



Karakteristik Butir Soal Penilaian Akhir Semester Mata Pelajaran Sejarah Kelas XI SMA Negeri 1 Pangkalpinang

Anis Faridah^{1*}

¹ SMA Negeri 1 Pangkalpinang; anisf99999@gmail.com

*Korespondensi

Dikirim: 28-11-2021; Diterima: 26-12-2021; Diterbitkan: 30-12-2021

Abstract: The problem behind this research is the development of final semester assessment questions for history subjects that do not go through the item analysis stages so that the quality of the items is unknown. The purpose of this study was to describe the characteristics of the end-of-semester assessment items for history class XI SMA Negeri 1 Pangkalpinang using a classical test theory approach. This research is quantitative descriptive. The research subjects were 138 students of class XI majoring in social studies. The results showed that the PAS questions for the class XI history subjects of SMA Negeri 1 Pangkalpinang were feasible to use. This is evidenced by the validity of the items in which there are 39 items (97.5%) that are proven to be empirically valid with a reliability coefficient of 0.818. In addition, there are 27 items (67.5%) that meet the criteria for the level of difficulty, discriminating power, and distractor functioning to be used directly to measure students' abilities without improvement. Meanwhile, as many as 12 items (30%) needed to be repaired, and 1 item (2.5%) was declared invalid so that it could not be used to measure the ability of students in historical subjects.

Keywords: characteristics of test items; classical test theory; final semester exam, history

Abstrak: Permasalahan yang melatarbelakangi penelitian ini adalah pengembangan soal penilaian akhir semester mata pelajaran sejarah yang tidak melalui tahapan analisis butir soal sehingga kualitas butir soal tidak diketahui. Tujuan penelitian ini adalah untuk mendeskripsikan karakteristik butir soal penilaian akhir semester mata pelajaran sejarah kelas XI SMA Negeri 1 Pangkalpinang menggunakan pendekatan teori tes klasik. Penelitian ini merupakan penelitian deskriptif kuantitatif. Subjek penelitian berjumlah 138 peserta didik kelas XI jurusan IPS. Hasil penelitian menunjukkan bahwa soal PAS mata pelajaran sejarah kelas XI SMA Negeri 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.

Kata Kunci: karakteristik butir soal; penilaian akhir semester; sejarah; teori tes klasik



Jurnal Fajar Historia is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Pendahuluan

Pembelajaran tidak dapat dipisahkan dengan penilaian. 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 tertulis di undang-undang nomor 14 tahun 2005 bahwa tugas guru selain mengajar adalah menilai dan mengevaluasi. Penilaian ialah penafsiran terhadap hasil suatu pengukuran tentang pencapaian peserta didik (Danni et al., 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 alat ukur yang berkualitas haruslah valid dan reliabel. Penjelasan tersebut membuktikan bahwa validitas dan reliabilitas menjadi syarat mutlak yang harus dimiliki oleh instrumen yang berkualitas. Dengan demikian, alat ukur kemampuan peserta didik perlu memenuhi syarat validitas dan reliabilitas sehingga dapat menginformasikan keadaan peserta didik secara akurat. Sebagaimana yang diungkapkan oleh Retnawati (2016) bahwa alat ukur yang berkualitas memberikan informasi yang akurat sehingga kesalahan pengukuran dapat diminimalisir.

Validitas didefinisikan sebagai keakuratan suatu alat ukur dalam mengukur kompetensi yang diinginkan (Mardapi, 2008; Rindermann & Baumeister, 2015). Ditegaskan pula oleh (Santoso, et al., 2019) bahwa validitas membahas terkait sejauh mana keakuratan suatu instrumen yang digunakan dalam pengukuran. Dengan demikian, validitas dapat diartikan sebagai keakuratan suatu alat ukur dalam mengukur variabel

atau kompetensi. Validitas suatu instrumen dapat diketahui melalui pengujian dengan menghubungkan skor butir dengan skor total (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 paket soal yang tingkatannya setara kemudian diberikan 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 berkategori baik haruslah memenuhi kriteria tingkat kesulitan, daya pembeda, dan keberfungsian distraktor.

Tingkat kesukaran butir soal merupakan keberfungsian suatu butir soal untuk mengidentifikasi siswa yang berkemampuan tinggi dan rendah (Danni et al., 2021). Butir yang tergolong terlalu mudah cenderung tidak akan efektif dalam mengidentifikasi kemampuan peserta didik begitu pula butir yang terlalu sulit. Indeks tingkat kesulitan butir soal dapat diperoleh melalui proporsi testee 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 akan cenderung semakin sukar, begitu pula sebaliknya apabila suatu butir soal semakin tinggi indeks kesukarannya maka cenderung semakin mudah.

Daya pembeda merupakan kecakapan butir soal untuk membedakan antara siswa berkemampuan tinggi dan rendah (Solichin, 2017). Indeks daya beda butir soal berada direntang 0,00 sampai dengan 1,00. Butir soal yang teridentifikasi memperoleh indeks kesukaran $>0,30$ dapat diterima sedangkan yang memperoleh indeks $<0,3$ dapat direvisi (Naga, 1992). Azwar (2016) menambahkan apabila suatu butir soal memperoleh indeks daya beda negatif maka hal tersebut 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 baik jika sedikitnya 5% peserta tes memilih distraktor tersebut (Mahjabeen et al., 2017).

Karakteristik butir soal menjadi standar penentuan butir soal yang baik (Sarjana, 2018). 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 yang layak dipakai adalah butir soal yang berkriteria baik. Fenomena tersebut terjadi di SMA Negeri 1 Pangkalpinang, khususnya pada pengembangan butir soal Penilaian Akhir Semester (PAS) mata pelajaran sejarah kelas XI. Oleh sebab itu, perlu dilakukan suatu tindakan untuk mengetahui kualitas soal. Butir soal PAS yang dikembangkan tanpa melalui tahapan analisis butir soal akan berdampak pada tidak diketahuinya kualitas butir soal sehingga hasil pengukuran menjadi diragukan kesahihannya dan diketahui mana butir yang dapat dipakai dalam pengukuran kemampuan siswa dan mana yang tidak layak digunakan.

Hasil penelitian yang dilakukan oleh Kholis (2017) menunjukkan bahwa tingkat kesukaran soal yang ada dalam buku Sejarah Kebudayaan Islam kurikulum 2013 kelas IX sebanyak 25% tergolong mudah, 35% tergolong sedang, dan 40% tergolong sukar. Penelitian tersebut fokus pada soal-soal yang ada di buku Sejarah Kebudayaan Islam kelas IX sedangkan pada penelitian ini membahas tidak hanya tingkat kesukaran soal melainkan daya pembeda dan keberfungsian distraktor juga menjadi karakteristik butir yang dianalisis. Selain itu pada penelitian ini butir soal yang dianalisis adalah butir soal buatan guru yang digunakan pada PAS mata pelajaran sejarah kelas XI di SMA Negeri 1 Pangkalpinang.

Selanjutnya, penelitian yang dilakukan oleh Yuslita, Zulfan, & Arifin (2016) membahas mengenai tingkat kesukaran dan daya beda butir soal mata pelajaran sejarah di SMA Negeri 5 Banda Aceh menunjukkan bahwa soal yang dikembangkan oleh guru sebanyak 60% tergolong jelek. Hal ini menunjukkan bahwa perlu adanya penguatan kompetensi guru dalam pembuatan butir soal. Selain itu, penelitian tersebut hanya membahas mengenai tingkat kesukaran dan daya beda sedangkan penelitian ini menganalisis tingkat kesukaran, daya beda, dan butir soal buatan guru mata pelajaran sejarah di SMA Negeri 1 Pangkalpinang.

Penelitian yang dilakukan oleh Putri & Ofianto (2019) mengenai efektivitas

penggunaan aplikasi Anajohn, Anates, dan Iteman untuk menganalisis butir soal. Hasil penelitian tersebut menunjukkan bahwa aplikasi yang efektif adalah Iteman. Hasil penelitian tersebut adalah membandingkan tiga aplikasi, sedangkan penelitian ini adalah untuk menganalisis karakteristik butir soal menggunakan salah satu dari tiga aplikasi tersebut. Akan tetapi hasil penelitian Putri & Ofianto (2019) tersebut menguatkan peneliti untuk menggunakan aplikasi Iteman guna menganalisis karakteristik butir soal PAS mata pelajaran sejarah di SMA Negeri 1 Pangkalpinang.

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

Metode Penelitian

Penelitian ini tergolong ke dalam penelitian deskriptif kuantitatif. Tujuan penelitian ini adalah untuk mendeskripsikan karakteristik butir soal PAS mata pelajaran sejarah kelas XI SMA Negeri 1 Pangkalpinang tahun ajaran 2020/2021. Subjek penelitian adalah siswa kelas XI jurusan IPS di SMA Negeri 1 Pangkalpinang yang berjumlah 138 orang. Teknik pengumpul data berupa teknik dokumentasi yaitu berupa respon jawaban siswa pada PAS mata pelajaran sejarah tahun ajaran 2020/2021 yang diperoleh dari pendidik. Soal yang dianalisis berbentuk pilihan ganda terdiri dari 40 butir soal dengan 5 alternatif jawaban. Karakteristik butir soal meliputi validitas, reliabilitas, tingkat kesukaran, daya beda, dan keberfungsian distraktor dianalisis menggunakan 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 guna 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
Daya beda	>0,90	Tolak
	>0,30	Terima

Karakteristik	Indeks	Putusan
	0,10 – 0,30	Revisi
	<0,10	Tolak
Keberfungsian distraktor	>0,05	Terima
	<0,05	Tolak

Hasil Penelitian

Hasil analisis butir soal menggunakan pendekatan teori tes klasik dengan bantuan *software* SPSS v.26 dan ITEMAN v.3 dijelaskan di bawah ini.

Validitas Butir Soal PAS Sejarah

Pembuktian validitas soal PAS mata pelajaran sejarah kelas XI SMA Negeri 1 Pangkalpinang secara empirik menggunakan respon jawaban peserta didik dengan bantuan *software* SPSS v.26 ditampilkan pada tabel 2.

Tabel 2. Ringkasan Validitas Soal PAS 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

Estimasi Reliabilitas Soal

Estimasi reliabilitas dilakukan menggunakan persamaan *cronbach alpha* yang dianalisis menggunakan bantuan *software* SPSSv.26. Adapun hasil estimasi reliabilitas soal PAS mata pelajaran sejarah kelas XI SMA Negeri 1 Pangkalpinang ditampilkan pada tabel 3.

Tabel 3. Hasil Estimasi Reliabilitas Soal PAS Sejarah

Cronbach's Alpha	N of Items
0,818	40

Tingkat Kesukaran Butir Soal

Tingkat kesukaran butir PAS Sejarah kelas XI SMA Negeri 1 Pangkalpinang dianalisis menggunakan pendekatan teori tes klasik dengan bantuan *software* ITEMAN v.3 yang diketahui melalui proporsi menjawab benar (*proportion correct*). Adapun hasil analisis tingkat kesulitan butir soal secara ringkas ditampilkan pada tabel 4.

Tabel 4. Indeks Tingkat Kesukaran Butir Soal PAS 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

Daya Beda Butir Soal

Ringkasan hasil analisis daya beda butir soal PAS mata pelajaran sejarah kelas XI SMA Negeri 1 Pangkalpinang menggunakan ITEMAN v.3 ditampilkan pada tabel 5.

Tabel 5. Indeks Daya Beda Butir Soal PAS 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

Keberfungsian Distraktor

Ringkasan hasil analisis keberfungsian distraktor butir soal PAS mata pelajaran sejarah kelas XI SMA Negeri 1 Pangkalpinang menggunakan ITEMAN v.3 yang diketahui melalui proporsi peserta tes memilih jawaban (*proportion endorsing*). Adapun hasil analisis keberfungsian distraktor ditunjukkan pada tabel 6.

Tabel 6. Indeks Keberfungsian Distraktor Butir Soal PAS 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)

Pembahasan

Validitas butir soal PAS mata pelajaran sejarah yang ditampilkan pada tabel 2 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 dapat dikatakan memenuhi kriteria validitas. Butir soal yang telah memenuhi kriteria validitas 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 2 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 layak untuk dipakai dalam mengukur kemampuan Sejarah siswa kelas XI SMA Negeri 1 Pangkalpinang.

Estimasi reliabilitas soal PAS mata pelajaran sejarah menggunakan persamaan *cronbach alpha* memperoleh koefisien alpha sebesar 0,818. Hasil tersebut menunjukkan bahwa koefisien reliabilitas soal yang diperoleh $> 0,70$ sehingga dapat dinyatakan bahwa soal PAS Sejarah kelas XI SMA Negeri 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.

Tingkat kesukaran menjadi salah kriteria butir soal yang berkualitas khususnya pada instrumen berbentuk tes. Tabel 4 menginformasikan bahwa dari 40 butir soal terdapat 38 butir (95%) yang memiliki indeks kesukaran di rentang 0,30 – 0,70. Berdasarkan sebaran tingkat kesukaran butir soal dapat disimpulkan bahwa 38 butir soal PAS Sejarah kelas XI SMA Negeri 1 Pangkalpinang memenuhi kriteria tingkat kesukaran butir soal yang baik. Sebagaimana yang ditegaskan oleh Susetyo (2015) bahwa indeks kesukaran di rentang 0,30 – 0,70 tergolong dalam tingkat kesukaran yang baik. Sedangkan menurut Heri Retnawati (2016) butir soal dengan indeks kesukaran di rentang 0,30 – 0,70 merupakan butir soal yang tingkat kesukarannya tergolong sedang. Butir soal yang memiliki tingkat kesulitan berkategori sedang akan lebih efektif untuk mengukur kemampuan siswa dibandingkan dengan butir yang berkategori mudah atau sulit (Danni et al., 2021).

Berdasarkan tabel 4 pula dapat diketahui bahwa ada 1 butir (2,5%) yang tergolong mudah yaitu butir 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 terdapat 95% butir yang layak untuk digunakan dalam mengukur kemampuan siswa yaitu butir 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 kesulitan soal PAS Sejarah kelas XI SMA Negeri 1 Pangkalpinang tergolong sedang dengan rata-rata indeks kesukaran sebesar 0,593.

Tabel 5 menginformasikan ada 37 butir soal (87,5%) dengan indeks daya pembeda $>0,30$. Dengan demikian dapat dikatakan bahwa ada 37 butir soal yang memenuhi kriteria daya pembeda. Butir soal yang memenuhi kriteria daya pembeda menandakan bahwa butir mampu 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 pembeda $<0,10$ sehingga dinyatakan gugur dan tidak layak dipakai guna mengukur kemampuan siswa. Butir soal yang dinyatakan gugur adalah butir soal nomor 19.

Tabel 6 menunjukkan bahwa ada distraktor yang tidak berfungsi. 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 pergantian 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 PAS Sejarah kelas XI SMA Negeri 1 Pangkalpinang ditinjau dari karakteristik butir berdasarkan pendekatan teori tes klasik. Adapun rangkuman putusan ditampilkan pada tabel 7.

Tabel 7. Rangkuman Putusan Butir Soal UAS Sejarah Kelas XI SMA Negeri 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

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dipaparkan, maka dapat disimpulkan bahwa soal yang dikembangkan oleh pendidik untuk PAS mata pelajaran sejarah kelas XI SMA Negeri 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 tingkat kesulitan, daya pembeda, dan keberfungsian distraktor maka sebanyak 27 butir soal (67,5%) diterima tanpa revisi sehingga layak dipakai langsung untuk mengukur kemampuan kognitif peserta didik pada mata pelajaran sejarah 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 layak untuk mengukur kemampuan siswa pada mata pelajaran sejarah. Dengan demikian, perlu adanya perbaikan dalam proses pengembangan butir soal khususnya melalui tahapan analisis butir soal agar soal yang dikembangkan oleh guru benar-benar berkualitas dan efektif untuk mengukur kemampuan peserta didik.

References

- 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 instrument 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>.
- Kholis, R. A. N. (2017). Analisis Tingkat Kesulitan (Difficulty Level) Soal Pada Buku Sejarah Kebudayaan Islam Kurikulum 2013. *Jurnal Pendidikan Agama Islam*, 14(2), 305–315. <https://doi.org/10.14421/JPAI.2017.142-07>
- 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 sekur pada pengukuran pendidikan*. Jakarta: Gunadarma.
- Putri, R. H., & Ofianto, O. (2019). Efektivitas Analisis Butir Menggunakan Anajohn, Anates Dan Iteman Studi Soal Usbn Pelajaran Sejarah Kota Padang. *Jurnal Kronologi*, 1(2), 1–11. Retrieved from <http://kronologi.pppj.unp.ac.id/index.php/jk/article/view/7>.
- Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The Development of an Instrument 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>.
- Sarjana, A. (2018). Penerapan Pembelajaran Kooperatif Berbasis Masalah dengan Model Jigsaw untuk Meningkatkan Prestasi Belajar Siswa Pada Mata Pelajaran Sejarah Indonesia Kelas XI di SMAN 4 Praya Tahun Pelajaran 2018-2019. *Fajar Historia: Jurnal Ilmu Sejarah Dan Pendidikan*, 2(2), 106–117. <https://doi.org/10.29408/FHS.V2I2.1370>.
- 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>.
- Yuslita, H., Zulfan, Z., & Arifin, A. (2016). Analisis Tingkat Kesukaran Soal Dan Daya Pembeda Soal Mata Pelajaran Sejarah Kelas XI Semester Ganjil di SMA Negeri 5 Banda Aceh Tahun Pelajaran 2015-2016. *JIM: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, 1(1), 131–138. Retrieved from <http://www.jim.unsyiah.ac.id/sejarah/article/view/1784>.
- 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>.