



Analisis kualitas butir soal pilihan ganda materi statistika SMA berdasarkan pendekatan teori tes klasik

Else Ona Sabat¹, Fransiska Gusman², Bakher Nenotaek³

¹ Pendidikan Matematika, Institut Pendidikan Soe, Soe

² Teknologi Informasi, Universitas Utpadaka Swastika, Tangerang Banten

³ Pendidikan Matematika, Universitas Nusa Cendana, Kupang

alseonasabat@gmail.com

Abstract

This study aims to evaluate the empirical quality of multiple-choice test instruments for Mathematics subjects in Statistics material at the high school (SMA) level. The analyzed instrument consisted of 20 multiple-choice questions that were tested on 25 students in class X of SMA PGRI Mnelalete. The research method used is quantitative descriptive with a Classical Test Theory approach. Data analysis was carried out with the help of Microsoft Excel software to calculate the parameters of the validity of the question item, the reliability of the instrument, the level of difficulty, and the differentiating power of the question. The results of the analysis showed that the reliability coefficient of the test was 0.84 (high category). Cumulatively, of the 20 questions analyzed, there were 14 questions (70%) that met the good quality criteria so that they were immediately accepted in the question bank, 4 questions (20%) required revisions to sentence construction or distractor functions, and 2 questions (10%) were discarded because they had insignificant discriminating power. Periodic evaluation of this spreadsheet-based empirical assessment instrument has proven to be effective in helping teachers in mapping the feasibility of measuring student learning outcomes.

Keywords: Question Item Analysis; Classical Test Theory; Statistics; Validity; Reliability.

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kualitas empiris instrumen tes pilihan ganda mata pelajaran Matematika pada materi Statistika tingkat Sekolah Menengah Atas (SMA). Instrumen yang dianalisis terdiri atas 20 butir soal pilihan ganda yang diujicobakan kepada 25 siswa kelas X SMA PGRI Mnelalete. Metode penelitian yang digunakan adalah deskriptif kuantitatif dengan pendekatan Teori Tes Klasik (Classical Test Theory). Analisis data dilakukan dengan bantuan perangkat lunak Microsoft Excel untuk menghitung parameter validitas butir soal, reliabilitas instrumen, tingkat kesukaran, dan daya pembeda soal. Hasil analisis menunjukkan bahwa koefisien reliabilitas tes sebesar 0,84 (kategori tinggi). Secara akumulatif, dari 20 butir soal yang dianalisis, terdapat 14 butir soal (70%) memenuhi kriteria kualitas baik sehingga langsung diterima dalam bank soal, 4 butir soal (20%) memerlukan revisi pada konstruksi kalimat atau fungsi distraktor, dan 2 butir soal (10%) dibuang karena memiliki daya pembeda yang tidak signifikan. Evaluasi berkala terhadap instrumen penilaian empiris berbasis spreadsheet ini terbukti efektif membantu guru dalam memetakan kelayakan alat ukur hasil belajar siswa.

Kata Kunci: Analisis Butir Soal; Teori Tes Klasik; Statistika; Validitas; Reliabilitas.

1. PENDAHULUAN

Evaluasi hasil belajar merupakan salah satu komponen krusial dalam sistem pembelajaran matematika di tingkat Sekolah Menengah Atas (SMA). Melalui proses evaluasi yang tepat, pendidik dapat memperoleh data yang akurat mengenai pencapaian kompetensi siswa serta efektivitas strategi mengajar yang telah diterapkan. Salah satu materi kompetensi esensial dalam kurikulum matematika SMA adalah statistika, yang mencakup penyajian data, ukuran pemusatan data (mean, median, modus), dan ukuran penyebaran data (simpangan baku, varians). Karakteristik materi statistika menuntut siswa tidak hanya terampil dalam berhitung secara prosedural, melainkan juga harus memiliki kemampuan penalaran logis serta interpretasi data yang kuat.

Mengingat pentingnya kedudukan materi ini, instrumen evaluasi yang digunakan untuk mengukur kemampuan siswa harus memiliki derajat validitas dan reliabilitas yang tinggi. Dalam praktik pembelajaran sehari-hari di sekolah, instrumen tes objektif seperti pilihan ganda (*multiple choice*) sering menjadi pilihan utama guru karena objektivitas penilaiannya yang tinggi serta efisiensi waktu dalam proses koreksi hasil. Namun, permasalahan yang kerap muncul di lapangan menunjukkan bahwa sebagian besar instrumen tes tersebut disusun dan diujikan langsung kepada siswa tanpa melalui tahapan uji coba dan analisis karakteristik butir soal secara empiris. Akibatnya, tingkat kedalaman materi dan akurasi pengukuran instrumen tersebut sering kali dipertanyakan.

Untuk mengatasi permasalahan tersebut, pendekatan Teori Tes Klasik (*Classical Test Theory*) menawarkan solusi metodologis yang sangat praktis dan teruji untuk menganalisis parameter empiris dari sebuah tes. Melalui Teori Tes Klasik, kualitas sebuah alat ukur dapat dinilai secara komprehensif melalui indikator utama, yaitu validitas butir soal, reliabilitas tes, tingkat kesukaran (*difficulty index*), dan daya pembeda (*discrimination index*). Analisis parameter ini tidak selalu membutuhkan perangkat lunak psikometri yang rumit, pemanfaatan program *spreadsheet* seperti Microsoft Excel dengan formula statistik standarnya sudah sangat memadai bagi guru untuk melakukan pengolahan data nilai secara mandiri dan akurat.

Oleh karena itu, penelitian ini bertujuan untuk menganalisis kualitas butir soal pilihan ganda pada materi statistika SMA berdasarkan pendekatan Teori Tes Klasik, yang berfokus pada parameter tingkat kesukaran, daya pembeda, validitas, dan reliabilitas soal. Melalui penelitian ini, diharapkan dapat diperoleh gambaran empiris mengenai kualitas instrumen penilaian sehingga dapat menjadi acuan dalam penyusunan evaluasi pembelajaran yang lebih berkualitas dan objektif di masa mendatang.

2. METODE PELAKSANAAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif. Penelitian deskriptif kuantitatif dalam konteks ini bertujuan untuk memberikan

gambaran secara sistematis, faktual, dan akurat mengenai karakteristik setiap butir soal matematika tanpa melakukan manipulasi terhadap subjek penelitian.

Subjek uji coba dalam penelitian ini adalah 25 siswa kelas X SMA PGRI Mnelalete. Jumlah subjek disesuaikan dengan ukuran satu rombongan belajar dan secara umum dinilai representatif untuk keperluan analisis butir soal tahap awal dalam teori tes klasik, di mana ukuran sampel kecil (kisaran 25–30 responden) sudah memadai untuk mengidentifikasi butir soal yang cacat, valid, atau memerlukan revisi sebelum diujikan pada kelompok sampel yang lebih besar. Validasi instrumen dilakukan melalui expert judgment secara mandiri oleh penulis selaku dosen pengampu mata kuliah kajian matematika SMA yang memiliki keahlian di bidang pendidikan matematika. Telaah difokuskan pada kesesuaian butir soal dengan indikator pencapaian kompetensi materi statistika SMA, tingkat kognitif serta kebenaran kunci jawaban dan tata bahasa yang benar agar tidak membingungkan siswa. Objek penelitian yang dianalisis berupa lembar jawaban siswa dari instrumen tes pilihan ganda materi statistika yang terdiri atas 20 butir soal dengan 5 pilihan jawaban (A, B, C, D, E). Skor jawaban diatur dengan sistem biner, di mana jawaban benar diberi skor 1 dan jawaban salah diberi skor 0. Proses pengolahan data dilakukan sepenuhnya menggunakan bantuan Microsoft Excel. Langkah-langkah analisis kuantitatif dengan pendekatan Teori Tes Klasik dijabarkan sebagai berikut:

Uji Validitas Butir Soal

Validitas butir soal dihitung dengan menggunakan rumus korelasi *Point-Biserial* (r_{pbi}) karena data skor berupa data dikotomi (0 dan 1) dengan skor total kontinu. Rumus korelasi *Point-Biserial* adalah:

$$r_{pbi} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

Keterangan:

r_{pbi} : koefisien validitas butir soal;

M_p : rata – rata skor total siswa yang menjawab benar;

M_t : rata – rata skor total seluruh peserta;

S_t : simpangan baku skor total;

p : proporsi peserta yang menjawab benar;

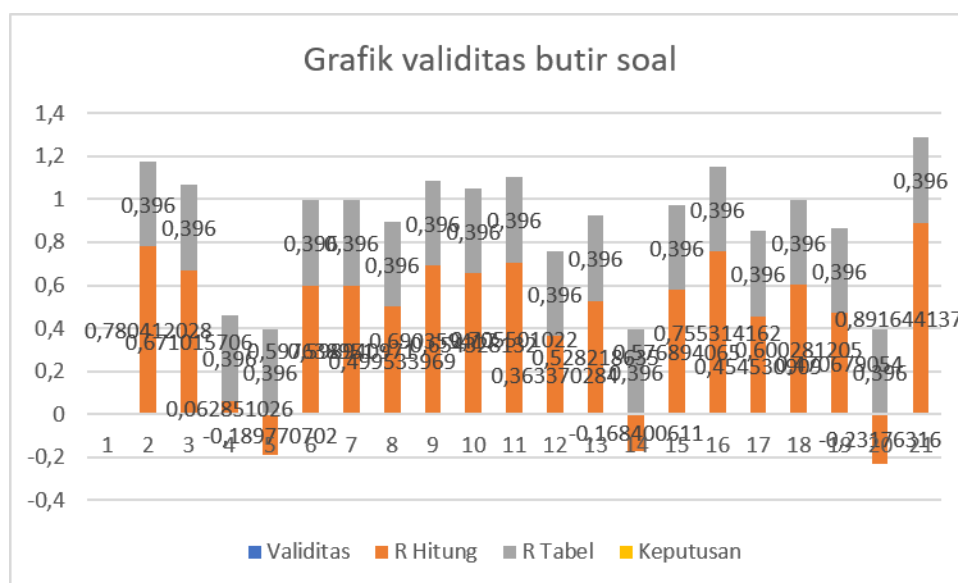
q : proporsi peserta yang menjawab salah ($q = 1 - p$).

Kriteria pengambilan keputusan:

Jika $r_{hitung} \geq r_{tabel}$, maka butir soal dinyatakan valid

Jika $r_{hitung} < r_{tabel}$, maka butir soal dinyatakan tidak valid

Berikut adalah grafik validitas butir soal pilihan ganda materi statistika.



Gambar 1. Grafik validitas butir soal

Uji Reliabilitas Instrumen

Reliabilitas instrumen ditentukan menggunakan rumus Kuder-Richardson 20 (KR-20) karena instrumen penelitian berupa tes dengan skor dikotomi (0 = salah dan 1 = benar). Perhitungan reliabilitas dilakukan menggunakan Microsoft Excel dengan rumus sebagai berikut:

$$r_{11} = \frac{k}{k-1} \left(1 - \frac{\sum pq}{S_t^2} \right)$$

Keterangan:

r_{11} : koefisien reliabilitas instrumen (KR-20);

k : jumlah butir soal;

p : proporsi peserta yang menjawab benar pada setiap butir;

q : proporsi peserta yang menjawab salah pada setiap butir ($q = 1 - p$);

$\sum pq$: jumlah hasil perkalian p dan q dari seluruh butir soal;

S_t^2 : varians skor total peserta.

Kriteria pengambilan keputusan:

Interpretasi koefisien reliabilitas dapat menggunakan kriteria pada tabel berikut:

Tabel 1. Interpretasi koefisien reliabilitas

Nilai	
Koefisien	Interpretasi
0,80 - 1,00	Sangat tinggi
0,60 - 0,79	Tinggi
0,40 - 0,59	Cukup
0,20 - 0,39	Rendah
0,00 - 0,19	Sangat rendah

Uji Tingkat Kesukaran (TK)

Tingkat kesukaran butir soal ditentukan menggunakan indeks kesukaran (P), yaitu perbandingan antara jumlah peserta didik yang menjawab benar suatu butir soal dengan jumlah seluruh peserta yang mengikuti tes. Perhitungan dilakukan menggunakan Microsoft Excel dengan rumus:

$$P = \frac{B}{N}$$

P : indeks kesukaran butir soal;

B : Jumlah siswa yang menjawab benar pada suatu butir soal;

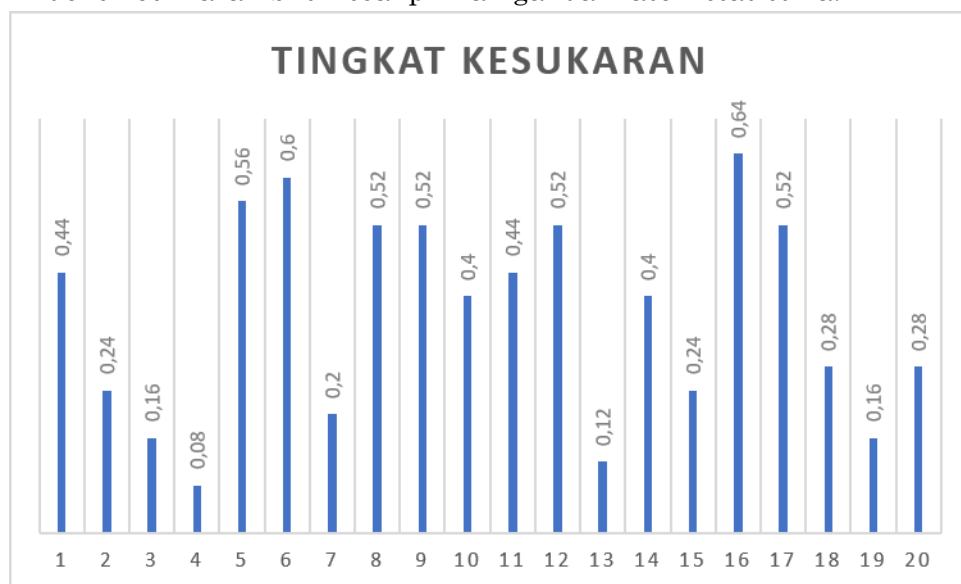
N : jumlah seluruh peserta yang mengikuti tes.

Kriteria tingkat kesukaran sebagai berikut:

Tabel 2. Kriteria Tingkat Kesukaran

Nilai Indeks Kesukaran (P)	Kategori
$0,00 \leq P \leq 0,30$	Sukar
$0,31 \leq P \leq 0,70$	Sedang
$0,71 \leq P \leq 1,00$	Mudah

Berikut grafik indeks kesukaran butir soal pilihan ganda materi statistika.



Gambar 2. Grafik tingkat kesukaran butir soal

Uji Daya Pembeda (DP)

Daya pembeda butir soal digunakan untuk mengetahui kemampuan suatu butir soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan kemampuan rendah. Perhitungan dilakukan dengan membagi peserta didik ke dalam dua kelompok, yaitu **kelompok atas** (27% peserta dengan skor total tertinggi) dan **kelompok bawah** (27% peserta dengan skor total terendah). Indeks daya pembeda (D) dihitung menggunakan rumus:

$$D = P_A - P_B$$

Keterangan:

$P_A = \frac{B_A}{J}$ = proporsi jawaban benar kelompok atas;

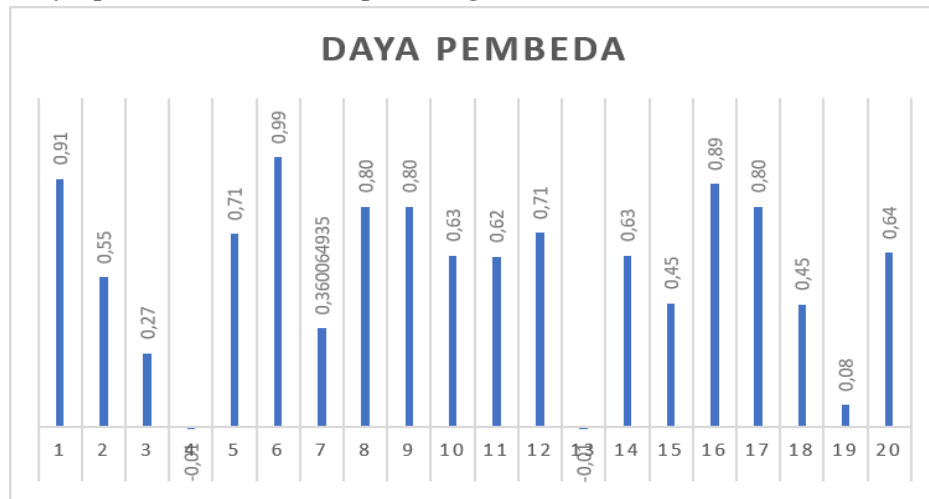
$P_B = \frac{B_B}{J}$ = proporsi jawaban benar kelompok bawah.

Kriteria daya pembeda sebagai berikut:

Tabel 3. Kriteria Daya Pembeda

Nilai D	Kategori
$D < 0,20$	Jelek
$0,20 \leq D < 0,40$	Cukup
$0,40 \leq D < 0,70$	Baik
$0,40 \leq D \leq 1,00$	Sangat baik
$D < 0$	Sangat jelek (dibuang atau direvisi)

Berikut grafik daya pembeda butir soal pilihan ganda materi statistika.



Gambr 3. Grafik daya pembeda butir soal

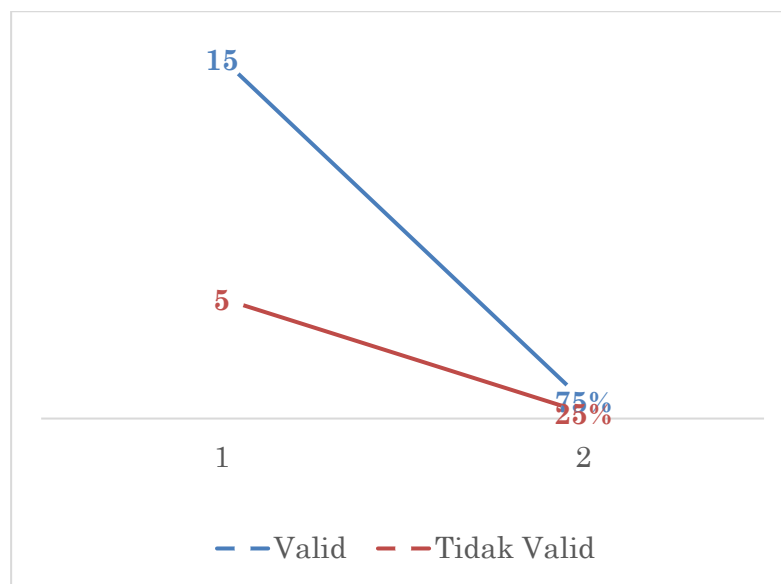
3. HASIL DAN PEMBAHASAN

3.1 Hasil

Berdasarkan hasil analisis data lembar jawaban dari 25 siswa pada 20 butir soal statistika menggunakan Microsoft Excel, diperoleh rekapitulasi data empiris parameter Teori Tes Klasik sebagai berikut.

3.1.1 Hasil uji Validitas

Hasil analisis validitas butir soal disajikan dalam diagram dibawah ini:



Gambar 4. Hasil uji validitas butir soal

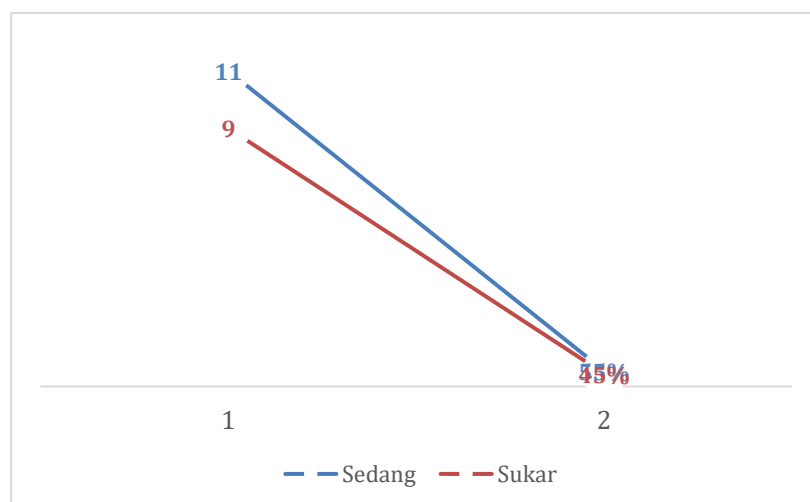
Terdapat 15 butir soal (75%) yang valid dan 5 butir soal (25%) tidak valid karena nilai r_{hitung} lebih kecil dari r_{tabel} 0,396 (yaitu soal nomor 3, 4, 11, 13 dan 19).

3.1.2 Hasil uji Reliabilitas

Pengujian reliabilitas instrumen secara menyeluruh menggunakan rumus KR-20, diperoleh koefisien reliabilitas (r_{11}) sebesar 0,84. Nilai ini berada di atas ambang batas standar kelayakan ($\geq 0,70$), yang berarti instrumen tes statistika ini memiliki konsistensi internal yang tinggi untuk mengukur kemampuan siswa secara stabil.

3.1.3 Hasil uji Tingkat Kesukaran (TK)

Hasil analisis reliabilitas dapat dilihat pada diagram dibawah ini:

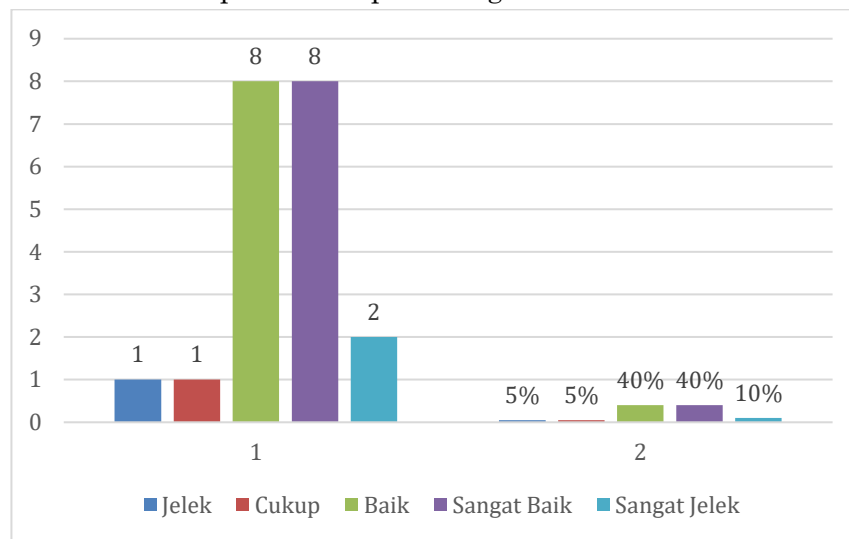


Gambar 5. Hasil uji tingkat kesukaran

Distribusi tingkat kesukaran soal terdiri atas 11 butir soal (55%) kategori Sedang, 9 butir soal (45%) kategori Sukar.

3.1.4 Hasil uji daya pembeda (DP)

Hasil analisis reliabilitas dapat dilihat pada diagram dibawah ini:



Gambar 6. Hasil uji Daya Pembeda

Kualitas daya pembeda soal menunjukkan terdapat 8 butir soal (40%) kategori Sangat Baik, 8 butir soal (40%) kategori Baik, 1 butir soal (5%) kategori Cukup, dan 1 butir soal (5%) kategori Jelek dan 2 butir soal (10%) kategori Sangat Jelek.

Secara keseluruhan, analisis gabungan menghasilkan keputusan akhir berupa 14 butir soal (70%) langsung diterima, 4 butir soal (20%) harus direvisi, dan 2 butir soal (10%) ditolak atau dibuang.

3.2 Pembahasan

Analisis karakteristik butir soal menggunakan Teori Tes Klasik memberikan perspektif penting terkait kelayakan fungsional soal di dalam kelas. Temuan dari data tersebut dibahas dalam poin-poin berikut:

3.2.1 Butir Soal Kategori Diterima

Sebagian besar instrumen, yakni sebanyak 14 butir soal, berada dalam kondisi ideal (valid, tingkat kesukaran sedang, dan daya pembeda di atas 0,30). Karakteristik soal dengan tingkat kesukaran sedang sangat efektif untuk menghindari kejenuhan pada siswa, sekaligus memberikan kontribusi linier yang optimal terhadap nilai reliabilitas keseluruhan tes. Kelompok soal ini sangat layak disimpan langsung ke dalam bank soal.

3.2.2 Butir Soal Kategori Direvisi

Terdapat 4 butir soal (nomor 3, 4, 13, dan 19) yang masuk dalam kategori direvisi. Masalah utama kelompok ini terletak pada indeks daya pembeda yang berada pada rentang sangat jelek dan jelek ($D < 20$ - $D < 0$). Sebagai contoh, pada soal nomor 3 dan 19, tingkat kesukaran tergolong sukar. Soal yang tergolong terlalu sukar menyebabkan sebagian besar siswa tidak mampu menjawab dengan benar sehingga skor tes menjadi rendah dan kemampuan siswa kurang dapat diukur secara optimal. Kondisi ini juga dapat menurunkan motivasi serta kepercayaan diri peserta didik dalam mengerjakan tes.

Selain itu, apabila soal yang terlalu sukar tidak mampu membedakan siswa berkemampuan tinggi dan rendah, maka daya pembeda soal menjadi rendah sehingga kualitas butir soal menurun. Meskipun demikian, soal yang sukar tidak selalu harus dibuang, karena apabila memiliki daya pembeda yang baik dan sesuai dengan tujuan pengukuran, misalnya untuk mengukur kemampuan berpikir tingkat tinggi, soal tersebut masih layak dipertahankan.

3.2.3 Butir Soal Kategori Dibuang

Sebanyak 2 butir soal dinyatakan tidak layak dan harus dikeluarkan dari instrumen (nomor 4 dan 13). Faktor penyebab kegagalan fungsi pengukuran soal ini adalah Soal nomor 4 dan 13 tergolong Sukar dan memiliki daya pembeda yang sangat jelek, bahkan dua soal ini bernilai negatif ($-0,01$). Nilai daya pembeda negatif menunjukkan sebuah anomali psikometrik di mana siswa dari kelompok bawah justru lebih banyak menjawab benar dibandingkan siswa dari kelompok atas. Hal ini biasanya dipicu oleh kunci jawaban yang salah cetak, teks soal yang bias/ambigu sehingga membingungkan siswa pintar, atau materi soal yang belum diajarkan secara tuntas di kelas sehingga siswa menebak secara acak.

4. SIMPULAN

Berdasarkan hasil analisis kuantitatif menggunakan Microsoft Excel dengan pendekatan Teori Tes Klasik pada instrumen pilihan ganda materi statistika SMA, dapat disimpulkan bahwa instrumen ini memiliki kualitas keseluruhan yang memadai dengan koefisien reliabilitas tinggi sebesar 0,84. Analisis individual butir soal menunjukkan tingkat kelayakan yang bervariasi: 70% butir soal dinyatakan berkualitas baik dan siap digunakan kembali (Diterima), 20% butir soal membutuhkan penyempurnaan struktur kalimat atau opsi pilihan (Direvisi), dan 10% butir soal tidak layak dan harus dieliminasi dari instrumen (Ditolak). Penggunaan spreadsheet terbukti mempermudah guru melakukan diagnosis psikometrik sederhana demi menjaga mutu evaluasi pembelajaran secara berkelanjutan.

5. UCAPAN TERIMA KASIH

Terimakasih kepada mahasiswa/i semester VI Angkatan 10 program studi Pendidikan Matematika Institut Pendidikan Soe dan siswa/i kelas X SMA PGRI Mnelalete yang telah berpartisipasi dalam melakukan observasi lapangan mata kuliah Kajian Matematika SMA di SMA PGRI Mnelalete, Kecamatan Amanuban Barat, Kabupaten Timor Tengah Selatan, NTT.

6. REKOMENDASI

Bagi Peneliti selanjutnya disarankan untuk mengembangkan analisis evaluasi kualitas ini tidak hanya sebatas parameter kuantitatif Teori Tes Klasik, melainkan dikombinasikan dengan analisis kualitatif dari para pakar (validitas isi) serta uji coba fungsi efektivitas setiap opsi distraktor secara mendalam.

7. REFERENSI

- Asria, L. (2023). *Analisis Butir Soal Penilaian Tengah Semester (PTS) Matematika Kelas XI Berdasarkan Teori Klasik Item Analysis of Mathematics Mid-Semester Assessment for Class XI Based on Classical Theory*. 4(1), 1–11.
- Farikhah, L. (n.d.). *Analisis butir soal matematika pada instrumen uji coba materi segitiga*. 71–78.
- Jumini, S., Madnasri, S., Cahyono, E., & Parmin, P. (2023). *Analisis Kualitas Butir Soal Pengukuran Literasi Sains Melalui Teori Tes Klasik Dan Rasch Model*. 758–765.
- Leseman, P. M., & Kroesbergen, E. H. (2019). *Enriching Mathematics Education with Visual Arts : Effects on Elementary School Students ` Ability in Geometry and Visual Arts*.
- Matematika, P. P., Indonesia, U. P., No, J. S., Sukasari, K., Bandung, K., & Barat, J. (2024). *Gunung Djati Conference Series , Volume 41 (2024) Conferences Series Learning Class Analisis Tingkat Kesukaran Soal Matematika Pada Kemampuan Penyelesaian Masalah Matematis Siswa Kelas VII Pendidikan Matematika , UIN Sunan Gunung Djati Bandung Jl . Cimencrang Kecamatan Gedebage , Kota Bandung , Jawa Barat 1 Prodi Gunung Djati Conference Series , Volume 41 (2024) Conferences Series Learning Class*. 41, 96–106.
- Mukhlisa, N., Hasan, K., & Zainal, Z. (n.d.). *PKM Pelatihan Analisis Butir Soal dengan Program ITEMAN (Item and Test Analysis) dan Microsoft Excel di UPTD SD Negeri 35 Parepare*.
- Nurhalimah, S., Hidayati, Y., Rosidi, I., Madura, U. T., & Soal, K. (2022). *HUBUNGAN ANTARA VALIDITAS ITEM DENGAN DAYA PEMBEDA DAN*. 4(3), 249–257.
- Pelajaran, M., Sma, K., & Kabupaten, D. I. (2015). *No Title*. 3(1), 35–43.
- Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). *Validitas and Reliabilitas*. 06(02), 10967–10975.
- Santoso, A., Kartianom, K., & Kassymova, G. K. (2020). *Kualitas butir bank soal statistika (Studi kasus : Instrumen ujian akhir mata kuliah statistika Universitas Terbuka)*. 6(2), 165–176.
- Suharsimi Arikunto. (2019). *Dasar - Dasar Evaluasi Pendidikan (Edisi 3)* (3rd ed.). Bumi Aksara.