

## Analisis Kualitas Butir Soal Ujian Semester Pelajaran Sejarah Kelas XII di Sma Negeri 1 Aceh Barat Daya pada Masa Pandemi Covid-19

Irma Yusriani<sup>(1)</sup>, Mawardi<sup>(2)</sup>, Muhjam Kamza<sup>(3)</sup>

<sup>1,2,3</sup>Universitas Syiah Kuala, Darussalam, Banda Aceh, Indonesia  
Email: [irma.yusriani99@gmail.com](mailto:irma.yusriani99@gmail.com), [mawardiumar@unsyiah.ac.id](mailto:mawardiumar@unsyiah.ac.id),  
[muhjam.kamza@unsyiah.ac.id](mailto:muhjam.kamza@unsyiah.ac.id),

**Abstrak:** Penelitian ini bermaksud untuk mendapati kualitas pertanyaan-pertanyaan ujian semester pelajaran sejarah kelas XII SMA Negeri 1 Aceh Barat Daya pada masa pandemi covid-19 yang ditinjau dari segi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Penelitian ini merupakan penelitian deskriptif kuantitatif. Subjek dalam penelitian ini adalah siswa kelas XII sebanyak 149 siswa dan arah dalam pengkajian ini ialah pertanyaan yang terdiri dari 30 butir, kunci jawaban serta jawaban siswa. Cara pengumpulan data yang digunakan ialah dokumentasi. Pemecahan data dikerjakan dengan berbantuan aplikasi Anates V.4.09. Pengkajian ini memperlihatkan hasil data secara keseluruhan soal ujian semester pelajaran sejarah kelas XII SMA Negeri 1 Aceh Barat Daya pada masa pandemi covid-19 belum berkualitas baik karena 56,6% soal yang dibuat oleh guru masih belum baik, selebihnya soal yang berkualitas baik hanya 20% dan yang kurang baik 23,3%.

---

### Tersedia Online di

[http://journal.unublitar.ac.id/pendidikan/index.php/Riset\\_Konseptual](http://journal.unublitar.ac.id/pendidikan/index.php/Riset_Konseptual)

---

### Sejarah Artikel

Diterima pada : 16-09-2021

Disetujui pada : 30-10-2021

Dipublikasikan pada : 31-10-2021

---

### Kata Kunci:

Analisis Butir Soal, Sejarah, Pandemi,

### DOI:

[http://doi.org/10.28926/riset\\_konseptual.v5i4.412](http://doi.org/10.28926/riset_konseptual.v5i4.412)

---

## PENDAHULUAN

Kegiatan belajar mengajar dapat disebut kombinasi antara kegiatan belajar oleh peserta didik dan mengajar oleh pendidik. Artinya peserta didik dan pendidik mempunyai peran dan kegunaannya masing-masing. Pada kegiatan ini terjadi pengiriman ilmu dari pendidik ke peserta didik. Idealnya siswa menerima setiap ilmu pengetahuan yang disampaikan oleh guru (Santoso & Subagyo, 2017).

Pada kegiatan belajar mengajar ada banyak hal yang mesti diamati oleh pendidik, salah satunya ialah evaluasi pembelajaran. Evaluasi dapat diartikan sebagai kegiatan pengumpulan data guna melihat seberapa besar tujuan pendidikan telah tercapai (Arikunto, 2009). Kegiatan evaluasi dimaksudkan untuk mendiagnosis kesulitan belajar siswa dan hasil diagnosis tersebut nantinya dapat menjadi acuan untuk memperbaiki kekurangan dalam proses belajar mengajar sebelumnya (Thoha, 2003). Evaluasi pembelajaran siswa bisa dilakukan dalam bentuk tes dan non tes (Wahidmurni et al., 2010). Ujian semester merupakan kegiatan evaluasi pembelajaran dalam bentuk tes. Tes menjadi suatu hal untuk menentukan sebesar apa potensi individu dalam merespon pertanyaan yang disuguhkan kepadanya (Mardapi, 2008). Ujian semester merupakan tes yang dilaksanakan dalam ujung semester baik ganjil maupun genap dengan semua materi yang telah dipelajari, maka dari itu dibutuhkan soal tes yang berkualitas baik.

Berlandaskan hasil tanya jawab dengan kepala sekolah SMA Negeri 1 Aceh Barat Daya, kegiatan pembelajaran dari akhir tahun ajaran 2019/2020 diterapkan secara daring (dalam jaringan), sebab menimbang situasi dan kondisi yang ada akibat pandemi covid-19 (coronavirus disease 2019). Pembelajaran daring ini dilakukan sesuai dengan Surat Edaran Gubernur Provinsi Aceh Nomor 440/4989 Tahun 2020 Tentang Pelaksanaan Kegiatan Belajar Mengajar di Rumah. Pembelajaran daring

berlangsung sampai awal tahun ajaran baru 2020/2021, selanjutnya penerapan sistem pembelajaran kembali berubah menjadi sistem shift yang berarti pembelajaran dilakukan secara bergantian antara pembelajaran tatap muka dan daring. Sistem shift ini diterapkan karena kondisi pandemi yang sudah mereda dan keadaan sudah memasuki masa new normal (tatanan normal baru). Proses pelaksanaannya sistem shift sudah dirancang sedemikian rupa agar dalam satu minggu kegiatan pembelajaran hanya boleh diikuti oleh 50% siswa dalam satu kelas, sedangkan 50% nya lagi akan mengikuti pembelajaran secara daring. Rancangan ini diterapkan guna mematuhi protokol kesehatan sesuai dengan surat edaran yang dikeluarkan oleh Gubernur Provinsi Aceh Nomor 440/7715 Tahun 2020 Tentang Pencegahan Penyebaran Covid-19 Dalam Tatanan Normal Baru, yang kemudian disesuaikan dengan daerah dan sekolah masing-masing.

Namun dalam pelaksanaannya, sistem shift ini memiliki beberapa kendala yang dialami oleh guru dan juga siswa. Pada pembelajaran tatap muka, kendala yang dialami tertuju pada jam pelajaran yang dikurangi, biasanya satu jam pelajaran berlangsung selama 45 menit dikurangi menjadi 25 sampai 30 menit, hal ini membuat guru sedikit kesulitan menyampaikan materi ajar karena keterbatasan waktu yang dimiliki. Sebaliknya, pada saat pembelajaran daring, kendala-kendala yang dialami tertuju pada tidak adanya alat penunjang pembelajaran seperti gadget/laptop bagi sebagian kecil siswa, keterbatasan kuota internet, dan gangguan jaringan. Hal-hal tersebut menyebabkan siswa sulit untuk menerima dan menyerap materi pelajaran yang berikan guru, sehingga hal seperti ini nantinya akan berdampak pada kegiatan evaluasi hasil belajar, terutama ujian semester. Melihat kondisi yang seperti ini peneliti akhirnya tertarik untuk membuat analisis atas butir soal ujian semester pada masa pandemi guna melihat keunggulan soal tersebut, apakah sudah baik atau belum.

Kualitas suatu soal dapat diketahui dengan cara membuat analisis terhadap soal tersebut. Analisis butir soal dapat dikatakan sebuah aktivitas guna menguji mutu suatu soal sehingga karakteristik soal tersebut dapat diketahui (Sudjana, 2009; Yusrizal, 2016). Maksud analisis butir soal ialah guna mengetahui manakah pertanyaan tes yang baik, kurang baik, dan yang buruk, maka kedepannya bisa diperbaiki (Daryanto, 2010; Purwanto, 2004). Adapun cara yang dilakukan untuk menganalisis butir soal yaitu dengan mencari validitas, reliabilitas, taraf kesukaran, daya pembeda, dan efektivitas pengecoh. Soal yang baik ialah soal yang terbukti valid, memiliki tingkat reliabilitas yang tinggi, taraf kesukaran yang sedang, daya pembeda yang tinggi dan pengecoh harus berfungsi dengan baik.

## METODE

Pendekatan yang dilakukan dalam penelitian ini ialah pendekatan kuantitatif dengan jenis penelitiannya yaitu deskriptif. Penelitian deskriptif merupakan penelitian yang dimaksudkan untuk eksplorasi dan klarifikasi mengenai suatu fenomena atau kenyataan dengan cara mendeskripsikan sejumlah variabel yang diteliti, pengolahan dan analisis data biasanya menggunakan pengolahan statistik yang bersifat deskriptif (Mulyadi, 2011).

Penelitian ini menggunakan tabel Issac dan Michael untuk menentukan jumlah sampel. Sampel berasal dari populasi siswa kelas XII SMA Negeri 1 Aceh Barat Daya yang terdiri dari 8 kelas dengan jumlah keseluruhan siswa sebanyak 257 orang. Dengan menggunakan tabel tersebut, diperoleh jumlah sampel sebanyak 149 siswa dari jumlah populasi dengan tingkat kesalahan yang dikehendaki adalah 5%. Teknik pengambilan sampel sendiri menggunakan teknik probability sampling dengan jenis simple random sampling, yang dalam artian setiap anggota populasi berkesempatan untuk terpilih sebagai anggota sampel dan pemilihan dilakukan secara acak (Sugiyono, 2018).

Metode pengumpulan data yang digunakan peneliti pada penelitian ini ialah dokumentasi. Dokumentasi merupakan metode pengumpulan data yang bersumber dari data tertulis (Arikunto, 2013). Pada penelitian ini metode ini dipakai guna

memperoleh data berupa soal, kunci jawaban dan lembar jawaban siswa. Setelah data didapati maka selanjutnya dilakukan analisis data yang berupa uji validitas, reliabilitas, taraf kesukaran, daya pembeda, dan efektivitas pengecoh. Analisis dilakukan menggunakan bantuan aplikasi Anates V.4.09, yang merupakan program aplikasi analisis soal oleh Drs. Karno M.Pd. dan Yusuf Wibisono, S.T. keunggulan aplikasi ini terletak pada bahasa yang digunakan yaitu bahasa Indonesia sehingga memudahkan pengguna dalam mengaplikasikannya (Arif, 2015).

### HASIL dan PEMBAHASAN

Hasil analisis butir soal ujian semester ganjil pelajaran sejarah kelas XII SMA Negeri 1 Aceh Barat Daya pada masa pandemi covid-19 dilihat dari segi validitas, reliabilitas, taraf kesukaran, daya pembeda, dan efektivitas pengecoh dapat dilihat pada uraian berikut:

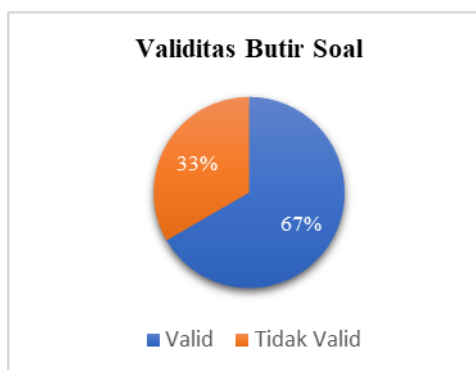
a) Validitas

Butir soal yang ingin diketahui validitasnya dapat digunakan teknik korelasi sebagai teknik analisisnya. Untuk mengetahui validitas butir soal objektif dapat menggunakan rumus korelasi point biserial, lalu hasil perhitungan tersebut dicocokkan dengan nilai r tabel product moment dengan taraf signifikan 5%. Suatu pertanyaan tes bisa disebut valid jika  $r$  hitung  $\geq r$  tabel product moment.

**Tabel 1.** Distribusi Butir Soal Berdasarkan Validitas

Indeks Validitas	Butir Soal	Jumlah	Persentase
$\geq 0,1609$ Valid	1, 2, 6, 10, 11, 12, 13, 14, 16, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30	20	66,6%
$< 0,1609$ Tidak Valid	3, 4, 5, 7, 8, 9, 15, 17, 18, 26	10	33,3%

Jumlah siswa kelas XII SMA Negeri 1 Aceh Barat Daya yang menjadi sampel (N) berjumlah 149 siswa, dan jumlah variabel yang dikorelasikan (nr) adalah 2. Maka dari itu  $n-2$  dari 149 adalah 147, sehingga diperoleh angka r tabel sebesar 0,1609. Berdasarkan hasil analisis yang telah dilakukan, diketahui jumlah pertanyaan tes yang dinyatakan valid sebanyak 20 butir dan pertanyaan tes yang tidak valid sebanyak 10 butir. Distribusi validitas dapat dilihat pada tabel 1 dan gambar 1. Pengujian terhadap validitas memperlihatkan bahwa validitas soal tersebut tergolong tinggi yaitu 66,6%, hal ini sesuai dengan tabel interpretasi koefisien korelasi validitas yang menyatakan bahwa besar koefisien korelasi antara 0,600-0,800 adalah tinggi.



**Gambar 1.** Diagram Butir Soal Berdasarkan Validitas

b) Reliabilitas

Menurut hasil analisis reliabilitas terhadap 30 butir soal, diketahui bahwa soal tersebut memiliki reliabilitas yang cukup dengan hasil perhitungan yang didapati

sebesar 0,40, hal ini sesuai dengan tabel interpretasi nilai reliabilitas yang menunjukkan bahwa besar nilai r antara 0,40-0,69 tergolong cukup.

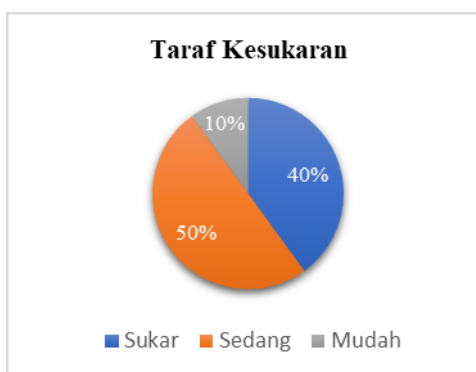
c) Taraf Kesukaran

Pengujian terhadap taraf kesukaran terhadap 30 butir soal, didapati bahwa soal dalam kategori sukar berjumlah 12 butir, soal dalam kategori sedang berjumlah 15 butir, dan soal dalam kategori mudah berjumlah 3 butir. Distribusi tingkat kesukaran 30 butir soal dapat dilihat pada tabel 2 dan gambar 2.

Menurut pengujian tersebut, terlihat bahwa pertanyaan tes tersebut masih belum seimbang. Sebaiknya skala antara taraf kesukaran menyebar dengan seimbang yaitu 25% pertanyaan sukar, 50% pertanyaan sedang dan 25% pertanyaan mudah; atau 20% pertanyaan sukar, 60% pertanyaan sedang dan 20% pertanyaan mudah; atau 15% pertanyaan sukar, 70% pertanyaan sedang dan 15% pertanyaan mudah (Arifin, 2016).

**Tabel 2.** Distribusi Pertanyaan Tes Berdasarkan Taraf Kesukaran

Taraf Kesukaran	Butir Soal	Jumlah	Persentase
Sukar 0,00 – 0,30	3, 4, 6, 10, 12, 15, 18, 22, 26, 27, 29, 30	12	40%
Sedang 0,31 – 0,70	2, 5, 7, 8, 9, 11, 13, 14, 16, 19, 20, 21, 23, 24, 28	15	50%
Mudah 0,71 – 1,00	1, 17, 25	3	10%



**Gambar 2.** Diagram Pertanyaan Tes Berdasarkan Taraf Kesukaran

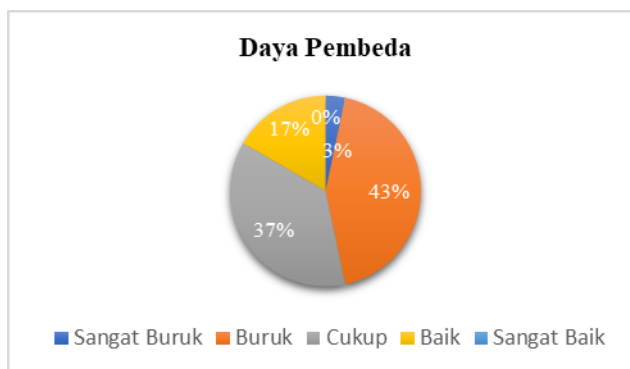
d) Daya Pembeda

Daya pembeda merupakan kemampuan butir soal untuk membedakan antara kelompok atas dan kelompok bawah. Soal yang baik ialah soal yang memiliki daya pembeda yang tinggi. Berlandaskan pengujian yang sudah dilakukan, terlihat 1 pertanyaan tes bertanda negatif yang berarti soal tersebut sangat buruk karena tanda negatif muncul apabila sebuah pertanyaan tes terbalik menunjukkan kualitas siswa, yaitu kelompok atas dikatakan bodoh dan kelompok bawah dikatakan pandai (Arikunto, 2009). Pertanyaan tes yang mempunyai daya pembeda yang buruk sebanyak 13 butir, daya pembeda yang cukup 11 butir, daya pembeda yang baik 5 butir. Pengujian terhadap 30 butir pertanyaan tes yang sudah dilakukan, tidak ditemukan pertanyaan tes yang sangat baik daya pembedanya, hal ini bisa dilihat pada tabel 3 dan gambar 3.

**Tabel 3.** Distribusi Pertanyaan Tes Berdasarkan Daya Pembeda

Daya Pembeda	Butir Soal	Jumlah	Persentase
Sangat buruk (Negatif)	17	1	3,3%
Buruk ( $D < 0,20$ )	3, 4, 5, 8, 9, 10, 12, 15, 18, 22, 26, 27, 29	13	43,3%

Daya Pembeda	Butir Soal	Jumlah	Persentase
Cukup (0,20 – 0,40)	1, 2, 6, 7, 13, 20, 21, 23, 24, 25, 30	11	36,6%
Baik (0,41 – 0,70)	11, 14, 16, 19, 28	5	16,6%
Sangat baik (0,70 – 1,00)	–	0	0%



**Gambar 3.** Diagram Butir Soal Berdasarkan Daya Pembeda

e) Efektivitas Pengecoh

Opsi Jawaban pada pertanyaan tes yang berbentuk pilihan ganda disebut berfungsi dengan baik jika dipilih secara merata oleh peserta tes yang menjawab salah. Butir soal yang mempunyai pengecoh sangat baik jika seluruh pengecoh tersebut berfungsi, butir soal disebut baik apabila hanya ada satu pengecoh yang tidak berfungsi, butir soal disebut buruk jika ditemui dua pengecoh yang tidak berfungsi, dan butir soal yang disebut mempunyai pengecoh yang sangat buruk jika tiga atau seluruh pengecoh yang tidak berfungsi (Hamzah, 2014).

**Tabel 4.** Distribusi Butir Soal Berdasarkan Pengecoh

Kriteria Pengecoh	Butir Soal	Jumlah	Persentase
Sangat baik	2, 6, 8, 14, 28, 29	6	20%
Baik	4, 7, 10, 11, 16, 18, 24	7	23,3%
Buruk	3, 5, 9, 13, 19, 23, 25, 27	8	26,6%
Sangat Buruk	1, 12, 15, 17, 20, 21, 22, 26, 30	9	30%

Berdasarkan hasil analisis yang sudah dilakukan, didapati bahwa pertanyaan tes yang mempunyai pengecoh sangat baik sebanyak 6 butir, pengecoh yang baik sebanyak 7 butir, pengecoh yang buruk 8 butir dan pengecoh yang sangat buruk 9 butir. Distribusi 30 butir soal berdasarkan efektivitas pengecoh bisa dilihat pada tabel 4 dan gambar 4.



**Gambar 4.** Diagram Butir Soal Berdasarkan Pengecoh

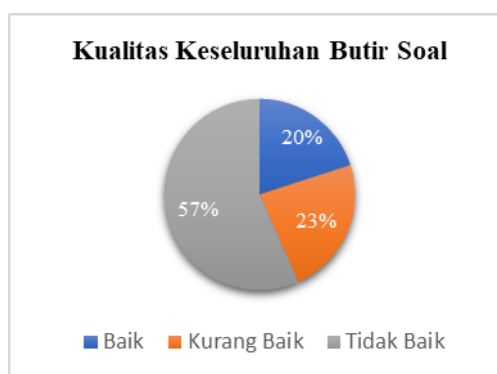
Selepas melakukan analisis per kriteria, selanjutnya pertanyaan tes tersebut dianalisis secara keseluruhan untuk dapat menentukan kualitas pertanyaan tes yang dipakai. Cara menentukan kualitas suatu soal dapat dilihat dari beberapa

pertimbangan penentuan kualitas soal yaitu: (1) Pertanyaan tes dapat disebut mempunyai kualitas yang baik jika pertanyaan tes itu melengkapi empat kriteria yaitu soal tersebut tergolong kategori valid, taraf kesukaran sedang, daya pembeda dan efektivitas pengecoh tergolong kategori baik sekali, baik, dan cukup. Kemudian soal ini layak disimpan ke bank soal; (2) suatu soal dapat disebut mempunyai kualitas kurang baik jika pertanyaan tes itu mencukupi tiga dari empat kriteria, dengan demikian pertanyaan tes tersebut harus direvisi; (3) Pertanyaan tes dikatakan mempunyai kualitas yang tidak baik jika soal itu hanya mampu mencukupi dua kriteria atau kurang, maka dari itu soal itu dapat dibuang dan digantikan dengan soal yang baru.

Berdasarkan hasil perhitungan keseluruhan kualitas pertanyaan tes, diketahui bahwa pertanyaan tes yang baik dan memenuhi semua kriteria berjumlah 6 butir, soal yang cukup baik atau hanya memenuhi tiga kriteria sebanyak 7 butir, dan pertanyaan tes yang tidak baik sebanyak 17 butir. Adapun distribusi butir soal secara keseluruhan bisa dilihat pada tabel 5 dan gambar 5.

**Tabel 5.** Hasil Analisis Keseluruhan Butir Soal

Kualitas Butir Soal	Butir Soal	Jumlah	Persentase
Soal yang baik	2, 11, 14, 16, 24, 28	6	20%
Soal yang kurang baik	6, 7, 13, 19, 20, 21, 23	7	23,3%
Soal yang tidak baik	1, 3, 4, 5, 8, 9, 10, 12, 15, 17, 18, 22, 25, 26, 27, 29, 30	17	56,6%



**Diagram 5.** Diagram Analisis Kualitas Butir Soal

Penyebab kegagalan suatu soal ialah karena tidak terpenuhinya satu atau lebih kriteria yang sudah ditetapkan, akibatnya kualitas soal tersebut menjadi kurang baik atau bahkan tidak baik. Penyebab kegagalan butir soal dapat dilihat pada tabel 6.

**Tabel 6.** Penyebab Kegagalan Butir Soal

Penyebab	Butir Soal	Jumlah	Persentase
Validitas (Tidak valid)	3, 4, 5, 7, 8, 9, 15, 17, 18, 26	10	33,3%
Daya pembeda (Buruk dan sangat buruk)	3, 4, 5, 8, 9, 10, 12, 15, 17, 18, 22, 26, 27, 29	14	46,6%
Tingkat kesukaran	1, 3, 4, 6, 10, 12, 15, 17, 18, 22, 25, 26, 27, 29, 30	15	50%
Efektivitas pengecoh (Buruk dan sangat buruk)	1, 3, 5, 9, 12, 13, 15, 17, 19, 20, 21, 22, 23, 25, 26, 27, 30	17	56,6%

Berlandaskan hasil perhitungan yang terdapat dalam tabel 6, bisa dilihat bahwa penyebab kegagalan terbesar terletak pada efektivitas pengecoh soal yaitu sebanyak 56,6%, hal ini menunjukkan bahwa pengecoh pada butir-butir soal tersebut masih belum berfungsi dengan baik karena tidak dipilih merata oleh peserta tes. Penyebab kegagalan kedua terletak pada tingkat kesukaran yaitu sebanyak 50% soal masih terlalu sukar dan mudah. Penyebab kegagalan ketiga ialah daya pembeda soal yang



masih buruk sebanyak 46,6%, hal ini menandakan bahwa soal tersebut belum mampu membedakan antara siswa kelompok atas dan kelompok bawah. Penyebab kegagalan terakhir ialah validitas, terdapat 33,3% soal yang tidak valid, hal ini berarti butir-butir soal itu belum bisa mengukur apa yang seharusnya diukur.

Uraian tersebut merupakan penyebab kegagalan butir soal yang ditelaah dan dibuktikan melalui uji statistic. Selain itu ada juga penyebab lainnya yang didapati dari kegiatan tanya jawab dengan pengampu mata pelajaran sejarah kelas XII SMA Negeri 1 Aceh Barat Daya. Berdasarkan hasil wawancara tersebut diketahui bahwa sebelum adanya pandemi covid-19, soal-soal ujian disusun oleh pihak MGMP (Musyawarah Guru Mata Pelajaran). Ketika adanya pandemi covid-19 maka soal-soal ujian disusun oleh guru mata pelajaran di tiap-tiap sekolah sesuai dengan jenjang kelas yang diajarkan. Hal ini diterapkan sesuai dengan kebijakan dari Dinas Pendidikan dengan tujuan agar tidak terjadinya perkumpulan atau kerumunan dimasa pandemi covid-19. Sesuai dengan hasil wawancara, ternyata soal yang dibuat oleh guru untuk diberikan kepada siswa tidak dianalisis terlebih dahulu. Sebagaimana yang telah disampaikan oleh guru mata pelajaran sejarah bahwa sebenarnya yang menjadi patokan nilai dan kenaikan kelas pada kurikulum 2013 bukan hanya terletak pada ujian semester saja, namun lebih cenderung kepada nilai tugas-tugas harian, sedangkan untuk kelulusan patokan nilai dilihat dari hasil ujian akhir sekolah (UAS). Maka dari itu analisis soal biasanya dilakukan hanya untuk soal UAS saja, sementara untuk ujian semester tergantung kepada guru mata pelajaran masing-masing mau melakukan analisis terhadap soal tersebut atau tidak.

### KESIMPULAN

Setelah dilakukan analisis maka diperoleh kesimpulan bahwa jika dilirik dari sudut per kriteria maka pertanyaan tes tersebut mempunyai validitas yang tinggi dengan persentase sebesar 66,6%, reliabilitas tergolong cukup karena koefisien reliabilitasnya hanya 0,40, tingkat kesukaran soal belum seimbang karena besarnya persentase antara soal yang sukar dan mudah tidak seimbang, soal yang sukar lebih besar yaitu 40% dan mudah hanya 10%, daya pembeda soal tergolong baik, karena butir soal yang memenuhi kriteria sebesar 53% dan efektivitas pengecoh masih belum berfungsi dengan baik dikarenakan sebesar 56,6% soal masih memiliki pengecoh yang buruk. Berdasarkan hasil analisis secara keseluruhan, maka didapati bahwa soal tersebut belum berkualitas baik, hal ini dikarenakan 56,6% soal yang dibuat oleh guru masih tidak baik, selebihnya soal yang berkualitas baik hanya 20% dan yang kurang baik 23,3%.

Dalam kegagalan butir soal terdapat dua faktor penyebab yaitu faktor internal dan faktor eksternal. Faktor internal dibuktikan dari hasil uji statistik dengan hasil yang didapati yaitu terdapat 4 penyebab kegagalan butir soal, yang paling utama ialah efektivitas pengecoh yang buruk sebesar 56,6%, selanjutnya tingkat kesukaran yang masih tergolong sukar dan mudah sebesar 50%, daya pembeda yang buruk sebesar 46,6%, dan soal yang tidak valid sebesar 33,3%. Adapun faktor eksternal yang dibuktikan dari hasil wawancara dengan guru mata pelajaran sejarah, hasil wawancara membuktikan bahwa ternyata soal tersebut belum dianalisis dahulu sebelum diujikan kepada peserta tes.

### DAFTAR RUJUKAN

- Arif, M. (2015). Penerapan Aplikasi Anates Bentuk Soal Pilihan Ganda. *Eduatic - Scientific Journal of Informatics Education*, 1(1), 1–9. <https://doi.org/10.21107/edutic.v1i1.398>
- Arifin, Z. (2016). *Evaluasi Pembelajaran prinsip, teknik dan prosedur*. PT Remaja Rosdakarya.
- Arikunto, S. (2009). *Dasar-Dasar Evaluasi Pendidikan*. PT Bumi aksara.
- Arikunto, S. (2013). *Prosedur Penelitian Suatu Pendekatan Praktik*. Rineka Cipta.

- Daryanto, H. (2010). *Evaluasi Pendidikan* (6th ed.). Rineka Cipta.
- Hamzah, A. (2014). *Evaluasi Pembelajaran Matematika*. Rajawali Pers.
- Mardapi, D. (2008). *Teknik Penyusunan Tes dan Non Tes*. Mitra Cendikia.
- Mulyadi, M. (2011). Penelitian Kuantitatif Dan Kualitatif Serta Pemikiran Dasar Menggabungkannya [Quantitative and Qualitative Research and Basic Rationale to Combine Them]. *Jurnal Studi Komunikasi Dan Media*, 15(1), 128.
- Purwanto, N. (2004). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. PT Remaja Rosdakarya.
- Santoso, H. B., & Subagyo, S. (2017). PENINGKATAN AKTIFITAS DAN HASIL BELAJAR DENGAN METODE PROBLEM BASIC LEARNING (PBL) PADA MATA PELAJARAN TUNE UP MOTOR BENSIN SISWA KELAS XI DI SMK INSAN CENDEKIA TURI SLEMAN TAHUN AJARAN 2015/2016. *TAMAN VOKASI*, 5(1), 40. <https://doi.org/10.30738/jtvok.v5i1.1428>
- Sudjana, N. (2009). *Penilaian Hasil Proses Belajar*. PT Remaja Rosdakarya.
- Sugiyono. (2018). *Metode Penelitian Kuantitatif*. Penerbit Alfabeta.
- Thoha, M. C. (2003). *Teknik Evaluasi Pendidikan*. PT Raja Grafindo.
- Wahidmurni, Mustikawan, A., & Ridho, A. (2010). *Evaluasi Pembelajaran Kompetensi dan Praktik*. Nuha Literia.
- Yusrizal. (2016). *Pengukuran dan Evaluasi Hasil dan Proses Belajar*. Penerbit Pale Media Prima.