



Analisis Butir Soal Ujian Tengah Semester Mata Kuliah Probabilitas: Bagaimana Kualitasnya?

Dita Aldila Krisma^{1*}, Syita Fatih 'Adna²

^{1,2}Program Studi Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Tidar

*e-mail: dita.aldila@untidar.ac.id

Diserahkan: 27/01/2023; Diterima: 03/04/2023; Diterbitkan: 30/04/2023

Abstrak. Butir soal dalam suatu instrumen tes untuk asesmen pembelajaran perlu diketahui kualitasnya melalui karakteristik butir soal. Penelitian ini bertujuan untuk menentukan karakteristik butir soal pada instrumen tes Ujian Tengah Semester (UTS) mata kuliah Probabilitas sehingga dapat diketahui butir soal yang sulit, butir soal yang membedakan kemampuan peserta tes, dan informasi kemampuan mahasiswa dalam menguasai materi yang diujikan sebagai dasar materi pertemuan setelah UTS. Penelitian ini merupakan penelitian deskriptif dengan pendekatan kuantitatif. Teknik sampling yang digunakan adalah *purposive sampling* sehingga sampel yang dipilih adalah mahasiswa sebanyak 124 mahasiswa semester 3 tahun ajaran 2022/2023 yang mengambil mata kuliah Probabilitas. Teknik pengumpulan data yang dilakukan adalah dokumentasi instrumen tes dan jawaban tes mahasiswa. Data yang diperoleh diolah menggunakan program Quest, dianalisis berdasarkan teori tes klasik, dan selanjutnya dideskripsikan. Instrumen tes yang dianalisis sebanyak 15 butir soal pilihan ganda dengan 5 pilihan jawaban yang memiliki 1 jawaban benar. Tingkat kesulitan butir soal pada soal UTS Probabilitas yaitu sedang sebanyak 4 butir soal atau sebesar 33,33% dan sulit sebanyak 11 butir soal atau sebesar 66,67%. Berdasarkan karakteristik estimasi reliabilitas, estimasi reliabilitas subjek dengan koefisien 0,67 (tinggi) dan estimasi reliabilitas butir soal dengan koefisien 0,89 (sangat tinggi). Distraktor pada semua butir soal sebanyak 15 butir dapat berfungsi. Setelah mendapatkan data berdasarkan karakteristiknya, peneliti dapat mengetahui materi yang sulit bagi peserta tes, memberi masukan kepada peserta tes tentang kemampuannya dalam menguasai materi pada materi yang diujikan, dan pembuat soal dapat memperbaiki penyajian materi pada materi yang sama untuk semester berikutnya.

Kata kunci: Butir soal, instrumen tes, kualitas, ujian tengah semester

Abstract. The quality of the items in a test instrument for learning assessment needs to be known through the characteristics of the items. This study aims to determine the characteristics of the items on the Mid Semester Examination (UTS) test instrument for the Probability course so that difficult items can be identified, items that differentiate the abilities of test takers, and information on students' ability to master the material being tested as the basis for meeting material after UTS. This research is a descriptive research with a quantitative approach. The sampling technique used was purposive sampling so that the sample chosen was 124 students in semester 3 of the 2022/2023 academic year who took the Probability course. The data collection technique used was documentation of test instruments and student test answers. The data obtained was processed using the Quest program, analyzed based on classical test theory, and then described. The test instrument analyzed consisted of 15 multiple choice questions with 5 answer choices that had 1 correct answer. The level of difficulty of the items on the Probability UTS questions was moderate as many as 4 items or 33.33% and difficult as many as 11 items or 66.67%. Based on the characteristics of the estimated reliability, the estimated reliability of the subject with a coefficient of 0.67 (high) and the estimated reliability of the

items with a coefficient of 0.89 (very high). The distractor for all 15 items can function. After obtaining data based on its characteristics, researchers can find out material that is difficult for test takers, provide input to test takers about their ability to master the material on the material being tested, and the question maker can improve the presentation of material on the same material for the following semester.

Keywords: Question items, test instrument, quality, mid semester evaluation

Pendahuluan

Asesmen merupakan salah satu proses yang perlu dilakukan dalam pelaksanaan pembelajaran. Asesmen adalah serangkaian proses untuk mendapatkan informasi mengenai hasil pembelajaran yang diikuti pembelajar dan penilaian kemajuan pembelajaran sehingga dapat digunakan untuk menjelaskan seberapa baik kinerja pembelajar (Miller et al., 2009; Reynolds et al., 2009). Sejalan dengan pernyataan tersebut, (Nitko & Brookhart, 2011) menyatakan bahwa asesmen diartikan sebagai proses mengumpulkan informasi yang diperlukan dalam pembuatan aturan mengenai pembelajar, kurikulum, program, sekolah, dan kebijakan pendidikan. Alfarisa & Purnama (2019) menambahkan bahwa asesmen dilakukan untuk mengetahui ketercapaian tujuan pembelajaran dan bagaimana pembelajar mencapai tujuan tersebut.

Salah satu jenis asesmen yang dapat dilakukan yaitu memberikan tes kepada pembelajar (Fathin & Retnawati, 2020; Miller et al., 2009). Tes perlu diketahui kualitasnya karena diperlukan untuk mendapatkan hasil penilaian yang benar-benar menggambarkan pencapaian peserta tes (Kartowagiran, Munadi, Retnawati, & Apino, 2018). Butir-butir tes yang dikembangkan harus dapat mengukur kemampuan peserta tes secara akurat dan membedakan peserta tes mana yang memiliki rendah dan tinggi, serta memastikan bahwa butir-butir tersebut berfungsi sesuai dengan kegunaannya (bin Abd. Razak et al., 2012). Kualitas soal yang baik harus terdapat butir soal dalam kategori baik (Retnawati, 2016). Dengan demikian, analisis pada butir soal suatu tes dibutuhkan untuk mengetahui kualitasnya.

Salah satu tes yang dilakukan untuk menganalisis butir soal adalah tes klasik. Teori tes klasik diukur melalui proporsi jawaban benar pada setiap item (Wu et al., 2016). Teori tes klasik mengacu pada analisis hasil tes berdasarkan skor. Statistika yang dihasilkan dari teori tes klasik ini adalah tingkat kesulitan soal, daya beda, kesalahan pengukuran, dan reliabilitas (Wu et al., 2016; Wang et al., 2011). Teori tes klasik banyak digunakan karena tidak memerlukan responden yang besar (lebih dari 100) dan mudah diterapkan.

Hasil penelitian yang pernah dilakukan peneliti-peneliti sebelumnya, analisis butir soal dapat diidentifikasi dengan beberapa karakteristik diantaranya tingkat kesulitan disertai faktor yang berkontribusi terhadap sulitnya butir soal dan strategi yang dilakukan mengatasi soal-soal tes yang sulit. Identifikasi karakteristik butir soal yang digunakan oleh Damopolii (2016) adalah tingkat kesukaran, daya pembeda, kualitas pengecoh, validitas, dan reliabilitas pada soal ujian tengah semester mata kuliah biologi umum. Karakteristik butir soal menurut Nurhasanah & Ahmad (2017) meliputi tingkat kesukaran, validitas, dan reliabilitas pada soal ujian tengah semester mata kuliah geometri. Selanjutnya, karakteristik soal menurut Jusrianto, Zahir, & Megawati (2018) terdiri dari tingkat kesulitan dan daya beda pada soal



ujian tengah semester mata kuliah pengantar komputer. Selain itu, identifikasi karakteristik butir soal yang dilakukan oleh Elviana (2020) terdiri dari daya pembeda, tingkat kesukaran, validitas, reliabilitas, efektivitas pengecoh pada soal ujian akhir semester mata kuliah pendidikan Agama Islam. Dengan demikian, penulis dapat melakukan analisis untuk mengetahui kualitas butir soal adalah estimasi tingkat kesulitan, daya beda, estimasi reliabilitas subjek dan butir soal, serta efektivitas distraktor.

Tingkat kesulitan merupakan persentase atau proporsi peserta tes yang dapat menjawab benar suatu item tes (Wu et al., 2016 & Ashraf & Jaseem, 2020). Daya beda tujuannya adalah mengidentifikasi item untuk peserta ujian dengan skor tinggi memiliki probabilitas tinggi untuk menjawab dengan benar dan peserta ujian dengan skor rendah memiliki probabilitas rendah untuk menjawab dengan benar (Crocker & Algina, 2008). Dengan demikian, koefisien daya beda ini dapat digunakan untuk mengidentifikasi apakah suatu butir soal dapat digunakan untuk membedakan individu memiliki kemampuan tinggi atau tidak. Hal tersebut didukung oleh Ashraf & Jaseem (2020) bahwa daya beda mengukur seberapa baik suatu butir soal dapat menjadi pembeda kemampuan penguasaan materi yang diujikan peserta tes.

Berdasarkan uraian tersebut, butir soal dalam suatu instrumen tes untuk asesmen pembelajaran perlu diketahui kualitasnya. UTS merupakan salah satu asesmen pembelajaran sumatif. UTS ini digunakan sebagai evaluasi hasil belajar mahasiswa selama tengah semester setelah mengikuti perkuliahan. Soal UTS yang baik dapat menggambarkan kemampuan mahasiswa. Pembuatan soal dilakukan oleh tim dosen mata kuliah probabilitas. Soal yang diujikan dalam bentuk pilihan ganda dan telah dilakukan penilaian tiap mahasiswa dan rata-rata kelas namun belum dilakukan analisis karakteristiknya meliputi tingkat kesulitan, daya beda, reliabilitas, dan efektivitas distraktor. Untuk mengetahui kualitas soal tersebut, tiap butir soal UTS perlu diketahui karakteristiknya. Karakteristik yang dianalisis meliputi tingkat kesulitan, daya beda, estimasi reliabilitas subjek dan butir soal, serta efektivitas distraktor. Butir soal perlu diketahui kualitasnya melalui karakteristik tersebut karena melalui analisis tersebut dapat diketahui materi yang sulit bagi peserta tes sehingga dapat memberi masukan kepada peserta tes tentang kemampuannya dalam menguasai materi pada materi yang diujikan. Dengan demikian, mahasiswa dapat mempelajari kembali materi sulit yang digunakan sebagai dasar pada materi pembelajaran paruh dua atau setelah UTS. Selain itu, dosen dapat memperbaiki penyajian materi untuk pertemuan atau semester berikutnya. Dengan diketahui karakteristik ini maka diketahui kualitas instrumen tes yang dapat dipertimbangkan penggunaannya sebagai soal evaluasi pada semester berikutnya atau pengembangannya untuk bank soal serta peningkatan kualitas perencanaan, pelaksanaan dan penilaian pembelajaran.

Metode Penelitian

Metode penelitian ini adalah penelitian deskriptif kuantitatif. Anderson & Arsenault (2005) menyatakan bahwa penelitian deskriptif kuantitatif bertujuan untuk menggambarkan data secara keseluruhan dengan mengelompokkan dan menyajikan data dalam bentuk tabel atau gambar. Penelitian ini dipilih mengidentifikasi karakteristik butir soal Ujian Tengah Semester

(UTS) mata kuliah Probabilitas. Dengan kata lain, penelitian deskriptif kuantitatif ini menggambarkan data kualitas butir secara keseluruhan dengan mengelompokkan data berdasarkan tingkat kesulitan, daya beda, estimasi reliabilitas subjek dan butir soal, serta efektivitas distraktor. Teknik pengumpulan data yang dilakukan melalui tes dengan instrumen penilaian UTS Probabilitas. Soal UTS tersebut terdiri dari 15 butir soal pilihan ganda dengan 5 pilihan jawaban. Populasi pada penelitian ini adalah seluruh mahasiswa Pendidikan Matematika Universitas Tidar. Penelitian ini bertujuan untuk mengetahui karakteristik soal UTS mata kuliah Probabilitas. Untuk itu, teknik sampling penelitian ini adalah *purposive sampling* maka diperoleh sampelnya adalah 124 mahasiswa yang mengikuti UTS Probabilitas sesuai dengan jadwal UTS.

Data yang diperoleh, selanjutnya dilakukan analisis butir soal dengan teori tes klasik. Keunggulan dari teori tes klasik ini adalah kemudahan dalam memahami konsepnya dan menggunakannya (Mardapi, 1998). Berdasarkan hasil penelitian Mardapi (1998), indeks kehandalan perangkat tes yang diuji menggunakan tes klasik sebesar 0,79 dan dapat dinyatakan baik. Indeks kehandalan tes menurut teori tes klasik dihitung menggunakan formula Spearman-Brown dan Cronbach-alpha (Crocker & Algina, 1986). Analisis pada penelitian ini menggunakan program Quest. Program ini merupakan salah satu program statistik berbasis komputer. Program Quest lebih akurat dibandingkan program statistik lainnya untuk menganalisis butir soal (Maharani & Putro, 2020). Kelebihan program ini dapat menganalisis data politomus, dikotomis, dan kombinasi dikotomis dan politomus (Maharani & Putro, 2020; Retnawati, 2016). Karakteristik yang dianalisis meliputi tingkat kesulitan, daya beda, estimasi reliabilitas subjek dan butir soal, serta efektivitas distraktor. Setelah mendapatkan data berdasarkan karakteristiknya, peneliti dapat mengetahui materi yang sulit bagi peserta tes, memberi masukan kepada peserta tes tentang kemampuannya dalam menguasai materi pada materi yang diujikan, dan memperbaiki penyajian materi untuk pertemuan atau semester berikutnya.

Selanjutnya, diuraikan karakteristik yang dianalisis meliputi tingkat kesulitan, daya beda, estimasi reliabilitas subjek dan butir soal, serta efektivitas distraktor. Berikut kriteria tiap karakteristik. Table 1 merupakan kriteria tingkat kesulitan

Tabel 1. Kriteria Tingkat Kesulitan (Gooding 2009)

Kriteria	Kategori
$p < 0,3$	Sulit
$0,3 \leq p \leq 0,7$	Sedang
$p > 0,7$	Mudah

Berikutnya kriteria untuk daya beda tersaji pada Tabel 2.

Tabel 2. Kriteria Daya Beda (Ebel, & Frisbie, 1986)

Kriteria	Kategori
≥ 40	Baik
$0,30 \leq D \leq 0,39$	Cukup baik
$0,20 \leq D \leq 0,29$	Sedang
< 20	Rendah

Reliabilitas suatu tes dapat dituliskan dalam interval $-1,00 \leq r \leq +1,00$. Reliabilitas dengan koefisien tinggi artinya reliabilitas suatu tes tinggi dan jika koefisien rendah maka reliabilitas tes tersebut rendah. Selain itu, koefisien reliabilitas diharapkan bersifat positif. Tabel 3 berikut menyajikan kriteia koefisien reliabilitas.

Tabel 3. Kriteria Estimasi Reliabilitas (Guilford, 1956)

Kriteria	Kategori
$0,80 < r \leq 1,00$	Sangat tinggi
$0,60 < r \leq 0,80$	Tinggi
$0,40 < r \leq 0,60$	Sedang
$0,20 < r \leq 0,40$	Rendah
$-1,00 < r \leq 0,20$	Sangat rendah (tidak reliabel)

Berikutnya kriteria untuk efektivitas distraktor dapat disajikan pada Tabel 4 berikut.

Tabel 4. Kriteria Efektivitas Distraktor (Retnawati et al., 2017)

Kriteria	Kategori
$< 5\%$	Tidak berfungsi
$\geq 5\%$	Berfungsi

Hasil Penelitian dan Pembahasan

Tingkat Kesulitan

Karakteristik pertama untuk mendeskripsikan kualitas butir soal adalah tingkat kesulitan tiap butir soal. Karakteristik tingkat kesulitan diperoleh dari proporsi responden yang menjawab benar pada tiap butir soal. Dengan kata lain, tingkat kesulitan ini ditunjukkan dari proporsi yang menjawab benar. Tingkat kesulitan pada interval 0,3 - 0,7 termasuk dalam katgeori butir soal yang baik dengan kategori tingkat kesulitan kategori sedang. Butir soal yang sulit terletak pada koefisien kurang dari 0,3 dan butir soal yang mudah terletak pada koefisien lebih dair 0,7. Hasil estimasi tingkat kesulitan butir soal UTS Probabilitas tersaji pada Tabel 4.

Tabel 5. Hasil Estimasi Tingkat Kesulitan

No	Koefisien	Kategori	No	Koefisien	Kategori	No	Koefisien	Kategori
1	0,172	Sulit	6	0,122	Sulit	11	0,205	Sulit
2	0,430	Sedang	7	0,236	Sulit	12	0,212	Sulit
3	0,148	Sulit	8	0,402	Sedang	13	0,051	Sulit
4	0,220	Sulit	9	0,325	Sedang	14	0,207	Sulit
5	0,187	Sulit	10	0,207	Sulit	15	0,391	Sedang

Berdasarkan Tabel 5, dapat dilihat bahwa sebanyak 11 butir soal atau sebesar 66,67% termasuk kategori sulit dan 4 butir soal atau sebesar 33,33% termasuk dalam kategori sedang. Urutan butir soal dari yang tersulit sampai termudah adalah nomor 13, 6, 1, 3, 5, 11, 10, 14, 12, 4, 7, 9, 8, dan 2. Butir soal nomor 13 merupakan soal yang paling sulit. Berikut soal nomor 13.

Suatu serum kebenaran yang biasa diberikan kepada seorang tersangka, diketahui 90% dapat dipercaya bila ternyata tersangka tersebut memang bersalah dan 99% dapat dipercaya bila tersangka tersebut tidak bersalah. Dengan kata lain, 10% di antara yang bersalah dinyatakan tidak bersalah dan 1% di antara yang tidak bersalah dinyatakan bersalah oleh serum tersebut. Bila seorang tersangka diambil dari sejumlah tersangka yang hanya 5% di antaranya pernah melakukan kejahatan, dan serum itu menunjukkan bahwa ia bersalah. Keterangan tersebut dinyatakan dalam tabel berikut.

	Hasil yang ditunjukkan serum	
	Bersalah (S)	Tidak bersalah (S ^C)
Tersangka bersalah (T)	90%	10%
Tersangka tidak bersalah (T ^C)	1%	99%

Peluang bahwa sesungguhnya ia tidak bersalah adalah

- A. 0,174
- B. 0,257
- C. 0,512
- D. 0,742
- E. 0,825

Gambar 1. Butir Soal Nomor 13

Soal tersebut merupakan materi dengan topik distribusi binomial. Persentase mahasiswa yang menjawab benar pada butir soal nomor 13 sebesar 5,1%.

Butir soal nomor 2 merupakan soal yang paling mudah diantara 15 butir soal yang diberikan. Berikut soal nomor 2.

Banyak susunan huruf berbeda yang dapat disusun dari kata "MATEMATIKA" adalah ...

- A. 604.800
- B. 302.400
- C. 151.200
- D. 75.600
- E. 37.800

Gambar 2. Butir Soal Nomor 2

Topik materi pada butir soal nomor 2 adalah permutasi. Sebesar 43% peserta tes dapat memilih jawaban butir soal nomor 2 dengan benar.

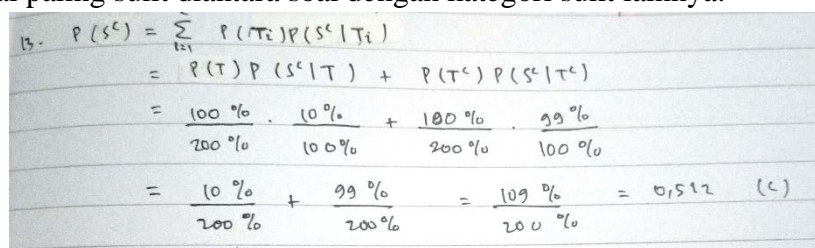
Pada penilaian hasil belajar, tingkat kesulitan ini sesuai untuk menunjukkan apakah pembelajar telah mempelajari konsep yang ditekankan atau belum (Ashraf and Jaseem 2020). Jika koefisien tingkat kesulitan semakin mendekati 0,00 maka butir tersebut semakin sulit dan jika koefisien tingkat kesulitan semakin mendekati 1,00 maka butir tersebut semakin mudah. Semakin besar proporsi peserta tes yang dapat menjawab benar maka butir soal tersebut semakin mudah (Ashraf & Jaseem 2020). Pada butir soal UTS Probabilitas

menunjukkan koefisien tingkat kesulitan soal lebih mendekati 0,00 yaitu sebanyak 11 butir soal pada kategori sulit dan koefisien tingkat kesulitan soal yang lebih mendekati 1,00 sebanyak 4 butir termasuk kategori sedang. Interpretasi tingkat kesulitan butir soal adalah koefisien proporsi tingkat kesulitan rendah menunjukkan bahwa butir soal terlalu sulit dan hanya sedikit peserta tes yang dapat menjawab benar atau jika tingkat kesulitan butir soal itu tinggi maka mayoritas peserta tes dapat menjawab soal dengan benar terlepas dari kemampuannya (Doust, Khan, & Al-Ghafri 2021). Dengan demikian, tidak banyak mahasiswa yang dapat menjawab soal dengan benar pada 11 butir soal UTS Probabilitas yang berkategori sulit.

Tingkat kesulitan berhubungan dengan proporsi jumlah butir soal. Kriteria tingkat kesulitan menurut Suparman (2020) menyatakan bahwa proporsi tingkat kesulitan tes adalah 15% mudah, 70% sedang dan 15% sulit untuk tes dengan tujuan evaluasi belajar. Berbeda dengan kriteria menurut Sudjana (2017) bahwa proporsi tingkat kesulitan butir soal yaitu 3-4-4 artinya 30% tingkat kesulitan rendah, 40% tingkat kesulitan sedang, dan 30% tingkat kesulitan sulit atau dengan proporsi 3-5-2. Sementara Rao, Prasad, Sajitha, & Permi (2016) menyatakan proporsi tingkat kesulitan butir soal yaitu 30:70 artinya 30% kategori sulit dan 70% kategori sedang. Pada butir soal UTS Probabilitas yang digunakan untuk evaluasi belajar menunjukkan bahwa sebanyak 73,33% butir soal termasuk pada tingkat sulit dan sebanyak 26,67% butir soal termasuk dalam kategori sedang. Proporsi tersebut memberikan hasil tingkat kesulitan yang terdistribusi tidak proporsional sehingga tidak sejalan dengan kriteria proporsi yang disebutkan.

Adanya butir-butir soal dengan kategori sulit yang terdistribusi tidak proporsional ini karena disebabkan oleh berbagai faktor. Tingginya proporsi butir soal dengan kategori sulit dikarenakan banyaknya mahasiswa yang salah menjawab soal. Hasil tingkat kesulitan tes yang terdistribusi tidak proporsional juga ditunjukkan pada hasil penelitian Fauzie, Pada, & Supriatno (2021) bahwa temuannya tidak selaras dengan teori, lebih tingginya proporsi soal yang sulit pada bank soal dapat disebabkan karena banyaknya siswa yang salah menjawab soal. Mahasiswa dapat menjawab salah pada butir-butir soal dapat disebabkan karena faktor kurangnya pemahaman konsep peserta tes terhadap materi. Hal tersebut didukung oleh Retnawati, Kartowagiran, et al. (2017) dan Star & Rittle-Johnson (2009) yang menyatakan bahwa kurangnya kemampuan dasar pembelajar dalam memahami konsep materi dapat memberi kontribusi pada tingkat kesulitan butir soal.

Berikut contoh jawaban penyelesaian oleh mahasiswa pada butir soal nomor 13 yang merupakan soal paling sulit diantara soal dengan kategori sulit lainnya.



$$\begin{aligned}
 13. \quad P(S^c) &= \sum_{i=1}^n P(T_i)P(S^c|T_i) \\
 &= P(T)P(S^c|T) + P(T^c)P(S^c|T^c) \\
 &= \frac{100\%}{200\%} \cdot \frac{10\%}{100\%} + \frac{100\%}{200\%} \cdot \frac{99\%}{100\%} \\
 &= \frac{10\%}{200\%} + \frac{99\%}{200\%} = \frac{109\%}{200\%} = 0,545 \quad (c)
 \end{aligned}$$

Gambar 3. Jawaban Mahasiswa Nomor 13

Pada soal nomor 13, langkah penyelesaian yang seharusnya dilakukan mahasiswa adalah mengetahui konsep materi untuk menyelesaikan soal tersebut. Konsep materi untuk menyelesaikan soal tersebut adalah materi teorema Bayes. Selanjutnya menggunakan rumus teorema Bayes yang dinyatakan dalam $P(T^c|S)$ untuk menghitung peluang bahwa sesungguhnya ia tidak bersalah. Namun, mahasiswa menyelesaikan soal tersebut menerapkan konsep peluang total sehingga dari langkah pertama yang ditulis sudah menunjukkan kesalahan kemampuan dasar berupa kesalahan konsep yang digunakan. Tipe kesalahan tersebut sejalan dengan tipe kesalahan yang dinyatakan oleh (Wahyuni & Nurhadi, 2018) bahwa kesalahan yang dilakukan pada kesalahan dasar (*basic error*) adalah kesalahan konsep dasar. Hal tersebut didukung Fauzie, Pada, & Supriatno (2021) bahwa faktor yang berkontribusi terhadap kesalahan dalam menjawab karena peserta tes belum menguasai konsep dan mereka mungkin menjawab tes dengan menebak.

Dalam hal ini, pengajar harus mengetahui kekuatan dan kelemahan dari setiap indikator yang dinilai. Oleh karena itu, pengajar harus menemukan cara untuk mengatasi masalah dan metode pengajaran yang lebih tepat. Selain itu, pendidik sebagai pembuat tes juga perlu mempertimbangkan urutan butir soal dengan tingkat kesulitan mudah, sedang, dan sulit karena hal tersebut dapat memberikan pengaruh terhadap hasil tes. Hal ini selaras dengan Kusumawati & Hadi (2018) dan Debeer & Janssen (2013) bahwa kurang berhasilnya peserta tes dalam menjawab soal tes dapat disebabkan karena urutan butir soal yang salah atau penempatan urutan butir soal yang kurang menguntungkan sehingga mempengaruhi hasil tes.

Daya Beda

Karakteristik kedua untuk mendeskripsikan kualitas butir soal adalah daya beda. Indeks daya beda ditunjukkan pada indeks korelasi poin biserial. Indeks daya beda pada butir soal yang baik yaitu lebih dari atau sama dengan 0,3. Indeks daya beda pada butir soal yang tidak baik yaitu kurang dari 0,3. Pada penelitian ini, daya beda untuk butir soal UTS Probabilitas tersaji pada Tabel 6.

Tabel 6. Hasil Analisis Daya Beda

No	Koefisien	Kategori	No	Koefisien	Kategori	No	Koefisien	Kategori
1	0,81	Baik	6	0,73	Baik	11	0,48	Baik
2	0,50	Baik	7	0,48	Baik	12	0,39	Baik
3	0,68	Baik	8	0,56	Baik	13	0,41	Baik
4	0,65	Baik	9	0,54	Baik	14	0,80	Baik
5	0,83	Baik	10	0,54	Baik	15	0,30	Baik

Berdasarkan Tabel 6, seluruh koefisien daya beda memberikan hasil koefisien yang positif. Pada hasil tersebut terdapat kelimabelas butir soal dengan kategori baik. Hasil tersebut menunjukkan bahwa soal UTS pada butir soal dengan kategori baik dapat diartikan dapat membedakan kemampuan peserta tes dengan kemampuan tinggi dan rendah. Butir soal nomor 5 termasuk dalam kategori baik dengan koefisien tertinggi diantara koefisien lainnya yaitu sebesar 0,83. Berikut soal nomor 5 disajikan pada Gambar 4.

Dua buah dadu dilambungkan sebanyak satu kali. Peluang munculnya jumlah kedua mata dadu 7 atau 11 adalah

- A. $\frac{4}{36}$
- B. $\frac{8}{36}$
- C. $\frac{12}{36}$
- D. $\frac{16}{36}$
- E. $\frac{18}{36}$

Gambar 4. Butir Soal Nomor 5

Berikutnya, soal nomor 1 juga memiliki daya beda baik dengan koefisien 0,81. Berikut soal nomor 1.

Banyaknya cara duduk pada sebaris 7 kursi dari 3 orang Sumatra, 2 orang Jawa, dan 2 orang Kalimantan, dengan mengelompokkan sesuai dengan asal pulau adalah

- A. 12
- B. 24
- C. 48
- D. 72
- E. 144

Gambar 5. Butir Soal Nomor 1

Jika indeks daya beda lebih dari atau sama dengan 0,3 maka butir soal dikatakan baik (Retnawati 2016). Butir soal pada UTS Probabilitas menunjukkan indeks daya beda lebih dari atau sama dengan 0,3 sehingga dapat dinyatakan bahwa daya beda butir soal tersebut baik. Butir soal yang baik ini dapat membedakan kemampuan peserta tes yang kemampuannya tinggi dan yang kemampuannya rendah. Hal tersebut didukung Cohen-Swerdlik (2009) dan Argianti & Retnawati (2020) yang menyatakan bahwa daya beda ini menunjukkan bahwa suatu butir soal dapat membedakan antara peserta tes yang memiliki skor tinggi dan rendah. Lebih lanjut, Khoshaim & Rashid (2016) menyatakan bahwa butir soal yang memiliki koefisien daya beda dengan kategori tinggi menunjukkan bahwa hanya peserta tes dengan skor atau nilai keseluruhan tinggi yang dapat menjawab butir soal tersebut dengan akurat dan tepat.

Reliabilitas

Karakteristik ketiga adalah estimasi reliabilitas subjek dan butir soal. Nilai reliabilitas subjek adalah 0,67 (tinggi) dan nilai reliabilitas butir soal adalah 0,89 (sangat tinggi). Reliabilitas menunjukkan kekonsistenan suatu alat ukur jika diujikan berulang-ulang kepada subjek yang berbeda atau dilakukan dalam waktu yang berbeda. Koefisien reliabilitas merupakan koefisien keajegan alat ukur atau tes (Retnawati, 2016). Lebih lanjut, Retnawati menyatakan bahwa jika hasil pengukuran mendekati keadaan peserta tes yang sesungguhnya maka tes tersebut dapat dikatakan reliabel. Koefisien reliabilitas subjek pada soal UTS

Probabilitas termasuk dalam kategori sedang dan koefisien reliabilitas butir soalnya termasuk dalam kategori tinggi. Artinya, butir soal UTS Probabilitas dengan kategori tinggi tersebut akan memberikan hasil yang sama atau mendekati sama jika digunakan untuk mengukur hal yang sama pada waktu berbeda. Hal tersebut sejalan dengan pernyataan (Retnawati 2016) bahwa suatu tes atau alat ukur dengan koefisien reliabilitas tinggi maka hasilnya sama atau mendekati sama ketika digunakan untuk mengukur hal yang sama pada waktu berbeda.

Semakin besar reliabilitas suatu instrumen, maka semakin kecil kesalahan pengukuran, demikian sebaliknya, semakin kecil reliabilitas skor, maka semakin besar kesalahan pengukurannya (Retnawati, 2016). Jika analisis data memberikan hasil reliabilitas kecil maka menunjukkan kesalahan pengukurannya semakin besar. Beberapa faktor dapat menjadi sebab terhadap hasil reliabilitas yang kecil yaitu karakteristik instrumen yang digunakan, teknik penyusunan dan pelaksanaan pengujian instrumen, kualitas butir instrumen, adanya kerjasama peserta tes saat mengisi atau menjawab instrumen, butir-butir instrumen yang meragukan, kondisi peserta tes selama mengisi instrumen, tempat penyelenggaraan tes yang kurang mendukung atau gabungan faktor-faktor tersebut (Retnawati 2016).

Pada soal UTS Probabilitas ini menunjukkan hasil yang tinggi maka kesalahan pengukurannya dapat dikatakan kecil. Hasil reliabilitas tinggi ini dapat didukung beberapa faktor. Faktor yang mendukung tersebut diantaranya pelaksanaan tes mengikuti aturan baku dapat mendukung hasil reliabilitas tinggi. Pada pelaksanaan pengambilan data, seluruh mahasiswa tertib dalam mengerjakan seperti menyelesaikan soal sesuai dengan durasi waktu yang diberikan. Selama mengerjakan, mahasiswa dalam kondisi baik atau tidak sakit. Antar mahasiswa tidak melakukan kerjasama selama mengerjakan UTS. Lingkungan penyelenggaraan pengambilan data kondusif dan mendukung seperti susunan kursi yang tertata rapi, mahasiswa membawa dan menggunakan alat tulis masing-masing selama mengerjakan, suhu ruang yang sesuai, suasana luar kelas tidak ramai, dan ada pengawas yang mengondisikan ruangan selama pelaksanaan UTS.

Efektivitas Distraktor

Karakteristik keempat adalah efektivitas distraktor. Distraktor dikatakan efektif jika lebih dari atau sama dengan 5% peserta tes memilih distraktor tersebut (Testa, Toscano, & Rosato, 2018). Tabel berikut menyajikan klasifikasi efektivitas butir soal.

Tabel 7. Efektivitas Distraktor

Kriteria	Kategori	Banyak butir soal
$\geq 5\%$	Distraktor berfungsi	15
$< 5\%$	Distraktor tidak berfungsi	0

Berikut contoh butir soal nomor 7 dengan distraktor yang berfungsi.

Peluang mahasiswa A dan B menerima beasiswa prestasi berturut-turut $\frac{9}{10}$ dan $\frac{11}{12}$.
Peluang mahasiswa A menerima beasiswa, tetapi mahasiswa B tidak menerima beasiswa adalah

A. $\frac{9}{120}$
B. $\frac{11}{120}$
C. $\frac{22}{120}$
D. $\frac{36}{120}$
E. $\frac{99}{120}$

Gambar 6. Butir Soal Nomor 7

Pada soal tersebut, jawaban yang benar adalah pilihan A. Mahasiswa yang menjawab B sebanyak 75 mahasiswa, C sebanyak 7 peserta tes, D sebanyak 4 peserta tes, dan E sebanyak 8 peserta tes. Poin biserial pilihan jawaban A adalah 0,48, pilihan jawaban B adalah -0,57, pilihan jawaban C -0,03, pilihan jawaban D 0,00, dan pilihan jawaban E adalah 0,32. Berdasarkan hasil tersebut menunjukkan bahwa poin biserial pada distraktor lebih kecil dari poin biserial jawaban yang benar maka dapat dikatakan bahwa distraktor tersebut berfungsi. Dengan demikian dapat disimpulkan bahwa tiap butir soal UTS Probabilitas memiliki distraktor yang baik dan berfungsi.

Contoh lain dari distraktor yang berfungsi adalah butir soal nomor 5. Berikut soalnya.

Peluang mahasiswa A dan B menerima beasiswa prestasi berturut-turut $\frac{9}{10}$ dan $\frac{11}{12}$.
Peluang mahasiswa A menerima beasiswa, tetapi mahasiswa B tidak menerima beasiswa adalah

A. $\frac{9}{120}$
B. $\frac{11}{120}$
C. $\frac{22}{120}$
D. $\frac{36}{120}$
E. $\frac{99}{120}$

Gambar 7. Butir Soal Nomor 5

Pada butir soal nomor 5, kunci jawabannya adalah B. Mahasiswa yang memilih jawaban B yang merupakan jawaban benar sebanyak 23 mahasiswa. Distraktor A, C, D, dan E ini berfungsi karena tiap distraktor ada yang memilih dan poin biserial bernilai negatif atau lebih kecil dari poin biserial pada kunci jawaban. Poin biserial pada distraktor secara berturut-turut adalah -0,06, -0,03, -0,45, dan -0,17. Pada A dipilih 20 mahasiswa, distraktor C dipilih 4

mahasiswa, distraktor D dipilih 50 mahasiswa, dan distraktor E dipilih 26 mahasiswa serta ada 1 mahasiswa yang tidak memilih jawaban.

Distraktor tiap butir soal pada instrumen UTS probabilitas dapat berfungsi. Pada tes pilihan ganda, distraktor yang berfungsi akan menghasilkan skor tes yang lebih reliabel terlepas dari jumlah distraktor (Syahrial and Haryanto, 2020). Kusaeri & Suprananto (2012) menyatakan bahwa pada item yang baik, akan ada distraktor yang dipilih oleh peserta tes. Lebih lanjut, Amelia & Kriswanto (2017) menyatakan bahwa distraktor yang terlihat jelas tidak mungkin dipilih akan memberikan kemudahan peserta tes membuat keputusan bahwa distraktor tersebut tidak perlu dipilih atau salah, sehingga mereka dapat memilih jawaban benar dengan menebak, sementara distraktor yang terlalu dekat dengan jawaban yang benar akan menyebabkan soal menjadi terlalu sulit. Fungsi distraktor dan tingkat kesulitan soal-soal tes saling terkait satu sama lain sehingga dalam menyusun perangkat tes keduanya saling terkait dan perlu dipertimbangkan (Santoso, Kartianom, & Kassymova, 2019).

Kualitas Soal UTS

Kualitas soal UTS ini dilihat dari empat karakteristik yaitu tingkat kesulitan, daya beda, estimasi reliabilitas subjek dan butir soal, serta efektivitas distraktor. Berdasarkan tingkat kesulitan diketahui bahwa sebanyak 11 butir soal atau sebesar 66,67% termasuk dalam kategori sulit dan 4 butir soal atau sebesar 33,33% termasuk dalam kategori sedang. Proporsi tingkat kesulitan pada soal UTS ini belum seimbang dikarenakan lebih banyak butir soal yang sulit, sementara akan lebih seimbang jika terdapat soal dengan kategori mudah, sedang, dan sulit. Menurut Arifin (2012), proporsi tingkat kesulitan pada tes, idealnya berdistribusi secara merata. Sementara Arikunto (2012) menyatakan bahwa soal yang tidak terlalu mudah dan tidak terlalu sulit merupakan soal yang baik.

Koefisien daya beda butir soal UTS pada kategori baik. Pada hasil tersebut, 15 butir soal yang diuji memberikan hasil dengan nilai yang positif sehingga termasuk kategori baik karena indeks daya beda pada butir soal lebih dari atau sama dengan 0,3. Koefisien daya beda pada butir soal UTS dengan kategori baik ini menunjukkan bahwa butir soal dapat membedakan kemampuan mahasiswa yang tinggi dan kemampuan mahasiswa yang rendah. Mahasiswa dengan kemampuan yang tinggi dapat menjawab soal-soal ujian dengan benar yang lebih banyak dibandingkan mahasiswa dengan kemampuan rendah. Hal ini didukung Argianti & Retnawati (2020) bahwa peserta tes yang memiliki kemampuan tinggi atau pintar pasti akan lebih mampu menjawab soal tes dibandingkan dengan yang kemampuan rendah.

Nilai reliabilitas subjek adalah 0,67 (tinggi) dan nilai reliabilitas butir soal adalah 0,89 (sangat tinggi). Butir soal UTS ini dapat dikatakan reliabel sehingga dapat menunjukkan apabila soal ini diujikan secara berulang kepada subjek yang berbeda atau dilaksanakan dalam waktu yang berbeda akan menunjukkan kekonsistenan, hasil pengukuran mendekati keadaan peserta tes yang sesungguhnya.

Distraktor tiap butir soal dapat berfungsi sehingga distraktor dikatakan efektif karena lebih dari atau sama dengan 5% tiap pilihan jawaban dipilih oleh oleh peserta tes. Distraktor yang berfungsi ini dapat menunjukkan bahwa pilihan jawaban pada butir soal memberikan pilihan yang masuk akal dan memberikan kemungkinan dipilih oleh peserta tes sehingga dapat menghasilkan skor tes yang lebih reliabel.

Setelah mendapatkan data berdasarkan karakteristiknya, peneliti dapat mengetahui materi yang sulit bagi peserta tes berdasarkan hasil analisis tingkat kesulitan, mengetahui kualitas butir soal untuk membedakan kemampuan mahasiswa sehingga dapat memberi masukan kepada peserta tes tentang kemampuannya dalam menguasai materi pada materi yang diujikan, dan memperbaiki penyajian materi untuk pertemuan atau semester berikutnya.

Kesimpulan dan Saran

Pada penelitian ini, kesimpulan yang dapat ditarik adalah butir soal UTS Probabilitas menunjukkan proporsi tingkat kesulitan butir soal pada kategori sulit dan sedang sehingga belum seimbang, koefisien daya beda pada kategori baik, perangkat tes ini reliabel, dan distraktornya dapat berfungsi. Analisis yang digunakan berdasarkan teori tes klasik menggunakan program Quest dapat diketahui kualitas butir soal dalam empat karakteristik yaitu tingkat kesulitan, daya beda, estimasi reliabilitas subjek dan butir soal, serta efektivitas distraktor. Tingkat kesulitan butir soal pada soal UTS Probabilitas terdapat dua kategori yaitu sedang sebanyak 4 butir soal atau sebesar 33,33% dan sulit sebanyak 11 butir soal atau sebesar 66,67%. Pada instrumen tes tersebut lebih banyak butir soal dengan tingkat kesulitan yang sulit. Karakteristik daya beda pada soal UTS Probabilitas termasuk dalam kategori baik sehingga tiap butir soal dapat digunakan untuk membedakan mahasiswa dengan kemampuan tinggi dan kemampuan rendah. Berdasarkan karakteristik estimasi reliabilitas memberikan hasil estimasi reliabilitas subjek dengan koefisien 0,67 (tinggi) dan estimasi reliabilitas butir soal dengan koefisien 0,89 (sangat tinggi). Reliabilitas butir soal dengan kategori tinggi ini ketika digunakan untuk mengukur hal yang sama atau dilaksanakan dalam waktu berbeda hasilnya sama atau mendekati sama. Untuk karakteristik efektivitas distraktor, distraktor pada semua butir soal sebanyak 15 butir dengan pilihan jawaban sebanyak 5 pilihan dan 1 jawaban diantaranya jawaban benar ini dapat berfungsi.

Diketahuinya kualitas soal melalui karakteristik tersebut, dapat diketahui butir soal yang sulit, butir soal yang dapat membedakan kemampuan peserta tes, dan informasi kemampuan mahasiswa dalam menguasai materi yang diujikan sebagai dasar materi pertemuan setelah UTS. Setelah mendapatkan data berdasarkan karakteristiknya, peneliti dapat mengetahui materi yang sulit bagi peserta tes, memberi masukan kepada peserta tes tentang kemampuannya dalam menguasai materi pada materi yang diujikan, dan pembuat soal dapat memperbaiki penyajian materi pada materi yang sama untuk semester berikutnya. Berdasarkan hasil penelitian ini, untuk meningkatkan kualitas pembelajaran, pembuat soal dapat mempertimbangkan proporsi tingkat kesulitan soal dan efektivitas distraktornya. Selain itu, pengajar dapat meningkatkan kualitas pembelajaran dengan memilih metode yang tepat untuk menanamkan konsep dan memperbaiki penyajian materi untuk pertemuan atau semester berikutnya. Peluang penelitian yang dapat dilakukan adalah menganalisis karakteristik butir soal berdasarkan teori modern.

Daftar Pustaka

Amelia, R. N. & Kriswantoro. (2017). Implementasi Item Response Theory sebagai Basis Analisis Kualitas Butir Soal dan Kemampuan Kimia Siswa Kota Yogyakarta. *Jurnal Kimia Dan Pendidikan Kimia*, 2(10), 1-12. <https://doi.org/10.20961/jkpk.v2i1.8512>

- Anderson, G., & Arsenault, N. (2005). *Fundamentals of Educational Research*. USA: Taylor & Francis e-Library.
- Argianti, A. & Retnawati, H. (2020). Characteristics of Math National-Standardized School Exam Test Items in Junior High School: What Must Be Considered?. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 24(2),156–165. doi: 10.21831/pep.v24i2.32547.
- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, S. (2012). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Ashraf, Z. A. & Jaseem, K. (2020). Classical and Modern Methods in Item Analysis of Test Tools. *International Journal of Research and Review*, 7(5), 397–403.
- Cohen-Swerdlik. (2009). *Educational Tests and Measurement: An Introduction*. USA: Mc Graw Hill.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Harcourt Brace Jovanovich.
- Debeer, D. & Rianne, J. (2013). Modeling Item-Position Effects within an IRT Framework. *Journal of Educational Measurement*, 50(2), 164–85. doi: 10.1111/jedm.12009.
- Doust, A.R, Waseem, A.K., & Al-Ghafri, M. (2021). An Item Analysis Study on TIMSS 2015 Mathematics Items of Omani and Iranian Students Comparison IRT and CDM Approaches. *International Journal of Mathematics Trends and Technology*, 67(9):87–95. doi: 10.14445/22315373/ijmtt-v67i9p510.
- Ebel, R. L., and D. A. Frisbie. (1986). *Essentials of Educational Measurement*. New Jersey: Prentice Hall.
- Fathin, A. Z. & Retnawati, H. (2020). How Good Are the Characteristics of the Mathematics National Standardized School Examination Items in Vocational School?. *Journal of Physics: Conference Series* 1581(1). doi: 10.1088/1742-6596/1581/1/012029.
- Fauzie, M., Pada, A. U. T., & Supriatno. (2021). Analysis of the Difficulty Index of Item Bank According to Cognitive Aspects during the Covid-19 Pandemic. *Jurnal Penelitian Dan Evaluasi Pendidikan* 25(2):162–69. doi: 10.21831/pep.v25i2.42603
- Gooding, S. (2009). Children's Difficulties with Mathematical Word Problems. *British Society for Research into Learning Mathematics*, 29, 31-36.
- Guilford, J. (1956). *Fundamental Statistic in Psychology and Education*. 3rd ed. New York: McGraw-Hill Book Company, Inc.
- Jusrianto, Zahir, A., & Megawati. (2018). Analisis Kualitas Tes Ujian Akhir Semester Mata Kuliah Pengetahuan Komputer. *Jurnal Studi Guru dan Pembelajaran* 1(1): 1-9.
- Kartowagiran, B., Munadi. S., Retnawati, H., & Apino, E. (2018). The Equating of Battery Test Packages of Mathematics National Examination 2013-2016. *SHS Web of Conferences* 42:00022. doi: 10.1051/shsconf/20184200022.
- Khoshaim, H.B. & Rashid, S. (2016). Assessment of the Assessment Tool: Analysis of Items in a Non-MCQ Mathematics Exam. *International Journal of Instruction*, 9(1):119–32. doi: 10.12973/iji.2016.9110a.
- Krisma, D. A., & Retnawati, H. (2020). The Final Semester Test of Mathematics Subject in Vocational High School, How Difficult Is It?. *Journal of Physics: Conference Series* 1581(1). doi: 10.1088/1742-6596/1581/1/012033.
- Kusaeri & Suprananto. (2012). *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu
- Kusumawati, M. & Hadi, S. (2018). An Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from Mathematics Assessments in Senior High School. *Research and Evaluation in Education* 4(1):70–78. doi: 10.21831/reid.v4i1.20202.
- Maharani, A. V., & Putro, N. H. P. S. (2020). Item Analysis of English Final Semester Test. *Indonesian Journal of EFL and Linguistics*, 5(2), 491. <https://doi.org/10.21462/ijefl.v5i2.302>



- Mardapi, D. (1998). Analisis Butir dengan Teori Tes Klasik dan Teori Respons Butir. *Jurnal Kependidikan*, 28(1): 25-34.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching*. New Jersey: Pearson.
- Nitko, A. J. & Brookhart, S. M. (2014). *Educational Assessment of Students Pearson New International Edition*. 6th ed. England: Pearson.
- Rao, C., Prasad, H. L.K., Sajitha, K., Permi, H., & Shetty, J. (2016). Item Analysis of Multiple Choice Questions: Assessing an Assessment Tool in Medical Students. *International Journal of Educational and Psychological Researches* 2(4):201. doi: 10.4103/2395-2296.189670.
- Retnawati, H., Arlinwibowo, J., & Sulistyaningsih, E. (2017). The Students' Difficulties in Completing Geometry Items of National Examination. *International Journal on New Trends in Education and Their Implications (IJONTE)* 8(4):1-14.
- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, Dan Psikometrian)*. Yogyakarta: Parama Publishing.
- Retnawati, H., Kartowagiran, B., Arlinwibowo, J., & Sulistyaningsih, E. (2017). Why Are the Mathematics National Examination Items Difficult and What Is Teachers' Strategy to Overcome It?. *International Journal of Instruction*, 10(3):257-76. doi: 10.12973/iji.2017.10317a.
- Reynolds, C. R., Livingston, R.B., & Wilson, V. (2009). *Measuerment and Assessment in Education*. 2nd ed. New Jersey: Pearson.
- Santoso, A., Kartianom, & Kassymova, G. K. (2019). Kualitas Butir Bank Soal Statistika (Studi Kasus: Instrumen Ujian Akhir Mata Kuliah Statistika Universitas Terbuka). *Jurnal Riset Pendidikan Matematika*, 6(2):165-76. doi: 10.21831/jrpm.v6i2.28900.
- Star, J. R., & Rittle-Johnson, B. (2009). It Pays to Compare: An Experimental Study on Computational Estimation. *Journal of Experimental Child Psychology*, 101:408-26.
- Sudjana, N. (2017). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.
- Suparman. (2020). Terjadi Item Has Perfect Score. 9, 83-104
- Syahrial & Haryanto. (2020). Function of Distractors in Mathematics Test Items on the Achievement Tests Based on the Rasch Model. (1):210-16. doi: 10.5220/0008519502100216.
- Testa, S., Toscano, A., & Rosato, R. (2018). Distractor Efficiency in an Item Pool for a Statistics Classroom Exam: Assessing Its Relation with Item Cognitive Level Classified According to Bloom's Taxonomy. *Frontiers in Psychology* 9(AUG):1-12. doi: 10.3389/fpsyg.2018.01585.
- Wahyuni, R., & Nurhadi, D. (2018). Tipe-Tipe Kesalahan Siswa Dalam Menyelesaikan Soal Matematika Pada Aturan Eksponen Dan Scaffoldingnya: Studi Kasus Di Smkn 11 Malang. *Teknologi Dan Kejuruan: Jurnal Teknologi, Kejuruan, Dan Pengajarannya* 41(2):173-86. doi: 10.17977/um031v41i22018p173.
- Wang, Q, Wang, Z., & Osterlind, S.J. (2011). Modeling the Effects of Home and Student Factors on Text Comprehension. *New Waves - Educational Research & Development* 14(1):22-23.
- Wu, M., Tam, H. P., & Jen, T. H. (2016). *Educational measurement for applied researchers*. Singapura: Springer