

ANALISIS KUALITAS BUTIR SOAL PENILAIAN AKHIR SEMESTER GASAL TAHUN 2021/2022 MENGGUNAKAN ITEMAN DI SMP NEGERI 1 KAMBERA

Damianus Nd. Marambaawang¹, Vidriana Oktoviana Bano², Riwa Rambu Hada Enda³
e-mail: umbundamungri@gmail.com, vidri.bano@unkriswina.ac.id, riwa@unkriswina.ac.id
^{1,2,3}Program Studi Pendidikan Biologi, FKIP, Universitas Kristen Wira Wacana Sumba,
Nusa Tenggara Timur, Indonesia

ABSTRAK

Pada data hasil ujian akhir semester gasal tahun pelajaran 2021/2022, terdapat banyak siswa SMPN 1 Kambera yang mendapatkan skor di bawah standar KKM. Mengacu pada hal itu penelitian ini bertujuan untuk mengevaluasi karakteristik setiap item soal yang dipakai dalam ujian akhir semester tersebut dengan menggunakan beberapa kriteria yang telah ditentukan. Penelitian ini menerapkan metode deskriptif kuantitatif. Sampel yang diambil adalah kelas VIIID. Top of Form Hasil analisis menggunakan ITEMAN menunjukkan bahwa 25% item tidak valid dan 75% item valid. Selain itu, reliabilitasnya mencapai tingkat yang tinggi yaitu "0,886". Berdasarkan analisis tingkat kesulitan, 17% soal dikategorikan sebagai sangat mudah, 14% dikategorikan sebagai mudah, dan 15% dikategorikan sebagai sedang. Terdapat juga 17% soal yang dinilai sangat sulit, dan 37% soal yang dinilai sulit. Berdasarkan level kognitif, terdapat 20 butir soal yang masuk dalam tingkat kesulitan tinggi (HOTS) dan 15 butir soal yang masuk dalam tingkat kesulitan rendah (LOWS). Dari segi daya pembeda, 88% butir soal memiliki daya pembeda yang baik, 3% item memiliki daya pembeda yang kurang baik, dan 9% item tidak memiliki daya pembeda yang baik. Dalam penelitian mengenai efektivitas distraktor, ditemukan 54% pilihan distraktor yang efektif, dan 46% pilihan distraktor tidak efektif.

Kata kunci: Analisis, Butir soal, Evaluasi, ITEMAN

ABSTRACT

In the data from the odd semester final exam results for the 2021/2022 school year, there are many students at SMPN 1 Kambera who score below the KKM standard. Referring to this, this study aims to evaluate the characteristics of each item of questions used in the final exam of the semester using several predetermined criteria. This study applies a quantitative descriptive method. The sample taken is class VIIID. The results of the analysis using ITEMAN show that 25% of the items are invalid and 75% of the items are valid. In addition, its reliability reaches a high level of "0.886". Based on the level of difficulty analysis, 17% of the questions were categorized as very easy, 14% were categorized as easy, and 15% were categorized as moderate. There were also 17% questions that were considered very difficult, and 37% questions that were considered difficult. Based on the cognitive level, there are 20 items that are included in the high difficulty level (HOTS) and 15 items that are included in the low difficulty level (LOWS). In terms of discriminating power, 88% of the items had good discriminating power, 3% of the items had poor discriminating power, and 9% of the items did not have good discriminating power. In research on distractor effectiveness, 54% of distractor choices were found to be effective, and 46% of distractor choices were ineffective.

Keywords: Analysis, Items, Evaluation, ITEMAN

PENDAHULUAN

Di era pendidikan abad ke-21 ini, guru dituntut untuk mengimplementasikan keterampilan berpikir tingkat tinggi atau yang dikenal juga sebagai Higher Order Thinking Skills (HOTS). Penerapan keterampilan berpikir tingkat tinggi bertujuan untuk merangsang kemampuan penalaran siswa di luar sekadar mengingat dan mengulang kembali informasi. Kemampuan ini lebih menekankan pada kemampuan menganalisis, membuat keputusan yang tepat, dan memecahkan suatu masalah (Asfiah, 2021).

Untuk menanggapi tatanan ini maka dibutuhkan kreativitas guru pada satuan pendidikan untuk menyusun instrumen yang baik dan efektif dalam mengevaluasi hasil belajar siswa. Soal-soal yang efektif dalam evaluasi adalah soal-soal yang dapat secara akurat menggambarkan sejauh mana tingkat pemahaman dan penguasaan siswa terhadap materi yang telah diajarkan. Dengan kata lain, soal tersebut harus mampu mengungkapkan tingkat pemahaman siswa secara tepat. Tes yang bermutu mempunyai persyaratan validitas, reliabilitas, objektivitas, kepraktisan, serta ekonomis (Muluki, 2020).

Merujuk pada kualitas instrumen yang digunakan pada evaluasi, mengkaji kembali tentang karakteristik instrumen yang dipakai dalam evaluasi perlu dilakukan agar bisa memahami karakteristik butir soal yang ada. (Fauziana & Wulansari, 2021) Mengungkapkan bahwa melakukan analisis butir soal memiliki potensi untuk meningkatkan kualitas butir soal melalui revisi atau penghapusan soal yang tidak efektif. Selain itu, analisis tersebut dapat memberikan informasi diagnostik tentang pemahaman siswa terhadap materi yang telah diajarkan. Berdasarkan hal tersebut, pelaksanaan analisis item soal dapat menjadi panduan untuk melakukan perbaikan terhadap soal yang digunakan. Soal-soal yang memiliki kualitas baik dapat dimasukkan ke dalam bank soal dan dapat digunakan kapan saja sebagai instrumen evaluasi untuk mengukur hasil belajar.

Saat ini berkat perkembangan teknologi datang bermacam program PC yang dapat dipakai guna menjalankan analisis butir soal. Salah satunya adalah ITEMAN. Berkah (2019) menerangkan bahwa program *Item and Test Analysis* (ITEMAN) ialah perangkat lunak (aplikasi) yang dibuat dengan bahasa program pc serta dicetuskan khusus guna analisis statistik butir soal. Program ini dibuat dengan pendekatan kajian statistik butir soal yang berfungsi guna memastikan mutu butir soal yang bersumber pada data empiris. Program iteman ini menjalankan analisis soal dengan cara klasik yaitu memakai rumus-rumus baku yang telah lumrah dipakai dalam menjalankan analisis soal. Hasil analisis dari program ITEMAN ini melingkupi tingkatan kesulitan, daya beda, serta statistik penyebaran respons, reliabilitas tes, kekeliruan pengukuran, serta distribusi skor (Hartini, 2021).

Dalam kutipan Huda & Wahyuni (2020) terdapat beberapa kegunaan dari program ITEMAN yaitu: “(1) Menganalisis data file jawaban butir soal yang dibuat melalui *manual entry* data atau mesin *scanner* dalam format ASCII. (2) Menskor dan menganalisis data soal pilihan ganda dan skala Likert untuk 30.000 siswa dan 250 butir soal. (3) Menganalisis sebuah tes yang terdiri dari 10 skala atau subtes. Program ini memberikan informasi kuantitatif tentang setiap butir, seperti daya pembeda, tingkat kesukaran, dan efektivitas distraktor. Selain itu, program ini juga dapat menghitung reliabilitas tes menggunakan metode KR-20/Alpha, *split-half*, dan *Spearman-Brown* (S-B). Program ITEMAN juga dapat menghitung standar *error of measurement*, *mean*, *variance*, standar *deviasi*, *skew*, *kurtosis* untuk jumlah skor pada jawaban benar, skor minimum dan maksimum, skor median, dan frekuensi distribusi skor”.

Berdasarkan data hasil ujian akhir semester gasal di kelas VIII tahun ajaran 2021/2022 di SMP Negeri 1 Kambera, ditemukan bahwa sekitar 65% siswa belum mencapai nilai standar KKM (Kriteria Ketuntasan Minimal). Temuan ini mengindikasikan bahwa terdapat banyak siswa yang mendapatkan skor di bawah standar KKM. Standar KKM yang ditetapkan di SMPN 1 Kambera adalah 73. Menurut data yang dihasilkan dari guru mata pelajaran menerangkan bahwa rendahnya hasil ujian siswa ini dipengaruhi lantaran masih rendahnya kemahiran siswa dalam menyelesaikan soal yang disajikan. Mengacu pada hal itu ini bakal jadi PR besar bagi guru untuk selalu meningkatkan kompetensi siswa tegasnya. Selepas dari itu butir soal yang dipakai itu belum dilakukan kajian terkait kualitasnya.

Berdasarkan hasil kajian tersebut di atas, peneliti memandang perlu untuk melakukan penelitian terkait kualitas butir soal ujian akhir semester gasal tahun ajaran 2021/2022 tersebut. Pelaksanaan kajian butir soal ini bermaksud untuk mendapati taraf validitas dan reliabilitas, taraf kesukaran serta sebaran dimensi proses kognitif, tingkat daya pembeda dan keberhasilan opsi pengecoh. Dalam penelitian ini, peneliti akan menggunakan ITEMAN untuk menganalisis kualitas item soal.

Dalam segi teori, hasil penelitian ini dapat digunakan sebagai sarana untuk meningkatkan kualitas evaluasi tes pilihan ganda guna mencapai standar evaluasi pendidikan yang optimal. Dan juga dapat bermanfaat secara praktis bagi guru pada tingkat pendidikan sebagai bahan masukan dalam penyusunan alat evaluasi yang lebih efektif. Data hasil penelitian ini dapat membantu mengumpulkan informasi yang spesifik tentang kualitas butir soal dan dapat dijadikan bahan refleksi bagi guru dalam mengembangkan proses pembelajaran selanjutnya

METODOLOGI PENELITIAN

Desain penelitian yang digunakan dalam studi ini adalah pendekatan kuantitatif deskriptif. Pemilihan desain kuantitatif dilakukan karena data yang dikumpulkan akan dianalisis dan dideskripsikan dalam bentuk angka. Peneliti akan menggambarkan dan menganalisis data terkait mutu item soal pada ujian akhir semester gasal mata pelajaran IPA tahun 2021/2022 di SMPN 1 Kambera.

Penelitian ini dilakukan pada bulan Desember tahun 2022 di SMPN 1 Kambera yang terletak di Jln. S. Parman, Kelurahan Lambanapu, Kecamatan Kambera, Kabupaten Sumba Timur, Provinsi Nusa Tenggara Timur. Populasi yang menjadi fokus penelitian ini adalah siswa kelas VIII pada penilaian akhir semester gasal di SMPN 1 Kambera, pada tahun ajaran 2021/2022 dengan jumlah total 155 siswa. Namun, untuk penelitian ini, sampel yang digunakan adalah kelas VIII D. Variabel yang diteliti dalam penelitian ini adalah kualitas butir soal mata pelajaran IPA yang digunakan pada ujian akhir semester gasal di SMPN 1 Kambera tahun 2021/2022. Bentuk soal tersebut adalah soal pilihan ganda yang disusun oleh guru IPA di SMPN 1 Kambera, dengan total jumlah soal sebanyak 35 nomor. Peneliti akan memberikan deskripsi mengenai kualitas butir soal berdasarkan kriteria-kriteria tertentu, seperti tingkat validitas, reliabilitas, tingkat kesukaran, dimensi kognitif, daya pembeda, dan keberfungsian distractor.

Metode pengumpulan data yang digunakan dalam penelitian ini menggunakan metodologi analisis dokumen. Peneliti mengumpulkan data penelitian untuk selanjutnya dilakukan analisis. Data yang digunakan terdiri dari kisi-kisi soal mata pelajaran IPA dan lembar soal dari guru, serta lembar jawaban siswa sebagai sumber data. Data tersebut digunakan oleh guru kelas VIII SMPN 1 Kambera pada ujian akhir semester gasal tahun pelajaran 2021/2022.

Untuk teknik analisis data pada analisis kualitas dimensi proses kognitif, peneliti menganalisis setiap item soal yang digunakan apakah soal itu berbasis HOTS atau LOWS. Langkah selanjutnya peneliti melakukan validitas, reliabilitas, skalabilitas, kesukaran, daya beda, dan efektivitas analisis pengecoh menggunakan ITEMAN. Interpretasi data dari hasil analisis ITEMAN selanjutnya dilakukan sesuai dengan kriteria dan asumsi yang telah ditentukan.

Beberapa kriteria yang ditentukan yaitu sebagai berikut:

1. Tingkat validitas

Tabel 1. Rumus dan kriteria validitas butir soal

Rumus	Harga/nilai koefisien korelasi (r)
$r_{xy} = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{(n\sum X^2 - (\sum X)^2)} \sqrt{(n\sum Y^2 - (\sum Y)^2)}}$	$0,8 \leq r \leq 1$ = sangat tinggi
	$0,6 \leq r \leq 0,79$ = tinggi
	$0,4 \leq r \leq 0,59$ = cukup
	$0,2 \leq r \leq 0,39$ = rendah
	$0,0 \leq r \leq 0,19$ = rendah sekali

Arikunto (dalam Taufik, Berlian, Fathia, & Rasyidin 2022:143)

2. Tingkat reliabilitas/konsistensi butir soal

Tabel 2. Rumus dan kriteria reliabilitas butir soal

Rumus	Nilai tabel koefisien korelasi
$r_{11} = \frac{n}{n-1} \left(1 - \frac{\sum x(n-x)}{n(SD)^2} \right)$	$0,00 \leq R_{11} < 0,20$ = kategori sangat rendah
	$0,21 \leq R_{11} < 0,40$ = kategori rendah
	$0,41 \leq R_{11} < 0,60$ = kategori sedang atau cukup
	$0,61 \leq R_{11} < 0,80$ = kategori tinggi
	$0,81 \leq R_{11} < 1,00$ = kategori sangat tinggi
Rostina (dalam Chasanah, Widodo & Suprpto, 2022:470)	

3. Analisis tingkat kesukaran

Tabel 3. Kriteria kesukaran butir soal

Rumus	Kriteria tingkat kesukaran
$p = \frac{\sum b}{N}$	00 - 0,20 = kategori sangat sukar, dibuang
	0,2 - 0,30 = kategori sukar, kurang baik, direvisi
	0,31 - 70 = kategori sedang, dan cukup baik
	0,71 - 90 = kategori mudah, kurang baik, direvisi
	0,91 - 1,00 = kategori sangat mudah, butir tes dibuang
(Widoyoko, 2018:177)	

4. Analisis daya beda

Tabel 4. Menunjukkan rumus dan kriteria daya pembeda butir soal

Rumus	Kriteria besarnya koefisien daya beda
$DP = \frac{B_a - B_b}{\frac{1}{2} N}$	0,51 - 1,00 = kategori sangat baik, dapat digunakan
	0,41 - 0,50 = kategori baik, dapat digunakan
	0,31 - 0,40 = kategori cukup baik, direvisi
	0,21 - 0,30 = kategori kurang baik, perlu pembahasan dan revisi
	0,00 - 0,20 = kategori tidak baik, dibuang
(Widoyoko, 2018:181)	

5. Analisis efektivitas pengecoh

Pada analisis Efektivitas Pengecoh menggunakan kriteria menurut (Arbiatin & Mulabbiah, 2020): 5% Diterima; <5% Direvisi; 0% Ditolak.

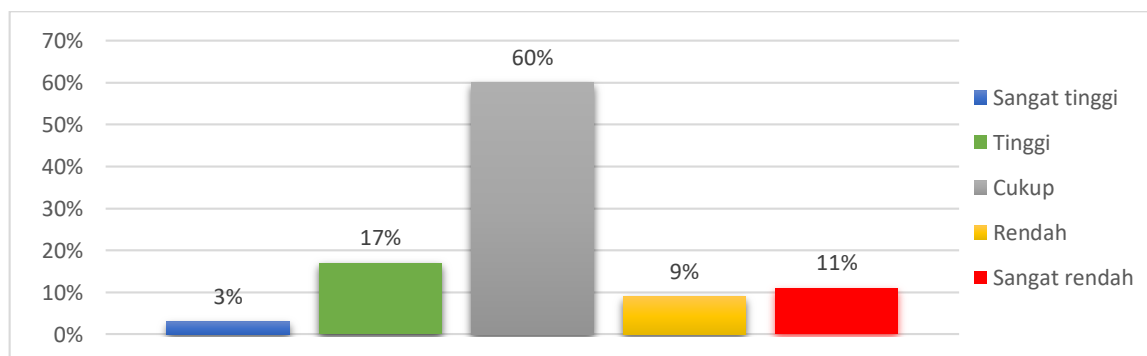
HASIL DAN PEMBAHASAN

Hasil Penelitian

Berikut adalah ringkasan hasil analisis butir soal menggunakan program Iteman, yang ditampilkan dalam bentuk tabel:

1. Hasil analisis validitas

Tabel 6. Persentase tingkat validitas soal



Dalam analisis validitas butir soal, terdapat karakteristik validitas yang mencakup kategori sangat tinggi, kategori tinggi, kategori sedang, kategori rendah, dan kategori sangat rendah. (Taufik et al., 2022).

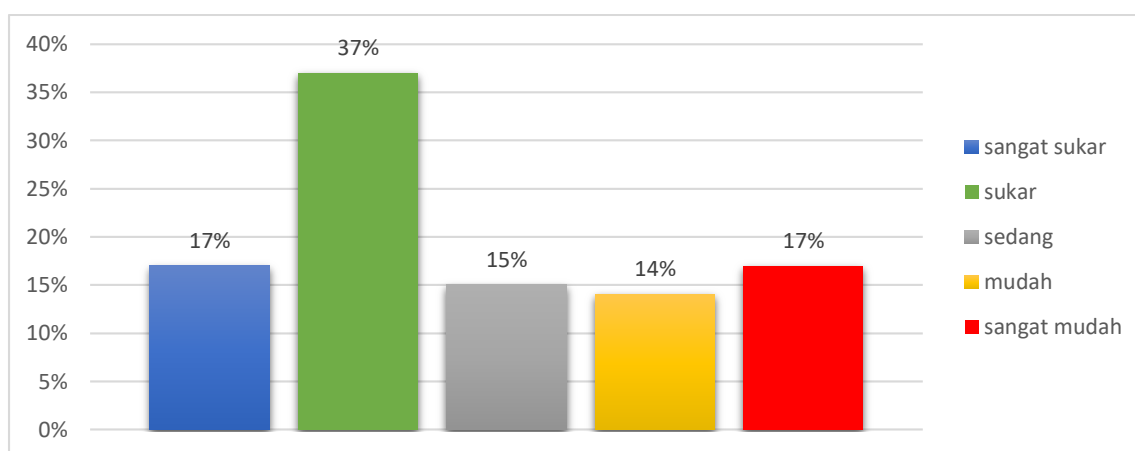
2. Hasil analisis reliabilitas

Tabel 7. Tingkat reliabilitas

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Score items	0.886	2.114	0.826	0.726	0.786	0.905	0.841	0.880

3. Hasil analisis tingkat kesukaran

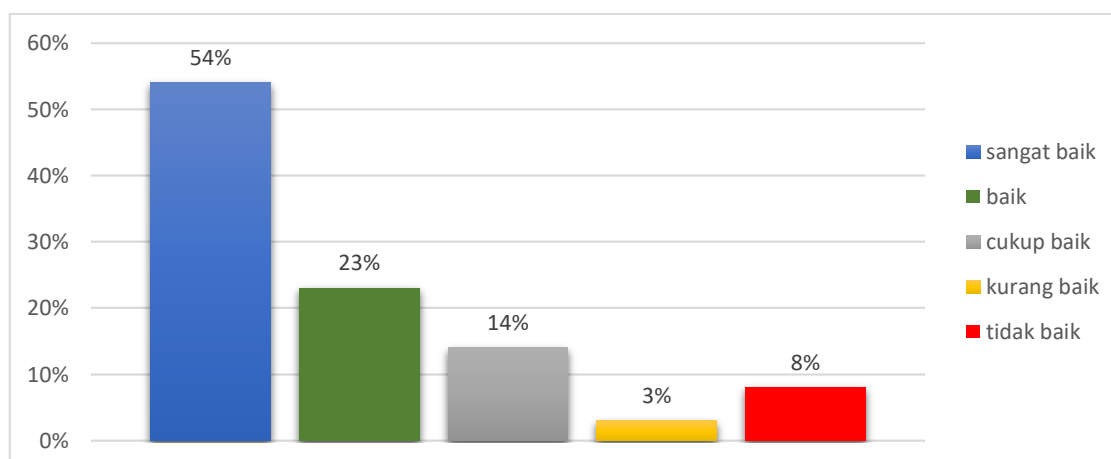
Tabel 8. Persentase tingkat kesukaran



Karakteristik tingkat kesulitan butir soal pada analisis tingkat kesulitan meliputi item kategori sulit dan sangat sulit, sedang, kategori mudah, dan sangat mudah, (Widoyoko, 2018:177).

4. Hasil analisis daya beda

Tabel 9. Persentase daya beda

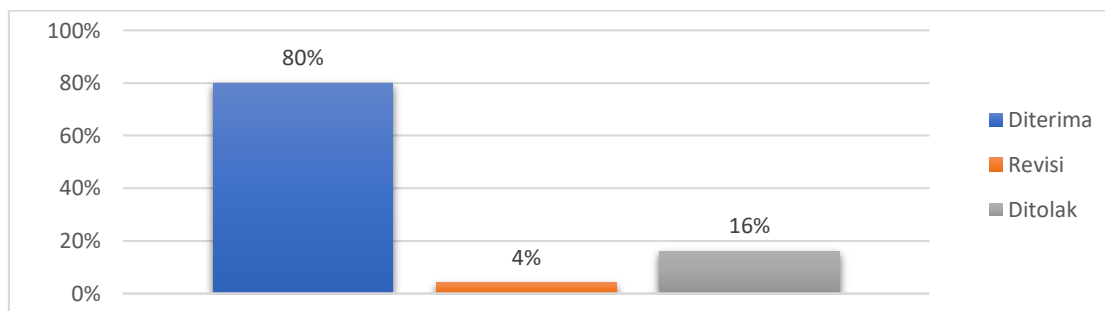


Untuk analisis daya pembeda butir soal, karakteristik daya pembeda berupa sangat baik, baik, cukup, kurang baik dan tidak baik atau buruk (Widoyoko, 2018)

5. Analisis Efektivitas Pengecoh

Analisis efektivitas pengecoh butir soal dilakukan dengan menggunakan karakteristik yang terdiri dari kategori "diterima" (sangat baik), "direvisi", dan "ditolak" (Arbiatin & Mulabbiyah, 2020).

Tabel 10. Persentase efektivitas pengecoh



Tabel 11. Opsi pengecoh yang tidak berfungsi dengan baik

Nomor soal	Keterangan opsi	
	Direvisi	Ditolak
1	B	D
4	B	C,D
5		A,B,C
13	D	
16	B	D
17		D
19		B,C,D
20	B,C	
26		C
28		B,C
29		B,D
30		B
32		B,D
34		D
35		B,D

PEMBAHASAN

1. Hasil analisis validitas

Pelaksanaan analisis validitas bertujuan untuk menentukan tingkat ketepatan dan ketelitian suatu instrumen (tes) dalam menjalankan fungsi ukurnya (Sabara Yusuf, Ahmad, Nusantara & Lamangajto, 2022). Dalam kajian (Purniasari, Masykuri, Sri & Airini, 2021:207) "Untuk menentukan tingkat validitas setiap soal, dapat menggunakan teknik perbandingan nilai korelasi point biserial yang dihasilkan dengan nilai r tabel pada taraf signifikansi 1% dan derajat kebebasan ($N-2$). Jika nilai korelasi point biserial (y_{pbis}) lebih tinggi daripada nilai r tabel, maka soal tersebut dapat dikategorikan sebagai valid".

Pada hasil analisis iteman nilai korelasi point biserial yang kemudian dikonsultasikan dengan kriteria validitas menurut (Taufik et al., 2022) pada tabel 1 diperoleh data tingkat validitas pada tabel 6 yaitu sebanyak 3% butir soal kategori sangat tinggi pada nomor (22). 17% butir soal kategori tinggi yaitu terdapat pada nomor (3, 14, 26, 27, 30, 31). 20% kategori cukup terdapat pada nomor (7, 12, 15, 21, 23, 33, 34). 46% butir soal kategori rendah yaitu terdapat pada nomor (1, 8, 9, 10, 11, 13, 16, 17, 18, 20, 24, 25, 28, 29, 30, 25). Dan 11% butir soal kategori rendah sekali yaitu terdapat pada nomor (2, 4, 5, 19).

Berdasarkan tingkat validitas data di atas butir soal yang terdapat pada nomor (3, 7, 12, 14, 15, 21, 22, 23, 26, 27, 30, 31, 33, 34) sudah cukup baik dan valid dalam evaluasi hal tersebut karena memiliki nilai tingkat validitas yang tinggi. Sementara butir soal pada nomor (1, 2, 4, 5, 6, 8, 9, 10, 11, 13, 16, 17, 18, 19, 20, 24, 25, 28, 29, 32, 35) kurang baik dan tidak valid karena memiliki tingkat validitas yang rendah dalam evaluasi. Butir soal yang valid merupakan butir soal yang baik sebab memiliki tingkat kecermatan yang tinggi dalam menjalankan fungsi ukurnya dan dapat disimpan pada bank soal. Sementara butir soal yang tidak memenuhi tingkat valid dapat dibuang atau perlu direvisi kembali untuk disimpan pada bank soal (Gusti, 2021).

2. Hasil analisis reliabilitas

Tujuan dari analisis reliabilitas adalah untuk mengevaluasi sejauh mana hasil pengukuran dapat diandalkan atau dipercaya. Menurut (Chasanah et al., 2022) Suatu hasil pengukuran yang dapat dipercaya terjadi ketika pengukuran dilakukan secara berulang terhadap kelompok subjek yang sama menghasilkan hasil pengukuran yang relatif konsisten selama aspek yang diukur pada subjek tersebut tidak mengalami perubahan. Dalam tulisan (Ikawati, Jayadi, & Hermians, 2022) metode yang dapat digunakan untuk mengukur tingkat keandalan pengujian (reliabilitas tes) adalah dengan nilai cronbach alpha pada output iteman. Biasanya, nilai reliabilitas 0,70 atau lebih menunjukkan bahwa hasil pengukuran dianggap reliabel. Menurut (Rizalia, Syam, & Yulita, 2021) rentang indeks reliabilitas adalah dari 0 hingga 1, di mana semakin tinggi koefisien reliabilitas suatu tes (mendekati 1), maka semakin tinggi pula tingkat keandalan atau ketepatan tes tersebut.

Pada hasil analisis program Iteman tabel 7, memiliki nilai reliabilitas yang tinggi yaitu nilai Alpha “0.886” yang artinya rata-rata soal yang digunakan reliabel. Hal ini menunjukkan bahwa berdasarkan perhitungan reliabilitas rata-rata soal yang digunakan dalam evaluasi memiliki konsistensi yang tinggi dalam melakukan pengukuran.

3. Hasil analisis tingkat kesukaran

Pelaksanaan analisis kesulitan item bertujuan untuk menentukan apakah pertanyaan yang diajukan termasuk sulit atau mudah. “Tingkat kesulitan suatu butir soal dinyatakan dalam bentuk persentase peserta tes yang menjawab dengan benar, yang juga disebut sebagai tingkat kesukaran soal (item difficulty). Nilai tingkat kesulitan butir soal dapat berkisar antara 0,00 hingga 1,0” Pahmi, Hudaya, dan Jay a (2023).

Dari data hasil analisis tingkat kesukaran butir soal dalam Tabel 3, ditemukan bahwa 17% soal tergolong sangat mudah, yaitu pada nomor (4, 5, 19, 28, 29, dan 32). Selain itu, 14% soal tergolong mudah, terdapat pada nomor (1, 2, 16, 17, 35). Terdapat juga 15% soal yang tergolong sedang pada nomor (6, 8, 10, 15, 20). Selanjutnya, 37% soal kategori sukar (3, 7, 11, 12, 18, 21, 22, 23, 24, 25, 27, 30, 34). 17% soal tergolong sangat sukar pada nomor (9, 14, 19, 26, 31, 33). Berdasarkan data tersebut, terlihat bahwa tidak ada keseimbangan antara tingkat kesukaran dari setiap butir soal yang digunakan. Soal-soal yang digunakan didominasi oleh soal-soal yang tergolong sukar. Menurut (Widoyoko, 2018) Idealnya, dalam evaluasi, soal-soal yang digunakan sebaiknya memiliki tingkat kesukaran yang seimbang. Hal ini dapat dicapai dengan menggunakan persentase 25% soal mudah, 50% soal sedang, dan 25% soal sulit.

Mengacu pada kriteria butir soal menurut (Widoyoko, 2018) tabel 3 butir soal dengan indeks kesukaran 00–0,20 (sangat sukar) dan 0,9–1,00 (sangat mudah) dikategorikan sebagai kelompok item yang tidak baik dalam evaluasi. Dengan demikian soal yang terdapat nomor (4, 5, 19, 28, 29, 32) dan (9, 14, 19, 26, 31, 33) ialah soal yang tidak baik untuk dipakai dalam evaluasi oleh karena itu harus dibuang. Menurut (Fatimah & Khairuddin, 2019) item soal yang sangat mudah dalam evaluasi tidak memberikan tantangan yang cukup bagi siswa untuk meningkatkan usaha mereka dalam memecahkannya. Di sisi lain, soal yang terlalu sulit dapat menyebabkan siswa menjadi putus asa dan kehilangan motivasi untuk mencoba lagi karena melebihi kemampuan mereka.

Butir soal yang terdapat pada nomor (1, 2, 16, 17, 35) dan (3, 7, 11, 12, 18, 21, 22, 23, 24, 25, 27, 30, 34) merupakan soal yang kurang baik dan perlu direvisi kembali untuk dapat disimpan pada bank soal. Butir soal tersebut berada pada indeks kesukaran 0,21 – 0,30 “Sukar” dan 0,71 - 90 “Mudah”.

¹Damianus Nd. M., ²Vidriana O. B., ³Riwa R. H. E. | Analisis Kualitas Butir Soal Ujian Akhir Semester Gasal Tahun 2021/2022 Menggunakan Iteman Di SMP Negeri 1 Kambera

Sementara butir soal pada nomor (6, 8, 10, 15, 20) berada pada taraf kesukaran 0,31 - 70 “sedang” ini berarti soal-soal tersebut tidak terlalu mudah dan juga tidak terlalu sulit. Dengan demikian item-item tersebut dapat dipakai untuk evaluasi hasil belajar siswa dan disimpan dalam bank soal tanpa perlu melakukan revisi. Kualitas butir soal yang digunakan adalah baik ketika tingkat kesukarannya seimbang, yaitu tidak terlalu sulit tetapi juga tidak didominasi oleh soal terlalu mudah.

Pada analisis dimensi proses kognitif, pada kisi-kisi soal yang dibuat oleh guru IPA di SMP Negeri 1 Kambera berkisar pada level C1-C4. Level C1 11% terdapat pada nomor (1, 13, 21, 29). Level C2 14% terdapat pada nomor (3, 4, 15, 20, 25). Level C3 17% (14, 16, 17, 19, 30, 32). Dan level C4 57% pada nomor (2, 5, 6, 7, 8, 9, 10, 11, 12, 18, 22, 23, 24, 26, 27, 28, 31, 33, 34, 35).

Berdasarkan data tersebut rata-rata soal yang digunakan didominasi soal pada level C4. Soal pada level C4 merupakan soal yang berbasis HOTS yaitu 57% (2, 5, 6, 7, 8, 9, 10, 11, 12, 18, 22, 23, 24, 26, 27, 28, 31, 33, 34, 35) sementara 43% lainnya (1, 3, 4, 13, 14, 15, 16, 17, 19, 20, 21, 25, 29, 30, 32) merupakan soal kategori LOTS. Pernyataan tersebut merujuk pada revisi Teori Taksonomi Bloom oleh Andersson dan Krathwohl. Teori ini mengelompokkan kemampuan berpikir menjadi dua tingkat, yaitu kemampuan berpikir tingkat rendah (Lower Order Thinking Skill atau LOTS) yang mencakup kemampuan mengingat (C1), memahami (C2), dan menerapkan (C3). Sementara kemampuan berpikir tingkat tinggi (Higher Order Thinking Skill atau HOTS) yang melibatkan kemampuan analisis sintesis (C4), evaluasi (C5), dan kreativitas atau mencipta (C6) (Supriadi, 2020).

Berdasarkan perbandingan tingkat kesukaran dengan ranah kognitif dari soal yang dipakai guru dalam evaluasi terlihat bahwa soal yang digunakan didominasi oleh butir soal yang sulit. Hal ini dikarenakan 54% butir soal dikategorikan sukar dalam analisis tingkat kesukaran: (3, 7, 9, 11, 12, 14, 18, 19, 21, 22, 23, 24, 25, 26, 27, 30, 31, 33, 34) dan 57% berada pada level C4. Dari data ini dapat dikatakan bahwa soal yang sulit mempengaruhi hasil belajar siswa. Dengan demikian, kreativitas guru dalam membuat butir soal perlu ditingkatkan. Disisi lain perlu adanya juga peningkatan mutu pembelajaran untuk meningkatkan hasil belajar siswa.

Dalam era pembelajaran abad ke-21, diperlukan untuk menerapkan keterampilan berpikir tingkat tinggi atau HOTS. Dengan demikian, dalam menjawab tuntutan ini, peran guru sangat penting dalam meningkatkan kualitas pembelajaran. Dalam mengevaluasi hasil belajar siswa, guru perlu memiliki kreativitas dalam menyusun soal yang sesuai dengan tuntutan abad 21. Tujuan dari penerapan kemampuan berpikir tingkat tinggi saat ini adalah untuk menguji kemampuan analitis siswa, bukan hanya kemampuan mengingat informasi secara pasif. Siswa perlu mampu mengolah dan menerapkan pemahaman mereka terhadap masalah yang dihadapi (Asfiah, 2021). Dalam studi yang dilakukan oleh (Asfiah, 2021:108) mengenai evaluasi dan pembuatan soal HOTS, terdapat beberapa tahapan yang harus diikuti, ialah: “1) Menganalisis Kompetensi Dasar yang dapat menjadi dasar penyusunan soal HOTS. 2) Menyusun kerangka pertanyaan. 3) Memilih stimulus yang menarik dan relevan dengan konteks. 4) Menulis item pertanyaan sesuai dengan kisi-kisi pertanyaan. 5) Membuat rubrik penilaian atau kunci jawaban”.

4. Hasil analisis daya pembeda

Dalam analisis daya pembeda, dilakukan evaluasi terhadap soal-soal tes untuk melihat sejauh mana kemampuannya dalam membedakan siswa berdasarkan prestasi mereka. Dalam kutipan (Paskalin & Susanti, 2020) mengemukakan dua manfaat dari daya pembeda butir soal, yaitu: “1) untuk meningkatkan kualitas setiap butir soal berdasarkan data empirisnya. Dengan melihat indeks daya pembeda, kita dapat mengetahui apakah suatu butir soal sudah baik, perlu direvisi, atau harus ditolak. 2) untuk mengetahui sejauh mana suatu butir soal mampu mendeteksi atau membedakan kemampuan siswa, baik mereka yang telah memahami materi yang diajarkan maupun yang belum”.

Berdasarkan data daya beda yang terdapat pada Tabel 8, ditemukan bahwa persentase sebanyak 54% dari keseluruhan butir soal menunjukkan tingkat daya beda yang sangat baik. Item soal yang termasuk dalam kategori ini adalah (3, 4, 7, 9, 12, 14, 15, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, dan 34). 23% dari item soal yang menunjukkan tingkat daya beda kategori baik. Butir soal tersebut dapat

¹Damianus Nd. M., ²Vidriana O. B., ³Riwa R. H. E. | Analisis Kualitas Butir Soal Ujian Akhir Semester Gasal Tahun 2021/2022 Menggunakan Iteman Di SMP Negeri 1 Kambera

ditemukan pada nomor-nomor berikut: (1, 6, 11, 13, 17, 24, 25, dan 35). Sebanyak 11% tergolong cukup baik, yaitu pada nomor: (10, 18, dan 20). Terdapat 3% item yang berada pada kategori daya pembeda kurang baik, yaitu pada nomor 8. Dan 9% butir soal yang tidak menunjukkan tingkat daya beda yang baik, yaitu pada nomor: (2, 5, dan 19).

Jika melihat kriteria daya beda butir soal berdasarkan (Widoyoko, 2018) yang tercantum dalam Tabel 4, dapat disimpulkan bahwa sebanyak 77% dari total butir soal telah menunjukkan tingkat daya beda yang baik dalam mengevaluasi hasil belajar dan dapat disimpan pada bank soal. Butir soal tersebut dapat ditemukan pada nomor-nomor berikut: (3, 4, 7, 9, 12, 14, 15, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 1, 6, 11, 13, 17, 24, 25, dan 35). Butir soal tersebut diklasifikasikan "Baik" dengan kisaran nilai daya beda antara 0,41-0,50, serta "Sangat baik" dengan kisaran nilai daya beda antara 0,51-1,00. Soal-soal dengan kategori tersebut efektif dalam membedakan antara siswa yang telah memahami materi dengan siswa yang belum memahami materi yang diajarkan.

Sebanyak 12% butir soal perlu adanya perbaikan berada pada kategori 0,31–0,40 “Cukup baik; 0,21–0,30 “kurang baik”, yaitu pada nomor (10, 18, 18, 20 dan 8). Soal-soal tersebut memiliki daya beda yang kurang baik dalam evaluasi sehingga perlu adanya pembahasan dan perlu direvisi untuk dapat disimpan pada bank soal. Sementara butir soal dengan kriteria 0,00 – 0,20 “Tidak baik” yaitu 9% butir soal (2, 5, 19) merupakan butir soal yang ditolak. Soal tersebut tidak adanya kemampuan untuk membedakan antara peserta tes yang telah menguasai materi dan yang belum dengan demikian soal tersebut tidak dapat digunakan dalam evaluasi. Butir soal yang ideal untuk digunakan dalam evaluasi hasil belajar harus mampu membedakan antara kelompok siswa yang memiliki prestasi tinggi dengan kelompok siswa yang memiliki prestasi rendah di antara semua peserta tes (Bano, Marambaawang & Njoeroemana, 2022).

5. Hasil analisis efektivitas pengecoh

Pelaksanaan analisis efektivitas opsi pengecoh bertujuan untuk mengetahui seberapa efektif pilihan dapat mengelabui peserta tes yang tidak mengetahui kunci jawaban yang tersedia. Menurut kutipan (Arbiatin & Mulabbayah, 2020) sebuah distraktor dianggap berfungsi jika proporsinya lebih dari 5%, sedangkan distraktor tidak berfungsi dengan baik jika proporsinya kurang dari 5%.

Berdasarkan hasil analisis butir soal pada tabel 10 ditemukan sebanyak 54% item telah memiliki opsi pengecoh yang berfungsi dengan baik (2, 3, 6, 7, 8, 9, 10, 11, 12, 14, 18, 21, 22, 23, 24, 25, 27, 31, 33). Sementara 46% lainnya beberapa opsi pengecoh belum berfungsi dengan baik yaitu terdapat pada nomor (1, 4, 5, 13, 15, 16, 17, 19, 20, 26, 28, 29, 30, 32, 34, 35). Setiap opsi pengecoh yang telah berfungsi dengan baik dapat disimpan pada bank soal, hal ini karena opsi tersebut cukup efektif dalam mengecoh siswa. Sementara opsi pengecoh yang belum berfungsi dengan baik dapat direvisi atau dibuang.

Berdasarkan Tabel 11, opsi pengecoh yang ditolak atau direvisi dikarenakan adanya perbedaan yang terlalu mencolok antara jawaban yang benar (kunci jawaban) dan jawaban yang salah. Hal ini menyebabkan banyak siswa yang mampu memilih jawaban yang benar. Opsi pengecoh yang efektif seharusnya mampu mengecoh peserta tes yang tidak menguasai materi yang diberikan dan dipilih oleh setidaknya 5% dari seluruh peserta yang mengikuti tes. Oleh karena itu setiap opsi pengecoh dalam melakukan evaluasi perlu dikaji ulang pilihan jawaban pada butir soal yang pengecohnya tidak berfungsi dengan baik agar pengecoh tersebut dapat berfungsi dengan baik (Fiska, Hidayati, Qomaria & Hadi, 2021).

KESIMPULAN

Dari hasil analisis kualitas item soal ujian akhir semester gasal tahun ajaran 2021/2022 di SMP Negeri 1 Kambera menggunakan program ITEMAN, dapat ditarik kesimpulan bahwa 75% dari item soal tersebut memiliki validitas yang baik, sementara 25% sisanya tidak valid. Secara keseluruhan, item soal yang digunakan memiliki tingkat reliabilitas yang tinggi, dengan nilai reliabilitas sebesar 0,886. Dalam analisis tingkat kesulitan item, ditemukan bahwa 17% dari item soal tergolong sangat mudah (sebaiknya dihilangkan), 14% tergolong mudah (perlu direvisi), 15% tergolong sedang (dapat disimpan

¹Damianus Nd. M., ²Vidriana O. B., ³Riwa R. H. E. | Analisis Kualitas Butir Soal Ujian Akhir Semester Gasal Tahun 2021/2022 Menggunakan Iteman Di SMP Negeri 1 Kambera dalam bank soal tanpa revisi), 37% tergolong sulit (perlu direvisi), dan 17% tergolong sangat sulit (sebaiknya dihilangkan).

Pada hasil analisis daya pembeda diperoleh data sebanyak 54% butir soal memiliki daya pembeda yang sangat baik dan 23% memiliki daya beda kategori baik (dapat disimpan pada bank soal). 11% butir memiliki daya beda kategori cukup baik dan 1% kurang baik (perlu adanya pembahasan dan revisi). 9% item tidak memiliki daya pembeda (harus dibuang). Pada hasil analisis efektivitas pengecoh ditemukan sebanyak 54% butir soal telah memiliki efektivitas pengecoh yang menjalankan fungsinya dengan baik sementara 46% lainnya beberapa opsi pengecoh kurang baik. Berdasarkan beberapa temuan butir soal yang belum memenuhi kriteria yang diharapkan hal ini menunjukkan perlukan adanya peningkatan kreativitas guru dalam menyusun butir yang baik agar bentuk soal yang digunakan lebih berkualitas dan tepat sasaran.

DAFTAR PUSTAKA

- Arbiatin, E., & Mulabbiyah. (2020). Analisis Kelayakan Butir Soal Tes Penilaian Akhir Semester Mata Pelajaran Matematika Kelas VI di SDN 19 Ampenan Tahun Pelajaran 2019/2020. *Jurnal PGMI*, 12(2), 146–171. <https://doi.org/10.20414/ELMIDAD.V12I2.2627>
- Asfiah, S. (2021). Implementasi Penilaian Berbasis High Order Thinking Skills Pada Mapel PAI Dalam Meningkatkan Berpikir Kritis Dan Kreatif Siswa di Tingkat SMP. *Quality: Journal of Empirical Research in Islamic Education*, 9(1), 103–120. <https://doi.org/http://dx.doi.org/10.21043/quality.v9i1.10136>
- Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis Kriteria Butir Soal Ujian Sekolah Mata Pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, 8(1), 145. <https://doi.org/10.32884/ideas.v8i1.660>
- Berkah, M. (2019). *Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Aqidah Kelas VIII SMP Muhammadiyah Ahmad Dahlan Metro Tahun Pelajaran 2018/2019*. Fakultas Tarbiyah dan Ilmu Keguruan INSTITUT AGAMA ISLAM NEGERI (IAIN) METRO.
- Chasanah, N., Widodo, W., & Suprpto, N. (2022). Pengembangan Instrumen Asesmen Literasi Sains Untuk Mendeskripsikan Profil Peserta Didik. *PENDIPA Journal of Science Education*, 6(2), 474–483. <https://doi.org/10.33369/pendipa.6.2.474-483>
- Fauziana, A., & Wulansari, A. D. (2021). Analisis Kualitas Butir Soal Ulangan Harian di Sekolah Dasar dengan Model Rasch. *Jurnal Ibriez : Jurnal Kependidikan Dasar Islam Berbasis Sains*, 6(1), 10–19. <https://doi.org/10.21154/IBRIEZ.V6I1.112>
- 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>
- Gusti, U. A. (2021). Analisis Kualitas Soal Uas Mata Pelajaran Ipa Di Smp Islam Khaira Ummah Kota Padang Kelas VIII Tahun Ajaran 2019/2020. *Biopedagogia*, 3(2), 81–87. <http://180.250.193.171/index.php/biopedagogia/article/view/2325>
- Hartini, H. (2021). Analisis Kualitas Tes dan Butir Soal Sejarah di SMAN 1 Praya Timur. *Jurnal Inovasi Pendidikan IPA*, 7(1), 34–45. <https://doi.org/10.21831/jipi.v7i1.38852>
- Huda, N., & Wahyuni, T. S. (2020). Penggunaan Aplikasi Item and Test Analysis (Iteman) Pada Soal Try Out UN IPA SMP Tahun 2019. *Jurnal Pembelajaran Sains*, 4(1), 2527–9157. <http://journal2.um.ac.id/index.php/jpsi/article/view/9738>
- Ikawati, H. D., Jayadi, A., & Hermians. (2022). Analisis Kualitas Tes dan Butir Soal Sejarah di SMAN 1 Praya Timur. *Jurnal Pendidikan Dan Konseling*, 4, 1349–1358. <https://doi.org/DOI:https://doi.org/10.31004/jpdk.v4i4.6489>
- Pahmi, S., Hudaya, C., & Jaya, A. (2023). Pengaruh Model Pembelajaran Tefa (Teaching Factory) Dalam Pembelajaran Fisika Untuk Meningkatkan Hasil Belajar Siswa SMK. *Jurnal Hasil Kajian, Inovasi, Dan Aplikasi Pendidikan Fisika*, 9(1), 1–15. <https://doi.org/DOI:https://doi.org/10.31764/orbita.v9i1.11718>
- Paskalin, G., & Susanti, M. M. I. (2020). Analisis Butir Soal Materi Gaya Kelas Iv Di Sd Negeri Kentungan. *DIDAKTIKA: Jurnal Pendidikan Sekolah Dasar*, 3(1), 23–32. <https://doi.org/10.21831/didaktika.v3i1.31431>
- Purniasari, L., Masykuri, M., Sri, D., & Ariani, R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata

- ¹Damianus Nd. M., ²Vidriana O. B., ³Riwa R. H. E. | Analisis Kualitas Butir Soal Ujian Akhir Semester Gasal Tahun 2021/2022 Menggunakan Iteman Di SMP Negeri 1 Kambera Pelajaran Kimia SMA N 1 Kutowinangun Tahun Pelajaran 2019/2020 Menggunakan Model Iteman dan Rasch. *Jurnal Pendidikan Kimia*, 10(2), 205–214. <https://doi.org/10.20961/JPKIM.V10I2.48244>
- Rizalia, S., Syam, A. N., & Yulita, N. (2021). Kajian Kualitatif Tes Hasil Belajar Buatan Guru Biologi MAN di Kota Kendari. *Journal of Natural Science and Integration*, 4(1), 105. <https://doi.org/10.24014/jnsi.v4i1.10410>
- Sabara, A., Yusuf, F. M., Ahmad, J., Nusantara, E., & Lamangajto, C. (2022). Validitas Perangkat Pembelajaran Berbasis Steam Materi Struktur Dan Fungsi Jaringan Tumbuhan Untuk Meningkatkan Penguasaan Konsep. *Jambura Edu Biosfer Journal*, 4(2), 76–85. <https://doi.org/10.34312/jebj.v4i2.14734>
- Supriadi, G. (2020). *Penilaian Berbasis Higher Order Thinking Skill (Hots)* (Rini (ed.); Agung Isti). Aswaja Pressindo.
- Taufik, A. N., Berlian, L., Fathia, A., & Rasyidin, E. Y. (2022). Pengembangan Instrument Tes Literasi Sains Pada Tema Yuk Makan Talas Beneng Untuk Calon Guru Ipa. *Jurnal Aksioma*, 11(2), 141–150. <https://doi.org/https://doi.org/10.22487/aksioma.v11i2.2452>
- Widoyoko, S. E. P. (2018). *Penilaian Hasil Pembelajaran Di Sekolah*. Pustaka Pelajar.