, *outfit ZSTD* dan *Point Measure Correlation (Pt-measure Corr)*) sedangkan pada CTT hanya berdasarkan nilai r hitung berdasarkan rumus korelasi Pearson (product Moment). Selain itu, hasil ringkasan validitas butir soal UAS Biologi Kelas X tersebut menerapkan aturan jika ada salah satu kriteria yang tidak terpenuhi maka soal dianggap tidak valid meskipun dua kriteria lainnya terpenuhi.

SUMMARY OF 41 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	MNSQ	ZSTD	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	13.5	40.0	-.91	.39	1.00	-.1	1.06	.0		
S.D.	7.0	.0	.96	.06	.14	1.0	.32	1.0		
MAX.	28.0	40.0	.98	.61	1.34	2.1	2.18	1.9		
MIN.	3.0	40.0	-2.81	.34	.67	-3.0	.59	-2.6		
REAL RMSE	.41	TRUE SD	.86	SEPARATION	2.11	Person RELIABILITY	.82			
MODEL RMSE	.40	TRUE SD	.87	SEPARATION	2.19	Person RELIABILITY	.83			
S.E. OF Person MEAN	= .15									

Person RAW SCORE-TO-MEASURE CORRELATION = .99  
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .85

Gambar 6. Reliabilitas Person

SUMMARY OF 40 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	MNSQ	ZSTD	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	13.9	41.0	.00	.39	.99	-.1	1.06	.0		
S.D.	6.1	.0	.91	.07	.19	1.2	.37	1.2		
MAX.	25.0	41.0	2.00	.62	1.51	3.4	2.40	3.0		
MIN.	3.0	41.0	-1.46	.34	.73	-2.0	.35	-1.9		
REAL RMSE	.41	TRUE SD	.81	SEPARATION	1.96	Item RELIABILITY	.79			
MODEL RMSE	.40	TRUE SD	.81	SEPARATION	2.05	Item RELIABILITY	.81			
S.E. OF Item MEAN	= .15									

UMEAN=.0000 USCALE=1.0000  
Item RAW SCORE-TO-MEASURE CORRELATION = -.99  
1640 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 1667.68 with 1560 d.f. p=.0289  
Global Root-Mean-Square Residual (excluding extreme scores): .4086  
Capped Binomial Deviance = .2208 for 1640.0 dichotomous observations

Gambar 7. Reliabilitas Item

Penentuan reliabilitas item dan reliabilitas person dengan Model Rasch dikategorikan sebagai berikut: lemah ( $< 0,67$ ); cukup ( $0,67-0,80$ ); bagus ( $0,81-0,90$ ); bagus sekali ( $0,91-0,94$ ); dan istimewa ( $>0,94$ ). Nilai reliabilitas person adalah  $0,82$  artinya konsistensi jawaban siswa termasuk kategori bagus sedangkan nilai reliabilitas item adalah  $0,79$  artinya reliabilitas soal termasuk kategori cukup.

Tabel 6. Perbandingan Hasil Analisis Butir Soal UAS Biologi Kelas X dengan CTT dan Model Rasch

Indikator	CTT	Rasch
Tingkat kesukaran	2 kategori: sedang (57,5%) dan sulit (42,5%)	3 kategori: mudah (15%); sedang (65%) dan sulit (20%)
Daya pembeda	5 kategori: sangat rendah (5%); rendah (22,5%); sedang (20%); tinggi (30%) dan sangat tinggi (22,5%)	3 kategori: rendah, sedang dan tinggi
Efektifitas Distraktor	Berfungsi (72,5%) dan tidak berfungsi (27,5%)	Berfungsi (85%) dan tidak berfungsi (15%)
Validitas	Valid (65%) dan tidak valid (35%)	Valid (50%) dan tidak valid (50%)
Reliabilitas	Sangat tinggi ( $r_{11}=0,855$ )	Reliabilitas item : $0,79$ (cukup) Reliabilitas person : $0,82$ (bagus)

## KESIMPULAN

Hasil analisis butir dengan menggunakan *classical test theory* (CTT) menunjukkan bahwa tingkat kesukaran soal terdapat dua kategori butir soal yaitu soal dengan kategori sedang (57,5%) dan sulit (42,5%). Daya pembeda terdapat lima kategori yaitu sangat rendah (5%), rendah (22,5%), sedang (20%), tinggi (30%) dan sangat tinggi (22,5%). Efektifitas distraktor yang berfungsi sebanyak 72,5% sedangkan yang tidak berfungsi 27,5%. Validitas butir soal yang tergolong valid terdapat 65% sedangkan yang tergolong tidak valid 35%. Secara keseluruhan berdasarkan analisis CTT, butir soal UAS Biologi Kelas X ini adalah reliabel dengan nilai  $r_{11} = 0,855$  (sangat tinggi). Hasil analisis butir soal dengan *item response theory* (IRT) menunjukkan bahwa tingkat kesukaran soal terdapat tiga kategori yaitu mudah (15%), sedang (65%) dan sulit (20%). Daya pembeda terdapat tiga kategori yaitu rendah, sedang dan tinggi. Efektifitas distraktor yang berfungsi sebanyak 85%

sedangkan yang tidak berfungsi 15%. Validitas butir soal seimbang antara valid dan tidak valid yaitu sama-sama 50%. Reliabilitas item sebesar 0,79 dengan kategori cukup sedangkan reliabilitas item sebesar 0,82 dengan kategori bagus.

### UCAPAN TERIMA KASIH

Ucapan terima kasih kepada Bapak Bowo Sugiharto dari Pendidikan Biologi FKIP Universitas Sebelas Maret Surakarta selaku pembimbing dalam analisis soal ini dan SMK Negeri 1 Lokop.

### DAFTAR PUSTAKA

- Alfarisa, F., & Purnama, D. N. (2019). Analisis butir soal ulangan akhir semester mata pelajaran ekonomi sma menggunakan rasch model. *Jurnal Pendidikan Ekonomi Undiksha*, 11(2), 366-374.
- Amalia, A. N., & Widayati, A. (2012). Analisis butir soal tes kendali mutu kelas XII SMA mata pelajaran ekonomi akuntansi di kota Yogyakarta tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, 10(1), 1-26. <https://doi.org/10.21831/jpai.v10i1.919>
- Arifin, Z. (2017). Kriteria instrumen dalam suatu penelitian. *Jurnal Theorems (the original research of mathematics)*, 2(1), 28-36.
- Nitko, A. J. (1996). *Educational assessment of students*. Prentice-Hall Order Processing.
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model rasch. *Indonesian Journal Of Educational Research and Review*, 3(1), 11-19. <https://doi.org/10.23887/ijerr.v3i1.24080>
- Fernanda, J. W., & Hidayah, N. (2020). Analisis kualitas soal ujian statistika menggunakan classical test theory dan rasch model. *Square: Journal of Mathematics and Mathematics Education*, 2(1), 49-60. <https://doi.org/10.21580/square.2020.2.1.5363>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12-23. <https://doi.org/10.21831/pep.v22i1.15609>
- Kurniawan, R. Y., Prakoso, A. F., Hakim, L., Dewi, R. M., & Widayanti, I. (2017). Pemberian pelatihan analisis butir soal bagi guru di kabupaten jombang: Efektif?. *Jurnal Pemberdayaan Masyarakat Madani (JPMM)*, 1(2), 179-193. <https://doi.org/10.21009/jpmm.001.2.03>
- Rahayu, R., & Djazari, M. (2016). Analisis kualitas soal pra ujian nasional mata pelajaran ekonomi akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 14(1), 84-94. <https://doi.org/10.21831/jpai.v14i1.11370>
- Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis kualitas butir soal tipe HOTS pada kompetensi sistem REM di sekolah menengah kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 95-104.
- Sitorus, A. P., Wurarah, M., & Gedoan, S. P. (2018). Analisis Butir Soal Ulangan Harian Buat Guru Mata Pelajaran Biologi Kelas X Sma N 1 Remboken. *JSME (Jurnal Sains, Matematika & Edukasi)*, 5(1), 49-54.
- Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. In 1st International Conference on Education Innovation (ICEI 2017) (pp. 38-42). Atlantis Press. <https://doi.org/10.2991/icei-17.2018.11>

- Sumintono, B. & Widhiarso, W. (2015). Aplikasi Permodelan RASCH pada Assessment Pendidikan. Cimahi: Trim Komunikata
- Septiana, N. (2016). Analisis butir soal ulangan akhir semester (UAS) biologi tahun pelajaran 2015/2016 kelas X dan XI pada MAN Sampit. *Edu Sains: Jurnal Pendidikan Sains dan Matematika*, 4(2), 115-121. <https://doi.org/10.23971/eds.v4i2.514>
- Yustika, A., Susatyo, E. B., & Nuswowati, M. (2014). Uji kriteria instrumen penilaian hasil belajar kimia. *Jurnal Inovasi Pendidikan Kimia*, 8(2), 1330- 1339. <https://doi.org/10.15294/jipk.v8i2.4438>

LAMPIRAN

No. Soal	Indikator										Validitas	Reliabilitas
	DI	Kategori	DP	Kategori	ED							
					A	B	C	D	E			
1	0,5122	Sedang	0,72727	Sangat Tinggi	22%	5%	12%	51%	10%	Valid	r <sub>11</sub> = 0,855 (Sangat Tinggi /Sangat Reliabel)	
2	0,34146	Sedang	0,81818	Sangat Tinggi	24%	12%	24%	34%	5%	Valid		
3	0,34146	Sedang	0,18182	Rendah	44%	34%	5%	10%	7%	Tidak Valid		
4	0,29268	Sulit	0,09091	Rendah	10%	5%	29%	12%	44%	Tidak Valid		
5	0,41463	Sedang	0,54545	Tinggi	20%	20%	10%	7%	41%	Valid		
6	0,21951	Sulit	0,18182	Rendah	22%	32%	7%	22%	17%	Tidak Valid		
7	0,17073	Sulit	0,45455	Tinggi	49%	20%	17%	5%	10%	Valid		
8	0,53659	Sedang	0,72727	Sangat Tinggi	10%	17%	54%	7%	12%	Valid		
9	0,58537	Sedang	0,54545	Tinggi	17%	10%	59%	0%	12%	Valid		
10	0,60976	Sedang	0,81818	Sangat Tinggi	7%	20%	2%	61%	10%	Valid		
11	0,21951	Sulit	0,09091	Rendah	37%	15%	22%	2%	22%	Tidak Valid		
12	0,29268	Sulit	0,54545	Tinggi	44%	5%	29%	2%	20%	Valid		
13	0,14634	Sulit	0,45455	Tinggi	37%	15%	22%	10%	15%	Valid		
14	0,4878	Sedang	0,63636	Tinggi	15%	17%	49%	10%	10%	Valid		
15	0,19512	Sulit	0,09091	Rendah	20%	5%	51%	10%	15%	Tidak Valid		
16	0,4878	Sedang	0,27273	Sedang	49%	12%	20%	12%	7%	Tidak Valid		
17	0,46341	Sedang	0,72727	Sangat Tinggi	24%	5%	7%	17%	46%	Valid		
18	0,5122	Sedang	0,54545	Tinggi	5%	5%	29%	51%	7%	Valid		
19	0,46341	Sedang	0,63636	Tinggi	22%	20%	5%	7%	46%	Valid		
20	0,12195	Sulit	0,27273	Sedang	24%	27%	22%	12%	12%	Tidak Valid		
21	0,07317	Sulit	0,27273	Sedang	34%	12%	7%	7%	39%	Valid		
22	0,34146	Sedang	0,72727	Sangat Tinggi	24%	12%	20%	34%	10%	Valid		
23	0,43902	Sedang	0,36364	Sedang	17%	32%	44%	2%	5%	Tidak Valid		
24	0,43902	Sedang	0,36364	Sedang	27%	7%	44%	5%	17%	Valid		
25	0,12195	Sulit	0,09091	Rendah	34%	15%	27%	12%	12%	Tidak Valid		
26	0,46341	Sedang	0,63636	Tinggi	29%	46%	12%	10%	2%	Valid		
27	0,31707	Sedang	0,45455	Tinggi	32%	2%	15%	44%	7%	Valid		
28	0,2439	Sulit	0,36364	Sedang	24%	24%	32%	17%	2%	Valid		
29	0,2439	Sulit	0,36364	Sedang	15%	12%	20%	27%	24%	Valid		
30	0,26829	Sulit	0	Sangat Rendah	29%	20%	27%	5%	20%	Tidak Valid		
31	0,5122	Sedang	0	Sangat Rendah	51%	15%	15%	7%	12%	Tidak Valid		
32	0,4878	Sedang	0,72727	Sangat Tinggi	0%	44%	7%	49%	0%	Valid		
33	0,36585	Sedang	0,36364	Sedang	22%	37%	15%	10%	17%	Valid		
34	0,41463	Sedang	0,72727	Sangat Tinggi	20%	7%	10%	22%	41%	Valid		
35	0,12195	Sulit	0,18182	Rendah	27%	12%	17%	10%	34%	Tidak Valid		
36	0,17073	Sulit	0,18182	Rendah	27%	15%	17%	29%	12%	Tidak Valid		
37	0,07317	Sulit	0,09091	Rendah	20%	7%	7%	44%	22%	Tidak Valid		
38	0,21951	Sulit	0,45455	Tinggi	34%	22%	7%	5%	29%	Valid		
39	0,31707	Sedang	0,72727	Sangat Tinggi	41%	12%	12%	2%	32%	Valid		
40	0,46341	Sedang	0,63636	Tinggi	46%	22%	15%	2%	15%	Valid		

## LAMPIRAN

### Keterangan:

- DI : *Difficulty Index* (tingkat kesukaran)
- DP : *Discriminating Power* (daya pembeda)
- ED : Efektifitas Distraktor
-  : ED tidak berfungsi (kunci jawaban < 5%)