

## Analisis Kualitas Tes dan Butir Soal Sejarah di SMAN 1 Praya Timur

Hastuti Diah Ikawati<sup>1</sup>, Agus Jayadi<sup>2</sup>, Hermansyah<sup>3</sup>

<sup>1,2,3</sup> Universitas Pendidikan Mandalika Mataram

Email: [hastutidiahikawati@undikma.ac.id](mailto:hastutidiahikawati@undikma.ac.id)<sup>1\*</sup>

### Abstrak

Kualitas tes sangat mempengaruhi akurasi tes itu sendiri. Oleh karenanya analisis kualitas tes merupakan suatu tahap yang harus ditempuh untuk mengetahui derajat kualitas suatu tes, baik tes secara keseluruhan maupun butir soal yang menjadi bagian dari tes. Dalam penilaian hasil belajar, tes diharapkan dapat menggambarkan sampel perilaku dan menghasilkan nilai yang objektif serta akurat. Tes yang digunakan guru harus memiliki kualitas yang baik dilihat dari berbagai segi. Tes hendaknya disusun sesuai dengan prinsip dan prosedur penyusunan tes, setelah digunakan perlu diketahui apakah tes tersebut berkualitas baik atau kurang baik. Adapun jenis penelitian ini adalah penelitian korelasi dengan menggunakan pendekatan kuantitatif dengan metode eksperimen semu. Untuk memperoleh data tentunya diperlukan suatu alat atau instrumen. Instrumen yang digunakan peneliti adalah observasi, wawancara, dan studi dokumentasi. Untuk memperoleh data melalui wawancara, peneliti menggunakan patokan wawancara yang dalamnya terdapat pertanyaan-pertanyaan untuk dijawab oleh partisipan. Kemudian untuk mengumpulkan data melalui observasi, peneliti menggunakan pedoman observasi. Dari hasil analisis menggunakan program ITEMAN, butir soal yang dikelompokkan baik berdasarkan tingkat kesukaran adalah butir 21 butir. Dari hasil analisis menggunakan program ITEMAN, butir soal yang dikategorikan sangat bagus dari segi daya pembeda adalah sebanyak 17 butir, sedangkan yang dikategorikan cukup bagus tidak perlu perbaikan adalah sebanyak 7 butir, yang dikategorikan kurang bagus dan perlu perbaikan adalah sebanyak 3 butir, dan yang dikategorikan jelek adalah sebanyak 2 butir. Berdasarkan hasil analisis menggunakan program ITEMAN, ada satu butir soal yang menunjukkan kemampuan peserta tes tersebut rendah, yaitu terdapat pada butir ke 26. Berdasarkan hasil perhitungan dengan menggunakan bantuan output item and Test Analysis program ITEMAN Version 3.0 MicroCAT Testing System, uji reliabilitas memperoleh hasil 0,467.

**Kata kunci:** *Kualitas, Tes, Butir Soal*

### Abstract

The quality of the test greatly affects the accuracy of the test itself. Therefore, analysis of test quality is a step that must be taken to determine the degree of quality of a test, both the test as a whole and the items that are part of the test. In the assessment of learning outcomes, the test is expected to describe a sample of behavior and produce an objective and accurate value. The test used by the teacher must have good quality in terms of various aspects. The test should be prepared in accordance with the principles and procedures for preparing the test, after using it, it is necessary to know whether the test is of good or poor quality. The type of this research is correlation research using quantitative approach with quasi-experimental method. To obtain data, of course, a tool or instrument is needed. The instruments used by the researcher were observation, interviews, and documentation studies. To obtain data through interviews, researchers used an interview benchmark in which there were questions to be answered by participants. Then to collect data through observation, researchers used observation guidelines. From the results of the analysis using the ITEMAN program, the items that are grouped well based on the level of difficulty are 21 items. From the results of the analysis using the

ITEMAN program, the items that are categorized as very good in terms of discriminating power are 17 items, while those categorized as good enough that do not need improvement are 7 items, those categorized as poor and need improvement are 3 items, and those categorized as bad is as much as 2 points. Based on the results of the analysis using the ITEMAN program, there is one item that shows the ability of the test taker is low, which is in item 26. Based on the results of calculations using the output item and Test Analysis program ITEMAN Version 3.0 MicroCAT Testing System, the reliability test obtained a result of 0.467 .

**Keywords:** *Quality, Test, Items*

## PENDAHULUAN

Tes merupakan alat ukur untuk mengetahui tingkat penguasaan siswa terhadap materi pelajaran. Tujuan dari kegiatan evaluasi adalah supaya mengetahui apakah tujuan yang dirumuskan sudah tercapai atau belum dan apakah dalam materi yang disampaikan atau diajarkan pada siswa sudah tepat, (Wirandani, dkk., 2019).

Kualitas tes sangat mempengaruhi akurasi tes itu sendiri. Oleh karenanya analisis kualitas tes merupakan suatu tahap yang harus ditempuh untuk mengetahui derajat kualitas suatu tes, baik tes secara keseluruhan maupun butir soal yang menjadi bagian dari tes. Dalam penilaian hasil belajar, tes diharapkan dapat menggambarkan sampel perilaku dan menghasilkan nilai yang objektif serta akurat. Tes yang digunakan guru harus memiliki kualitas yang baik dilihat dari berbagai segi. Tes hendaknya disusun sesuai dengan prinsip dan prosedur penyusunan tes, setelah digunakan perlu diketahui apakah tes tersebut berkualitas baik atau kurang baik.

Analisis kualitas tes berkaitan dengan pertanyaan apakah tes sebagai suatu alat ukur benar-benar mengukur apa yang hendak dan seharusnya diukur, sampai mana tes tersebut dapat diandalkan dan berguna. Kedua hal ini sekaligus merupakan karakteristik alat ukur yang baik. Mengingat urgensi dari soal sebagai alat untuk mengukur keberhasilan pembelajaran, sangat penting untuk diadakan penelitian tentang analisis butir soal, (Utomo:2018)

Penyusunan tes tidak hanya sekedar mengandalkan kemampuan penyusunan kalimat dalam setiap butir soal, tetapi harus memperhatikan kaidah-kaidahnya. Kaidah penyusunan tidak lain merupakan tahap-tahap yang perlu dilalui apabila menyusun suatu tes. Salah satu tahap tersebut adalah menyusun kisi-kisi tes.

Setiap kegiatan belajar harus diketahui sejauh mana proses belajar tersebut telah memberikan nilai tambah bagi kemampuan siswa. Guru diharuskan dapat mengembangkan instrumen evaluasi yang berkualitas untuk mengukur kemampuan siswa sebab tes ini tidak hanya sebagai seleksi peringkat, tetapi tes juga mampu memberikan informasi atau saran terhadap pengguna tes atau siswa untuk melakukan langkah apa ke depan yang harus dilakukan, (Pas dan Yusrizal, 2018).

Salah satu cara untuk melihat peningkatan kemampuan tersebut adalah dengan melakukan tes. Tes sebagai alat penilaian adalah pertanyaan-pertanyaan yang akan diberikan kepada siswa untuk mendapat jawaban dari siswa dalam bentuk lisan (tes lisan), dalam bentuk tulisan (tes tulisan), atau dalam bentuk perbuatan (tes tindakan).

Tes bidang studi IPS Sejarah di SMA N 1 Praya Timur digunakan soal yang disajikan dalam bentuk tes pilihan ganda. Tes tersebut merupakan tes obyektif pilihan ganda, jika tes berbentuk pilihan ganda alternatif jawaban telah disediakan, sehingga siswa dapat menebak jawaban atau mencari jalan pintas dengan jalan bertanya atau mencontek pada teman terdekat. Didalam setiap kerangka tes terdapat salah satu ciri butir yang dikenal sebagai tingkat kesukaran butir. Semakin sedikit peserta tes yang menjawab benar suatu butir, maka soal tersebut tergolong sukar. Demikian juga sebaliknya, jika jumlah peserta jauh lebih banyak yang menjawab benar, maka butir soal tersebut tergolong mudah.

Bilangan yang menunjukkan sukar dan mudahnya suatu soal disebut indeks kesukaran. Indeks kesukaran butir soal adalah bilangan yang menunjukkan sukar dan mudahnya soal, semakin tinggi indeks kesukaran butir maka soal semakin mudah (Arikunto, 2009). Soal yang baik adalah soal tidak terlalu mudah dan tidak terlalu sukar.

Kualitas tes ditentukan oleh kualitas item-itemnya. Sebuah tes yang berisi item berkualitas tinggi walaupun dalam jumlah yang sedikit akan jauh lebih berguna daripada sebuah tes yang berisi puluhan item berkualitas rendah. Item-item yang berkualitas rendah tidak saja akan menurunkan fungsi tes akan tetapi akan memberikan hasil pengukuran yang kurang baik pula. Oleh sebab itu, alat tes yang berkualitas mutlak diperlukan guna memperoleh hasil ukur secara valid dan reliabel. Dalam menyusun butir tes penting untuk memastikan reliabilitasnya karena menentukan homogenitas tes pada konstruk yang diukur (Adhi & Aly, 2018). Butir tes yang valid dibutuhkan untuk meningkatkan kualitas evaluasi dan mengembangkan bank soal yang sesuai (Ahmed & Moalwi, 2017; Musa et al., 2018).

Untuk melihat kelemahan dan kekurangan tes, dapat dilakukan dengan dua cara yaitu: 1) dengan analisis soal, 2) dengan menghitung validitas serta reliabilitas soal (Firmansyah & Sridadi, 2018). Berdasarkan hal tersebut, maka penting dilakukan uji kualitas tes mata pelajaran Sejarah di SMA N 1 Praya Timur. Adapun tujuan penelitian adalah untuk mengetahui a) tingkat kesukaran butir soal, b) daya pembeda butir soal, c) validitas soal, d) Uji reliabilitas butir soal. Dari serangkaian kegiatan tersebut akan dihasilkan satu paket alat ukur berupa tes yang berkualitas dilihat dari validitas dan reliabilitasnya.

## **METODE**

Penelitian ini dilaksanakan di SMA N 1 Praya Timur tahun 2022. Adapun jenis penelitian ini adalah penelitian korelasi dengan menggunakan pendekatan kuantitatif dengan metode eksperimen semu, penggunaan metode eksperimen dalam penelitian ini, dimaksudkan untuk mengetahui apakah ada analisis kualitas tes dan butir soal mata pelajaran sejarah.

### **1. Instrument Penelitian**

Untuk memperoleh data tentunya diperlukan suatu alat atau instrumen. Instrumen yang digunakan peneliti adalah observasi, wawancara, dan studi dokumentasi. Untuk memperoleh data melalui wawancara, peneliti menggunakan patokan wawancara yang dalamnya terdapat pertanyaan-pertanyaan untuk dijawab oleh partisipan. Kemudian untuk mengumpulkan data melalui observasi, peneliti menggunakan pedoman observasi. Sedangkan melalui dokumentasi dapat diperoleh data dari dokumen soal Ujian Sekolah (US) kelas XI SMA N 1 Praya Timur. (Wirandani, dkk., 2019).

### **2. Teknik Pengumpulan Data**

Berikut ini, Langkah-langkah (prosedur penelitian) dalam penggunaan lembar perekam data ini meliputi :

- a. Menghimpun soal Ujian Sekolah (US) mata pelajaran IPS Sejarah
- b. Memberikan kode pada setiap muatan soal, yaitu Koding Tahun Pelajaran, Koding Kompetensi dan Indikator SKL, Koding Nomor Halaman, dan Koding Nomor Soal
- c. Mengklasifikasikan tingkatan soal berdasarkan Taksonomi Bloom
- d. Mengisi kolom tingkatan kognitif dengan indikator yang sesuai
- e. Menghitung presentase tingkat proses kognitif serta tingkat pengetahuan soal berdasarkan taksonomi Bloom.
- f. Membuat kesimpulan dari persentase yang telah diperoleh dari hasil penelitian. (Mujib, dkk., 2018)

### **3. Teknik Analisis Data**

Penelitian ini berfokus pada analisis butir tes pilihan ganda untuk mengetahui kualitas tes mata pelajaran IPS Sejarah di SMA N 1 Praya Timur. Analisis butir tes mencakup: (1) tingkat kesukaran butir tes, (2) tingkat daya pembeda butir tes, (3) validitas butir tes dan (4) reliabilitas tes. Untuk mengukur tingkat kesukaran digunakan rumus perbandingan jumlah responden yang menjawab benar dengan jumlah responden. Untuk mengukur validitas dan daya pembeda soal digunakan uji Pearson product moment. Kemudian, uji Cronbach alpha digunakan untuk menentukan reliabilitas tes. Hasil uji tingkat kesukaran menempatkan butir tes pada kategori mudah, sedang dan sukar. Hasil uji daya pembeda mengklasifikasikan butir tes pada kategori baik dan unggul. Uji validitas membuktikan seluruh butir tes valid. Uji reliabilitas menyatakan bahwa tes yang dirancang reliabel. Dengan demikian butir tes yang dirancang tidak membutuhkan perbaikan dan dapat digunakan untuk mengukur keterampilan berpikir kritis dan pengetahuan prosedural pada grammar Bahasa Inggris. (Alfiana, dkk., 2021).

## **HASIL DAN PEMBAHASAN**

Untuk mencapai tujuan penelitian, maka data yang dikumpulkan berupa lembar jawaban peserta tes, dianalisis secara kuantitatif dan hasilnya sebagai berikut :

1. Tingkat Kesukaran Soal.  
Dari hasil analisis menggunakan program ITEMAN, butir soal yang dikelompokkan baik berdasarkan tingkat kesukaran adalah butir 21 butir.
2. Analisis Daya Pembeda  
Dari hasil analisis menggunakan program ITEMAN, butir soal yang dikategorikan sangat bagus dari segi daya pembeda adalah sebanyak 17 butir, sedangkan yang dikategorikan cukup bagus tidak perlu perbaikan adalah sebanyak 7 butir, yang dikategorikan kurang bagus dan perlu perbaikan adalah sebanyak 3 butir, dan yang dikategorikan jelek adalah sebanyak 2 butir.
3. Analisis Validitas  
Berdasarkan hasil analisis menggunakan program ITEMAN, ada satu butir soal yang menunjukkan kemampuan peserta tes tersebut rendah, yaitu terdapat pada butir ke 26.
4. Uji Reliabilitas.  
Berdasarkan hasil perhitungan dengan menggunakan bantuan output item and Test Analysis program ITEMAN Version 3.0 MicroCAT Testing System, uji reliabilitas memperoleh hasil 0,467.

Untuk memperoleh gambaran lebih lengkap tentang data yang terkumpul pada tahap pemberian tes mata pelajaran teknologi informasi dan komunikasi, maka beberapa statistic yang perlu mendapat perhatian seperti koefisien reliabilitas Alpha Cronbach untuk data dikotomi, standar kesalahan pengukuran atau standard error of measurement (SEM) dan bentuk distribusi atau analisis kemencengan distribusi. Kesalahan baku pengukuran atau standard error of measurement (SEM), ukuran ini sangat cocok digunakan untuk menafsirkan skor individu, dan seringkali ukuran SEM lebih berguna dari pada reliabilitas. Adapun SEM yang didapat dari hasil analisis adalah 1,642.

Bentuk distribusi dalam analisis memegang peranan penting dan berbentuk distribusi yang paling umum dijumpai adalah bentuk distribusi normal, sehingga akan banyak teknik statistic yang mendasarkan analisis pada distribusi ini. Bentuk distribusi normal merupakan distribusi simetris belahan kiri dan kanan sama, ukuran pemusatan berada pada angka yang sama, yaitu angka pemusatan mean, median, dan mode. Bila mean dan median berada pada sisi kanan atau mungkin juga sebelah kiri, posisi ini akan menunjukkan pada nilai skewness data. Jika angka skewness positif maka diartikan menjulur ke kanan, modus atau median berada di sebelah kanan mean atau dapat pula ditafsirkan dalam distribusi data lebih banyak angka rendah, sebaliknya jika modus atau median berada

di sebelah kiri distribusi angka skewness negative, berarti banyak angka bernilai tinggi dalam distribusi data yang dianalisis. Dalam hal ini data yang dihasilkan dalam penelitian ini adalah -0,821, dalam artian banyak angka bernilai tinggi dalam distribusi data yang dianalisis dan median berada di sebelah kiri.

Kurtosis juga merupakan suatu analisis bentuk distribusi, pada skewness bentuk distribusi menjulur ke kiri atau ke kanan jadi orientasi distribusi horizontal, sedangkan analisis kurtosis menunjukkan orientasi distribusi ke atas atau ke bawah dalam distribusi. Bentuk distribusi normal disebut juga mesokurtik, jika bentuk distribusi terlalu meruncing ke atas (leptokurtik) sedangkan distribusi terlalu ke bawah (platikurtik) berarti distribusi melandai. Jika angka kurtosis positif dapat diartikan bentuk distribusi leptokurtik, sebaliknya jika angka kurtosis negative diartikan berbentuk distribusi platikurtik atau distribusi cenderung merata. Pada data dengan angka kurtosis mendekati nol, diartikan data berdistribusi normal.

Deskripsi data hasil analisis data tes dengan bantuan program analisis butir berdasarkan teori klasik yaitu program perangkat lunak ITEMAN. Adapun hasil analisis data tes sebagai berikut :

Tabel 4.10 : Deskripsi Data Hasil Analisis Tes

Statistik	Nilai
N Butir	30
N Respon	513
Mean	23,842
Variance	5,061
Standar Deviasi	2,250
Skew	-0,821
Kurtosis	1,029
Minimum	14,000
Maximum	29,000
Median	24,000
Alpha	0,467
SEM	1,642
Mean P	0,795
Mean Item Total	0,247
Mean Biserial	0,508

## SIMPULAN

Dari data yang terkumpul melalui lembar jawaban peserta ujian, diseleksi, dihitung banyak peserta menjawab betul atau banyak penjawab pada masing-masing alternatif jawaban, kemudian dimasukkan kedalam rumus tertentu sehingga diperoleh nilai P dan nilai B.

Nilai P dan B untuk setiap butir soal ditentukan kualifikasi butir soal berdasarkan kriteria yang telah dikemukakan di atas sehingga dapat disimpulkan :

1. Butir soal yang dikelompokkan baik berdasarkan tingkat kesukaran adalah butir nomor 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 15, 16, 18, 19, 20, 22, 23, 24, 27, 29, dan 30. Banyak butir soal yang dikelompokkan soal baik ditinjau dari segi tingkat kesukaran soal adalah 21 butir.
2. Butir soal yang dikategorikan baik sekali dari segi daya pembeda soal adalah 1, 3, 6, 15, dan 25, sedangkan butir dikategorikan baik (cukup memuaskan) adalah butir 2, 4, 5, 8, 10, 11, 13, 16, 19, 21, 22, 27, 29, dan 30.

3. Dari penjelasan di atas tingkat kesukaran soal (baik), daya pembeda soal (baik sekali, cukup memuaskan) maka ditemukan 19 butir soal masih ditolerir untuk digunakan yaitu butir soal 1, 2, 3, 4, 5, 6, 8, 10, 11, 13, 15, 26, 19, 21, 22, 25, 27, 29, dan 30.
4. Berdasarkan kesimpulan butir soal yang kurang baik dari segi tingkat kesukaran soal, daya pembeda butir soal sebanyak 19 butir, kemudian dianalisis efektifitas distraktor ternyata hanya option yang tidak efektif hanya terdapat pada butir ke 26 option C dan butir ke 28 option A yang kurang baik dijawab oleh kurang dari 2,5% peserta tes.

#### DAFTAR PUSTAKA

- Wirandani, T., Ayu Cendra Kasih,, Latifah. ANALISIS BUTIR SOAL HOTS (HIGH ORDER THINKING SKILL) PADA SOAL UJIAN SEKOLAH KELAS XII MATA PELAJARAN BAHASA INDONESIA DI SMK AN-NAHL. *Parole (Jurnal Pendidikan Bahasa dan Sastra Indonesia)*, Volume 2 Nomor 4, Juli 2019.
- Utomo, B. Analisis Validitas Isi Butir Soal sebagai Salah Satu Upaya Peningkatan Kualitas Pembelajaran di Madrasah Berbasis Nilai-Nilai Islam. *Jurnal Pendidikan Matematika* Vol 2 No 1 (2018).
- Pasi, S., N., & Yusrizal. ANALISIS BUTIR SOAL UJIAN BAHASA INDONESIA BUATAN GURU MTsN DI KABUPATEN ACEH BESAR. *Master Bahasa Vol. 6 No.2; Mei 2018:195–202*.
- Arikunto, S.. 2009. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi. Aksara.
- Adhi, M. I., & Aly, S. M. (2018). Student Perception and Post-exam Analysis of One Best MCQs and One Correct MCQs: A Comparative Study. *Journal of the Pakistan Medical Association*, 68 (4), 570–575.
- Firmansyah, H., & Sridadi. (2018). Analisis butir soal tes pilihan ganda mata pelajaran pendidikan jasmani olahraga dan kesehatan kelas VIII SMPN 1 Wonosari Tahun Ajaran 2017/2018. *Jurnal Pendidikan Jasmani Kesehatan Dan Rekreasi*, 7 (12), 139–157.
- Arifin, Z. 2011. *Penelitian Pendidikan*. Bandung: Remaja Rosda.
- Sudjana, N. 2010. *Penelitian Hasil Proses Belajar Mengajar*. Bandung: PT. Remaja Rosdakarya
- Mujib, N., R., Anselmus J.E Toenlloe, Henry Praherdhiono. ANALISIS BUTIR SOAL UJIAN NASIONAL IPA SD/MI TAHUN 2015 SAMPAI 2017 BERDASARKAN TAKSONOMI BLOOM. *JKTP: Jurnal Kajian Teknologi Pendidikan*, Vol 1, No 2 (2018).
- Ahmed, I. A. M., & Moalwi, A. A. (2017). Correlation Between Difficulty and Discrimination Indices of MCQs Type A in Formative Exam in Anatomy. *Journal of Research & Method in Education*, 7(5), 28–43.
- Alfiana, H., Hari Karyono, Wawan Gunawan. ANALISIS BUTIR TES KETERAMPILAN BERPIKIR KRITIS DAN PENGETAHUAN PROSEDURAL GRAMMAR BAHASA INGGRIS. *JINOTEP (Jurnal Inovasi Teknologi Pembelajaran)* Vol 8 (1) (2021): 20-29.