

Karakteristik butir soal ujian akhir semester mata pelajaran sejarah

Anis Faridah^{a,1*}

^a Sekolah Menengah Atas Negeri 1 Pangkalpinang, Kota Pangkalpinang, Indonesia

¹ anisf99999@gmail.com*

* Korespondensi Penulis

ARTICLE INFO

Article history

Received December 26, 2021

Revised December 26, 2021

Accepted December 27, 2021

Available Online December 12, 2021

Keywords

Butir Soal

Sejarah

Teori Tes Klasik

Ujian Akhir Semester

Tingkat Kesukaran

ABSTRACT

This research is a study of quantitative descriptive. The purpose of this research is to describe the characteristics of final semester exam items for grade XI in the History subject at SMA N 1 Pangkalpinang using classical test theory approach. The research of the subject were 138 students of class XI in Social Sciences Major. The result of the research shows that final exam questions in the history subject class XI of SMA N 1 Pangkalpinang is proper to use. This shows that from the validity of the items which there are 39 items of questions (97,5%) which are proven empirically valid with a 0.818 reliability coefficient. Other than that, there are 27 items of questions (67,5%) that can fulfill the criteria for the difficulty level, distinguishing power, and distractor function so it can be used directly to measure the student's ability without correction. While 12 items of questions (30%) need to be fixed and 1 item of question (2,5%) is declared to be invalid so it can't be used to measure the student's ability in History Subject.

Penelitian ini merupakan penelitian deskriptif kuantitatif. Tujuan penelitian ini adalah mendeskripsikan karakteristik butir soal ujian akhir semester (UAS) mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang menggunakan pendekatan teori tes klasik. Subjek penelitian berjumlah 138 peserta didik kelas XI jurusan IPS. Hasil penelitian menunjukkan bahwa soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang telah layak digunakan. Hal ini dibuktikan dari validitas butir soal yang mana terdapat 39 butir soal (97,5%) terbukti valid secara empirik dengan koefisien reliabilitas sebesar 0,818. Selain itu terdapat 27 butir soal (67,5%) yang memenuhi kriteria tingkat kesukaran, daya beda, dan keberfungsian distraktor sehingga dapat digunakan langsung untuk mengukur kemampuan peserta didik tanpa perbaikan. Sedangkan sebanyak 12 butir soal (30%) perlu dilakukan perbaikan dan 1 butir soal (2,5%) dinyatakan gugur sehingga tidak dapat digunakan untuk mengukur kemampuan peserta didik pada mata pelajaran Sejarah.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Pendahuluan

Penilaian merupakan komponen penting dalam proses pembelajaran. Hasil penilaian dapat menginformasikan keberhasilan suatu pembelajaran. Sebagaimana ditegaskan oleh Mardapi (2016) bahwa hasil penilaian dapat digunakan untuk mengetahui keberhasilan pembelajaran. Oleh karena itu, pembelajaran dan penilaian tak dapat dipisahkan. Pembelajaran tanpa penilaian dapat menyebabkan keberhasilan suatu pembelajaran tidak terukur sehingga kemampuan peserta didik setelah mengikuti pembelajaran tidak diketahui.

Fungsi penilaian yang begitu penting dalam pembelajaran membuat guru harus mampu melakukan penilaian terhadap peserta didik untuk mengetahui capaian suatu pembelajaran. Sebagaimana yang tertuang pada undang-undang nomor 14 tahun 2005 tentang Guru dan Dosen bahwa tugas guru selain mengajar adalah menilai dan mengevaluasi. Penilaian merupakan penafsiran terhadap hasil suatu pengukuran tentang capaian peserta didik (Danni, Wahyuni, & Tauratiya, 2021). Sedangkan Stiggins & Chappuis (2012) mendefinisikan penilaian sebagai suatu rangkaian pengumpulan informasi capaian peserta didik. Berdasarkan definisi tersebut, dapat disimpulkan bahwa penilaian dilakukan setelah kegiatan pengukuran yang tujuannya adalah untuk mengumpulkan informasi terkait capaian peserta didik setelah mengikuti pembelajaran.

Penilaian tidak dapat dipisahkan dari kegiatan pengukuran. Suwandi (2010) menyatakan bahwa pengukuran dan penilaian merupakan kegiatan yang saling berkaitan. Oleh karena itu, guna melakukan penilaian terhadap capaian peserta didik secara tepat maka perlu melalui proses pengukuran yang baik. Pengukuran yang baik membutuhkan alat ukur atau instrumen yang berkualitas (Danni et al., 2021). Instrumen yang berkualitas akan menghasilkan informasi yang akurat.

Reynolds, Livingston, & Willson (2009) menyebutkan bahwa validitas dan reliabilitas merupakan syarat yang harus terpenuhi oleh instrumen yang berkualitas. Senada dengan hal tersebut, Danni & Tauratiya (2020) mengungkapkan bahwa instrumen yang berkualitas memiliki validitas dan reliabilitas yang baik. Penjelasan tersebut membuktikan bahwa validitas dan reliabilitas menjadi syarat mutlak yang harus dimiliki oleh instrumen yang berkualitas. Oleh karena itu, pengembangan instrumen untuk mengukur kemampuan peserta didik perlu memenuhi syarat validitas dan reliabilitas sehingga dapat menginformasikan keadaan peserta didik secara akurat. Sebagaimana yang diungkapkan oleh Retnawati (2016) bahwa instrumen yang berkualitas selalu menghasilkan informasi akurat sehingga kesalahan pengukuran dapat diminimalisir.

Validitas didefinisikan sebagai ketepatan suatu instrumen dalam mengukur kompetensi yang hendak diukur (Mardapi, 2008; Ramadhan, Mardapi, Prasetyo, & Utomo, 2019; Rindermann & Baumeister, 2015). Ditegaskan pula oleh (Santoso, Kartianom, & Kassymova, 2019) bahwa validitas membahas terkait sejauh mana keakuratan suatu instrumen yang digunakan dalam pengukuran. Berdasarkan definisi yang telah dipaparkan, dapat disimpulkan bahwa validitas merupakan ketepatan suatu instrumen dalam mengukur variabel atau kompetensi. Validitas suatu instrumen dapat diketahui melalui pengujian dengan menghubungkan skor butir dengan skor total menggunakan persamaan korelasi product moment (Yusup, 2018). Nilai koefisien validitas berada pada rentang -1,0 sampai dengan +1,0 (Fraenkel, Wallen, & Hyun, 2012). Sedangkan reliabilitas dapat diartikan sebagai keajegan hasil tes dari suatu instrumen yang diujikan berkali-kali pada siswa yang sama atau dari dua soal yang setara kepada responden yang sama (Sudjana, 2016). Pengestimasian reliabilitas dapat dilakukan menggunakan persamaan Cronbach alpha. Azwar (2015) menambahkan bahwa suatu instrumen dapat dikatakan memenuhi syarat reliabilitas apabila memiliki indeks reliabilitas $>0,7$. Apabila indeks reliabilitas suatu instrumen $<0,7$ maka butir dinyatakan gugur atau direvisi. Hal tersebut ditegaskan oleh Tavakol & Dennick (2011) bahwa instrumen yang memiliki reliabilitas rendah maka dinyatakan gugur atau direvisi.

Sebagai syarat instrumen yang berkualitas maka validitas dan reliabilitas mutlak harus terpenuhi oleh suatu instrumen termasuk perangkat soal buatan pendidik pada ujian akhir semester. Terpenuhinya kriteria validitas dan reliabilitas dapat memberikan jaminan bahwa instrumen memberikan informasi kemampuan peserta didik secara akurat. Selain validitas dan reliabilitas, terdapat pula beberapa kriteria instrumen yang berkualitas khususnya pada instrumen berbentuk tes/soal. Iskandar & Rizal (2018) menyebutkan bahwa kriteria butir soal yang baik meliputi tingkat kesukaran, daya beda, dan keberfungsian distraktor.

Tingkat kesukaran butir soal merupakan keberfungsian suatu butir soal dalam mengidentifikasi peserta didik yang berkemampuan tinggi dan rendah (Danni et al., 2021). Butir soal yang memiliki tingkat kesukaran tergolong terlalu mudah tidak akan efektif dalam mengidentifikasi kemampuan peserta didik begitu pula butir soal yang terlalu sulit. Indeks tingkat kesukaran dapat diketahui dari proporsi peserta tes menjawab benar (Mardapi, 1998). Retnawati (2016) menyatakan bahwa indeks tingkat kesukaran yang ideal berada direntang 0,30 – 0,70. Semakin rendah indeks kesukaran yang dimiliki maka butir soal cenderung semakin sulit, begitu pula sebaliknya semakin tinggi indeks kesukaran maka cenderung semakin sulit.

Daya beda butir soal merupakan kemampuan butir soal dalam membedakan peserta tes berkemampuan tinggi dan rendah (Solichin, 2017). Indeks daya beda butir soal berada direntang 0,00 sampai dengan 1,00. Butir soal yang memiliki indeks kesukaran $>0,30$ dapat diterima dan butir soal dengan indeks $<0,3$ dapat direvisi (Naga, 1992). Azwar (2016) menambahkan bahwa butir soal yang memiliki indeks daya beda negatif menunjukkan bahwa butir tersebut tidak berguna maka dapat dieliminasi.

Kriteria butir soal selanjutnya adalah keberfungsian distraktor. Keberfungsian distraktor merupakan kriteria yang melekat pada soal berbentuk pilihan ganda atau soal yang memiliki pilihan jawaban. Keberfungsian distraktor digunakan untuk mengetahui keefektifan alternatif jawaban dalam mengecoh peserta tes (Azwar, 2016). Pengembangan butir soal yang memiliki alternatif jawaban perlu memperhatikan kehomogenan jawaban pengecoh dengan jawaban yang benar sehingga butir soal sulit dijawab oleh peserta tes. Sebagaimana yang diungkapkan oleh Ratnaningsih & Isfarudi (2013) bahwa butir soal dengan alternatif jawaban hendaknya memiliki distraktor yang cenderung homogen sehingga soal sulit ditebak. Sedangkan alternatif jawaban/distraktor dapat dinyatakan berfungsi apabila dipilih sedikitnya 5% dari jumlah peserta tes (Mahjabeen et al., 2017).

Karakteristik butir soal menjadi standar penentuan butir soal yang baik. Fungsi yang begitu penting, sehingga mengharuskan setiap pengembangan instrumen tes maupun non tes baik dalam penelitian maupun pembelajaran harus memenuhi kriteria butir soal yang baik. Akan tetapi masih banyak pendidik yang mengembangkan soal tanpa melalui tahapan analisis butir soal sehingga kualitas soal buatan pendidik tidak diketahui. Hal ini dapat berdampak pada hasil pengukuran soal tersebut dapat diragukan keakuratannya. Sebagaimana yang ditegaskan oleh Retnawati & Hadi (2014) bahwa butir soal yang digunakan dalam pengukuran adalah butir soal yang ber kriteria baik. Fenomena tersebut terjadi di SMA N 1 Pangkalpinang, khususnya pada pengembangan butir soal UAS mata pelajaran Sejarah kelas XI. Butir soal UAS yang dikembangkan belum melalui analisis butir soal sehingga karakteristik butir soal belum diketahui. Oleh karena itu, perlu dilakukan suatu tindakan untuk mengetahui kualitas butir soal sehingga dapat diketahui mana butir soal yang dapat digunakan dalam pengukuran kemampuan peserta didik dan mana yang tidak dapat digunakan karena tidak memenuhi syarat.

Berdasarkan uraian literatur dan fenomena yang terjadi, maka penelitian ini bertujuan untuk mendeskripsikan karakteristik butir soal UAS mata pelajaran Sejarah kelas XI di SMA N 1 Pangkalpinang. Karakteristik butir soal berupa validitas, reliabilitas, tingkat kesukaran, daya beda, dan keberfungsian distraktor dianalisis menggunakan pendekatan teori tes klasik dengan bantuan software SPSS v.26 dan ITEMAN v.3.

2. Metode

Penelitian ini merupakan penelitian deskriptif kuantitatif. Tujuan penelitian ini adalah untuk menganalisis dan mendeskripsikan karakteristik butir soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang tahun ajaran 2020/2021. Subjek penelitian adalah peserta didik kelas XI jurusan IPS di SMA N 1 Pangkalpinang yang berjumlah 138 peserta didik. Teknik pengumpul data yang digunakan adalah teknik dokumentasi yaitu berupa respon jawaban peserta didik pada UAS mata pelajaran Sejarah tahun ajaran 2020/2021 yang diperoleh dari pendidik. Soal berbentuk pilihan ganda yang berjumlah 40 butir soal dengan 5 alternatif jawaban. Karakteristik butir soal meliputi validitas, reliabilitas, tingkat kesukaran, daya beda, dan keberfungsian distraktor dianalisis menggunakan pendekatan teori tes klasik dengan bantuan software SPSS v.26 dan ITEMAN v.3. Selanjutnya hasil analisis berupa indeks angka dibandingkan dengan kriteria berdasarkan pendekatan teori tes klasik kemudian diinterpretasikan untuk menentukan karakteristik butir soal. Kriteria penentuan putusan terkait karakteristik butir soal secara ringkas ditampilkan pada Tabel 1.

Tabel 1. Kriteria penentuan butir soal

Karakteristik	Indeks	Putusan
Validitas	$r_{hit} > r_{tabel}$ Sig. < 0,05	Valid
	$r_{hit} < r_{tabel}$ Sig. > 0,05	Tidak Valid
Reliabilitas	> 0,70	Reliabel
	< 0,70	Tidak Reliabel
Tingkat kesukaran	0,30 – 0,70	Terima
	0,10 – 0,29	Revisi
	0,70 – 0,89	Revisi
	< 0,10	Tolak
	> 0,90	Tolak
Daya beda	> 0,30	Terima
	0,10 – 0,30	Revisi
	< 0,10	Tolak
Keberfungsian distraktor	> 0,05	Terima
	< 0,05	Tolak

3. Hasil dan Pembahasan

3.1. Hasil

3.1.1. Validitas soal UAS Sejarah

Pembuktian validitas soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang secara empirik menggunakan respon jawaban peserta didik dengan bantuan *software* SPSS v.26 ditampilkan pada Tabel 1.

Tabel 1. Ringkasan Validitas Soal UAS Sejarah

Kriteria	No Butir
$r_{hit} > r_{tabel}$ Sig. < 0,05	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,20,21,22, 23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40
$r_{hit} < r_{tabel}$ Sig. > 0,05	19

3.1.2. Estimasi Reliabilitas

Estimasi reliabilitas dilakukan menggunakan persamaan *Cronbach alpha* yang dianalisis menggunakan bantuan *software* SPSSv.26. Adapun hasil estimasi reliabilitas soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang disajikan pada Tabel 2.

Tabel 2. Hasil Estimasi Reliabilitas Soal UAS Sejarah

Cronbach's Alpha	N of Items
0,818	40

3.1.3. Tingkat Kesukaran

Tingkat kesukaran butir soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang dianalisis menggunakan pendekatan teori tes klasik dengan bantuan *software* ITEMAN v.3 yang diketahui melalui proporsi menjawab benar (*proportion correct*). Adapun ringkasan hasil analisis tingkat kesukaran butir soal ditampilkan pada Tabel 3.

Tabel 3. Indeks Tingkat Kesukaran Butir Soal UAS Sejarah

Indeks	No Butir
0,30 – 0,70	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,20,21,22, 23,24,25,27,28,29,30,31,32,33,34,35,36,37,38,39,40
0,10 – 0,29	-

0,70 – 0,89	-
<0,10	19
>0,90	26

3.1.4. Daya Beda

Ringkasan hasil analisis daya beda butir soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang menggunakan ITEMAN v.3 ditampilkan pada Tabel 4.

Tabel 4. Indeks Daya Beda Butir Soal UAS Sejarah

Indeks	No Butir
>0,30	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,20,21,22,24,25,27,28,29,30,31,32,33,34,35,36,37,38,39,40
0,10 – 0,30	23,26
<0,10	19

3.1.5. Keberfungsian Distraktor

Ringkasan hasil analisis keberfungsian distraktor butir soal UAS mata pelajaran sejarah kelas XI SMA N 1 Pangkalpinang menggunakan ITEMAN v.3 yang diketahui melalui proporsi peserta tes memilih jawaban (*proportion endorsing*). Adapun hasil analisis ditampilkan pada Tabel 5.

Tabel 5. Indeks Keberfungsian Distraktor Butir Soal UAS Sejarah

Indeks	No Butir
>0,50	2,3,5,8,9,10,11,12,13,14,15,16,18,19,20,22,25,27,28,29,31,32,34,35,36,37,38,40
<0,50	1(A),4(B,C), 6(D),7(D,E), 17(A),21(B,C),24(C,D),26(A,C,D),30(B,D),33(A,E),39(A,D)

3.2. Pembahasan

Validitas butir soal UAS mata pelajaran Sejarah yang ditampilkan pada Tabel 1 menunjukkan bahwa 97,5% butir soal memperoleh $r_{hit} > r_{tabel}$ atau $Sig. < 0,05$ sehingga butir-butir tersebut dapat dinyatakan valid dan sisanya 2,5% dinyatakan tidak valid. Sebagaimana yang ditegaskan oleh Arikunto (2013) apabila $r_{hit} > r_{tabel}$ atau $Sig. < 0,05$ maka butir soal dinyatakan valid. Butir soal yang valid dapat memberikan informasi yang akurat sesuai dengan keadaan sebenarnya (Danni et al., 2021). Oleh karena itu butir yang tidak memenuhi syarat validitas dinyatakan gugur dan perlu dieliminasi. Tabel 1 juga menginformasikan bahwa butir soal no 19 setelah dilakukan pengujian secara empirik tidak memenuhi kriteria valid karena memperoleh r_{hitung} sebesar 0,08 dengan nilai signifikansi 0,186 ($sig. > 0,05$) sehingga dinyatakan gugur dan tidak dapat digunakan untuk mengukur kemampuan Sejarah peserta didik.

Estimasi reliabilitas pada soal UAS mata pelajaran Sejarah yang terdiri dari 40 butir soal menggunakan persamaan *cronbach alpha* memperoleh koefisien alpha sebesar 0,818. Hasil tersebut menunjukkan koefisien reliabilitas soal lebih besar dari 0,70 sehingga dapat dinyatakan bahwa soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang telah memenuhi kriteria reliabilitas. Azwar (2016) menegaskan bahwa soal yang memiliki koefisien reliabilitas $>0,70$ maka dinyatakan reliabel. Sejalan dengan hal tersebut, Fisher (2007) menambahkan bahwa koefisien reliabilitas di rentang 0,67 – 0,80 tergolong cukup. Oleh karena itu, perangkat soal mata pelajaran Sejarah buatan pendidik layak untuk digunakan pada ujian akhir semester karena memiliki konsistensi yang baik.

Selain validitas dan reliabilitas, tingkat kesukaran butir soal menjadi salah kriteria butir soal yang berkualitas khususnya pada instrumen berbentuk tes. Hasil analisis tingkat kesukaran butir yang ditampilkan pada Tabel 3 menunjukkan bahwa dari 40 butir soal terdapat 38 butir (95%) yang memiliki indeks kesukaran di rentang 0,30 – 0,70. Berdasarkan hasil tersebut dapat disimpulkan bahwa 38 butir soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang memenuhi kriteria tingkat kesukaran butir soal yang baik. Sebagaimana yang ditegaskan oleh Susetyo (2015)

bahwa indeks kesukaran yang berada pada rentang 0,30 – 0,70 tergolong dalam tingkat kesukaran yang baik. Sedangkan menurut Heri Retnawati (2016) butir soal yang memiliki indeks kesukaran di rentang 0,30 – 0,70 merupakan butir soal yang tingkat kesukarannya tergolong sedang. Butir soal yang memiliki tingkat kesukaran berkategori sedang akan lebih efektif dalam mengukur kemampuan peserta didik dibandingkan dengan butir soal yang tergolong mudah atau sulit (Danni et al., 2021).

Berdasarkan Tabel 3 pula dapat diketahui bahwa terdapat 1 butir soal (2,5%) yang tergolong mudah yaitu butir soal nomor 26 dengan indeks kesukaran sebesar 0,91. Hasil analisis menggunakan *software* ITEMAN v.3 diketahui bahwa terdapat 125 dari 138 peserta didik yang mampu menjawab butir soal nomor 26 dengan benar. Selain itu, terdapat 1 butir soal (2,5%) tergolong sulit yaitu butir nomor 19 yang diketahui hanya mampu dijawab dengan benar oleh 21 dari 138 peserta didik. Dengan demikian dapat disimpulkan bahwa terdapat 95% butir soal yang dapat digunakan untuk mengukur kemampuan peserta didik yaitu butir soal nomor 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40. Sedangkan sisanya sebanyak 2 butir soal (5%) dinyatakan gugur karena tidak memenuhi kriteria tingkat kesukaran yang baik, yaitu butir soal nomor 19 dan 26. Secara keseluruhan, tingkat kesukaran butir soal UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang tergolong sedang dengan rata-rata indeks kesukaran sebesar 0,593.

Tabel 4 menginformasikan bahwa terdapat 37 butir soal (87,5%) yang memiliki indeks daya beda $>0,30$. Berdasarkan hasil tersebut dapat dinyatakan bahwa apabila ditinjau dari daya beda butir soal maka terdapat 37 butir soal yang memenuhi kriteria daya beda butir soal. Dengan demikian butir soal tersebut dapat membedakan peserta didik yang berkemampuan tinggi dan rendah (Iskandar & Rizal, 2018). Sedangkan terdapat 2 butir soal (5%) memiliki indeks daya beda direntang 0,10 – 0,29 sehingga butir tersebut perlu dilakukan revisi agar dapat digunakan untuk mengukur kemampuan peserta didik. Adapun butir soal yang perlu direvisi adalah butir soal nomor 23 dan 26. Selain itu, terdapat pula 1 butir soal (2,5%) yang memiliki indeks daya beda $<0,10$ sehingga dinyatakan gugur dan tidak dapat digunakan untuk mengukur kemampuan peserta didik. Butir soal yang dinyatakan gugur adalah butir soal nomor 19.

Tabel 5 menunjukkan bahwa terdapat alternatif jawaban atau distraktor yang tidak berfungsi dengan baik. Hal ini ditunjukkan dari indeks proporsi peserta tes memilih jawaban (*proportion endorsing*) $<0,05$ yang artinya bermakna peserta tes yang terkecoh dengan distraktor tidak sampai 5%. Oleh karena itu, distraktor yang tidak berfungsi dengan baik perlu diperbaiki. Perbaikan distraktor dapat dilakukan melalui penggantian distraktor dengan alternatif jawaban lain (Wibawa, 2019). Mardapi (2008) menjelaskan bahwa dalam menyusun distraktor perlu memperhatikan kehomogenan atau kesamaan antara distraktor dengan kunci jawaban. Selain itu, distraktor juga harus logis agar memperbesar peluang peserta didik terkecoh.

Berdasarkan hasil dan pembahasan yang telah diuraikan, maka dapat dirangkum putusan terkait diterima atau ditolaknya butir soal UAS mata pelajaran kelas XI SMA N 1 Pangkalpinang ditinjau dari karakteristik butir yang baik berdasarkan pendekatan teori tes klasik. Adapun rangkuman putusan ditampilkan pada Tabel 6.

Tabel 6. Rangkuman Putusan Butir Soal UAS Sejarah Kelas XI SMA N 1 Pangkalpinang

Nomor Butir	Jumlah (%)	Putusan
2,3,5,8,9,10,11,12,13,14,15,16,18,20,22,25,27,28,29,31,32,34,35,36,37,38,40	27 (67,5%)	Terima
1,4,6,7,17,21,23,24,26,30,33,39	12 (30%)	Revisi
19	1 (2,5%)	Tolak

4. Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dipaparkan, maka dapat disimpulkan bahwa soal yang dikembangkan oleh pendidik untuk UAS mata pelajaran Sejarah kelas XI SMA N 1 Pangkalpinang telah layak digunakan. Hal ini ditunjukkan dari validitas butir soal yang mana terdapat 39 butir soal (97,5%) terbukti valid secara empirik, reliabilitas soal tergolong baik dengan koefisien sebesar 0,818. Apabila ditinjau dari karakteristik butir soal, maka sebanyak 27 butir soal (67,5%) diterima tanpa revisi sehingga dapat digunakan langsung untuk mengukur kemampuan peserta didik

karena telah memenuhi kriteria tingkat kesukaran, daya beda, dan keberfungsian distraktor. Sedangkan sebanyak 12 butir soal (30%) perlu dilakukan perbaikan atau revisi dan 1 butir soal (2,5%) dinyatakan gugur sehingga tidak dapat digunakan untuk mengukur kemampuan peserta didik pada mata pelajaran Sejarah.

Daftar Pustaka

- Arikunto, S. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Azwar, S. (2015). *Reliabilitas dan Validitas* (4th ed.). Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2016). *Konstruksi tes kemampuan kognitif*. Yogyakarta: Pustaka Pelajar.
- Danni, R., & Tauratiya, T. (2020). Analisis Kemampuan Berpikir Kritis Mahasiswa Program Studi Hukum Keluarga Islam IAIN Syaikh Abdurrahman Siddik Bangka Belitung. *Tarbawy : Jurnal Pendidikan Islam*, 7(1), 17–22. <https://doi.org/10.32923/tarbawy.v7i1.1191>
- Danni, R., Wahyuni, A., & Tauratiya, T. (2021). Item Response Theory Approach: Kalibrasi Butir Soal Penilaian Akhir Semester Mata Pelajaran Bahasa Arab. *Arabi : Journal of Arabic Studies*, 6(1), 93–104. <https://doi.org/10.24865/AJAS.V6I1.320>
- Fisher, W. (2007). Rating scale instrumen quality criteria. *Rasch Measurement Transactions*, 21(1), 1095.
- Fraenkel, J. L., Wallen, N. E., & Hyun, H. H. . (2012). *How to design and evaluate research in education* (8th ed.). New York: Mc Graw-Hill Company.
- Iskandar, A., & Rizal, M. (2018). Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12–23. <https://doi.org/https://doi.org/10.21831/pep.v22i1.15609>
- Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. *Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University*, 13(4), 310–315. Retrieved from <https://www.apims.net/index.php/apims/article/view/9>
- Mardapi, D. (1998). Analisis Butir dengan Teori Tes Klasik dan Teori Respons Butir. *Jurnal Kependidikan*, 28, 25–34.
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendikia Offset.
- Mardapi, D. (2016). *Pengukuran penilaian dan evaluasi pendidikan* (2nd ed.). Yogyakarta: Nuha Litera.
- Naga, D. S. (1992). *Pengantar teori sekor pada pengukuran pendidikan*. Jakarta: Gunadarma.
- Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The Development of an Instrumen to Measure the Higher Order Thinking Skill in Physics. *European Journal of Educational Research*, 8(3), 743. <https://doi.org/10.12973/eu-jer.8.3.743>
- Ratnaningsih, D. J., & Isfarudi, I. (2013). Analisis Butir Tes Objektif Ujian Akhir Semester Mahasiswa Universitas Terbuka Berdasarkan Teori Tes Modern. *Jurnal Pendidikan Terbuka Dan Jarak Jauh*, 14(2), 98–109. Retrieved from <http://jurnal.ut.ac.id/index.php/jptjj/article/view/402>
- Retnawati, H., & Hadi, S. (2014). Sistem bank soal daerah terkalibrasi untuk menyongsong era desentralisasi. *Jurnal Ilmu Pendidikan*, 20(2), 183–193. <https://doi.org/https://doi.org/10.17977/jip.v20i2.4615>
- Retnawati, Heri. (2016). *Validitas, reliabilitas dan karakteristik butir*. Yogyakarta: Nuha Medika.

- Reynolds, C. R., Livingston, R. B., & Willson, V. (2009). *Measurement and Assessment in Education* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Rindermann, H., & Baumeister, A. E. E. (2015). Validating the interpretations of PISA and TIMSS tasks: A rating study. *International Journal of Testing*, 15(1), 276–296. <https://doi.org/10.1080/15305058.2014.966911>
- Santoso, A., Kartianom, K., & Kassymova, G. K. (2019). Kualitas butir bank soal statistika (Studi kasus: Instrumen ujian akhir mata kuliah statistika Universitas Terbuka). *Jurnal Riset Pendidikan Matematika*, 6(2), 165–176. <https://doi.org/10.21831/jrpm.v6i2.28900>
- Solichin, M. (2017). Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes dan Validitas Ramalan dalam Evaluasi Pendidikan. *Dirasat: Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192–213. <https://doi.org/10.26594/DIRASAT.V2I2.879>
- Stiggins, R., & Chappuis, J. (2012). *Introduction to student invoved assessment for learning* (6th ed.). Boston: Addison Wesley.
- Sudjana, N. (2016). *Penilaian Hasil Belajar Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya.
- Susetyo, B. (2015). *Prosedur Penyusunan dan Analisis Tes*. Bandung: PT Refika Aditama.
- Suwandi, S. (2010). *Model Assesmen dalam Pembelajaran* (2nd ed.). Surakarta: Yuma Pustaka.
- Tavakol, M., & Dennick, R. (2011, June 27). Making sense of Cronbach's alpha. *International Journal of Medical Education*, Vol. 2, pp. 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Wibawa, E. A. (2019). Karakteristik Butir Soal Tes Ujian Akhir Semester Hukum Bisnis. *Jurnal Pendidikan Akuntansi Indonesia*, 17(1), 86–96. <https://doi.org/10.21831/JPAI.V17I1.26339>
- Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Tarbiyah : Jurnal Ilmiah Kependidikan*, 7(1), 17–23. <https://doi.org/10.18592/TARBIYAH.V7I1.2100>