

## ANALISIS BUTIR SOAL PENILAIAN AKHIR SEMESTER MATEMATIKA

Ziyana Endah Khairun Nisa<sup>1</sup>, Kana Hidayati<sup>2</sup>

<sup>1,2</sup> Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

\*Corresponding author. Jl. Colombo, 55281, Yogyakarta, Indonesia.

E-mail: [ziyanaendah.2021@student.uny.ac.id](mailto:ziyanaendah.2021@student.uny.ac.id)<sup>1)</sup>  
[kana@uny.ac.id](mailto:kana@uny.ac.id)<sup>2)</sup>

Received 14 April 2023; Received in revised form 25 August 2023; Accepted 27 September 2023

### Abstrak

Soal merupakan salah satu alat untuk mengukur keberhasilan dalam mencapai tujuan pembelajaran. Pentingnya penelitian analisis butir soal untuk mengetahui kualitas soal sudah dalam kategori baik atau tidak. Tujuan penelitian ini adalah untuk menentukan item kualitas penilaian akhir semester (PAS) matematika kelas VII tahun pelajaran 2021/2022 pada validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Jenis penelitian ini adalah analisis deskriptif dengan pendekatan kuantitatif. Objek penelitian ini adalah 69 jawaban siswa pada PAS matematika kelas VII SMP Rafirna tahun 2021/2022. Didapatkan kesimpulan bahwa soal penilaian akhir semester (PAS) matematika kelas VII tahun 2021/2022 jenis soal pilihan ganda dan uraian ialah 17 butir soal pilihan ganda (85%) dinyatakan valid, 3 soal (15%) tidak valid, sedangkan soal bentuk uraian yang berjumlah 5 butir, semuanya valid. Reliabilitas soal pilihan ganda maupun uraian, dengan *software IBM statistics 26* kemudian dikonversikan dengan rumus *alpha strata* diperoleh koefisiennya =  $0,795 > 0,70$ , sehingga soal pilihan ganda maupun uraian soal PAS dinyatakan *reliable*. Tingkat kesukaran butir soal pilihan ganda diperoleh 2 soal (10%) kategori mudah, 7 soal (35%) sedang, dan 11 soal (55%) sukar. Daya pembeda soal pada bentuk soal pilihan ganda diperoleh 4 soal (20%) pada kategori baik, 12 soal (60%) cukup, dan 4 soal (20%) tidak baik.

**Kata kunci:** Daya pembeda; tingkat kesukaran; reliabilitas, validitas.

### Abstract

Questions are a tool for measuring success in achieving learning objectives. The importance of research on item analysis is to find out whether the quality of the questions is in the good category or not. The purpose of this study was to determine the final of semester assessment (PAS) quality items for class VII mathematics for the 2021/2022 school year in terms of validity, reliability, level of difficulty, and appropriateness. This type of research is descriptive analysis with a quantitative approach. The object of this research is 69 students' answers to PAS mathematics class VII Rafirna Middle School in 2021/2022. It was concluded that the end of semester assessment questions (PAS) for class VII mathematics in 2021/2022 were of the type of multiple choice questions and descriptions, namely 17 multiple choice questions (85%) were declared valid, 3 questions (15%) were invalid, while the questions in the form of descriptions totaling 5 items, all of which are valid. The reliability of multiple choice and essay questions using *IBM statistics 26 software* was then converted using the *alpha strata formula* to obtain a coefficient =  $0.795 > 0.70$ , so that the multiple choice questions and essay questions were declared reliable. The difficulty level of the multiple choice questions was 2 questions (10%) in the easy category, 7 questions (35%) in the medium category, and 11 questions (55%) in the difficult category. The appropriateness of questions in the form of multiple choice questions was found to be 4 questions (20%) in the good category, 12 questions (60%) sufficient, and 4 questions (20%) not good.

**Keywords:** Appropriateness; level of difficulty; reliability; validity.



This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

## PENDAHULUAN

Pembelajaran merupakan kegiatan inti yang dilaksanakan dalam sebuah lembaga pendidikan, khususnya pada lembaga formal. Salah satu komponen pembelajaran adalah evaluasi pembelajaran. Salah satu bentuk dari evaluasi yang biasa dilaksanakan oleh sekolah adalah penilaian akhir semester. Evaluasi adalah suatu proses yang sistematis untuk menentukan atau membuat keputusan sampai sejauh mana tujuan-tujuan pengajaran telah dicapai siswa (Syahputra et al., 2020).

Istilah penilaian merupakan alih bahasa dari istilah *assessment*. Menurut Maemonah (2018) istilah asesmen diartikan sebagai penilaian proses, kemajuan, dan hasil belajar siswa (*outcomes*). Sistem penilaian dalam pembelajaran, khususnya pada pembelajaran dengan kurikulum 2013 yang masih berlaku saat ini adalah mengacu pada Peraturan Kementerian Pendidikan dan Kebudayaan Indonesia Nomor 53 Tahun 2015 dan Standar penilaian Pendidikan menurut PP Nomor 57 Tahun 2021 (RI, 2021). Bahar et al. (2021) mengungkapkan bahwa salah satu komponen vital dalam proses kegiatan belajar mengajar yang harus dikuasai oleh guru adalah penilaian.

Muluki (2020) mengungkapkan bahwa tes merupakan salah satu alat ukur yang sering digunakan dalam proses pembelajaran baik dalam bentuk tes uraian maupun pilihan ganda. Instrumen tes yang baik adalah instrumen tes yang memenuhi beberapa standar kriteria tes dalam kategori baik atau cukup. Dalam suatu sekolah, penilaian umum yang rutin dilakukan oleh guru maupun instansi salah satunya adalah penilaian akhir semester (PAS). Soal yang digunakan untuk memperoleh data nilai siswa dari PAS sekolah-

sekolah wilayah Kabupaten Batang yang berlaku saat ini, soal tes yang digunakan biasanya menggunakan instrumen tes yang telah dibuat tim soal dari MGMP setempat. Sejalan yang dikatakan oleh Buckley-Walker & Lipscombe (2021) bahwa dalam konteks penilaian, biasanya sekolah menggunakan tim guru untuk merancang penilaian kelas yang sesuai dengan tahun/tingkat kelas.

Penelitian terkait analisis butir soal ulangan akhir semester pernah dilakukan oleh beberapa peneliti sebelumnya. Hasil penelitian Bahar et al. (2021) tentang “analisis karakteristik soal ujian sekolah” menunjukkan bahwa dari 30 soal yang telah dibuat dan diselesaikan oleh siswa, 27 menunjukkan valid, sedangkan 3 soal menunjukkan tidak valid. Hasil estimasi reliabilitas menunjukkan kategori baik, tingkat kesulitan dalam kategori sedang, dan daya beda menunjukkan kategori baik. Penelitian Hasnah (2017) tentang “analisis kualitas soal matematika ujian sekolah kelas XII IPA” menghasilkan bahwa pada aspek tingkat kesukaran soal dari model logistik 1, 2, dan 3 parameter secara berturut-turut adalah 77,5%, 62,5%, dan 52,5% berada pada kategori buruk.

Selanjutnya, penelitian oleh Khasanah et al. (2023) menunjukkan hasil analisis kualitas soal ulangan harian Matematika yang meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas pengecoh adalah terdapat 47% soal dalam kategori tidak valid, reliabilitas dalam kategori tinggi, tingkat kesukaran soal dominan pada kategori mudah sebesar 53% sehingga perlu diperbaiki, sebanyak 13% daya pembeda soal maupun efektivitas pengecoh dalam kategori buruk. Penelitian Susanto et al. (2015) terkait “analisis validitas,

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

reliabilitas, tingkat kesukaran, dan daya beda pada butir soal ujian akhir semester ganjil matematika” menghasilkan kesimpulan bahwa terdapat 45% soal dalam kriteria tidak valid, reliabilitas dalam kategori baik, tingkat kesukaran butir soal dalam kategori kurang baik, dan 15% tergolong jelek sekali berdasarkan analisis daya beda.

Berdasarkan penelitian-penelitian sebelumnya terkait analisis butir soal matematika dapat disimpulkan bahwa soal ulangan harian yang telah dibuat guru mata pelajaran maupun soal ujian hasil musyawarah guru matematika belum sepenuhnya dalam kategori baik, sehingga analisis karakteristik butir soal seharusnya perlu dianalisis agar soal yang disebarluaskan kepada siswa sudah memenuhi karakteristik soal yang baik. Lebih lanjut dikatakan oleh Muluki et al. (2020) bahwa kualitas soal banyak yang tidak diketahui secara pasti dikarenakan analisis butir soal terkait dengan materi, konstruksi soal, bahasa, validitas, reliabilitas, dan analisis butir soal yang terdiri dari tingkat kesukaran, daya pembeda, dan distraktor soal sering dikatakan rendah.

Kegiatan analisis butir soal cukup jarang dilakukan oleh guru karena beberapa faktor, misalnya soal belum sepenuhnya menggambarkan tingkat pencapaian kompetensi peserta didik yang sesungguhnya, selain itu faktor lainnya adalah kegiatan analisis butir jarang dilakukan karena dianggap rumit dan memakan waktu yang cukup lama (Yusuf, 2021).

Selanjutnya, beberapa penelitian sebelumnya yang dikaji adalah analisis soal pada jenjang SMA dan pada materi-materi tertentu, sedangkan untuk analisis karakteristik soal pada tingkat SMP dan pada materi matematika semester gasal tahun ajaran 2021/2022

masih terbatas. Di wilayah Kabupaten Batang khususnya, pada tingkat SMP, penelitian terkait analisis kualitas butir soal untuk soal ulangan maupun soal ujian sangat terbatas, sehingga penelitian terkait analisis kualitas butir soal matematika pada soal ulangan akhir semester yang dibuat oleh musyawarah guru matematika perlu dikaji untuk mengetahui kualitas butir soal yang diselesaikan oleh siswa sehingga nantinya bisa menjadi salah satu referensi untuk tindak lanjut terkait analisis kualitas soal matematika di Kabupaten Batang.

Oleh karena itu, tujuan penelitian ini untuk menganalisis instrumen soal penilaian akhir semester (PAS) kelas VII tahun pelajaran 2021/2022 mata pelajaran matematika terkait dengan 4 kriteria instrumen butir soal yaitu : validitas butir soal, reliabilitas, tingkat kesukaran butir soal, dan daya pembeda butir soal.

## **METODE PENELITIAN**

Jenis penelitian yang digunakan dalam penelitian ini adalah deskriptif kuantitatif dengan teknik dokumentasi. Analisis deskriptif adalah analisis yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul (Tilaar & Hasriyanti, 2019). Subjek dalam penelitian ini adalah siswa kelas VII SMP Takhassus Rafirna Batang tahun pelajaran 2021/2022 sebanyak 69 siswa. Instrumen penelitian ini adalah tes objektif berupa pilihan ganda dan menjodohkan dan soal esai.

Soal dalam penilaian akhir semester (PAS) kelas VII di SMP Takhassus Alquran Rafirna merupakan soal yang dibuat oleh Musyawarah Guru Mata Pelajaran (MGMP) Dinas Pendidikan dan Kebudayaan Kabupaten Batang yang terdiri dari 4 tipe soal, soal

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

tipe A berupa pilihan ganda sebanyak 10 butir yang masing-masing memiliki 4 opsi jawaban, soal tipe B berupa soal Benar/Salah, soal tipe C berupa soal menjodohkan yang masing-masing memiliki 4 opsi jawaban, dan soal tipe D berupa 5 soal uraian. Dalam penelitian ini, yang akan dianalisis adalah soal pada tipe A, C dan D yaitu soal tipe pilihan ganda sebanyak 10 butir dan tipe C sebanyak 10 butir, dan sebanyak 5 butir soal uraian. Materi yang disajikan dalam soal PAS Genap yaitu perbandingan, aritmetika sosial, garis dan sudut, segiempat dan segitiga, dan penyajian data.

Pengumpulan data dilakukan melalui dokumentasi berupa lembar soal PAS, kunci jawaban, dan lembar jawaban siswa. Data dianalisis menggunakan beberapa rumus validitas, reliabilitas, tingkat kesukaran, dan daya beda dengan menggunakan *microsoft excel* dan *software IBM statistics 26 for windows*. Fiska et al. (2021) mengatakan bahwa aspek yang perlu diperhatikan dalam mengetahui kualitas butir soal yang baik, dimulai dari validitas, daya beda, tingkat kesukaran, efektivitas pengecoh dan reliabilitas pada setiap butir soal. Jumlah sampel dalam penelitian ini adalah 69 lembar jawaban siswa.

### **Analisis Validitas**

Bukti validitas dari suatu penilaian kelas di sekolah merupakan konsep penting karena akan dijadikan sebagai keputusan penting berdasarkan penilaian yang dilakukan guru terhadap siswa (Buckley-Walker & Lipscombe, 2021). Validitas butir soal dihitung pada masing-masing jenis soal (pilihan ganda dan uraian) dengan bantuan *software IBM statistic 26 for windows* dengan melihat hasil perhitungan pada *pearson correlation* ( $r$  hitung), selanjutnya hasil

$r$  hitung dibandingkan dengan tabel *r product moment (rtabel)* dengan  $N = 69$  untuk menentukan kriteria setiap butir soal pada masing-masing jenis soal. Validitas butir soal ditentukan dengan menggunakan rumus korelasi *product moment pearson* dimana skor pada masing-masing butir soal dikaitkan dengan skor keseluruhan (Khasanah et al., 2023). Hartono et al. (2023) mengungkapkan bahwa prediksi validitas merupakan korelasi pearson yang biasa antar skor tes dan skor kriteria validasi. Skor tes yang diperoleh berupa pilihan ganda dan menjodohkan yang masing-masing memiliki 4 pilihan jawaban dihitung dengan menggunakan skor benar dinilai 1 dan skor untuk jawaban salah adalah 0. Sedangkan skor tes bentuk soal uraian dihitung dengan memperhatikan kontribusi bobot skor untuk tiap jenis tes.

Uji validitas butir soal baik jenis soal pilihan ganda maupun uraian dihitung dengan menggunakan *software IBM SPSS 26 for windows*, dengan  $N = 69$  dan taraf signifikansi 0,05 atau 95% dengan kriteria keputusan adalah:

- a. Jika  $r$  hitung  $>$   $r$  tabel, maka butir soal dikatakan valid
- b. Jika  $r$  hitung  $\leq$   $r$  tabel, maka butir soal dikatakan tidak valid.

### **Analisis Reliabilitas**

Reliabilitas tes digunakan untuk menilai konsistensi tindakan tes yang hasilnya menunjukkan kejagan (Khasanah et al., 2023). Setelah mengetahui butir soal yang valid, selanjutnya dilakukan pengujian reliabilitas. Reliabilitas butir soal yang dihitung adalah reliabilitas komposit. Reliabilitas tes pada soal pilihan ganda dan uraian pada butir soal yang valid dihitung dengan bantuan *software IBM statistics 26 for windows*, kemudian dikonversikan dengan rumus *alpha*

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

strata untuk menghitung reliabilitas komposit dari jenis soal pilihan ganda dan uraian. Model yang digunakan adalah *alpha Cronbach*, kemudian hasil perhitungan dengan *alpha Cronbach* ditentukan berdasarkan kriteria keputusan koefisien alpha yang  $> 0,70$  dikatakan reliabilitas dalam kategori baik (Fiska et al., 2021).

Garis besar prosedur perhitungan koefisien reliabilitas alpha berstrata diperuntukkan bagi sebuah tes tunggal yang berisi kelompok-kelompok butir yang homogen. Namun, itu juga dapat digunakan untuk menghitung reliabilitas skor komposit yang terdiri dari masing-masing tes yang diujikan secara terpisah bila ukuran reliabilitasnya dari masing-masing tes itu dapat diestimasi dan setiap komponen diberikan bobot yang sama. Prosedur yang digunakan untuk menghitung reliabilitas skor komposit, yaitu :

- Mengestimasi ukuran reliabilitas tiap-tiap komponen atau strata.
- Menghitung varians dari komponen dan skor komposit.
- Menggunakan persamaan 4) untuk menghitung reliabilitas skor komposit.
- Menghitung *Standard Error Measurement* (SEM) skor komposit.

Berdasarkan nilai alpha cronbach, rentangan nilai koefisien alpha berkisar antara 0 (tanpa reliabilitas) sampai dengan 1 (reliabilitas sempurna). Menurut Budiastuti & Bandur (2018) untuk menentukan nilai koefisien alpha sebagai berikut:

- 0 artinya tidak memiliki reliabilitas (*no reliability*)
- $R > 0,70$  artinya reliabilitas dapat diterima (*acceptable reliability*)
- $R > 0,80$  artinya reliabilitas baik (*good reliability*)

- 0,90 artinya reliabilitas sangat baik (*excellent reliability*)
- 1 artinya reliabilitas sempurna (*perfect reliability*)

### Analisis Tingkat Kesukaran

Selanjutnya adalah tingkat kesukaran soal. Tingkat kesukaran soal adalah representasi numerik dari tingkat kesulitan suatu soal tes (Fietri et al., 2021). Untuk menghitung indeks kesukaran butir soal adalah mengoreksi lembar jawaban siswa pada tipe soal A dan C, dan untuk jawaban benar diberi skor 1, sedangkan untuk jawaban yang salah diberi skor 0. Angka indeks kesukaran dihitung menggunakan rumus (1):

$$P_i = \frac{\sum X_i}{N} \quad (1)$$

Keterangan:

- $P_i$  : tingkat kesukaran butir soal  
 $\sum X_i$  : jumlah peserta tes yang menjawab benar  
 $N$  : jumlah peserta tes

Sedangkan tingkat kesukaran ditentukan menggunakan kriteria pada Tabel 1.

Tabel 1. Kriteria tingkat kesukaran

Nilai $P$	Kategori
$p < 0,30$	Sukar
$0,30 \leq p \leq 0,70$	Sedang
$p > 0,70$	Mudah

Sumber: Iskandar & Rizal (2018)

Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Siswa tidak akan meningkatkan kemampuan mereka untuk memecahkan masalah jika soal terlalu mudah. Sebaliknya, jika pertanyaannya terlalu sukar, siswa akan kehilangan harapan dan motivasi untuk mencoba lagi karena tampaknya tidak mungkin tercapai (Ambarwati & Ismiyati, 2021).

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

### Analisis Daya Pembeda

Selanjutnya, untuk menentukan butir soal memiliki daya pembeda dalam kategori baik, sedang, atau rendah, digunakan rumus (2):

$$D = \frac{\sum X_A}{N_A} - \frac{\sum X_B}{N_B} \quad (2)$$

Keterangan:

- $D$  : Indeks daya beda butir soal  
 $\sum X_A$  : Banyaknya peserta tes yang menjawab benar pada kelompok atas  
 $\sum X_B$  : Banyaknya peserta tes yang menjawab benar pada kelompok bawah  
 $N_A$  : Banyaknya peserta tes pada kelompok atas  
 $N_B$  : Banyaknya peserta tes pada kelompok bawah

Adapun kriteria yang digunakan adalah tercantum dalam Tabel 2.

Tabel 2. Kriteria indeks daya beda butir

Indeks Daya Beda	Kriteria Butir
$0,40 \leq D \leq 1,0$	Sangat Baik
$0,3 \leq D < 0,4$	Baik
$0,2 \leq D < 0,3$	Cukup
$0,0 < D \leq 0,2$	Tidak Baik
$D \leq 0,0$	Sangat Tidak Baik

Sumber: Iskandar & Rizal (2018)

## HASIL DAN PEMBAHASAN

Data hasil tes siswa yang telah diperoleh dianalisis secara kuantitatif untuk mengetahui karakteristik butir soal yang dihasilkan diantaranya: validitas, reliabilitas, tingkat kesukaran soal, daya beda soal, pada soal penilaian akhir semester (PAS) kelas VIII mata pelajaran matematika tahun pelajaran 2021/2022 jenis soal pilihan ganda dan uraian.

### 1. Analisis Validitas

Uji validitas butir soal baik jenis soal pilihan ganda maupun uraian dihitung dengan menggunakan *software*

*IBM SPSS 26 for windows*, dengan  $N = 69$  dan taraf signifikansi 0,05.

Uji validitas dihitung pada masing-masing jenis soal untuk mengetahui valid tidaknya soal pada masing-masing butir di setiap jenis soal.

Hasil perhitungan validitas tiap butir soal jenis soal pilihan ganda dengan *software IBM statistic for windows* disajikan pada Tabel 3, kemudian dibandingkan dengan  $r$  tabel pada  $N = 69$  diperoleh  $r_{tabel} = 0,235$ .

Tabel 3. Validitas butir soal pilihan ganda

Nomor Butir Soal	Nilai Pearson Correlation	R tabel	Keterangan
1A	0,357	0,235	Valid
2A	0,382	0,235	Valid
3A	0,230	0,235	Tidak Valid
4A	0,266	0,235	Valid
5A	0,207	0,235	Tidak Valid
6A	0,400	0,235	Valid
7A	0,143	0,235	Tidak Valid
8A	0,241	0,235	Valid
9A	0,527	0,235	Valid
10A	0,399	0,235	Valid
1C	0,300	0,235	Valid
2C	0,374	0,235	Valid
3C	0,293	0,235	Valid
4C	0,292	0,235	Valid
5C	0,421	0,235	Valid
6C	0,463	0,235	Valid
7C	0,386	0,235	Valid
8C	0,335	0,235	Valid
9C	0,398	0,235	Valid
10C	0,508	0,235	Valid

Berdasarkan Tabel 3 terlihat bahwa dari 20 soal pilihan ganda yang dianalisis, terdapat 17 soal (85%) pada kriteria valid, dan 3 soal (15%) pada kriteria tidak valid. Artinya, dari 20 soal tersebut, soal yang tidak valid yaitu pada soal nomor 3A, 5A, dan 7A harus dibuang atau diganti dengan soal yang lain. Menurut Anita et al. (2018) faktor yang mempengaruhi validitas adalah faktor internal dan eksternal, faktor internal tes seperti arahan tes yang

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

disusun dengan makna tidak jelas. Sedangkan faktor eksternal antara lain waktu pengerjaan tidak cukup, sehingga siswa menjadi gugup dan tegang karena tergesa-gesa bahkan bisa melakukan kecurangan.

Selanjutnya, dilakukan uji validitas butir soal untuk jenis soal uraian untuk mengetahui valid atau tidaknya setiap butir soal pada jenis soal uraian (Tabel 4).

Tabel 4. Validitas soal uraian

Nomor Butir Soal	Nilai Pearson Correlation	R tabel	Keterangan
1	0,807	0,235	Valid
2	0,863	0,235	Valid
3	0,718	0,235	Valid
4	0,764	0,235	Valid
5	0,624	0,235	Valid

Berdasarkan Tabel 4 diperoleh bahwa kelima butir soal uraian (100%) dalam soal penilaian akhir semester kelas VII tahun pelajaran 2021/2022 mata pelajaran matematika dikatakan valid karena semua nilai *pearson correlation* pada setiap butir soal lebih dari nilai pada  $r_{tabel}$  dengan  $N = 69$ .

## 2. Analisis Reliabilitas

Uji reliabilitas yang digunakan dalam penelitian ini adalah dengan menggunakan *Coefficient alpha* (*Cronbach's Alpha*). Reliabilitas dalam penelitian ini adalah menggunakan uji reliabilitas skor komposit yang dihitung berdasarkan reliabilitas strata, yang masing-masingnya diperlakukan sebagai sebuah subtes tunggal. Reliabilitas dihitung pada masing-masing jenis soal, kemudian dihitung menggunakan rumus alpha berstrata.

Koefisien *alpha* berstrata, merupakan kasus khusus yang dapat digunakan untuk menghitung reliabilitas komposit dengan berdasarkan rumus (3) Feldt dan Brenan:

$$r_{STRAT,\alpha} = 1 - \frac{\sum_{j=1}^k \sigma_i^2 (1 - r_i)}{\sigma_c^2} \quad (3)$$

Keterangan:

$r_{STRAT,\alpha}$  : Koefisien reliabilitas alpha berstrata

$\sigma_i^2$	:	Varians butir pada komponen ke-i
$r_i$	:	Reliabilitas komponen ke-i
$\sigma_c^2$	:	Varians skor komposit (skor total tes)

Koefisien alpha berstrata di atas diperkenalkan oleh Cronbach yang berguna untuk mengestimasi reliabilitas instrumen yang terdiri dari beberapa kategori. Sama seperti koefisien Alpha, koefisien Alpha berstrata adalah pengukuran *internal consistency* dengan melibatkan komponen-komponen tes.

Berdasarkan pengujian, diperoleh hasil reliabilitas komposit yang disajikan pada Tabel 5.

Tabel 5. Statistik deskriptif hasil pengukuran soal PAS

Dimensi	Varians	Reliabilitas
Pilihan Ganda	0,002	0,636
Uraian	0,121	0,798
<b>Skor Total</b>	<b>0,123</b>	

$$r_{STRAT} = 1 - \frac{0,002(1-0,636)+0,121(1-0,798)}{0,123} = 0,795$$

Hasil perhitungan uji reliabilitas dengan bantuan *software IBM statistics for windows* metode diperoleh nilai pada *Alpha Cronbach* dan dengan perhitungan manual dengan  $r_{STRAT}$  adalah 0,795 kemudian mengacu pada penentuan koefisien nilai alpha, maka diperoleh interpretasi bahwa instrumen tes yang berupa 17 butir pilihan ganda dan 5 butir soal uraian adalah reliabel karena koefisien nilai *alpha* = 0,795 > 0,70.

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

### 3. Analisis Tingkat Kesukaran

Berikut adalah hasil perhitungan tingkat kesukaran soal untuk setiap butir soal pada jenis soal pilihan ganda dan uraian. Hasil Perhitungan dan Kriteria Tingkat Kesukaran Soal pada Setiap Butir Jenis Soal Pilihan Ganda disajikan pada Tabel 6.

Tabel 6. Hasil perhitungan dan kriteria tingkat kesukaran soal pada setiap butir jenis soal pilihan ganda

Nomor Butir Soal	Indeks Kesukaran	Kategori
1A	0,15942	Sukar
2A	0,2899	Sukar
3A	0,27536	Sukar
4A	0,21739	Sukar
5A	0,11594	Sukar
6A	0,14493	Sukar
7A	0,24638	Sedang
8A	0,44928	Sukar
9A	0,28986	Sukar
10A	0,27536	Sukar
1C	0,76812	Mudah
2C	0,78261	Mudah
3C	0,14493	Sukar
4C	0,50725	Sedang
5C	0,44928	Sedang
6C	0,3913	Sedang
7C	0,52174	Sedang
8C	0,36232	Sedang
9C	0,42029	Sedang
10C	0,30435	Sukar

Berdasarkan tabel 6, indeks kesukaran soal pilihan ganda berada pada kriteria mudah, sedang, dan sukar. butir soal dengan tingkat kesukaran mudah sebaiknya direvisi dan dikaji ulang menjadi lebih baik dan lebih kompleks agar peserta didik yang mengikuti tes bisa menjawab setiap butir soal dengan baik dan benar. soal dengan kategori sedang artinya soal tersebut tidak terlalu mudah dan tidak terlalu sukar. Soal dalam kategori sedang sebaiknya disimpan dalam bank soal. Hal ini sejalan dengan yang

dikemukakan oleh Amiruddin et al. (2020) bahwa soal dalam kategori sedang artinya soal tersebut dapat menstimulus siswa untuk berpikir dalam menyelesaikan setiap permasalahan yang diangkat pada butir soal tersebut, sehingga soal-soal tersebut sebaiknya disimpan di bank soal dan dapat dipergunakan kembali dalam tes-tes berikutnya.

Sedangkan soal dalam kategori sukar menurut Purwanti (2014) menyarankan bahwa butir soal kategori sukar harus diadakan perbaikan dengan mengganti soal dimana sebagian peserta didik mampu menjawab soal tersebut karena kemungkinan sebagian besar siswa telah memahami materi yang ditanyakan. Selanjutnya Kurniasi et al. (2020) juga mengungkapkan bahwa butir soal dengan tingkat sukar merupakan soal yang materinya butuh pemahaman tinggi dan perlu ketelitian dalam perhitungan. Soal yang baik adalah soal yang tingkatnya tidak terlalu mudah atau tidak terlalu sukar karena soal yang terlalu mudah tidak bisa merangsang siswa untuk berusaha keras memecahkan soal.

Tabel 7. Analisis tingkat kesukaran butir jenis soal pilihan ganda

Kriteria	Butir Soal	Jumlah	Persentase
Mudah	1C, 2C	2	10%
Sedang	7A, 4C, 5C, 6C, 7C, 8C, 9C	7	35%
Sukar	1A, 2A, 3A, 4A, 5A, 6A, 8A, 9A, 10A, 3C, 10C	11	55%

Berdasarkan Tabel 7, terlihat bahwa soal penilaian akhir semester (PAS) kelas VII mata pelajaran matematika tahun pelajaran 2021/2022 yang terdiri dari 10 soal pilihan ganda

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

(tipe A) dan 10 soal menjodohkan (tipe C) yang masing-masing memiliki 4 opsi jawaban, sehingga total butir soal yang dianalisis sebanyak 20 butir soal ditemukan 2 butir soal atau sekitar 10% berada pada kategori mudah, 7 soal atau sekitar 35% berada pada kategori sedang, dan 11 soal atau sekitar 55% berada pada kategori sukar.

Pada butir soal nomor 1C dan 2C (jenis soal menjodohkan) merupakan butir soal yang paling mudah, karena pada kedua butir soal tersebut berasal dari materi yang mudah dipahami yaitu berupa soal untuk melihat gambar yang disajikan sehingga tidak menstimulus siswa untuk berpikir dan menyelesaikan suatu permasalahan.

Butir soal nomor 7A, 4C, 5C, 6C, 7C, 8C, dan 9C merupakan butir soal yang berada dalam kategori tingkat kesukaran sedang, artinya butir-butir soal tersebut tidak terlalu mudah dan tidak terlalu sulit. Butir soal dalam kategori sedang bisa dipertahankan ataupun dijadikan acuan dalam membuat soal-soal untuk penilaian selanjutnya. Menurut Amiruddin et al. (2020) mengungkapkan bahwa butir soal yang memiliki tingkat kesukaran sedang sebaiknya disimpan dalam bank soal sehingga dapat dimanfaatkan kembali untuk tes-tes berikutnya. Sedangkan untuk butir soal sukar terdapat beberapa hal yang dapat dijadikan tindak lanjut yaitu: (1) butir soal tersebut dibuang dan tidak dikeluarkan lagi dalam tes atau penilaian yang akan datang, (2) diteliti kembali, dilacak dan ditelusuri sehingga dapat diketahui faktor apa yang menyebabkan butir item yang bersangkutan sulit dijawab oleh tester.

Selanjutnya, setelah dilakukan perbaikan, butir-butir soal tersebut dapat dikeluarkan lagi pada tes atau penilaian yang akan datang, (3) butir

soal yang sukar sebaiknya digunakan dalam tes seleksi yang sifatnya ketat karena kondisi tersebut mendukung penggunaan soal yang sukar dengan asumsi bahwa tester memiliki kemampuan yang memadai dan akan lolos dalam seleksi yang diadakan (Susanto et al., 2015). Menurut Hanna & Retnawati (2022) mengungkapkan bahwa soal tes yang baik seharusnya tidak memiliki pertanyaan dengan tingkat kesulitan tertinggi ataupun terendah.

Pada butir soal nomor 1A, 2A, 3A, 4A, 5A, 6A, 8A, 9A, 10A, 3C, 10C atau berjumlah 11 soal dengan persentase 55% merupakan butir soal yang berada pada kategori sukar. Butir soal nomor 9A, 10A, dan 10C merupakan butir soal yang berasal dari materi yang kurang dipahami oleh siswa, selain itu membutuhkan pemahaman yang tinggi untuk bisa menyelesaikannya. Untuk butir soal nomor 3A dan 4A merupakan butir soal yang membutuhkan ketelitian dan kecermatan dalam menyelesaikannya. Butir soal sukar yang diperoleh adalah 11 butir soal atau 55% yang artinya soal sukar pada jenis soal pilihan ganda memiliki jumlah paling banyak dari soal pada kategori mudah dan sedang sehingga soal-soal tersebut perlu dikaji kembali untuk mengetahui apa penyebab soal-soal tersebut dianggap sulit oleh siswa.

Berdasarkan kriteria keputusan untuk kesukaran butir soal, diketahui bahwa soal yang dianggap baik atau butir soal berada pada kategori sedang adalah jika indeks kesukaran butir soal berada pada nilai 0,30 – 0,70, sehingga dapat dikatakan bahwa soal penilaian akhir semester kelas VII SMP tahun pelajaran 2021/2022 mata pelajaran matematika yang digunakan belum memiliki kualitas baik ditinjau dari

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

tingkat kesukaran soal jenis pilihan ganda karena butir soal yang berada pada kategori sedang hanya berjumlah 7 butir dari 20 butir soal pilihan ganda atau 35% saja. Selain itu, hasil analisis butir soal pada tingkat kesukaran soal diperoleh bahwa soal pada kategori sukar memiliki persentase yang paling tinggi dibandingkan dengan kategori mudah dan sukar.

Tabel 8. Indeks kesukaran pada masing-masing butir soal uraian

Nomor Butir Soal	Indeks Kesukaran	Kriteria
1	0,271739	Sukar
2	0,242754	Sukar
3	0,26087	Sukar
4	0,228261	Sukar
5	0,233696	Sukar

Berdasarkan hasil pengujian pada Tabel 8 indeks kesukaran butir soal jenis soal uraian, yang berjumlah soal diperoleh bahwa kelima soal tersebut berada pada kategori sukar. Sehingga, perlu dikaji kembali untuk mengetahui apa yang menyebabkan kelima butir soal tersebut dianggap sukar oleh siswa.

Soal uraian yang dikerjakan siswa dianalisis oleh guru mata pelajaran berdasarkan rubrik penskoran yang telah dibuat oleh tim MGMP Matematika. Hasil jawaban siswa dinilai dari rentang 0 sampai 5. Berdasarkan hasil yang diperoleh terkait dengan indeks kesukaran soal pada bentuk uraian diperoleh bahwa kelima soal uraian dalam kategori sukar, yang berarti bahwa sebagian besar siswa mendapat skor yang rendah untuk soal bentuk uraian. Berdasarkan telaah yang telah dilakukan, banyak siswa yang tidak mampu menjawab soal berdasarkan permasalahan yang diberikan, siswa belum mampu menerapkan konsep dalam matematika yang digunakan untuk menyelesaikan soal, selain itu,

terdapat lembar jawab siswa yang kosong atau tidak menjawab soal uraian. Selanjutnya, faktor waktu pengerjaan juga menjadi faktor siswa mendapat skor yang rendah. Saat dilakukan wawancara terhadap beberapa siswa, diperoleh bahwa sebagian besar siswa merasa kekurangan waktu dalam mengerjakan soal matematika.

#### 4. Analisis Daya Pembeda

Berikut hasil pengujian indeks daya beda pada setiap butir soal jenis soal pilihan ganda dan uraian. Indeks daya beda pada masing-masing butir soal pilihan ganda disajikan pada Tabel 9.

Tabel 9. Indeks daya beda pada masing-masing butir soal pilihan ganda

Nomor Butir Soal	Indeks Kesukaran	Kategori
1A	0,2	Cukup
2A	0,28571	Cukup
3A	0,2	Cukup
4A	0,2	Cukup
5A	0	Tidak baik
6A	0,22857	Cukup
7A	0,085714	Tidak baik
8A	0,22857	Cukup
9A	0,31429	Baik
10A	0,28571	Cukup
1C	0,22857	Cukup
2C	0,25714	Cukup
3C	0,171428	Tidak baik
4C	0,141857	Tidak baik
5C	0,2	Cukup
6C	0,25714	Cukup
7C	0,28571	Cukup
8C	0,3714286	Baik
9C	0,31429	Baik
10C	0,34286	Baik

Berdasarkan hasil perhitungan yang telah diperoleh terkait dengan indeks daya beda soal disimpulkan bahwa dari 16 soal dengan rincian 10 soal pilihan ganda, sebanyak 20%, 2 butir soal, mempunyai daya pembeda

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

yang tidak baik, yang artinya soal tersebut tidak dapat digunakan ataupun direvisi, melainkan dengan menggantinya dengan soal yang baru. Butir soal yang memiliki daya pembeda rendah masih dapat diperbaiki (Firmansyah, 2018). Soal yang mempunyai tingkat daya beda rendah disarankan untuk diganti namun masih bisa digunakan dengan direvisi. Sedangkan untuk tingkat daya beda kategori baik, bisa langsung digunakan. Susanto et al. (2015) mengungkapkan bahwa soal yang memiliki daya pembeda yang baik hendaknya dimasukkan dan dicatat dalam buku bank soal. Butir soal tersebut dapat dikeluarkan lagi dalam tes selanjutnya karena kualitasnya sudah cukup memadai.

Tabel 10. Distribusi Indeks Daya Pembeda Butir Soal Jenis Soal Pilihan Ganda

Kriteria	Butir Soal	Jumlah	Persentase
Sangat Baik		0	0%
Baik	9A, 8B, 9B, 10B	4	20%
Cukup	1A, 2A, 3A, 4A, 6A, 8A, 10A, 1B, 2B, 5B, 6B, 7B	12	60%
Tidak Baik	5A, 7A, 3B, 4B	4	20%

Dari hasil perhitungan (Tabel 10), terdapat 4 butir soal dalam kriteria tidak baik pada pengujian daya beda soal, 2 soal pada pilihan ganda dan 2 soal pada bentuk menjodohkan.

Berdasarkan Tabel 10, diperoleh 4 butir soal atau 20% berada pada kategori butir soal yang baik

berdasarkan daya pembeda soal. Hal ini membuktikan bahwa keempat butir soal tersebut memiliki pembeda yang baik, yaitu dapat membedakan siswa pada kelompok atas dan kelompok bawah.

Butir soal pada nomor 1A, 2A, 3A, 4A, 6A, 8A, 10A, 1B, 2B, 5B, 6B, 7B atau sebanyak 12 butir soal berada pada kategori butir soal yang memiliki daya pembeda dalam kriteria cukup. Artinya bahwa soal-soal dalam kategori cukup pada daya pembeda soal sudah mampu membedakan tingkat daya pembeda siswa antara kelompok atas dengan kelompok bawah. Pada butir soal nomor 5A, 7A, 3B, 4B atau sebanyak 4 butir soal dalam kategori daya pembeda yang tidak baik sehingga butir-butir soal tersebut sebaiknya tidak digunakan lagi untuk pengujian ataupun tes selanjutnya.

Berdasarkan kriteria daya pembeda terdapat 4 butir soal atau sekitar 20% yang memiliki daya pembeda tidak baik dan sebaiknya untuk pengujian atau tes selanjutnya tidak digunakan lagi. Menurut Iskandar & Rizal (2018) dalam penelitiannya menemukan bahwa dari segi daya beda terdapat 4 butir soal (5%) memiliki daya beda sangat baik yang berarti bahwa butir soal tersebut mampu membedakan antara peserta tes yang pandai dan peserta tes yang kurang pandai, 20% yang memiliki daya beda pada kategori baik, 11 butir (13,75%) memiliki daya beda cukup baik yang berarti harus melewati tahap revisi sehingga dapat digunakan kembali, dan 49 butir (61,25%) memiliki daya beda yang tidak baik sehingga harus dibuang. Selanjutnya, data hasil analisis indeks daya beda soal tersaji pada Tabel 11.

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

Tabel 11. Indeks daya beda pada masing-masing butir soal uraian

Nomor Butir Soal	Indeks Daya Beda	Kriteria
1	0,17857	Tidak Baik
2	0,13571	Tidak Baik
3	0,07142	Tidak Baik
4	0,07142	Tidak Baik
5	0,08214	Tidak Baik

Berdasarkan Tabel 11 terlihat bahwa hasil pengujian daya pembeda butir soal jenis soal uraian sebanyak 5 butir, diperoleh kelima soal tersebut berada pada kategori tidak baik, sehingga butir-butir tersebut sebaiknya tidak digunakan lagi untuk tes-tes atau pengujian berikutnya, dengan kata lain butir-butir tersebut harus dibuang dan tidak dapat disimpan di bank soal. Sejalan dengan pendapat Hartono et al. (2023) bahwa untuk item soal dengan indeks daya pembeda sangat kecil bahkan negatif sebaiknya tidak digunakan lagi.

Faktor rendahnya indeks daya pembeda soal diantaranya adalah dipengaruhi oleh rendahnya skor siswa pada soal uraian. Daya pembeda merupakan kemampuan butir soal untuk membedakan siswa yang berkemampuan tinggi atau memahami materi dengan siswa yang berkemampuan rendah atau kurang memahami materi (Muluki et al., 2020). Rendahnya indeks daya pembeda yang diperoleh pada hasil penelitian terhadap soal uraian artinya kelima soal tersebut tidak mampu membedakan siswa berkemampuan tinggi dengan siswa berkemampuan rendah. Hal ini dikarenakan skor yang diperoleh siswa sebagian besar adalah rendah.

## KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan, dapat diperoleh kesimpulan yaitu terdapat 17 butir soal pilihan

ganda (85%) dinyatakan valid, dan 3 butir soal (15%) dinyatakan tidak valid, sedangkan pada butir soal bentuk uraian yang berjumlah 5 butir, semuanya dinyatakan valid. Reliabilitas soal baik pada bentuk soal pilihan ganda maupun bentuk soal uraian. Tingkat kesukaran butir soal untuk bentuk pilihan ganda diperoleh 2 soal (10%) dalam kategori mudah, 7 soal (35%) dalam kategori sedang, dan 11 soal (55%) dalam kategori sukar. Sehingga tingkat kesukaran butir soal pada soal bentuk pilihan ganda kurang baik, karena keseimbangan soal yang mudah, sedang, dan sukar belum proporsional. Soal yang sukar jumlahnya melebihi setengah dari jumlah soal pilihan ganda. Sedangkan untuk tingkat kesukaran pada soal bentuk uraian yang berjumlah 5 butir, semuanya dalam kategori sukar. Daya pembeda soal pada bentuk soal pilihan ganda diperoleh 4 soal (20%) pada kategori baik, 12 soal (60%) pada kategori cukup, dan 4 soal (20%) pada kategori tidak baik.

Berdasarkan kajian penelitian yang telah dilakukan, beberapa saran penelitian diantaranya yaitu: (1) TIM pembuat soal matematika dan MGMP guru matematika setempat masih perlu meningkatkan kualitas soal karena belum sepenuhnya soal yang diujikan baik dari validitas, reliabilitas, tingkat kesukaran, dan daya pembeda; (2) Pihak sekolah perlu meningkatkan kemampuan para guru dalam melakukan analisis butir soal dan akan lebih baik jika diadakan pelatihan bagi para guru sehingga soal akan lebih berkualitas.

## DAFTAR PUSTAKA

- Ambarwati, Y. F., & Ismiyati, I. (2021). Analisis Butir Soal Pilihan Ganda Ulangan Akhir Semester Genap Mata Pelajaran Kearsipan. *Measurement In Educational*

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

- Research (Meter)*, 1(2), 64–75.  
<https://doi.org/10.33292/meter.v1i2.144>
- Amiruddin, K., Mania, S., Nina Ichiana, N., Farham Majid, A., Tarbiyah dan Keguruan, F., & Islam Negeri Alauddin Makassar, U. (2020). Analysis of Final School Examination Questions for Mathematics Subjects. *Alauddin Journal of Mathematics Education Journal Homepage*, 2(2), 207–217. <http://journal.uin-alauddin.ac.id/index.php/ajme>
- Anita, A., Tyowati, S., & Zuldafrial, Z. (2018). Analisis Kualitas Butir Soal Fisika Kelas X Sekolah Menengah Atas. *Edukasi: Jurnal Pendidikan*, 16(1), 35–47. <https://doi.org/10.31571/edukasi.v16i1.780>
- Bahar, R., Istiyono, E., Widihastuti, W., Munadi, S., Nuryana, Z., & Fajaruddin, S. (2021). Analisis Karakteristik Soal Ujian Sekolah Hasil Musyawarah Guru Matematika Di Tasikmalaya. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(4), 2660. <https://doi.org/10.24127/ajpm.v10i4.4359>
- Buckley-Walker, K., & Lipscombe, K. (2021). Validity and The Design of Classroom Assessment in Teacher Teams. *Australian Educational Researcher*, 49(2), 425–444. <https://doi.org/10.1007/s13384-021-00437-9>
- Budiastuti, D., & Bandur, A. (2018). Validitas dan Reliabilitas Penelitian. In *Mitra Wacana Media*. [www.mitrawacanamedia.com](http://www.mitrawacanamedia.com)
- Fietri, W. A., Zulyusri, & Violita. (2021). Analisis Butir Soal Biologi Kelas XI Madrasah Aliyah Sakinah Kerinci Menggunakan Program Komputer Anates 4.0 For Windows. *Natural Science: Jurnal Penelitian Bidang IPA Dan Pendidikan IPA*, 17(2), 28–35. <https://doi.org/10.35791/agrsosek.17.2.2021.33834>
- Firmansyah, H. (2018). Analisis Butir Soal Tes Pilihan Ganda Mata Pelajaran Penjas-Orkes Kelas VII SMPN 1 Wonosari. *Yogyakarta: Universitas Negeri Yogyakarta*, 39.
- Fiska, J. M., Hidayati, Y., Qomaria, N., & Hadi, W. P. (2021). Analisis Butir Soal Ulangan Harian Ipa Menggunakan Software Anates Pada Pendekatan Teori Tes Klasik. *Natural Science Education Research*, 4(1), 65–76. <https://doi.org/10.21107/nser.v4i1.8133>
- Hanna, W. F., & Retnawati, H. (2022). Analisis Kualitas Butir Soal Matematika Menggunakan Model Rasch Dengan Bantuan Software Quest. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 11(4), 3695. <https://doi.org/10.24127/ajpm.v11i4.5908>
- Hartono, W., Hadi, S., Rosnawati, R., & Retnawati, H. (2023). Exploration of Student' Cognitive Mathematics Ability Diagnostic Instruments: Validity, Reliability, and Item Characteristics. *Pegem Journal of Education and Instruction*, 13(3), 386–394. <https://doi.org/10.47750/pegegog.13.03.39>
- Hasnah. (2017). Analisis Kualitas Soal Matematika Ujian Sekolah Kelas XII IPA SMA Negeri di Watansoppeng Berdasarkan Teori

DOI: <https://doi.org/10.24127/ajpm.v12i3.7575>

- Respon Butir. *PEP Educational Assessment*, 1(1), 27–33.
- 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>
- Khasanah, I., Fuady, A., & Sunismi. (2023). Analisis Soal Ulangan Harian Matematika Bentuk Pilihan Ganda. *Mathema Journal*, 5(2), 110–125.
- Kurniasi, E. R., Y, Y., & Karennisa, F. (2020). Analisis Soal Ulangan Harian Matematika Kelas IX SMP Negeri 1 Toboali. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, 12(1), 43–52. <https://doi.org/10.37640/jip.v12i1.276>
- Maemonah. (2018). Asesmen Pembelajaran. In *PGMI Press UIN Suka*. PGMI Press UIN Suka.
- Muluki, A., Bundu, P., & Sukmawati, I. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV MI Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86. <https://doi.org/10.23887/jisd.v4i1.23335>
- Purwanti, M. (2014). Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan Microsoft Office Excel 2010. *Jurnal Pendidikan Akuntansi Indonesia*, 12(2). <https://doi.org/10.21831/jpai.v12i2.2710>
- RI, P. (2021). Peraturan Pemerintah Republik Indonesia Nomor 57 Tahun 2021 Tentang Standar Nasional Pendidikan. *Standar Nasional Pendidikan*. [https://jdih.kemdikbud.go.id/arsip/Salinan\\_PP\\_Nomor\\_57\\_Tahun\\_2021.pdf](https://jdih.kemdikbud.go.id/arsip/Salinan_PP_Nomor_57_Tahun_2021.pdf)
- Susanto, H., Rinaldi, A., & Novalia. (2015). Analisis Validitas Reabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika. *Aljabar: Jurnal Pendidikan Matematika*, 6(2), 2013.
- Syahputra, A. T., Nurjannah, N., & Arsyam, M. (2020). Pemberian Skor Dan Sistem Penilaian Dalam Pembelajaran. *Jurnal Pendidikan*, 2(1), 1–8.
- Tilaar, A. L. F., & Hasriyanti, H. (2019). Analisis Butir Soal Semester Ganjil Mata Pelajaran Matematika pada Sekolah Menengah Pertama. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, 8(1), 57–68. <https://doi.org/10.15408/jp3i.v8i1.13068>
- Yusuf, R. (2021). Analisis Kualitas Butir Soal Ujian Nasional Mata Pelajaran Matematika Sekolah Menengah Pertama. *Journal of Didactic Mathematics*, 1(3), 158–164. <https://doi.org/10.34007/jdm.v1i3.417>