

CLASSICAL TEST THEORY: ANALISIS BUTIR SOAL ULANGAN HARIAN MATERI TURUNAN KELAS XI BERBANTUAN SOFTWARE ITEMAN V4.3

Cahya Mar'a Saliha Sumantri ¹⁾
Universitas Nahdlatul Ulama Pasuruan ¹⁾
Cahyamss17@gmail.com ¹⁾

Abstrak

Tes merupakan alat ukur untuk mengukur pengetahuan, keterampilan siswa dalam menyelesaikan suatu masalah. Penyusunan instrumen tes harus memenuhi standar agar dapat dijadikan sebagai tolak ukur kemampuan siswa. Salah satu cara untuk mendapatkan tes hasil belajar yang baik adalah melalui proses kegiatan analisis butir soal. Untuk itu, penelitian ini bertujuan untuk mendeskripsikan hasil analisis butir soal ulangan harian materi turunan siswa kelas XI dengan menggunakan *classical test theory* untuk menentukan karakteristik butir soal suatu tes yaitu tingkat kesukaran, daya pembeda, distribusi pengecoh, dan reliabilitas tes berbantuan software ITEMAN V4.3. Penelitian ini menggunakan metode deskriptif kuantitatif. Sampel penelitian ini berjumlah sebanyak 20 siswa kelas XI SMA. Analisis data menggunakan ITEMAN dan SPSS Statistics 27. Berdasarkan penelitian yang telah dilakukan, menunjukkan bahwa kualitas butir soal ulangan harian dengan bab Turunan untuk siswa kelas XI masuk dalam kategori baik untuk tingkat kesukaran. Hal ini diketahui berdasarkan karakteristik butir soal menunjukkan bahwa tingkat kesukaran butir 0.43 dimana masuk kategori sedang (baik). Sedangkan, daya beda butir soal masuk kategori perlu direvisi, dan 70% pengecoh sudah berfungsi dengan baik. Tetapi, untuk estimasi reliabilitas instrument yang dihasilkan guru mendapatkan hasil -0.747 dimana merupakan hasil yang tidak reliabel atau tidak konsisten.

Kata kunci: classical test theory; analisis butir soal; software iteman

Abstract

Tests are a measuring tool to measure students' knowledge and skills in solving a problem. The preparation of test instruments must meet standards so that they can be used as a benchmark for student ability. One way to get a good learning outcome test is through the process of item analysis activities. For this reason, this study aims to describe the results of the analysis of daily test items of derivative material for grade XI students by using the classical test theory to determine the characteristics of the question items of a test, namely the level of difficulty, discriminating power, distribution of tricks, and reliability of the test assisted by ITEMAN V4.3 software. This study uses a quantitative descriptive method. The sample of this study amounted to 20 students in grade XI of high school. Data analysis using ITEMAN and SPSS Statistics 27. Based on the research that has been conducted, it shows that the quality of the daily test question items with derivative chapters for grade XI students is in the good category for the level of difficulty. This is known based on the characteristics of the question items showing that the difficulty level of item is 0.43 which is in the category of medium (good). Meanwhile, the differentiation of question items in the category needs to be revised, and 70% of the deceivers have functioned properly. However, for the estimation of the reliability of the instruments produced by the teacher, the result was -0.747 which is an unreliable or inconsistent result.

Keywords: classical test theory; item test analysis; iteman software

Article Info

Received date:

Revised date:

Accepted date:

PENDAHULUAN

Standar penilaian pendidikan menurut Permendikbudristek Nomor 21 Tahun 2022 Pasal 1 adalah kriteria minimal mengenai mekanisme penilaian hasil belajar peserta didik. Sedangkan, penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengetahui kebutuhan belajar dan capaian perkembangan atau hasil belajar peserta didik. Hal tersebut berkaitan dengan tujuan pendidikan yang akan dicapai oleh para pendidik di sekolah yaitu keberhasilan guru dalam menyampaikan materi dan daya serap peserta didik menerima materi.

Sistem Pendidikan Nasional yang tercantum dalam Undang-undang Republik Indonesia Nomor 20 Tahun 2003 menyatakan bahwa evaluasi dilakukan dalam rangkai pengendalian mutu pendidikan secara nasional sebagai bentuk akuntabilitas penyelenggaraan pendidikan kepada pihak-pihak yang berkepentingan, yaitu peserta didik, lembaga, dan program pendidikan (Sukardi, 2009, p.1). Menurut Ralph Tyler, Cronbach, dan Stufflebeam, evaluasi merupakan sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagaimana tujuan pendidikan sudah tercapai serta proses evaluasi bukan sekedar mengukur sejauh mana tujuan tercapai, tetapi digunakan untuk membuat keputusan (Arikunto, 2021, p.3).

Asesmen adalah sebuah proses yang dilakukan oleh guru untuk mengambil suatu keputusan terhadap siswa. Semakin berkualitas asesmen, maka akan semakin mudah guru memahami kelebihan maupun kelemahan dari siswa (Irmayta, Rudibyani, & Efkar, 2018). Menurut Maemonah (2018: 3) asesmen adalah istilah sebuah proses yang ditempuh untuk mendapatkan informasi yang digunakan dalam rangka membuat keputusan-keputusan mengenai para peserta didik, kurikulum, program-program, dan kebijakan pendidikan, metode atau instrumen pendidikan lainnya oleh suatu badan, lembaga, organisasi atau institusi resmi yang menyelenggarakan suatu aktivitas tertentu.

Mengukur adalah membandingkan sesuatu dengan satu ukuran dan bersifat kuantitatif. Menilai adalah mengambil suatu keputusan terhadap sesuatu dengan ukuran baik buruk dan bersifat kualitatif. Oleh karena itu kegiatan evaluasi meliputi kegiatan mengukur dan menilai (Arikunto, 2021, p.3). Menurut Purwanti (2014), alat ukur evaluasi dapat berupa tes atau non tes. Tes merupakan alat ukur untuk mengukur pengetahuan, keterampilan siswa dalam menyelesaikan suatu masalah. Penyusunan instrumen tes harus memenuhi standar agar dapat dijadikan sebagai tolak ukur kemampuan siswa (Setiyawan & Wijayanti, 2020). Sedangkan menurut Suwanto (2022) tes prestasi hasil belajar adalah tes yang digunakan untuk mengukur pengetahuan dan

pemahaman siswa terhadap materi pelajaran. Standar evaluasi yang dikatakan berkualitas baik berupa instrumen tes yang paling tidak memenuhi syarat valid dan reliabel, disamping itu juga harus memperhatikan daya pembeda dan tingkat kesukaran (Iskandar & Rizal, 2017).

Berdasarkan hasil penelitian yang telah dilakukan oleh Mustaqim dan Sulisti (2024) disimpulkan bahwa daya pembeda soal PAS genap mata pelajaran matematika peminatan kelas XII IPA 1 SMAN 10 Pontianak paling banyak berada pada kategori cukup, yang berarti kualitas soal tersebut mampu membedakan dengan cukup baik peserta didik yang sudah memahami dan yang belum. Sedangkan, untuk tingkat kesukaran, paling banyak berada pada kategori mudah, yang berarti soal kurang dapat merangsang siswa untuk mempertinggi usaha dalam memecahkan suatu permasalahan yang diberikan. Hasil penelitian lain oleh Gunawan & Asria (2023) dapat disimpulkan bahwa hasil analisis kriteria tingkat indeks kesukaran menunjukkan 3 butir soal (15%) berada pada kategori sukar, 12 butir soal (60%) kategori sedang, dan 5 butir soal (25%) kategori mudah. Analisis realibilitas soal berada pada kategori baik dan analisis efektivitas pengecoh menunjukkan 3 butir soal memiliki efektivitas pengecoh pada kategori sangat baik.

Pengembangan tes yang dilakukan oleh guru belum dianalisa secara benar dan akurat apakah tes tersebut benar dapat mengukur kemampuan tertentu. Menurut hasil observasi yang dilakukan oleh Sinaga (2016), Pembuatan perangkat tes tersebut tidak mengikuti langkah-langkah pembuatan tes yang telah ada, misalnya tidak diawali dengan pembuatan kisi-kisi soal dan tidak diketahui karakteristik masing-masing butirnya karena tidak dianalisis, baik secara teoritik maupun secara empirik dan hal ini disebabkan karena guru kesulitan dalam membuat atau menyajikan soal yang bervariasi, lambat dalam pembuatan serta kurangnya waktu untuk melakukan analisa empirik.

Salah satu cara untuk mendapatkan tes hasil belajar yang baik adalah melalui proses kegiatan analisis butir soal (Ariandani, 2012). Menurut Sueadi (2016) pentingnya menganalisis butir soal yaitu dapat memberikan informasi secara terinci tentang keadaan masing-masing butir soal seperti kekuatan dan kelemahan butir soal, spesifikasi soal secara lengkap dan masalah yang terkandung dalam soal seperti kesalahan dalam pembuatan kunci jawaban, soal terlalu sukar atau mudah. Ada dua cara untuk melakukan analisis butir soal, yaitu klasik dan modern. Menurut Retnawati (2016:118) Pendekatan klasik melibatkan penelaahan butir soal melalui informasi dari jawaban siswa untuk meningkatkan kualitas

butir soal yang relevan dengan menggunakan teori tes klasik dan analisis karakteristik butir berdasarkan teori tes klasik (*classical test theory*) dapat dilakukan dengan program Excel, terutama untuk analisis tingkat kesulitan butir. Teori tes klasik merupakan pendekatan yang paling umum dan sering diaplikasikan di dunia pendidikan termasuk dunia penelitian (Susdelina et al., 2018). Sedangkan, Pendekatan modern menggunakan Winstep untuk melakukan analisis butir soal. Winsteps adalah perangkat lunak berbasis Windows yang membantu komputasi Rasch model, terutama di bidang evaluasi pendidikan, survei sikap, dan analisis skala (Aqida, 2024).

Aspek yang sangat menentukan kualitas butir soal dalam pendekatan teori tes klasik adalah tingkat kesukaran dan daya pembeda soal. Menurut Mardapi (2016) kriteria tingkat kesukaran butir soal yang baik apabila memiliki indeks kesukaran butir antara 0,30 sampai dengan 0,70. Butir soal dikatakan memiliki daya beda yang baik jika daya diskriminasinya minimal 0,3, sedangkan aitem yang memiliki indeks daya beda dibawah 0,3 menunjukkan bahwa daya beda soal yang rendah yang membutuhkan revisi butir soal (Azwar, 2012). Pengecoh berfungsi sebagai pengidentifikasi peserta tes yang berkemampuan tinggi. Menurut Nisa (2022) pengecoh dikatakan berfungsi efektif apabila banyak dipilih oleh peserta tes yang berasal dari kelompok bawah, sebaliknya

apabila pengecoh itu banyak dipilih oleh peserta tes yang berasal dari kelompok atas, maka pengecoh itu tidak berfungsi sebagaimana mestinya.

Reliabilitas tes menunjuk pada pengertian apakah suatu tes dapat mengukur secara konsisten sesuatu yang akan diukur dari waktu ke waktu (Himawan & Nurgiyantoro, 2022). Reliabilitas tes pilihan ganda pada teori tes klasik diestimasi menggunakan formula Alpha Ceonbach (KR-20) karena skor jawaban tes tersebut bersifat dikotomi (Nisa, 2022). Arikunto (2016), menyatakan bahwa uji reliabilitas menggunakan rumus Alpha Cronbach mempunyai klasifikasi reliabilitas: 0,00 – 0,20 (sangat rendah), 0,21 – 0,40 (rendah), 0,41 – 0,70 (sedang), 0,71 – 0,90 (tinggi), dan 0,91 1,00 (sangat tinggi).

Berdasarkan latar belakang masalah dan kajian teori yang telah dipaparkan sebelumnya, penelitian ini memiliki tujuan untuk mendeskripsikan hasil analisis butir soal ulangan harian materi turunan siswa kelas XI dengan menggunakan *classical test theory* untuk menentukan karakteristik butir soal suatu tes yaitu tingkat kesukaran, daya pembeda, distribusi pengecoh, dan reliabilitas tes berbantuan software ITEMAN V4.3.

METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah jenis deskriptif kuantitatif. Menurut Best (Samsu, 2021: 117) penelitian deskriptif adalah metode penelitian yang berusaha menggambarkan objek atau subjek yang diteliti sesuai dengan apa adanya. Analisis deskriptif kuantitatif digunakan untuk mendeskripsikan hasil perolehan skor pada software ITEMAN V4.3. Subjek yang digunakan adalah siswa kelas XI suatu SMA di wilayah Kota Malang. Data yang dianalisis merupakan hasil jawaban Ulangan Harian bab Turunan. Ukuran sampel yang diambil sebanyak 20 siswa. Instrumen tes yang digunakan adalah soal Ulangan Harian dengan jumlah butir soal sebanyak 5 soal pilihan ganda, dimana setiap soal memiliki lima pilihan jawaban yaitu A, B, C, D, dan E. Berikut kisi-kisi soal Ulangan Harian berdasarkan materi Turunan kelas XI yang disajikan pada Tabel 1 berikut.

Tabel 1. Kisi-kisi Soal Ulangan Harian

Kompetensi Dasar	Indikator Pencapaian Kompetensi Dasar	No. Butir
4.8 Menyelesaikan masalah yang berkaitan dengan turunan fungsi aljabar	4.8.2 menyelesaikan masalah kehidupan sehari-hari yang berkaitan dengan konsep turunan	5
	4.8.3 Menentukan turunan suatu fungsi dengan menggunakan aturan-aturan turunan	1,2

4.9 Menggunakan turunan pertama fungsi untuk menentukan titik maksimum, titik minimum, dan selang kemonotonan fungsi, serta kemiringan garis singgung kurva, persamaan garis singgung, dan garis normal kurva berkaitan dengan masalah kontekstual	4.9.1 Menentukan gradien suatu garis singgung dengan menggunakan konsep turunan dan menentukan persamaannya	4
	4.9.3 Menentukan titik stasioner, kecekungan, kemonotonan serta titik belok suatu fungsi dengan menggunakan konsep turunan	3

HASIL DAN PEMBAHASAN

Analisis kuantitatif instrumen tes dilakukan untuk mengetahui tingkat kesukaran, daya pembeda, keefektifan pengecoh dan reliabilitas instrumen. Analisis kuantitatif instrumen tes ini dilakukan dengan bantuan *Software Iteman V.4.3*. Adapun hasil analisis instrumen tersebut adalah sebagai berikut:

1. Penggunaan ITEMAN V4.3

Data yang diinput adalah sebagai berikut:

Kunci Jawaban - Notepad						
File	Edit	Format	View	Help		
Butir1	D	5	1	Y	M	
Butir2	E	5	1	Y	M	
Butir3	E	5	1	Y	M	
Butir4	E	5	1	Y	M	
Butir5	C	5	1	Y	M	

Gambar 1. Data Input Kunci Jawaban ITEMAN

Jawaban Siswa - Notepad	
File	Edit
Format	View
Help	
Abimu	DECEB
Amalia	DACBD
Attaariq	DECEA
Aufaa	DBEBC
Azuhra	CEEE
Cintya	CEEEE
Defry	EEAEB
Denny	CECE
Fadlih	CDED
Fadmillah	AAEDB
Fensan	CEEDE
Khotika	DAABA
Kintani	DEEE
Korunia	ABEE
Malika	DEBC
Nadin	EDCEB
Rio	EECEA
Salsabila	CCEDC
Syam	CEED
Vianzia	CEDEC

Gambar 2. Data Input Jawaban Siswa Notepad

2. Output ITEMAN V4.3

Hasil *output* berdasarkan proses di ITEMAN:

Tabel 2. *Output* di ITEMAN

Specification	Value	Specification	Value
Number of examinees	20	Total Items	5
Scored Items	5	Pretest Items	0
Multiple Choice Items	5	Polytomous Items	0
Number of domains	1	External scores	No
Minimum P	0,00	Maximum P	1,00
Minimum item mean	0,00	Maximum item mean	15,00
Minimum item correlation	0,00	Maximum item correlation	1,00
ITEMAN 3.0 Header	No	Exclude omits from option statistics	No
Number of ID columns	10	ID begins in column	1
Responses begin in column	11	Omit character	O
Not Admin character	N	Produce quantile tables	Yes
Correct for spuriousness	Yes	Produce quantile plots	Yes
Save data matrix	No	Include omit codes in matrix	N/A

Include Not Admin codes in matrix	N/A	Include scaled scores for	N/A
Scaling function	N/A	Scaled score setting 1	N/A
Scaled score setting 2	N/A	Dichotomous Classification	No
Classify based on	N/A	Cutpoint	N/A
Low group label	Low	High group label	High
Data is delimited by	N/A	Test for DIF	No
Group status is in column	N/A	Ability levels for DIF	N/A
Group 1 code	N/A	Group 2 code	N/A
Group 1 label	N/A	Group 2 label	N/A

Tabel 3. Summary Statistics di SPSS

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
Scored Items	5	2,150	0,875	1	4	0,430	-0,201

Interpretasi hasil analisis kuantitatif terhadap tingkat kesukaran, daya pembeda, pengecoh, dan reliabilitas dapat dilihat pada penjelasan berikut ini:

1. Tingkat kesukaran

Analisis tingkat kesukaran butir soal pilihan ganda dengan menggunakan program *Iteman Analysis* (Iteman). Tingkat kesukaran butir soal ditunjukkan pada kolom *Alpha*. Menurut Allen & Yen (1979) tingkat kesukaran yang dapat memaksimalkan informasi yang diperlukan dari tes yaitu perbedaan kemampuan siswa berkisar antara 0,3 sampai 0,7. Jika *p* bernilai 0,5 maka item soal memberikan informasi variabilitas maksimum. Jadi kriteria 0,3 sampai 0,7 adalah yang terbaik. Oleh karena itu dapat disimpulkan butir dengan tingkat kesukaran

dibawah 0,3 dianggap butir soal yang sukar, sedangkan jika indeksnya diatas 0,7, butir soal tersebut dianggap mudah (Allen & Yen, 1979: 121-122). Sehingga berdasarkan Allen & Yen (1979: 121) kriteria nilai p dirangkum dalam tabel berikut ini.

Tabel 4. Kategori Tingkat Kesukaran Soal

Kriteria	Kategori
$p < 0,3$	Sukar
$0,3 \leq p \leq 0,7$	Sedang (Baik)
$p > 0,7$	Mudah

Adapun tingkat kesukaran dari setiap butir soal disajikan pada tabel berikut:

Tabel 5. Tingkat Kesukaran Tiap Butir Soal

Nomor Butir Soal	Tingkat Kesukaran	Kategori
1	0.350	Sedang
2	0.600	Sedang
3	0.500	Sedang
4	0.500	Sedang
5	0.200	Sedang
Rata-rata tingkat kesukaran	0.43	Sedang

Berdasarkan Tabel 5 diatas dapat dilihat bahwa tingkat kesukaran keseluruhan soal berada pada kategori sedang karena tercakup di dalamnya hanya kategori sedang. Hal itu bisa saja terjadi karena siswa sudah memahami dengan baik indikator pencapaiannya.

2. *Item Discrimination Index* (Indeks Daya Pembeda)

Selanjutnya analisis yang akan dibahas adalah diskriminasi item. Diskriminasi item merujuk pada seberapa baik item dapat membedakan antara peserta tes dimana dapat membedakan gagasan apa yang akan diukur dalam tes. Sebagai contoh, apabila tes didesain untuk mengukur kemampuan membaca, item discrimination menunjukkan kemampuan item untuk membedakan antara individu yang bagus membaca dan yang kurang bagus dalam membaca (Reynolds, 2010:151). Pada analisis butir dalam penelitian ini, nilai koefisien korelasi point biserial terdapat pada *item statistic-row format* dalam kolom *Total Rpbis*. Koefisien korelasinya secara manual untuk suatu butir tes ditentukan dengan rumus sebagai berikut.

$$r_{pbis} = \frac{\bar{X}_i - \bar{X}}{S_x} \sqrt{\frac{p}{1-p}}$$

keterangan:

r_{pbis} = koefisien korelasi point biserial

$\bar{X}_i$ = rerata skor X untuk peserta tes yang menjawab benar butir tersebut

$\bar{X}$ = rerata skor X

S_x = standar deviasi dari skor X

p = proporsi peserta tes yang menjawab benar butir tersebut

Kriteria baik tidaknya butir soal menurut Ebel dan Frisbie (1991: 232) dapat diperhatikan pada Tabel 6 berikut.

Tabel 6. Kriteria Butir Soal

Indeks Diskriminasi Butir	Evaluasi Butir
$\geq 0,40$	Butir soal sangat baik (<i>Very Good</i>)
$0,3 - 0,39$	Butir soal baik, tetapi kemungkinan perbaikan atau peningkatan (<i>Reasonably Good</i>)
$0,2 - 0,29$	Butir soal yang <i>marginal</i> (sedang), biasanya membutuhkan perbaikan (<i>marginal</i>)
$< 0,19$	Butir soal jelek, dibuang atau diperbaiki melalui revisi (<i>Poor</i>)

Berikut hasil analisis dari Iteman dengan memperhatikan *Total Rpbis*, yaitu:

Tabel 7. Hasil Indeks Daya Pembeda Butir Soal

Nomor Butir Soal	Daya Beda (<i>Point Biser</i>)	Kategori
1	-0.212	<i>Poor</i>
2	0.063	<i>Poor</i>
3	-0.395	<i>Poor</i>
4	-0.194	<i>Poor</i>
5	-0.267	<i>Poor</i>

3. Efektifitas Pengecoh

Berdasarkan hasil analisis butir soal pilihan ganda dengan menggunakan *Program Iteman Versi 4.3*, diperoleh karakteristik setiap alternatif atau pilihan jawaban berupa kunci jawaban dan pengecoh pada setiap butir soal. Efektifitas pengecoh yang dimaksud ditunjukkan oleh tingkat kesukaran (*Option Prop*) dari masing-masing pilihan jawaban.

Distraktor dikatakan efektif apabila ada peserta tes yang memilih distraktor yang sudah dibuat. Selain itu distraktor yang

efektif harus bisa menarik peserta tes lebih banyak dari peserta tes yang berada pada kelompok bawah. Apabila distraktor tidak dipilih oleh peserta tes maka distraktor perlu untuk diganti.

Tingkat kesukaran pengecoh yang baik apabila $TK \geq 0,05$ yaitu pengecoh tersebut dipilih paling sedikit 5% dari seluruh peserta tes.

Berikut ini merupakan hasil analisis pengecoh dengan menggunakan *software Iteman* yang ditunjukkan oleh *option prop* dari masing-masing pilihan jawaban.

Gambar 3. Hasil Analisis Pengecoh Butir Soal

Item	Kunci Jawaban (yang berwarna merah) dan distraktor	Option prop	Keterangan
1	A B C D E	0.100 0.000 0.400 0.350 0.150	Distraktor berfungsi Distraktor tidak berfungsi Distraktor berfungsi * Distraktor berfungsi
2	A B C D E	0.150 0.100 0.050 0.100 0.600	Distraktor berfungsi Distraktor berfungsi Distraktor kurang berfungsi Distraktor berfungsi *
3	A B	0.100 0.050	Distraktor berfungsi Distraktor kurang berfungsi
	C D E	0.300 0.050 0.500	Distraktor berfungsi Distraktor kurang berfungsi *
4	A B C D E	0.000 0.150 0.050 0.150 0.500	Distraktor tidak berfungsi Distraktor berfungsi Distraktor kurang berfungsi Distraktor berfungsi *
5	A B C D E	0.150 0.200 0.200 0.150 0.100	Distraktor berfungsi Distraktor berfungsi * Distraktor berfungsi Distraktor berfungsi

Sehingga, dapat disimpulkan untuk distraktor yang tidak berfungsi perlu diganti, sedangkan distraktor yang kurang berfungsi perlu diperbaiki.

4. Reliabilitas Instrumen Tes

Menurut Allen & Yen (1979: 72) menyatakan bahwa suatu tes dikatakan reliabel jika skor amatan memiliki korelasi

yang tinggi dengan skor sebenarnya. Skor amatan merupakan skor hasil pengukuran dari suatu instrumen tes sedangkan skor sebenarnya adalah skor yang tidak bisa diketahui oleh manusia tetapi manusia bisa mengestimasi reliabilitas. Untuk mengestimasi reliabilitas maka dapat menggunakan reliabilitas konsistensi internal dengan koefisien alpha. Perhitungan dapat menggunakan Excel maupun SPSS.

Reliability Statistics	
Cronbach's Alpha ^a	N of Items
-.747	5

Gambar 4. Reliabilitas SPSS

Dari Gambar 4 di atas terlihat bahwa nilai estimasi reliabilitasnya adalah -0.747 , dimana hal ini berdasarkan Sujarweni (2014), menyatakan pengambilan keputusan dalam uji reliabilitas adalah sebagai berikut:

- Jika nilai Cronbach's Alpha > 0.60 maka instrumen tes dinyatakan reliabel atau konsisten.
- Jika nilai Cronbach's Alpha < 0.60 maka instrumen tes dinyatakan tidak reliabel atau tidak konsisten.

Oleh karena itu, hasil yang tertera pada Gambar 4 yaitu -0.747 dimana < 0.60 maka dapat disimpulkan instrumen tes tersebut tidak reliabel atau tidak konsisten.

PENUTUP

Hasil penelitian ini menunjukkan bahwa kualitas butir soal ulangan harian dengan bab Turunan untuk siswa kelas XI di salah satu Sekolah Tingkat Atas di Kota Malang masuk dalam kategori baik untuk tingkat kesukaran. Hal ini diketahui berdasarkan karakteristik butir soal menunjukkan bahwa tingkat kesukaran butir 0.43 dimana masuk kategori sedang (baik). Sedangkan, daya beda butir soal masuk kategori perlu direvisi, dan 70% pengecoh sudah berfungsi dengan baik. Tetapi, untuk estimasi reliabilitas instrument yang dihasilkan guru mendapatkan hasil -0.747 dimana merupakan hasil yang tidak reliabel atau tidak konsisten.

Saran untuk penelitian selanjutnya dapat menganalisis butir soal secara modern menggunakan *Item Respon Theory (IRT)* untuk mendapatkan kualitas soal yang lebih baik lagi. Serta memperbaiki butir-butir soal tersebut sekaligus pengecohnya.

DAFTAR PUSTAKA

- Allen, M.J. & Yen, W.M. (1979). *Introduction to Measurement Theory*. Monterey: Brooks/Cole Publishing Company
- Ariandani. (2012). Analisis butir soal matematika buatan guru yang digunakan untuk tes semester II kelas X SMA Negeri 8 Pekanbaru Tahun Pelajaran 2011/2012. *Jurnal Aksiomatik*, 18-24
- Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan* (2nd ed.). Jakarta: Bumi Aksara.

- Arikunto, Suharsimi. (2021). *Dasar-dasar Evaluasi Pendidikan Edisi 3*. Jakarta: Bumi Aksara.
- Aqida, Debbi Silvia. (2024). Penggunaan Rasch Model dalam Menganalisis Butir Soal Matematika Kelas 5 Sekolah Dasar. *Indonesian Journal of Islamic Elementary Education*, 4 (1), 91-101.
- Azwar, S. (2012). *Reliabilitas dan validitas edisi 4*. Yogyakarta: Pustaka Pelajar.
- Gunawan, & Lailatul A. (2023). Analisis Butir Soal Penilaian Tengah Semester (PTS) Matematika Kelas XI Berdasarkan Teori Klasik. *MATH LOCUS: Jurnal Riset dan Inovasi Pendidikan Matematika*, 4 (1), 1-11.
- Himawan, R. & Burhan, N. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN (Analysis of exercise items for odd semester end of semester Indonesian language subjects class VIII SMPN 1 Bambanglipuro Bantul using the ITEMAN program). *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, 8(!), 160-180.
- Irmayta, E., Rudibyani, R. B., & Efkar, T. (2018). Pengembangan instrumen asesmen pengetahuan pada materi asam basa arrhenius. *Jurnal Pendidikan dan Pembelajaran Kimia*, 7(1), 63–76.
- Iskandar, A., & Muhammad, R. (2017). Analisis Kualitas Soal Di Perguruan Tinggi Berbasis Aplikasi Tap. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 12–23.
- Maemonah. (2018). *Asesmen Pembelajaran*. Yogyakarta: PGMI Press UIN Suka.
- Mardapi, Djemari. (2016). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Parama Publishing.
- Mustaqim & Hidayu S. (2024). Analisis Butir Soal PAS Matematika Peminatan: Daya Pembeda, Tingkat Kesukaran, dan Kualitas Pengecoh. *Al-'Adad: Jurnal Tadris Matematika*, 3 (1), 44-56.
- Nisa, Chairun. (2022). Analisis Karakteristik Butir Soal Tes Ujian Kenaikan Kelas Pelajaran Matematika dengan Program Iteman. *YASIN: Jurnal Pendidikan dan Sosial Budaya*, 2(6), 862-870.
- Republik Indonesia. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia No. 21 Tahun 2022 tentang Standar Penilaian Pendidikan pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Jakarta.
- Republik Indonesia. 2003. *Undang-undang Republik Indonesia No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional*. Kementerian Pendidikan dan Kebudayaan. Jakarta.
- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, Dan Psikometrian)*. Yogyakarta: Parama Publishing.
- Samsu. (2021). *Metode Penelitian: Teori dan Aplikasi Penelitian Kualitatif, Kuantitatif, Mixed Methods, Serta Research & Development*. Jambi: PUSAKA.
- Setiawan, R.A. & Palupi, S.W.(2020). Analisis Kualitas Instrumen untuk Mengukur Kemampuan Pemecahan Masalah Siswa Selama Pembelajaran Daring di Masa Pandemi. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika, dan Statistika*, 1 (2), 130-139.
- Sinaga, Nurul Afni. (2016). Pengembangan Tes Kemampuan Pemecahan Masalah dan Penalaran Matematika Siswa SMP Kelas VIII. *PYTHAGORAS: Jurnal Pendidikan Matematika*, 11 (2), 169-181.
- Suaedi, M. (2016). Analisis terhadap kualitas butir soal buatan guru IPA kelas VIII MTs Negeri Slawi tahun pelajaran 2015/2016. *Jurnal Kamil*, 1(3), 1-15.
- Sukardi. (2009). *Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Susdelina, S., Perdana, S. A., & Febrian, F. (2018). Analisis Kualitas Instrumen Pengukuran Pemahaman Konsep Persamaan Kuadrat Melalui Teori Tes Klasik Dan Rasch Model. *Jurnal Kiprah*, 6(1), 41–48.

CEJou Vol X No X, Bulan Tahun
Cahya Mar'a Saliha Sumantri

- Suwarto (2022). Karakteristik tes ilmu pengetahuan alam. *Jurnal Pendidikan*, 31(1), 109-120.
- Sujarweni, V.W. (2014). SPSS untuk Penelitian. Yogyakarta:Pustaka Baru Press.